

Variability in prior expectations explains biases in confidence reports

Pablo Tano^{1*}, Florent Meyniel², Mariano Sigman³ and Alejo Salles⁴

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).

¹Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales. Instituto de Cálculo. Buenos Aires, Argentina.

²Cognitive 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. ³Universidad Torcuato Di Tella, Almirante Juan Saenz Valiente 1010, C1428BJJ Buenos Aires, Argentina. ⁴CONICET-Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales. Instituto de Cálculo. Buenos Aires, Argentina. *e-mail: pablotanoretamales@gmail.com

26 A quantitative approach to confidence helps formalize the concept and unify its different
27 manifestations. In statistical decision theory, the normative definition of decision confidence is the
28 probability of a certain choice being correct [7-10]. Many models of how confidence emerges in the
29 brain have been proposed, such as accumulators [2, 11-13], drift diffusion [3], and attractor models
30 [14, 15]. In these models, confidence is interpreted as an algorithmic construction from variables
31 available in the decision-making process, rather than a readout from a Bayesian representation of
32 knowledge [15], using confidence metrics such as the difference between decision variables, post-
33 decision evidence and reaction time combined with evidence [3, 13, 15, 17, 18]. Depending on the
34 parameter values, these models can be close approximations to confidence as the probability of the
35 decision being correct or instead show systematic deviations from optimality.

36 The normative account of human decision confidence was recently put to test in [19], which shows
37 that a statistical definition of confidence indeed predicts various aspects of human confidence
38 reports, in both perceptual and knowledge-based tasks. This study is part of a resurgence of
39 rationality as a paramount human trait, which came about with the realization that probabilities are
40 the proper language in contexts of uncertainty such as those we encounter in everyday life [20, 21,
41 22]. The Bayesian rationality program came a long way in explaining human behavior in a wide
42 range of higher cognitive domains, such as intuitive physics [23], intuitive psychology [24, 25], or
43 causal inference [26]. Multimodal sensory integration remains a classic illustration of the flexibility
44 and optimality of our inference mechanisms [27, 28].

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

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

57 agents with different prior beliefs will be relatively overconfident about the accuracy of their
58 estimators [42]. In this work, we build on this idea, performing and modeling a wagering
59 experiment to ask whether the apparent irrationality of confidence reports can be reconciled with
60 the statistical account of confidence if participants' biases on prior expectations are taken into
61 account. In line with the heuristics and biases program, it has been thoroughly documented that
62 humans have a bias for optimism [43-45], for which even neural mechanisms have been described
63 [46]. Recently, quantitative modeling has shown that in some contexts this bias can lead to
64 counterintuitive behaviors, such as optimists being sometimes more conservative in their choices
65 [47]. Independently of the nature of this phenomenon, we encode the expectation bias (be it
66 optimistic or pessimistic) as the prior entering the Bayesian inference mechanism, and postulate
67 that it is necessary and sufficient to fully explain the otherwise irrational biases in confidence.

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

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

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

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

91 Results

92 Predictions from our model.

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

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

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

126 Decisions in gameplay, on the other hand, must face a tension between maximizing the immediate
127 reward by selecting the machine that apparently pays more (a behavior known as exploitation) and
128 exploring alternatives to more accurately learn the payoff associated to all machines and optimize
129 future decisions. This exploration-exploitation tradeoff implies that in some situations the rational
130 action (for long-term maximization of reward) would be to play the machine with the lowest payoff
131 so far. Indeed, when comparing two decisions with the same machine history, humans chose the
132 machine with the lowest payoff in $(12\pm 4)\%$ of gameplay decisions, and only $(4\pm 3)\%$ when they are
133 asked which machine has the highest payoff so far (expressed as mean \pm s.d. , $t=8.83$, $p<0.0001$
134 $(n=17)$). A heuristic solution to model this tradeoff is to more often try a specific machine the more
135 it seems to pay than the other. We formalized this decision strategy with a sampling process
136 characterized by the value of d and a parameter σ (adjusted and fixed to a single value for all
137 subjects) which favours exploration by introducing uncertainty over the estimated value of the
138 perceived difficulty d , increasing the probability of exploring the machine with the lowest payoff as
139 d increases (see Model Details in *Methods*). We emphasize that the normative account of confidence
140 in the 'which machine is better' decision depends only on the learning component and the decision
141 made by the participant, that is, it is independent of the just mentioned process that may have given
142 rise to the decision.

143
144 With the Bayesian solution to the learning component, one can expect two distinct features in the
145 confidence reports when comparing high and low values of the prior bias b (*optimists* and
146 *pessimists*, respectively). First, pessimists should in general have higher confidence than optimists,
147 particularly when machines pay more (prediction #2). Although it may appear counter-intuitive
148 that pessimistic people are more confident, this can be easily understood: expecting less, whenever
149 they find a machine that pays somewhat well, they are very confident that this machine is indeed

150 the best one. This in turn is amplified in high generosity scenarios because better machines are
151 played more, so the beliefs of the unexplored machine remains dominated by the pessimistic prior.
152 Optimists on the other hand expect more, so playing a good machine does not separate its reward
153 distribution that much from the prior, and thus confidence is lower. This effect is illustrated in Fig.
154 1.

155
156 A second, subtler effect, is related to the change in confidence as a function of the difficulty of the
157 task for agents with highly definite prior biases ($w > 2$). We first introduce some useful nomenclature
158 for this purpose. Selecting the best machine is difficult when both have paid similarly so far. We
159 manipulated the *generative difficulty* of a block experimentally by systematically varying the
160 difference between the real reward rates of machines. However, given that subjects experience
161 those reward rates only through noisy observations, subjects may experience a difficulty level that
162 departs from the generative difficulty level that we manipulated experimentally. This is particularly
163 clear when a machine is left unexplored. To quantify objectively the difficulty of choices that
164 subjects were exposed to, we computed the *unbiased difficulty*, namely, the difficulty that would be
165 perceived by a subject who is unbiased (neither optimistic nor pessimistic). Specifically, it is
166 computed as one minus the difference in mean reward rates at the moment of the report, as
167 estimated by an agent with a non-informative prior distribution (i.e. with b and w fixed to 1 and 2).
168 In the long run, if all options are repeatedly explored, prior biases (when any) will fade out and the
169 *perceived difficulty*, which takes into account the prior bias of each subject, will be approximately
170 equal to the unbiased difficulty.

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

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

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

189
190 Importantly, the two last predictions pertaining confidence are clearly falsifiable. This is because
191 since the optimism levels for participants are estimated from independent data (i.e. the gameplay),
192 the model for confidence has no adjustable parameters: its predictions are inescapable.
193 Additionally, these predictions are caused by the Bayesian representation of knowledge, not by the
194 specific confidence readout from beliefs' distributions. Therefore, similar effects of the prior bias
195 can be predicted not only on the computation of statistical confidence, but also on other metrics like
196 the distance between the means, the estimated payoff of the most generous machine, or the average
197 payoff of both machines, to name a few. The critical aspect is that the estimation of the underlying
198 distribution should take into account the subject's prior expectations.

199

200 Optimistic and pessimistic behaviors in gameplay

201 Participants played a two armed bandit game. Their task was to maximize the total reward. We
202 varied the reward rates of machine only in distinct blocks, each comprising 16 trials (in total,
203 36,720 decisions and 765 confidence reports). There were three types of blocks: those with a
204 question asked mid-block about which machine pays more, and the report of the continuous
205 associated confidence level ('confidence' blocks), and as a control, blocks with only the 'which'
206 question ('which' block) and blocks without any question ('no' blocks), see Fig. 2 and the *Methods*
207 section for details.

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

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

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

244 Optimism inferred from gameplay behavior explains the bias in confidence

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

247 confidence reports for all payoff settings of the '*confidence*' type blocks after sorting participants
248 into optimistic ($n=10$) and pessimistic ($n=7$), according to whether their prior bias b (fitted only
249 from gameplay, without the use of the confidence reports) was greater or smaller than $\frac{1}{2}$. The other
250 parameters (w and σ) showed no significant difference between the two groups. As predicted by
251 our model, optimistic participants were overall less confident in their answer to the question
252 "which machine pays more?" than the pessimistic participants and the difference grows with the
253 generosity of the task (Fig. 5 and Fig. S2a in Supplementary Information). The average reported
254 confidence for optimists and pessimists in all blocks was 0.48 ± 0.03 vs. 0.55 ± 0.03 respectively, and
255 0.60 ± 0.06 vs. 0.81 ± 0.03 in high generosity (>0.65) blocks (expressed as mean \pm s.e.m.). The average
256 confidence reported by participants entered a two-way ANOVA with group (optimists and
257 pessimists) and generosity condition (high and low) as between- and within-subjects factors,
258 respectively. The ANOVA showed a main effect of the optimistic/pessimistic group ($F(1,15)= 5.445$,
259 $p<0.04$) and a significant interaction between group and generosity ($F(1,15)= 8.385$, $p<0.02$).
260 Furthermore, although the average unbiased difficulty was approximately the same for all
261 participants, the average reported confidence per participant was not. Indeed, it decreased with the
262 b value adjusted for that participant in the same way as the increase in the average difficulty
263 perceived by an optimal agent with that b value (see Fig. S2b in Supplementary Information).

264
265 We now turn to our third prediction, that confidence reports in pessimists and optimists should
266 depend differently on the difficulty experienced in the task. The results, shown in Fig. 5, are striking:
267 optimists and pessimists show very different confidence patterns, as predicted by the model.
268 The isoconfidence lines are mostly vertical for optimistic agents, signaling a direct dependence with
269 task difficulty. For pessimists, however, the isoconfidence lines rotate from vertical to horizontal as
270 machine generosity increases, indicating a weaker dependence with the unbiased difficulty in high
271 generosity situations. To further emphasize the importance of taking into account participants'
272 prior beliefs, we compare the predictions of our model with those made by a model with a non-
273 informative prior for every participant: without accounting for participants' biases, the model
274 would show approximately vertical isoconfidence lines for both groups in Fig. 5. Additionally,
275 without separating optimists from pessimists, human confidence report would be overall judged as
276 overconfident in high generosity situations and underconfident in low generosity situations when
277 compared with the model with a non-informative prior (see Fig. S4 in Supplementary Information).
278 However, this confidence bias is precisely what is expected by the normative model after including
279 the b and w values fitted from the decision part of the task.

280

281 The Effect of Reporting Confidence

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

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

300 Of course, our decision model as-is is incapable of explaining this effect, since it has no information
301 regarding whether the confidence was reported or not. However, we can model the effect of the
302 confidence report by adding a small *ad hoc* feature to the framework developed. When asked for
303 confidence, the model increases its decision variance σ (see *Methods*) by a factor manually fitted to
304 0.05 times one minus the reported confidence. With this extra parameter, we are able to reproduce
305 the post-report behavior (see Fig. 6).

306 Discussion

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

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

318
319 Taking into account participants' prior expectations can explain some apparently irrational
320 behavior in various contexts [42]. For instance, we can revisit the result of Tversky and Griffin [33]
321 in much the same way. Their experiment consisted of informing participants that a coin is biased,
322 yielding one outcome 60% of time. Subjects were not informed if the bias was toward tails or heads.
323 Outcomes of this mysterious coin were shown, and participants were asked to decide if the bias was
324 toward tails or heads, and report their confidence in that decision. Subjects were typically
325 overconfident when they observed a few tosses with a strong imbalance and underconfident when
326 they observed many tosses with a moderate imbalance, by comparison with a mathematical model
327 informed that the bias is exactly 60% (a delta function), as the authors reported. If instead we
328 model this prior with some uncertainty (with a Beta distribution centered on 60%), then both
329 humans and the normative model display the same pattern of overconfidence and underconfidence,
330 and the magnitude of the confidence bias increases with the uncertainty of the prior (see
331 Supplementary Information and Fig. S3). Although we did not reproduce Griffin and Tversky's
332 experiment, and therefore cannot claim that participants' prior beliefs in that task are indeed better
333 explained by a more permissive prior, a similar logic operates in our gambling task. If we ignore
334 participants' prior bias from gameplay, and use the 'reasonable' non-informative prior instead, they
335 would be overall misjudged as overconfident (underconfident) in high (low) generosity situations
336 (see Fig. S4). We showed that participants' prior beliefs can be major determinants in the analysis
337 of rationality in the human sense of confidence, and seemingly reasonable (but erroneous)
338 assumptions, like a non-informative prior would be in our experiment, can lead to serious mistakes
339 when judging rationality.

340
341 Taking into account participants' prior beliefs appears key to evaluating the rationality of their
342 behavior in our task. Such an approach should nevertheless avoid two pitfalls. The first is circularity

343 (*assuming* a difference by appealing to priors rather than *explaining* it), which can be a limitation of
344 Bayesian models. Indeed, in principle, any behavior can be accounted for by a particular set of prior
345 beliefs [48, 54]. The second is overfitting (improving the goodness-of-fit by resorting to more free
346 parameters). Our approach avoids these pitfalls by using a parsimonious, general Bayesian model
347 that accounts for both choice and confidence, resulting in a parameter-free confidence model.
348 Indeed, we assumed that these two aspects of behavior (choice and confidence) are exclusively
349 affected by the same set of prior beliefs. Similarly to cross-validation methods, where data are
350 divided into “training and test” sets [55] in order to limit the complexity of the model (i.e. the
351 number of free parameters), our experimental design is divided into “training and test” tasks
352 (choices and confidence report, respectively). Behavior in the training task is used to learn -and pin
353 up- the prior parameters for each individual. If the resulting parameter-free model for the test task
354 predicts human data accurately, it is likely to generalize well to future data, limiting the risk of
355 overfitting [56]. A different way to learn the prior beliefs of participants is by iterated learning [57,
356 58], in which responses given in one trial affect the data shown in the next. It can be shown that, if
357 certain conditions are met, responses are eventually sampled from their prior distribution. In
358 principle, this method could be implemented in our environment by a repetition of various
359 experiments in which a participant judges the payoff of a machine based on the observed outcomes
360 in that trial, and the real payoff of the machine in the next trial would be equal to the estimated one
361 by the participant in the previous trial (starting with a random payoff in the first trial). As in our
362 approach, prior beliefs are then used to make parameter-free predictions in the test task and
363 evaluate the generalization potential of the model.

364
365 We chose to display the confidence in a two dimensional plane instead of along a one-dimensional
366 quantity (e.g. confidence vs. difficulty) following Aitchison *et. al.* [29], who argue that there is always
367 the freedom of reparameterizing confidence reports by a monotonous function, which in one
368 dimension would allow us to trivially explain any observed human confidence pattern by a suitable
369 such transformation. When plotting the results along two independent variables this
370 reparameterization is no longer possible, and the arising pattern of isoconfidence lines is now a
371 robust indicator of the participants' behavior. Specifically, there are (at least) two kinds of
372 optimality. The two dimensional representation only analyzes the transformation of incoming data
373 into an internal representation from which confidence is read out as a continuous variable (first
374 type of optimality), separating it from the mapping of this continuous variable onto some external
375 scale in order to report it (second type of optimality). Since the first type of optimality is

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

387
388 Although confidence reports have traveled a winding road in the psychology and neuroscience
389 literature, recent work is settling in on a statistically normative account of confidence [9, 19]. This
390 study contributes to this view, showing how traits that have been traditionally seen as irrational
391 can be in fact understood as differences in prior expectations. This further fuels the view of humans
392 as rational animals, and signs off another success of the Bayesian rationality program [59].

393
394 The importance of principled, quantitative and robust behavioral models is not only theoretical, but
395 also practical. Here, the availability of a successful model for confidence allowed us to study a
396 further, 'higher order' phenomenon: how reporting confidence affects later behavior. This question,
397 despite its simplicity, seems to have been overlooked in the literature so far. To our knowledge, we
398 present the first contribution in this direction. We report that probing subjects for a confidence
399 report increases explorative behavior in subsequent trials, as if subjects relied less on their prior
400 experience. An accurate model of confidence is useful here to disprove alternative low-level
401 explanations such as that the time spent answering the confidence question washes out the carried
402 knowledge representation. Indeed, by telling low and high confidence regimes apart through the
403 use of the model, we are able to show that the change in behavior is specific to low confidence
404 situations, hence ruling out an explanation in term of forgetting.

405
406 However, our model for gameplay choices does not *explain* this increased explorative behavior. In
407 the context studied, participants should have a fairly robust idea of the machine's payoffs by the
408 time they get to the confidence question, such that increasing exploration at that stage proves

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

418
419 Our results also illustrate that principled quantitative models prove particularly informative when
420 they predict non-trivial, maybe even counterintuitive behaviors [47]: a pessimistic agent should
421 yields higher confidences than an optimistic one. This is particularly relevant in relation to the
422 "irrational" optimism bias, by which we tend to expect more from the world than what it actually
423 gives us [43]. This bias towards high expectations would thus mean that we should typically display
424 *underconfidence* with respect to an unbiased agent, according to the aforementioned relation.
425 However, we found that a given individual can display both *under* and *overconfidence*, with a
426 tendency to the latter in most domains [62]. Here, we propose an explanation in a given task,
427 showing that apparently irrational confidence judgements can be simply understood as varying
428 prior biases. How general the form of this connection is, and, how optimism and overconfidence
429 may coexist are interesting new avenues for research.

430

431

432 Methods

433 Experiment Details

434 A total of 18 adult participants played a two armed bandit game for which they were asked to
435 maximize total reward. One participant was excluded from the analysis for obtaining a total reward
436 consistent with random play. Each participant completed 135 blocks of 16 trials, giving a total of
437 36720 individual decisions and 765 confidence reports. Participants were informed they were
438 going to play a series of unrelated blocks in each of which the payoff of the machines was unknown

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

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

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

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

456 Model Details

457 *Bayesian knowledge update.* Observers begin each block with a prior distribution $Beta(ps, pf)$ for the
458 reward probability of each of the two machines, where ps and pf encode fictitious prior successes
459 and failures, respectively. A natural reparameterization of this distribution is by using its mean
460 $b=ps/(ps+pf)$, which is a prior measure of the expected payoff, and $w=ps+pf$, which encodes the
461 weight of prior evidence. Due to the conjugacy between the beta prior and the binomial likelihood
462 assumed for the rewards, the posterior distribution after experiencing s successes and f failures
463 results in a $Beta(ps+s, pf+f)$. Intuitively, low values of the prior mean b correspond to a pessimistic
464 perspective, while high values of this parameter represent a more optimistic take.

465

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

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

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

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

475
476 *Decision-making model and motivation.* The confidence report depends only on the learning
477 component, whose optimal solution is presented above. By contrast, the optimal solution for
478 deciding which machine to play given the experience so far in order to maximize future rewards is
479 more difficult. However, this is a well studied problem that has been solved in finite-horizon bandit
480 problems by dynamic programming -looking at all possible outcomes from the last trial to the
481 current one-, an approach that requires an amount of calculations that grows exponentially with the
482 number of remaining trials [52, 53].

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

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

502 all the variation in the behavior of different participants can be accounted for only by a different
503 prior mean b (Figs. 3, 4 and S1). This assumption agrees with the statistical analysis of the data,
504 which shows that the model that varies only b was a better model when compared to any other
505 model with one, two or three degrees of freedom ($x_p > 0.69$ -exceedance probability-, see below).
506 Additionally, it is not possible to cover the entire spectrum of behavioral summaries seen in Figs. 3,
507 4 and S1 just by varying w or σ , and neither w nor σ are as consistent as b for each participant
508 across different environmental conditions.

509

510 Analysis Details

511

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

522 We can check the statistical validity of this assertion by performing a permutation test in the
523 following manner. For each participant, we compute the variance of the values of b that best
524 represent his or her behavior in each of the 10 different conditions *separately*. For example, if a
525 participant is best represented by an optimistic behavior in some conditions, but by a pessimistic
526 behavior in others, their value of the variance will be high. A within-subject variance that is smaller
527 than the between-subject variance indicates that the optimistic/pessimistic difference pertains to
528 the group level rather than for individual subjects. Therefore, the test consisted on performing
529 100,000 random permutations between the participants' labels across conditions, and measuring
530 the variance for each surrogate across the 10 conditions. For every permutation, the calculated
531 variance for all surrogates was higher than the variance of all participants without permuting,
532 yielding $p < (1/100,000)$.

533

534 *Two-dimensional presentation of the results.* Since the iso confidence lines in Fig. 5 do not change if
535 we reparameterize confidence by a monotonous function, and therefore neither does the prediction
536 #3 of our model, we chose to display the data as was reported by the participant in the continuous
537 confidence bar between 0 and 1, without any calibration function. If we, for example, assign
538 quantiles to the answers of each subject, then prediction #2 of the model (namely, pessimists report
539 higher confidence than optimists) is partially opaqued. However, since the strength of this
540 asymmetry between optimists and pessimists increases when one option is played more frequently
541 than the other, we still see, after reparametrization, that optimists and pessimists' confidence varies
542 in a different way with the generosity of the task (which, in practice, is proportional to the
543 difference between the exploitation of different options), both for humans and for the normative
544 model. This is shown in Fig. S2a (see Supplementary Information).

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

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

558
559 *Model comparison.* The exceedance probability (xp) of a model quantifies the probability that this
560 model is more frequent than the others (within the tested set) in the general population of subjects.
561 We computed exceedance probabilities from the model evidence using the software developed by
562 [65]. The evidence for each model and each subject was calculated by integrating each model's
563 mean likelihood over its parameters, under the i.i.d data assumption. The integral was
564 approximated by a sum over a discrete grid. Grid points (10 points for b and w , 8 for σ) were spaced
565 linearly for parameters b and w and exponentially for σ (which corresponds to a non-informative
566 prior in log-space, a natural choice for variance parameters). The final result depends on the

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

573

574 *Data availability.* The behavioural data are available here:

575 https://figshare.com/articles/Behavioral_data/4788823

576 Acknowledgements

577 We would like to thank Falk Lieder, Tania Lombrozo and Tom Griffiths for insightful conversations,
578 Melisa Bentivegna for help with early data collection, and the organizers of the Subjective
579 Confidence Workshop at Les Treilles where some preliminary results were first presented. This
580 work was partly funded by grants CONICET PIP 11220130100384CO and UBACYT
581 20020130200202BA. FM is funded by the French center for atomic energy (CEA). M.S. is
582 sponsored by CONICET and the James McDonnell Foundation 21st Century Science Initiative in
583 Understanding Human Cognition – Scholar Award.

584

585 Author contributions

586 P.T. performed the experiment, analysed the data, interpreted the results and wrote the manuscript.

587 F.M. provided analytical tools, interpreted the results and edited the manuscript.

588 M.S. interpreted the results and edited the manuscript.

589 A.S. designed the study, performed the experiment, analysed the data, interpreted the results, and
590 wrote the manuscript.

591 Additional information

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

593

594
595
596

597 References

- 598 1. Plous, Scott. The psychology of judgment and decision making. McGraw-Hill Book Company, 1993.
599
- 600 2. Kepecs, Adam, et al. "Neural correlates, computation and behavioural impact of decision confidence."
601 Nature 455.7210 (2008): 227-231.
602
- 603 3. Kiani, Roozbeh, Leah Corthell, and Michael N. Shadlen. "Choice certainty is informed by both evidence
604 and decision time." Neuron 84.6 (2014): 1329-1342.
605
- 606 4. Persaud, Navindra, Peter McLeod, and Alan Cowey. "Post-decision wagering objectively measures
607 awareness." Nature neuroscience 10.2 (2007): 257-261.
608
- 609 5. Lak, A. et al. Orbitofrontal cortex is required for optimal waiting based on decision
610 confidence. Neuron 84, 190–201 (2014).
611
- 612 6. Kepecs, Adam, and Zachary F. Mainen. "A computational framework for the study of confidence in
613 humans and animals." Philosophical Transactions of the Royal Society of London B: Biological Sciences
614 367.1594 (2012): 1322-1337.
615
- 616 7. Berger, James O. Statistical decision theory and Bayesian analysis. Springer Science & Business
617 Media, 2013.
618
- 619 8. Baranski, Joseph V., and William M. Petrusic. "The calibration and resolution of confidence in
620 perceptual judgments." Perception & Psychophysics 55.4 (1994): 412-428.
621
- 622 9. Pouget, Alexandre, Jan Drugowitsch, and Adam Kepecs. "Confidence and certainty: distinct
623 probabilistic quantities for different goals." Nature neuroscience 19.3 (2016): 366-374.
624
- 625 10. Peterson, Cameron R., and Lee R. Beach. "Man as an intuitive statistician." Psychological bulletin
626 68.1 (1967): 29.
627
- 628 11. Yu, Shuli, Timothy J. Pleskac, and Matthew D. Zeigenfuse. "Dynamics of postdecisional processing of
629 confidence." Journal of Experimental Psychology: General 144.2 (2015): 489.
630
- 631 12. Navajas, Joaquin, Bahador Bahrami, and Peter E. Latham. "Post-decisional accounts of biases in
632 confidence." Current Opinion in Behavioral Sciences 11 (2016): 55-60.
633
- 634 13. Paz, Luciano, et al. "Confidence through consensus: a neural mechanism for uncertainty monitoring."
635 Scientific reports 6 (2016).
636
- 637 14. Wei, Ziqiang, and Xiao-Jing Wang. "Confidence estimation as a stochastic process in a
638 neurodynamical system of decision making." Journal of neurophysiology 114.1 (2015): 99-113.

- 639
640 15. Insabato, Andrea, et al. "Confidence-related decision making." *Journal of neurophysiology* 104.1
641 (2010): 539-547.
642
643 16. Meyniel, Florent, Mariano Sigman, and Zachary F. Mainen. "Confidence as bayesian probability: From
644 neural origins to behavior." *Neuron* 88.1 (2015): 78-92.
645
646 17. Pleskac, Timothy J., and Jerome R. Busemeyer. "Two-stage dynamic signal detection: a theory of
647 choice, decision time, and confidence." *Psychological review* 117.3 (2010): 864.
648
649 18. Drugowitsch, Jan, Rubén Moreno-Bote, and Alexandre Pouget. "Relation between belief and
650 performance in perceptual decision making." *PloS one* 9.5 (2014): e96511.
651
652 19. Sanders, Joshua I., Balázs Hangya, and Adam Kepecs. "Signatures of a statistical computation in the
653 human sense of confidence." *Neuron* 90.3 (2016): 499-506.
654
655 20. Jones, Matt, and Bradley C. Love. "Bayesian fundamentalism or enlightenment? On the explanatory
656 status and theoretical contributions of Bayesian models of cognition." *Behavioral and Brain Sciences*
657 34.04 (2011): 169-188.
658
659 21. Anderson, John Robert. *The adaptive character of thought*. Psychology Press, 1990.
660
661 22. Jaynes, Edwin T. *Probability theory: The logic of science*. Cambridge university press, 2003.
662
663 23. Sanborn, Adam N., Vikash K. Mansinghka, and Thomas L. Griffiths. "Reconciling intuitive physics and
664 Newtonian mechanics for colliding objects." *Psychological review* 120.2 (2013): 411.
665
666 24. Goodman, Noah D., et al. "Intuitive theories of mind: A rational approach to false belief." *Proceedings*
667 *of the twenty-eighth annual conference of the cognitive science society*. 2006.
668
669 25. Ullman, Tomer D., Noah D. Goodman, and Joshua B. Tenenbaum. "Theory learning as stochastic
670 search in the language of thought." *Cognitive Development* 27.4 (2012): 455-480.
671
672 26. Tenenbaum, Joshua B., and Thomas L. Griffiths. "Theory-based causal inference." *Advances in*
673 *neural information processing systems* (2003): 43-50.
674
675 27. Ernst, Marc O., and Martin S. Banks. "Humans integrate visual and haptic information in a statistically
676 optimal fashion." *Nature* 415.6870 (2002): 429-433.
677
678 28. Deroy, Ophelia, Charles Spence, and Uta Noppeney. "Metacognition in Multisensory Perception."
679 *Trends in Cognitive Sciences* 20.10 (2016): 736-747.
680
681 29. Aitchison, Laurence, et al. "Doubly bayesian analysis of confidence in perceptual decision-making."
682 *PLoS Comput Biol* 11.10 (2015): e1004519.
683
684 30. Bar-Tal, Yoram, Anat Sarid, and Liat Kishon-Rabin. "A test of the overconfidence phenomenon using
685 audio signals." *The Journal of general psychology* 128.1 (2001): 76-80.
686
687 31. Björkman, Mats, Peter Juslin, and Anders Winman. "Realism of confidence in sensory discrimination:

- 688 The underconfidence phenomenon." *Attention, Perception, & Psychophysics* 54.1 (1993): 75-81.
689
- 690 32. Camerer, Colin, and Dan Lovallo. "Overconfidence and excess entry: An experimental approach." *The*
691 *American Economic Review* 89.1 (1999): 306-318.
692
- 693 33. Griffin, Dale, and Amos Tversky. "The weighing of evidence and the determinants of confidence."
694 *Cognitive psychology* 24.3 (1992): 411-435.
695
- 696 34. Kvidera, Sara, and Wilma Koutstaal. "Confidence and decision type under matched stimulus
697 conditions: Overconfidence in perceptual but not conceptual decisions." *Journal of Behavioral Decision*
698 *Making* 21.3 (2008): 253-281.
699
- 700 35. Moore, Don A., and Paul J. Healy. "The trouble with overconfidence." *Psychological review* 115.2
701 (2008): 502.
702
- 703 36. Olsson, Henrik, and Anders Winman. "Underconfidence in sensory discrimination: The interaction
704 between experimental setting and response strategies." *Attention, Perception, & Psychophysics* 58.3
705 (1996): 374-382.
706
- 707 37. Gigerenzer, Gerd, Ulrich Hoffrage, and Heinz Kleinbölting. "Probabilistic mental models: a
708 Brunswikian theory of confidence." *Psychological review* 98.4 (1991): 506.
709
- 710 38. Mamassian, Pascal. "Overconfidence in an objective anticipatory motor task." *Psychological Science*
711 19.6 (2008): 601-606.
712
- 713 39. Graziano, Martin, and Mariano Sigman. "The spatial and temporal construction of confidence in the
714 visual scene." *PLoS One* 4.3 (2009): e4909.
715
- 716 40. Ma, Wei Ji, and Mehrdad Jazayeri. "Neural coding of uncertainty and probability." *Annual review of*
717 *neuroscience* 37 (2014): 205-220.
718
- 719 41. Griffiths, Thomas L., et al. "How the Bayesians got their beliefs (and what those beliefs actually are):
720 comment on Bowers and Davis (2012)." (2012): 415.
721
- 722 42. Van den Steen, Eric. "Overconfidence by Bayesian-rational agents." *Management Science* 57.5
723 (2011): 884-896.
724
- 725 43. Sharot, Tali. "The optimism bias." *Current Biology* 21.23 (2011): R941-R945.
726
- 727 44. Sharot, Tali, Christoph W. Korn, and Raymond J. Dolan. "How unrealistic optimism is maintained in
728 the face of reality." *Nature neuroscience* 14.11 (2011): 1475-1479.
729
- 730 45. Shepperd, James A., et al. "A primer on unrealistic optimism." *Current directions in psychological*
731 *science* 24.3 (2015): 232-237.
732
- 733 46. Sharot, Tali, et al. "Neural mechanisms mediating optimism bias." *Nature* 450.7166 (2007): 102-105.
734
- 735 47. Lefebvre, Germain, et al. "Behavioural and neural characterization of optimistic reinforcement
736 learning." *Nature Human Behaviour* 1 (2017): 0067.

- 737
738 48. Bowers, Jeffrey S., and Colin J. Davis. "Bayesian just-so stories in psychology and neuroscience."
739 Psychological bulletin 138.3 (2012): 389.
740
741 49. Macready, William G., and David H. Wolpert. "Bandit problems and the exploration/exploitation
742 tradeoff." IEEE Transactions on evolutionary computation 2.1 (1998): 2-22.
743
744 50. Cohen, Jonathan D., Samuel M. McClure, and J. Yu Angela. "Should I stay or should I go? How the
745 human brain manages the trade-off between exploitation and exploration." Philosophical Transactions of
746 the Royal Society of London B: Biological Sciences 362.1481 (2007): 933-942.
747
748 51. Daw, Nathaniel D., et al. "Cortical substrates for exploratory decisions in humans." Nature 441.7095
749 (2006): 876-879.
750
751 52. Gittins, John, Kevin Glazebrook, and Richard Weber. Multi-armed bandit allocation indices. John
752 Wiley & Sons, 2011.
753
754 53. Steyvers, Mark, Michael D. Lee, and Eric-Jan Wagenmakers. "A Bayesian analysis of human
755 decision-making on bandit problems." Journal of Mathematical Psychology 53.3 (2009): 168-179.
756
757 54. Diaconis, Persi, and David Freedman. "On the consistency of Bayes estimates." The Annals of
758 Statistics (1986): 1-26.
759
760 55. Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. The elements of statistical learning. Vol. 1.
761 Springer, Berlin: Springer series in statistics, 2001.
762
763 56. Myung, In Jae. "The importance of complexity in model selection." Journal of mathematical
764 psychology 44.1 (2000): 190-204.
765
766 57. Lewandowsky, Stephan, Thomas L. Griffiths, and Michael L. Kalish. "The wisdom of individuals:
767 Exploring people's knowledge about everyday events using iterated learning." Cognitive Science 33.6
768 (2009): 969-998.
769
770 58. Griffiths, Thomas L., and Michael L. Kalish. "Language evolution by iterated learning with Bayesian
771 agents." Cognitive science 31.3 (2007): 441-480.
772
773 59. Oaksford and Chater. "Bayesian Rationality. The Probabilistic Approach to Human Reasoning."
774 Oxford University Press (2007).
775
776 60. Denison, Stephanie, et al. "Rational variability in children's causal inferences: The sampling
777 hypothesis." Cognition 126.2 (2013): 285-300.
778
779 61. Gonzalez, Aaron, et al. "Is that your final answer? The effects of neutral queries on children's
780 choices." CogSci. 2012.
781
782 62. Salles, Alejo, et al. "The metacognitive abilities of children and adults." Cognitive Development 40
783 (2016): 101-110.
784
785 63. Shinnars, Pete (2011). PyGame - Python Game Development. Retrieved from

786 <http://www.pygame.org>.

787

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

790

791 65. Stephan, Klaas Enno, et al. "Bayesian model selection for group studies." Neuroimage 46.4 (2009):
792 1004-1017.

793

794

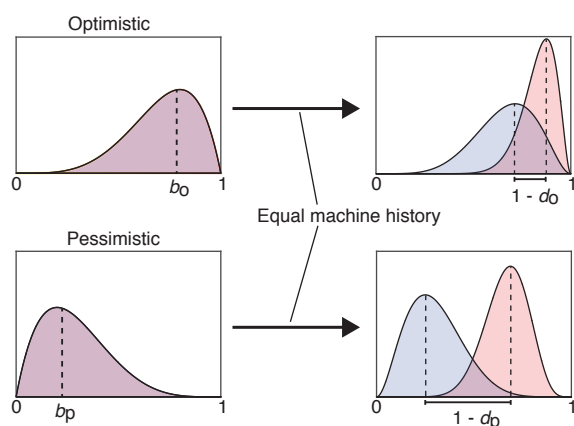


Figure 1. A probabilistic model of the task, pessimists should report more confidence than optimists. Initially (left), the prior belief distribution over the payoff of both machines lies in high values for optimists (high b_0) and low values for pessimists (low b_p). In situations in which one option 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 ($d_p < d_0$), and a higher confidence report for pessimists than optimists.

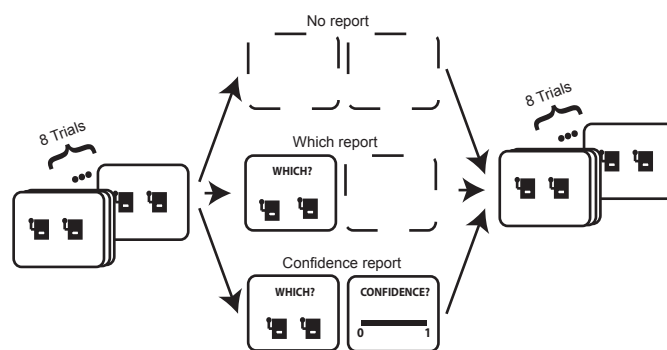


Figure 2. The bandit gambling task. Each block consisted of 16 trials. Depending on the type of block being played, the block was played without interruption (No report); the participant was required to choose which machine had the higher nominal payoff (Which report); or same as the Which report plus 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.

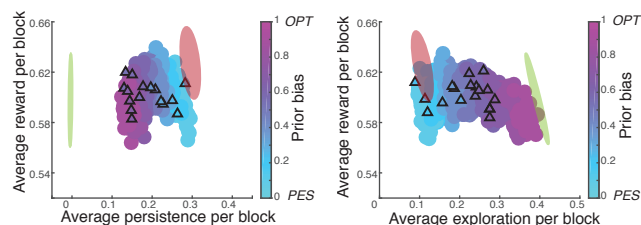


Figure 3. A range of behaviors accounted for by the optimism level. Average persistence per block corresponds to the proportion of trials in which a machine is chosen immediately after a no-reward trial in that machine. Average exploration per block is the proportion of trials in which there is a change in machine choice. Average reward per block is to the proportion of trials in which a reward was obtained. Participants are shown with black-edged 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.

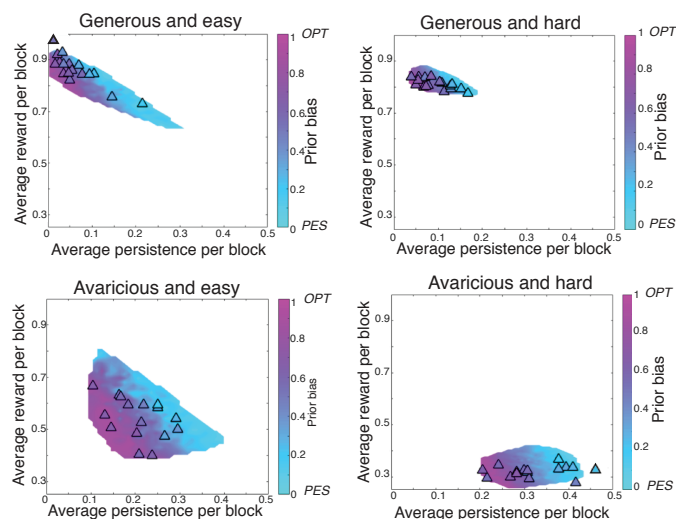


Figure 4. The optimism level is an idiosyncratic trait, stable across conditions. Average rewards vs. average persistence in different regimes. Legend as in previous figure. Each panel shows the averaged results from 10 to 15 different blocks, all belonging to a different regime (Generous and Easy, Generous 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*).

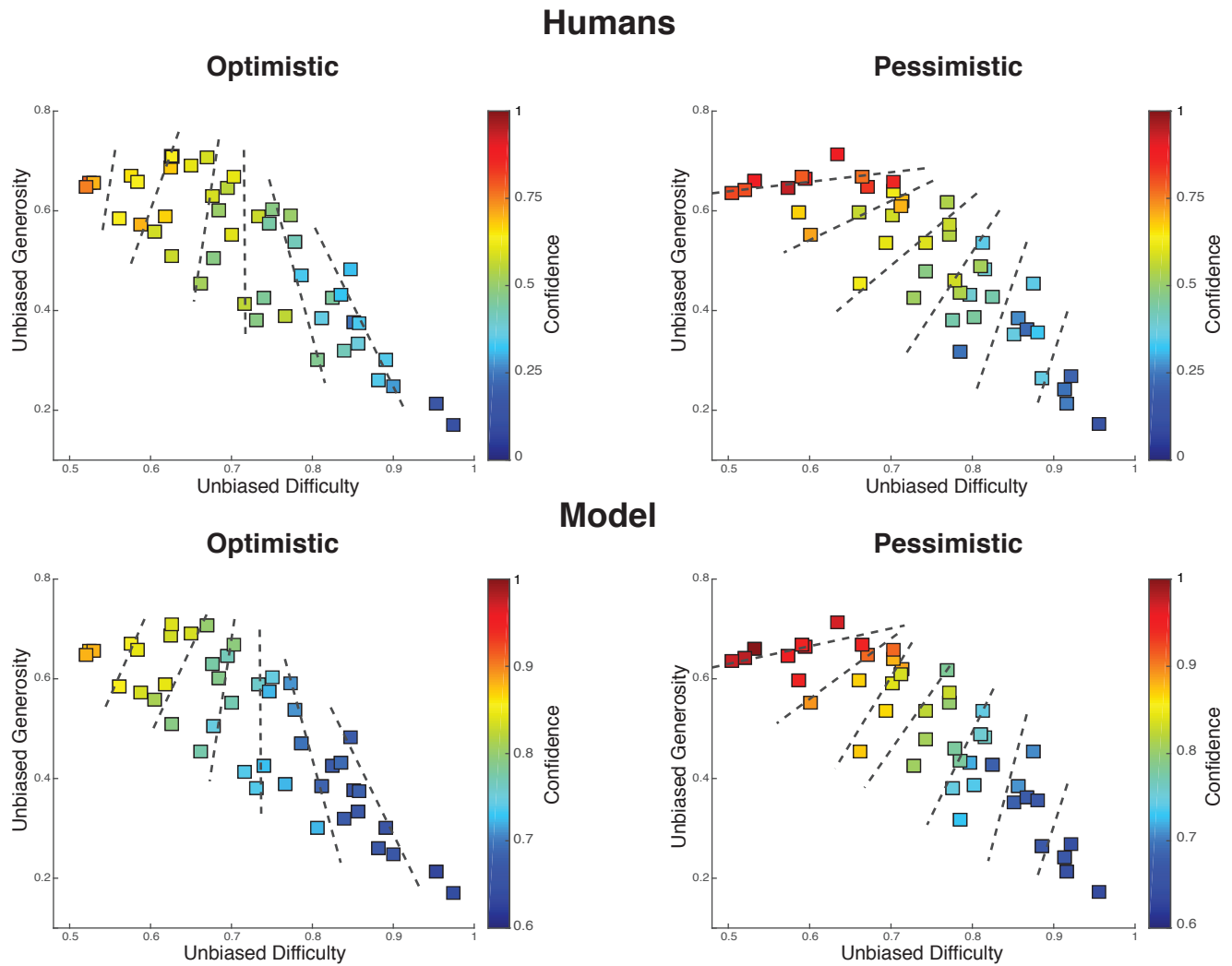


Figure 5. Over- and under-confidence are explained by the prior optimism level measured in gameplay. Confidence levels reported by humans and the normative model as a function of the unbiased difficulty and unbiased generosity on the block. Unbiased difficulty (generosity) is computed as the distance (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 confidence 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.

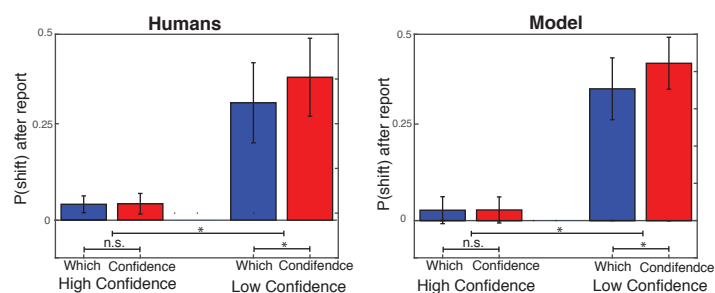


Figure 6. Probing confidence induces a deconfirmation bias in subsequent choices. In low confidence situations, the probability of shifting to the other option after the report is bigger in blocks in which confidence is reported ('confidence') than in blocks in which only the best machine is reported ('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 ($n=17$).