

1 **Explicit representation of confidence informs future**  
2 **value-based decisions**

3

4 **Tomas Folke<sup>1</sup>, Catrine Jacobsen<sup>2</sup>, Stephen M. Fleming<sup>3</sup> and Benedetto De Martino<sup>4, 1\*</sup>**

5 1. Department of Psychology, University of Cambridge Downing Street, Cambridge CB2 3EB, United Kingdom

6 2. University of Copenhagen, Institute of Food and Resource Economics, Copenhagen, Denmark

7 3. University College London, Wellcome Trust Centre for Neuroimaging, 12 Queen Square, WC1N 3BG London, United  
8 Kingdom

9 4. Institute of Cognitive Neuroscience, 17-19 Queen Square, WC1N 3AR, London, United Kingdom

10 \* Correspondence should be addressed to benedettodemartino@gmail.com

## 11 Abstract

12

13 Humans can reflect on previous decisions and report variable levels of confidence. But why  
14 maintain an explicit representation of confidence for choices that have already been made and  
15 therefore cannot be undone? Here we show that an explicit representation of confidence is  
16 harnessed for subsequent changes of mind. Specifically, when confidence is low, participants are  
17 more likely to change their minds when the same choice is presented again, an effect that is most  
18 pronounced in participants with greater fidelity in their confidence reports. Furthermore, we show  
19 that choices reported with high confidence follow a more consistent pattern (fewer transitivity  
20 violations). Finally, by tracking participants' eye movements we demonstrate that lower-level gaze  
21 dynamics can track uncertainty but do not directly impact changes of mind. Taken together, these  
22 results suggest that an explicit and accurate representation of confidence has a positive impact on  
23 the quality of future value-based decisions.

24

## 25 Introduction

26

27 As we navigate through life we are constantly faced with choices that require us to assign and  
28 compare the values of different options or actions. Some of these value-based choices seem  
29 relatively straightforward ('What should I eat for lunch?') and others less so ('Which job offer should  
30 I take?'). No matter how simple or complex these choices are, they are often accompanied by a  
31 sense of confidence in having made the right choice. Recent work has shown that it is possible to  
32 behaviourally and computationally dissociate a value estimate ('*How much do I like something?*')  
33 from internal fluctuations in confidence ('*How sure am I?*'). For example, at a behavioural level it  
34 has been shown that confidence shares only a limited amount of variance with value, and instead  
35 reflects an assessment of choice accuracy<sup>1</sup>. This relation between value and confidence is neatly  
36 accounted for computationally by assuming that confidence emerges from the dynamics of noisy  
37 accumulators in an evidence-accumulation framework<sup>1, 2, 3, 4</sup>. More recently, Lebreton and

38 colleagues have shown that confidence may be an inherent property of value estimation, sharing a  
39 quadratic relationship with a linear rating of value<sup>5</sup> (see also the work of Barron and colleagues<sup>6</sup>).  
40 But what is the function of confidence? Why maintain an explicit representation of confidence when  
41 a choice has already been made and therefore cannot be undone?

42

43 According to one view, confidence can be thought of as a by-product of a stochastic accumulation  
44 process implemented in the ventromedial prefrontal cortex (vmPFC) during value comparison.  
45 Previous work indicates the brain constructs an explicit representation of confidence that underpins  
46 verbal reports<sup>7, 8</sup>. A range of studies suggests that the rostralateral prefrontal cortex represents  
47 confidence in both value-based and perceptual decisions<sup>1, 9, 10, 11</sup>. Explicit representations of  
48 confidence allow individuals to communicate the strength of their beliefs to others, facilitating group  
49 decisions<sup>12, 13</sup>, but may play little role in one's own decision process.

50

51 An alternative view is that explicit representations of confidence are critical for guiding one's own  
52 future behaviour<sup>14</sup>. Work in perceptual decision-making has revealed commonalities between  
53 mechanisms supporting confidence construction and error-monitoring<sup>15, 16</sup>, suggesting changes of  
54 mind may be informed by confidence<sup>4</sup>. However, it is unknown whether confidence is harnessed  
55 over a longer timescale to guide future choices. Here we aim to test the hypothesis that an explicit  
56 (and well-tuned) representation of confidence in a recent choice can guide a decisions maker's  
57 choice when faced by the same (or a similar) decision again. To test this hypothesis, we presented  
58 participants with the same set of choices more than once during the course of two experiments and  
59 tested which factors were associated with a change of mind. We then investigated how confidence  
60 related to the degree of internal consistency in their patterns of choice. Choice consistency can be  
61 quantified by measuring the degree of transitivity across choices. Here we introduced a novel  
62 method for tagging choices as conforming to or violating transitivity. Using this method we were  
63 able to show that explicit representations of confidence are associated with more consistent

64 patterns of choice as a consequence of changes of mind. Finally, we directly contrasted the effect  
65 of explicit confidence reports with lower-level markers of uncertainty that we gathered using eye  
66 tracking, revealing that changes of mind were specifically associated with explicit reports of  
67 confidence.

68

## 69 Results

70 We collected data in two experiments in which hungry participants made choices between food  
71 items (which they could consume later) while their eye movements were monitored. In the first  
72 experiment the twenty-eight participants included in the study were shown high-definition pictures  
73 of *two* snacks and were asked to choose the preferred one (Figure 1 A). In a second experiment  
74 twenty-four participants chose their preferred snack among *three* snacks available in each trial  
75 (Figure 1 D). After making each choice, participants reported their degree of confidence in having  
76 made the ‘correct’ choice, which in this design equates to choosing the higher valued item. The  
77 value for individual items was elicited using a standard incentive compatible BDM-method<sup>17</sup>. The  
78 experimental procedure we used was adapted (with modifications) from a task we developed  
79 previously<sup>1</sup> (see methods for more details).

80

### 81 ***Relation between confidence and choice***

82 In line with a wealth of previous research<sup>18, 19, 20, 21, 22</sup> we found that the difference in value between  
83 the two items (constructed from values elicited through an incentive-compatible bidding procedure,  
84 BDM) was a reliable predictor of participants’ choices in both experiments (hierarchical logistic  
85 regression; *Experiment 1*:  $z = 11.48$ ,  $p < .0001$ , Figure 1C and F; *Experiment 2*:  $z = 6.66$ ,  $p < .0001$ ,  
86 Figure 1B and E). Note that in the three-choice design (Experiment 2) DV was calculated as the  
87 difference between the value of the reference item and the average of the two other available  
88 options (following Krajbich and Rangel<sup>23</sup>). In the supplemental materials (S1) we additionally report  
89 the result of a multinomial logistic regression model in which the value of each option was inputted

90 independently and therefore does not require a priori specification of DV. This analysis yielded the  
91 same pattern of results. In both studies we also identified a significant negative interaction between  
92 the summed value of all options (SV) and value difference (DV) (*Experiment 1*:  $z=-3.08$ ,  $p<.005$ ;  
93 *Experiment 2*:  $z=-2.84$ ,  $p<.005$ ), indicating that DV had a stronger influence on choice when item  
94 values were low, compared to when items were high in value (Figure 1 C and F). To our  
95 knowledge this effect has not been reported before but is consistent with the Weber–Fechner law  
96 in sensory perception in which the resolution of precepts diminishes for stimuli of greater  
97 magnitude. The effect is also compatible with the notion of normalization<sup>24, 25, 26</sup>. Confidence,  
98 unlike DV, was not in itself a predictor of choice (right or left item) but instead correlated with  
99 choice accuracy, with a steeper slope relating DV to choice when confidence was high, as found  
100 previously<sup>1</sup> (Fig. 1B, E; *Experiment 1*:  $z=7.43$ ,  $p<.0001$ ; *Experiment 2*:  $z=5.82$ ,  $p<.0001$ ).

101

102 Using eye tracking we measured the dynamics of eye movements between items during the  
103 choice, both the total amount of time participants spent looking at each item and how frequently  
104 gaze shifted back and forth between items (see supplementary materials, section S2). Replicating  
105 previous studies<sup>23, 27</sup> we found that the difference in dwell time (DDT) was a robust predictor of  
106 choice in both two-option and three-option experiments (*Experiment 1*:  $z=4.95$ ,  
107  $p<.0001$ ; *Experiment 2*:  $z=9.81$ ,  $p<.0001$ ; Figure 1 C and F).

108

109 For a full list of fitted models and their respective BIC scores see the supplementary materials  
110 (section S3).

## 111 **Figure 1**

### 112 ***Factors that contribute to confidence***

113

114 We next investigated which variables contributed to subjective confidence during value-based

115 choice. Our previous work shows an interrelationship between absolute difference in value (IDVI),

116 response time (RT) and confidence (i.e. participants are more confident when IDVI is high and their  
117 choice are faster)<sup>1</sup>. These findings are in line with the conceptual relation between confidence,  
118 strength of evidence (indexed by IDVI in the value-based framework) and decision time<sup>3, 28</sup>. We  
119 observed this same relation in the current study. In both experiments we found that IDVI was a  
120 significant predictor of confidence (*Experiment 1*:  $t=13.43$   $p<.0001$ ; *Experiment 2*:  $t=7.46$ ,  $p<.0001$ ).  
121 We also found that RT was a negative predictor of confidence (*Experiment 1*:  $t=-10.01$ ,  $p<.0001$ ;  
122 *Experiment 2*:  $t=-7.53$ ,  $p<.0001$ ). Additionally, we found that summed value positively predicted  
123 confidence, meaning that participants tended to be more confident when the options were all high  
124 in value (*Experiment 1*:  $t=3.50$ ,  $p<.005$ ; *Experiment 2*:  $t=4.80$ ,  $p<.0001$ ). This finding indicates that  
125 overall value might boost confidence, despite paradoxically making choices *less* accurate. More  
126 broadly these findings highlight how evidence and confidence, though related, play partially  
127 independent roles in the decision making process. Note that all of the predictors analysed in this  
128 section were entered into the same hierarchical linear regression; therefore all the effects hold  
129 when controlling for the other variables reported.

130

131 We also hypothesized that lower-level features of information sampling may reflect an individual's  
132 explicit confidence reports. To test this idea, we constructed a novel measure that captured  
133 uncertainty in information-sampling behaviour. This new measure, which we label "gaze shift  
134 frequency" (GSF), indexes how frequently gaze shifted back and forth among the options  
135 presented on the screen. This measure is independent of difference in dwell-time (*Experiment 1*  
136  $r=-.02$ , *Experiment 2*  $r=.04$ ): for a constant allocation of time between the options (e.g. 3 seconds  
137 for the left-hand option and 5 seconds for the right-hand option) one may shift fixation only once  
138 (switching from left to right after 3 seconds have elapsed, for example; low gaze shift frequency) or  
139 shift many times between the two options (high gaze shift frequency). We found that GSF was a  
140 robust negative predictor of confidence in both experiments (*Experiment 1*:  $t=-3.67$ ,  $p<.005$ ;  
141 *Experiment 2*:  $t=-8.94$ ,  $p<.0001$ ) see Figures 2 A and B. In other words, in trials in which

142 participants shifted their gaze more often between the available options their confidence was  
143 lower, even after accounting for changes in IDVI and RT. The four-way relationship between IDVI,  
144 RT, GSF and confidence is plotted in Figures 2 C and D. Correlation tables can be found in the  
145 supplementary materials (S4).

146

## 147 **Figure 2**

148

### 149 ***Confidence predicts change of mind***

150 In both experiments participants saw the same exact choice sets on more than one occasion. In  
151 Experiment 1 each pair was presented twice; in Experiment 2 each triad was presented three times  
152 (counterbalancing for different spatial locations). This design allowed us to determine factors  
153 affecting changes of mind when the same choice is encountered again. Note that the way we  
154 define change of mind in this study is different from how it is often defined in perceptual decision-  
155 making, as a reversal in an ongoing motor plan due to additional processing of sensory  
156 information<sup>4, 15, 29, 30</sup>. The hypothesis we sought to test was that an explicit report of confidence in  
157 an initial choice at time  $t$  would influence behaviour when the same decision was presented again  
158 at a future time  $t_{future}$ . In a hierarchical logistic regression, lower confidence at time  $t$  was indeed  
159 associated with increased changes of mind at time  $t_{future}$  in both experiments (*Experiment 1*:  $z=-$   
160  $6.70$ ,  $p<.0001$ ; *Experiment 2*:  $z=-5.71$ ,  $p<.0001$ ). The effect of confidence in predicting change of  
161 mind remained robust after controlling for several other factors that might correlate with the stability  
162 of a choice such as IDVI and RT. Because IDVI correlated positively with confidence (see the  
163 previous section and S4) we checked the covariance matrices and Variance Inflation Factors  
164 (VIFs) to ensure that these correlations did not influence the interpretation of our findings. Both the  
165 covariances and the VIFs were below standard thresholds, allowing straightforward interpretation  
166 of coefficients (see S5). Furthermore, to rule out the possibility that the effect we observed was  
167 driven by the presence of fast errors that were later corrected by the participant, we reanalysed the

168 data excluding all trials that were faster than each participant's mean response time. This analysis  
169 produced comparable results (See S6). Notably GSF, itself a correlate of confidence, did not  
170 predict change of mind when included in the regression analysis (Fig. 3 A and B, *coefficients in*  
171 *blue*), even when excluding reported confidence from the regression analysis (see supplementary  
172 materials, section S7a). Together, these results suggest that a low-level (and possibly implicit)  
173 representation of uncertainty indexed by GSF is insufficient to trigger a future change of mind. On  
174 the contrary, individuals may use an explicit representation of uncertainty (expressed through  
175 confidence) to reverse their initial decision when the same (or a similar) choice is presented again.  
176  
177 We next harnessed individual differences in metacognition to provide a more stringent test of this  
178 hypothesis. We reasoned that the impact of confidence on changes of mind would be more  
179 prominent in participants who have enhanced metacognitive skills, i.e. those whose explicit  
180 confidence ratings more accurately track the level of uncertainty underlying their decision process.  
181 In order to test this hypothesis we calculated an individual index of metacognitive sensitivity by  
182 computing the difference in slope between psychometric functions fitted to high and low confidence  
183 trials<sup>1, 31, 32</sup>. We then ran a logistic regression to predict changes of mind at time  $t_{future}$  using  
184 confidence measured at time  $t$ . In line with our initial hypothesis, we were able to show that the  
185 impact of confidence on changes of mind (here the negative coefficient of confidence predicting  
186 change of mind) is stronger in those subjects with greater metacognitive accuracy ( $r = -0.35$ ,  
187  $p = 0.01$ ) (Figure 3 C).

## 188 **Figure 3**

### 189 **Link between confidence and choice transitivity**

190 In the analyses presented above we established a link between an explicit representation of  
191 confidence and future changes of mind. However, these analyses are agnostic to the quality of the  
192 decisions that emerge as a consequence of changes of mind. Not all choices are born equal; some  
193 are more consistent than others, which is formally captured by the notion of transitivity. A transitive



194 ranking is characterized by the following structure: if an option A is preferred over option B and  
195 option B is preferred over option C, then it follows that A should be preferred over C (i.e.  $A > B$  and  
196  $B > C$  then  $A > C$ ). Transitivity is a normative prescription in utility theory<sup>33</sup>; however, failures of  
197 transitivity are commonly observed in human choices and represent a prominent violation of  
198 economic rationality and, more generally, of logical consistency<sup>34, 35</sup>. In order to test the relation  
199 between confidence and transitivity we found the (idiosyncratic) preference ranking of items that  
200 led to the lowest number of transitivity violations for each subject. Finding an optimal ranking of  
201 choice sets with more than a handful of items is extremely complex; however, a number of efficient  
202 algorithms that approximate a numerical solution have been developed for pairwise comparisons.  
203 In our study we used the Minimum Violations Ranking (MVR) algorithm<sup>36</sup> that minimizes the  
204 number of inconsistencies in the ranking of items conditional on each participant's choices. This  
205 method is conceptually similar to other methods based on revealed preferences such as Afriat's  
206 efficiency index<sup>37, 38</sup>. The MVR algorithm provides an optimal ranking of items for each participant  
207 so that we could tag choices violating this ranking, hereafter labelled transitivity violations (TV).  
208 Because most of these methods are not suited for ternary choice the analyses presented in this  
209 section were performed only on data collected for the experiment using binary choice (Experiment  
210 1). An alternative way to assess choice quality is to compute the choice ranking using BDM and  
211 test whether participants chose the item with the highest ranking. This method gives qualitatively  
212 similar results to those reported below (see S8).

213

214 After ordering the participants' choices according to the MVR algorithm, 4.5% of all decisions were  
215 classified as transitivity violations. We then split the dataset into trials in which participants reported  
216 high confidence and trials in which they reported low confidence (median split). A dramatic  
217 reduction in transitivity violations was observed in high confidence trials (16% of transitivity  
218 violations) in comparison to low confidence trials (84% of transitivity violations). (Figure 4 A) While  
219 these results are consistent with previous evidence provided in this paper and elsewhere<sup>1</sup>, note

220 that we did not rely on BDM value estimates (collected post-choice), instead relying only on  
221 subjects' choices to generate the optimal ranking. In other words, the link between confidence and  
222 the quality of a value-based decision is robust to the method used to elicit preference. In order to  
223 statistically quantify the relation between confidence and transitivity violations on a trial-by-trial  
224 basis (while accounting for other factors that may result in violations of transitivity) we constructed  
225 a set of hierarchical logistic regression models. We found that absolute difference in value (IDVI)  
226 was a robust negative predictor of TV ( $z=-6.59$ ,  $p<0.0001$ ; Figure 4 B) such that participants were  
227 more likely to violate transitivity when items were closer in value. Critically, this same model  
228 showed that even when IDVI was accounted for, confidence was a negative predictor of transitivity  
229 violations ( $z=-6.75$ ,  $p<0.0001$ ). In other words, participants were less confident during those trials in  
230 which they went against their best-fitting preference order. Finally, both response time ( $z=2.55$ ,  
231  $p=.01$ ) and summed value ( $z=2.55$ ,  $p=.01$ ) positively predicted transitivity violations, such that trials  
232 in which the value of both options was higher and/or in trials in which their responses were slower,  
233 participants' choices were more likely to result in transitivity violations. Similar to the change of  
234 mind analysis, eye tracking variables did not reliably predict transitivity violations (GSF=-1.74,  
235  $p=.08$ ; IDDTI  $z=-0.47$ ,  $p=.64$ ) (Figure 5B). Note that this was still true when reported confidence  
236 was excluded from the regression analysis (see supplementary materials, section S7b).

237

238 Finally, we examined whether intersubject variability in metacognitive ability affected transitivity  
239 violations. We reasoned that if a well-calibrated, explicit representation of uncertainty plays a role  
240 in guiding future decisions, participants with greater metacognitive ability would show a decrease in  
241 the number of transitivity violations when the same option was presented a second time. In line  
242 with this hypothesis we observed that greater metacognitive ability was associated with a marked  
243 reduction in transitivity violations between the first and second presentation of the same choice  
244 ( $\beta=0.85$ ,  $SE=0.42$ ,  $z(26)=2.03$ ,  $p<.05$ ; Figure 4 C). We also confirmed that this effect was not  
245 due to a relationship between metacognition and choice instability: the total number of transitivity

246 violations was unrelated to metacognitive accuracy ( $\beta=-1.83$ ,  $SE=1.61$ ,  $z(26)=-1.14$ ,  $p=0.25$ ).  
247 Together these analyses show that a more accurate explicit representation of confidence is  
248 associated with more optimal choices when participants are given the opportunity to change their  
249 minds.

250

## 251 **Figure 4**

252

## 253 **Discussion**

254 What is the advantage of explicitly representing one's confidence in value-based decision-making?

255 Most experimental setups elicit confidence after a decision has been made and cannot be  
256 changed. Our hypothesis was that an explicit representation of confidence might serve an  
257 important role in decision-making by signalling the need to explore different alternatives when the  
258 same (or a similar) choice is presented again.

259

260 Value-based decisions are often perceptually unambiguous (i.e. a banana is noticeably different  
261 from an apple) and most of the uncertainty is contingent on a number of internal processes such as  
262 memories or homeostatic states that are often difficult to manipulate experimentally. For example,  
263 a choice between two food items might be affected both by uncertainty about the tastes of the  
264 items and by uncertainty about one's own level of hunger. In order to take advantage of this  
265 information, a decision-maker should be able to correctly monitor uncertainty that arises from the  
266 different constitutive computations. A wealth of work has shown that humans can introspect on  
267 their choice process and report their level of confidence, an ability that has been associated with  
268 the psychological concept of metacognition. However, the functions of these explicit  
269 representations of confidence (as opposed to implicit markers of uncertainty such as decision time)  
270 have remained unclear. Furthermore, individuals show wide variations in how accurately they can  
271 track and report fluctuations in uncertainty (i.e. metacognitive accuracy).

272

273 In two independent experiments we showed that confidence reports (elicited after a value-based  
274 decision) reliably predicted a change of mind when the same choice was presented again. This  
275 effect is robust after controlling for other factors associated with the difficulty of a decision, such as  
276 difference in value and reaction time. Furthermore, intersubject variability in metacognitive  
277 accuracy modulated the degree to which confidence predicted change of mind: confidence was a  
278 stronger predictor of change of mind in participants with better metacognitive abilities. Critically,  
279 and in contrast to our findings on explicit confidence reports, a lower-level marker of uncertainty  
280 (GSF) did not predict subsequent changes of mind, suggesting that an explicit representation of  
281 uncertainty expressed through confidence is important for guiding future choices. Instead, we  
282 suggest that gaze shift frequency can be considered an ingredient that agents use to construct a  
283 subjective sense of certainty, together with decision time and strength of evidence (cf. 3). An  
284 alternative interpretation of our results is that gaze shift frequency does not contribute directly to  
285 subjective confidence but reflects an agent's attempt to gather more information to adaptively  
286 reduce uncertainty (a situation in which confidence would be low and reaction time slow). Future  
287 work is required to distinguish between these two hypotheses. A further methodological appeal of  
288 GSF as a trial-by-trial measure of uncertainty is that it can be easily gathered in animals. Recent  
289 years have seen a resurgence of interest in studying uncertainty and confidence using animal  
290 models<sup>39</sup>. This promising line of work relies heavily on the development of experimental paradigms  
291 (such as opt-out or post decision wagering) to measure the fluctuation in uncertainty during a  
292 decision process. GSF (which can be measured in rodents by tracking head movements) may  
293 prove a useful tool to monitor, on a trial-by-trial basis, internal fluctuations in uncertainty and its  
294 relation to the neural encoding of decision time and strength of evidence.

295

296 Tracking the level of decision uncertainty is helpful in guiding behaviour in a number of contexts;  
297 for example, in guiding learning <sup>40</sup>, in deciding whether to explore a new alternative or stick with the

298 current one<sup>41, 42</sup>, or in evaluating an alternative course of action<sup>18</sup>. At the neural level, the  
299 rostralateral prefrontal (RLPFC) cortex and frontopolar cortex have been shown to play key roles in  
300 tracking trial-by-trial evolution of uncertainty<sup>43, 44, 45</sup> and modulating uncertainty-driven behaviours<sup>18,</sup>  
301 <sup>41, 42, 46, 47, 48</sup>. At the same time, the RLPFC and frontal pole have also been shown (using a number  
302 of different methods) to play a key role in enabling metacognitive abilities<sup>1, 10, 11, 14, 32</sup>. It is therefore  
303 possible that these two processes are linked anatomically and computationally: individuals whose  
304 prefrontal cortex more closely tracks the trial-by-trial evolution of uncertainty might also have more  
305 accurate explicit representations of confidence. In turn, superior metacognitive abilities might  
306 confer the advantage of knowing how uncertain one's choice was and therefore guide future  
307 behavioural strategies, such as uncertainty-driven exploration<sup>42</sup> or changes of mind. Since we did  
308 not collect neural measures in this study we cannot test this hypothesis directly, but our findings  
309 provide a foundation for future studies of the neurobiology of changes of mind.

310

311 Another question we sought to address was: are changes of mind associated with more optimal  
312 decisions? In value-based decisions, the difference between correct decisions is often murky  
313 since value is a subjective construct. However, when people make a series of value-based choices  
314 across a set of options, their pattern of decisions is characterized by a variable degree of internal  
315 consistency. In experiment 1 we used a recently developed algorithm to find an optimal ranking of  
316 items that produced the lowest number of transitivity violations for each individual. In this way we  
317 identified when participants' decisions were inconsistent with their overall (idiosyncratic) pattern of  
318 decisions. Violations of transitivity are a paradigmatic example of irrationality in economic choice  
319 since they are easy to exploit. For example, when individual preferences are not transitive, it is  
320 possible to construct a choice set in which each decision appears fair on its own, but, when  
321 combined together, guarantees a loss (a phenomenon known as a Dutch book or arbitrage in  
322 finance)<sup>49</sup>. Here we showed that choices made with high confidence are overall more transitive and  
323 therefore more optimal according to the normative prescriptions of utility theory. Noticeably, this

324 effect is robust after controlling for the absolute difference in value and reaction time. This finding  
325 suggests that individuals can monitor and report that a given decision was noisier and therefore  
326 more likely to result in a decision inconsistent with their overall preference patterns, establishing  
327 confidence as a correlate of choice accuracy without relying on the BDM procedure to derive  
328 independent estimates of subjective utility. This result also resonates with the well-established  
329 finding in perceptual decision making that people are able to detect and signal errors as soon as  
330 they respond<sup>16, 50</sup> and with the proposal that confidence can facilitate cognitive control<sup>51</sup>. Here, we  
331 suggest that a similar process might operate in value-based decisions, in which errors can be  
332 thought of as choices that are at odds with one's overall preferences. Consistent with this proposal,  
333 we found that individuals who have a more accurate representation of confidence (greater  
334 metacognitive ability) were more likely to move towards a more internally consistent decision-  
335 making pattern over time. Our work sheds light on the reasons for an explicit representation of  
336 confidence in human decision-making. It explores value-based choices (aka economic choices) by  
337 borrowing methods and concepts from perceptual decision-making<sup>52</sup>. Similar to perceptual  
338 decision-making, we found that the same 'strength of evidence' in value (i.e. IDVI) is accompanied  
339 by a variable level of uncertainty that is represented explicitly as confidence. We suggest these  
340 representations play a functional role not only in allowing confidence to be shared with others, but  
341 also in guiding our own future choices. Taken together our results show that an explicit and  
342 accurate representation of confidence can have a positive impact on the quality of future value-  
343 based decisions.

344

## 345 **Methods**

346

### 347 **Experimental Procedures**

348 **Experiment 1:** Participants were required to make binary choices between 16 common  
349 snack items. Participants were asked to choose between each combination of the items (N  
350 = 120) twice, counterbalanced across the left-right spatial configurations (total number of

351 choices = 240). After each choice, participants indicated their confidence in their decision  
352 on a continuous rating scale. Neither choices nor confidence ratings were time constrained.  
353 Trial order was randomized with the only constraint being that the same pair was never  
354 repeated in subsequent trials. Participants' eye movements were recorded throughout this  
355 task.

356

357 At the end of the experiment, one choice from this phase was played out and the subject  
358 had the opportunity to buy the chosen item by means of an auction administered according  
359 to the Becker-DeGroot-Marschak (BDM) procedure: The experimenter randomly extracted  
360 a price from a uniform distribution (£0 to £3)—the 'market price' of that item. If the  
361 participant's bidding price (willingness-to-pay) was below the market price, no transaction  
362 occurred. The computer-generated value was drawn to a precision greater than 2 decimals  
363 to avoid the possibility of a tie but was rounded to pennies in the event of a transaction. If  
364 the subject's bidding price was above the market price, the participant bought the snack  
365 item at the market price<sup>17</sup>. At the end of the experiment, participants had to remain in the  
366 lab for an additional hour. During this hour, the only food they were allowed to eat was the  
367 item purchased in the auction, if any. At the end of the waiting period participants were  
368 debriefed and thanked for their participation. Participants were paid £25 for their time,  
369 deducting the cost of the food item, if they bought any. Both tasks were programmed using  
370 MATLAB 8.0 (MathWorks) running the Psychophysics toolbox (<http://psycho toolbox.org>) as  
371 well as the Eyelink toolbox extensions<sup>53, 54</sup>. The procedure of this experiment was approved  
372 by the UCL Research Ethics Committee (Project ID: 3736/004).

373

374 **Experiment 2:** Participants gave their willingness to pay for 72 common snack food items  
375 on a scale ranging from £0-£3, in a BDM procedure<sup>17</sup> similar to the one in experiment 1.  
376 Next they completed a choice task where, in each trial, they had to pick their favourite item

377 out of three options. The triplets presented in the choice task were tailored for each  
378 participant from their willingness-to-pay ratings. The items were divided into high-value and  
379 low-value sets by a median split. The 36 high-value items were randomly combined into 12  
380 high-value triplets; this procedure was mirrored to generate 12 low-value triplets. The high-  
381 value and low-value items were then mixed to generate medium value triplets, with 12  
382 triplets consisting of two high-value items and one low-value item, and 12 triplets with the  
383 reverse ratio. This resulted in 48 unique triplets, with counterbalanced spatial configurations  
384 (total trials =144), split into three blocks. Each triplet was shown once in each block, the  
385 presentation order inside blocks was randomized with the constraint that the triplet that  
386 ended one block was never shown first in the next block.

387

388 In the subsequent choice task, the triplets were presented inside 3 squares in an  
389 equidistant 2x2 grid (one randomly-determined position on the grid was left empty). We  
390 used a gaze-contingent paradigm in which the items were only visible when the participant  
391 fixated inside one of the squares, so that the participant could only see one item at a time.  
392 They had unlimited time to make up their mind and could make as many fixations as they  
393 wished. After each choice, participants indicated their confidence in their decision on a  
394 visual analogue rating scale without any time constraints. Participants' eye movements  
395 were recorded throughout the choice task. Both the choice task and the willingness to pay  
396 procedure were programmed in Experiment Builder version 1.10.1640, SR-Research,  
397 Ontario.

398

399 Following the choice task, an auction based on the BDM-ratings took place (see experiment  
400 1). After the auction, participants had to remain in the lab for an additional hour as in  
401 experiment 1. At the end of the waiting period participants were debriefed and thanked for  
402 their participation. Participants were paid £15 for their time, deducting the cost of the food



403 item, if they bought any. The procedure of this experiment was approved by the University  
404 of Cambridge Psychology Research Ethics Committee (Application number: Pre2014.113).

405

#### 406 **Exclusion Criteria**

407 Because the aim of the experiment was to explore the relationship between confidence  
408 and value, it was essential that we had enough measurement sensitivity in both the  
409 confidence scale and in the value scale (the BDM ratings), and that participants' choices  
410 reflected their stated preferences. We therefore excluded participants if any of the  
411 following criteria were met:

- 412 1. Participants used less than 25% of the BDM Scale.
- 413 2. Participants gave exactly the same BDM rating for more than 25% of the items.
- 414 3. Participants used less than 25% of the confidence scale.
- 415 4. Participants gave exactly the same confidence rating for more than 25% of their choices.
- 416 5. Participant choices did not correspond to their BDM ratings (When predicting choices from  
417 differences in value, the DV coefficient deviated more than 2 SD from the experimentwise  
418 mean).

419

#### 420 **Participants**

421 **Experiment 1:** 30 participants took part in the study. One participant did not complete the  
422 task and one participant was excluded because the BDM estimates were poor predictors of  
423 his choice (failed criterion 5). Thus 28 participants were included in the analysis (13  
424 females, age: 19-73). All participants were required to fast for four hours prior to taking part  
425 in the experiment. Blood glucose levels were taken to test their adherence to this criterion  
426 (mean glucose level = 83.57mg/dl, sd = 10.90mg/dl; by comparison, the mean fasting blood  
427 glucose levels for adults is 86.4mg/dl<sup>55</sup>). All participants gave informed consent prior to  
428 participating in this experiment.

429

430 **Experiment 2:** 30 participants completed the study. Of these 30, three were excluded due  
431 to a limited range in their BDM ratings (failed criterion 2). An additional three participants  
432 were excluded for a limited range in their use of the confidence scale (failed criterion 4). 24  
433 participants were included in the main analyses (17 females, age: 21-38). All participants  
434 were required to fast for four hours prior to doing the experiment. All participants gave  
435 informed consent prior to participating in this experiment.

436

437 Sample size was determined a-priori. A power estimation was based on previously  
438 published work that used a similar experimental setup<sup>26</sup>. We implemented a fixed sample  
439 stopping rule set a-priori (N=30). Statistical inferences were conducted only after all data  
440 were collected. If a participant did not fulfil one of the exclusion criteria (decided before data  
441 collection) would have been excluded from the analysis without replacement.

442

#### 443 **Eye Trackers**

444 For experiment 1, eye gaze was sampled at 250 Hz with a head-mounted SR Research  
445 EyeLink II eye-tracker (SR-Research, Ontario). For experiment 2, eye movements were  
446 recorded at 1000Hz with an EyeLink 1000 Plus eye-tracker (SR-Research, Ontario).

447

#### 448 **Preparation of the Eye-tracking Data**

449 **Experiment 1:** Areas of Interest (AI) were defined by splitting the screen in half, creating  
450 two equal sized areas. Fixations in the left AI were assumed to be directed towards the left  
451 snack item, and vice versa. We constructed two variables from the eye tracking data: the  
452 difference in dwell time between the two AIs (DDT), and gaze shift frequency (GSF). DDT  
453 was calculated by subtracting the total dwell time on the left side from the total dwell time  
454 on the right side. GSF was calculated as the number of times participants shifted their  
455 gaze from one AI to the other during each trial.

456 **Experiment 2:** AIs were pre-defined by the 3 squares that participants had to fixate to view  
457 the items (given the gaze-contingent design). We derived two variables from the eye  
458 tracking data: the total dwell time in each AI for a given trial, and GSF. Following  
459 experiment 1, GSF measured the number of fixations in one AI immediately followed by a  
460 fixation in another AI. To ensure that participants paid attention, we excluded trials where  
461 participants had not fixated on every option available at least once. 13 trials out of 3457  
462 were excluded from the analysis for this reason.

463

#### 464 **Hierarchical Models**

465 All hierarchical analyses reported in the results section were conducted using the *lme4*  
466 package (version 1.1-7<sup>56</sup>) in R. For the linear models degrees of freedom and p-values  
467 were obtained using the Kenward-Roger approximation, as implemented in the *pbkrtest*  
468 package<sup>57</sup>. For the choice models (Figures 1 C and F) we ran two hierarchical logistic  
469 regressions: In Experiment 1 we predicted the log odds ratio of picking the right-hand  
470 option on a given trial; for Experiment 2 we predicted the log odds ratio of picking the  
471 reference item. The reference item was determined as the first item encountered according  
472 to reading order in Latin languages (i.e. the upper left item for the trials when an item was  
473 presented in that position and the upper right item for the remaining trials). Fixed effect  
474 confidence intervals were estimated by multiplying the standard errors by 1.96<sup>58</sup>. Because  
475 these confidence intervals are estimates that do not take the covariance between  
476 parameters into account<sup>59</sup> they should not be interpreted too strictly, but rather serve to give  
477 the reader a sense of the precision of the fixed effect coefficients. Note that all predictors  
478 reported are z-scored on the participant level, and that all models allow for random slopes  
479 at the participant level. For completeness we report coefficients from the full model, while  
480 noting that this model is not in every case the most parsimonious. For a comprehensive list

481 of models tested and a formal model comparison using BIC scores see supplementary  
482 materials, section S3.

483

484 Note that the regression models for confidence in experiment 1 had issues converging. We  
485 addressed these issues by square root transforming the IDVI predictor. Notably, for the  
486 individual difference analyses investigating change of mind and transitivity we did not  
487 implement hierarchical models, but unpooled (individual-level) models. The rationale behind  
488 this choice was that for both analyses we were interested in studying between-subjects  
489 variation (Figure 3 C and Figure 4 C) that could be potentially affected by the shrinkage of  
490 parameters towards the group mean that is characteristic of hierarchical models<sup>60</sup>.

#### 491 **Data availability & Code availability**

492 The data and the code for the analyses presented in this article can be found at the BDM  
493 Lab GitHub page: <https://github.com/BDMLab>

494 The data can also be found on figshare: <https://dx.doi.org/10.6084/m9.figshare.3756144.v2>

## 495 **References**

- 496 1. De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. Confidence in value-based choice.  
497 *Nature Neuroscience* **16**, 105–110 (2013) <http://doi.org/10.1038/nn.3279>
- 498 2. Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. Neural correlates, computation and  
499 behavioural impact of decision confidence. *Nature* **455**, 227-231 (2008)  
500 [doi:10.1038/nature07200](https://doi.org/10.1038/nature07200)
- 501 3. Kiani, R., Corthell, L., & Shadlen, M. N. Choice Certainty Is Informed by Both Evidence and  
502 Decision Time. *Neuron* **84**, 1329–1342 (2014) <http://doi.org/10.1016/j.neuron.2014.12.015>
- 503 4. Van den Berg, R., Anandalingam, K., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M.  
504 A common mechanism underlies changes of mind about decisions and confidence. *Elife* **5**,  
505 e12192 (2016) <http://dx.doi.org/10.7554/eLife.12192>
- 506 5. Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. Automatic integration of confidence in

- 507 the brain valuation signal. *Nature Neuroscience* **18**, 1159–1167 (2015)
- 508 <http://doi.org/10.1038/nn.4064>
- 509 6. Barron, H. C., Garvert, M. M., & Behrens, T. E. Reassessing VMPFC: full of confidence?. *Nature*
- 510 *neuroscience* **18**, 1064–1066 (2015)
- 511 7. Meyniel, F., Sigman, M., & Mainen, Z. F. Confidence as Bayesian Probability: From Neural
- 512 Origins to Behavior. *Neuron* **88**, 78–92 (2015) <http://doi.org/10.1016/j.neuron.2015.09.039>
- 513 8. Pouget, A., Drugowitsch, J., & Kepecs, A. Confidence and certainty: distinct probabilistic
- 514 quantities for different goals. *Nature Neuroscience* **19**, 366–374 (2016)
- 515 <http://doi.org/10.1038/nn.4240>
- 516 9. Fleming, S. M., Huijgen, J., & Dolan, R. J. Prefrontal Contributions to Metacognition in
- 517 Perceptual Decision Making. *Journal of Neuroscience* **32**, 6117–6125 (2012)
- 518 <http://doi.org/10.1523/JNEUROSCI.6489-11.2012>
- 519 10. Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. Relating Introspective Accuracy
- 520 to Individual Differences in Brain Structure. *Science* **329** 1541–1543 (2010)
- 521 <http://doi.org/10.1126/science.1191883>
- 522 11. Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E., & Lau, H. Theta-burst
- 523 transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual
- 524 awareness. *Cognitive Neuroscience* **1**, 165–175 (2010)
- 525 <http://doi.org/10.1080/17588921003632529>
- 526 12. Bahrami, B., Olsen, K., Bang, D., Roepstorff, A., Rees, G., & Frith, C. What failure in collective
- 527 decision-making tells us about metacognition. *Philosophical Transactions of the Royal Society*
- 528 *B: Biological Sciences* **367**, 1350–1365 (2012) <http://doi.org/10.1098/rstb.2011.0420>
- 529 13. Bang, D., Fusaroli, R., Tylén, K., Olsen, K., Latham, P. E., Lau, J. Y. F., et al. Does interaction
- 530 matter? Testing whether a confidence heuristic can replace interaction in collective decision-
- 531 making. *Consciousness and Cognition* **26**, 13–23 (2014)
- 532 <http://doi.org/10.1016/j.concog.2014.02.002>

- 533 14. Lau, H., & Rosenthal, D. Empirical support for higher-order theories of conscious awareness.  
534 *Trends in cognitive sciences* **15**, 365-373 (2011)
- 535 15. Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. Changes of mind in decision-making.  
536 *Nature* **461**, 263–266 (2009) <http://doi.org/10.1038/nature08275>
- 537 16. Yeung, N., & Summerfield, C. Metacognition in human decision-making: confidence and error  
538 monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 1310–  
539 1321 (2012) <http://doi.org/10.1098/rstb.2011.0416>
- 540 17. Becker, G. M., DeGroot, M. H., & Marschak, J. Measuring utility by a single-response  
541 sequential method. *Behavioral Science* **9**, 226–232 (1964)
- 542 18. Boorman, E. D., Behrens, T. E. J., Woolrich, M. W., & Rushworth, M. F. S. How Green Is the  
543 Grass on the Other Side? Frontopolar Cortex and the Evidence in Favor of Alternative Courses  
544 of Action. *Neuron* **62**, 733–743 (2009) <http://doi.org/10.1016/j.neuron.2009.05.014>
- 545 19. FitzGerald, T. H. B., Seymour, B., & Dolan, R. J. The Role of Human Orbitofrontal Cortex in  
546 Value Comparison for Incommensurable Objects. *Journal of Neuroscience* **29**, 8388–8395  
547 (2009) <http://doi.org/10.1523/JNEUROSCI.0717-09.2009>
- 548 20. Lebreton, M., Jorge, S., Michel, V., Thirion, B., & Pessiglione, M. An Automatic Valuation  
549 System in the Human Brain: Evidence from Functional Neuroimaging. *Neuron* **64**, 431–439  
550 (2009) <http://doi.org/10.1016/j.neuron.2009.09.040>
- 551 21. Levy, I., Lazzaro, S. C., Rutledge, R. B., & Glimcher, P. W. Choice from Non-Choice:  
552 Predicting Consumer Preferences from Blood Oxygenation Level-Dependent Signals Obtained  
553 during Passive Viewing. *Journal of Neuroscience* **31**, 118–125 (2011)  
554 <http://doi.org/10.1523/JNEUROSCI.3214-10.2011>
- 555 22. Plassmann, H., O'Doherty, J. P., & Rangel, A. Appetitive and Aversive Goal Values Are  
556 Encoded in the Medial Orbitofrontal Cortex at the Time of Decision Making. *Journal of*  
557 *Neuroscience* **30**, 10799–10808 (2010) <http://doi.org/10.1523/JNEUROSCI.0788-10.2010>
- 558 23. Krajbich, I., & Rangel, A. Multialternative drift-diffusion model predicts the relationship between

- 559 visual fixations and choice in value-based decisions. *Proceedings of the National Academy of*  
560 *Sciences of the United States of America* **108**, 13852–13857 (2011)  
561 <http://doi.org/10.1073/pnas.1101328108>
- 562 24. Carandini, M., & Heeger, D. J. Normalization as a canonical neural computation. *Nature*  
563 *Reviews Neuroscience* **13**, 51-62 (2011) <http://doi.org/10.1038/nrn3136>
- 564 25. Louie, K., & Khaw, M. W. Normalization is a general neural mechanism for context-dependent  
565 decision making. *Proceedings of the National Academy of Sciences of the United States of*  
566 *America* **110**, 6139-6144 (2013) <http://doi.org/10.1073/pnas.1217854110/-/DCSupplemental>
- 567 26. Soltani, A., De Martino, B., & Camerer, C. A Range-Normalization Model of Context-Dependent  
568 Choice: A New Model and Evidence. *PLoS Computational Biology* **8**, e1002607 (2012)  
569 <http://doi.org/10.1371/journal.pcbi.1002607.t001>
- 570 27. Krajbich, I., Armel, C., & Rangel, A. Visual fixations and the computation and comparison of  
571 value in simple choice. *Nature Neuroscience* **13**, 1292–1298 (2010)
- 572 28. Kiani, R., & Shadlen, M. N. Representation of Confidence Associated with a Decision by  
573 Neurons in the Parietal Cortex. *Science* **324**, 759–764 (2009)  
574 <http://doi.org/10.1126/science.1169405>
- 575 29. Bronfman, Z. Z., Brezis, N., Moran, R., Tsetsos, K., Donner, T., & Usher, M. Decisions reduce  
576 sensitivity to subsequent information. *Proceedings of the Royal Society B: Biological Sciences*  
577 **282**, 20150228 (2015) <http://doi.org/10.1098/rspb.2015.0228>
- 578 30. Moran, R., Teodorescu, A. R., & Usher, M. Post choice information integration as a causal  
579 determinant of confidence: Novel data and a computational account. *Cognitive Psychology* **78**,  
580 99–147 (2015) <http://doi.org/10.1016/j.cogpsych.2015.01.002>
- 581 31. de Gardelle, V., & Mamassian, P. Does Confidence Use a Common Currency Across Two  
582 Visual Tasks? *Psychological Science* **25**, 1286–1288 (2014)  
583 <http://doi.org/10.1177/0956797614528956>
- 584 32. Fleming, S. M., & Lau, H. C. How to measure metacognition. *Frontiers in Human*

- 585 *Neuroscience*, **8**, 1-9 (2014) <http://dx.doi.org/10.3389/fnhum.2014.00443>
- 586 33. Neumann, Von, J., Morgenstern, O., Rubinstein, A., & Kuhn, H. W. Theory of games and  
587 economic behavior. (Princeton Univ Press, 2007)
- 588 34. Camerer, C. F., & Ho, T.-H. Violations of the betweenness axiom and nonlinearity in  
589 probability. *Journal of Risk and Uncertainty* **8**, 167–196 (1994)  
590 <http://doi.org/10.1007/BF01065371>
- 591 35. Loomes, G., Starmer, C., & Sugden, R. Observing violations of transitivity by experimental  
592 methods. *Econometrica: Journal of the Econometric Society* **59**, 425-439 (1991)  
593 <http://doi.org/10.2307/2938263>
- 594 36. Pedings, K. E., Langville, A. N., & Yamamoto, Y. A minimum violations ranking method.  
595 *Optimization and Engineering*, **13**, 349–370 (2011) <http://doi.org/10.1007/s11081-011-9135-5>
- 596 37. Afriat, S. N. Efficiency estimation of production functions. *International Economic Review*, 568-  
597 598 (1972)
- 598 38. Varian, H. R. Goodness-of-fit in optimizing models. *Journal of Econometrics* **46**, 125-140  
599 (1990)
- 600 39. Kepecs, A., & Mainen, Z. F. A computational framework for the study of confidence in humans  
601 and animals. *Philosophical Transactions of the Royal Society of London. Series B, Biological*  
602 *Sciences* **367**, 1322–1337 (2012) <http://doi.org/10.1098/rstb.2012.0037>
- 603 40. Meyniel, F., Schlunegger, D., & Dehaene, S. The Sense of Confidence during Probabilistic  
604 Learning: A Normative Account. *PLoS Computational Biology* **11**, e1004305 (2015)  
605 <http://doi.org/10.1371/journal.pcbi.1004305>
- 606 41. Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. Cortical substrates for  
607 exploratory decisions in humans. *Nature* **441** 876–879 (2006)  
608 <http://doi.org/10.1038/nature04766>
- 609 42. Badre, D., Doll, B. B., Long, N. M., & Frank, M. J. Rostrolateral prefrontal cortex and individual  
610 differences in uncertainty-driven exploration. *Neuron* **73**, 595-607 (2012)



- 611 <http://doi.org/10.1016/j.neuron.2011.12.025>
- 612 43. Koechlin, E., & Hyafil, A. Anterior Prefrontal Function and the Limits of Human Decision-  
613 Making. *Science* **318**, 594–598 (2007) <http://doi.org/10.1126/science.1142995>
- 614 44. Rushworth, M. F. S., & Behrens, T. E. J. Choice, uncertainty and value in prefrontal and  
615 cingulate cortex. *Nature Neuroscience* **11**, 389–397 (2008) <http://doi.org/10.1038/nn2066>
- 616 45. Yoshida, W., & Ishii, S. Resolution of Uncertainty in Prefrontal Cortex. *Neuron* **50**, 781–789  
617 (2006) <http://doi.org/10.1016/j.neuron.2006.05.006>
- 618 46. Donoso, M., Collins, A., & Koechlin, E. Foundations of human reasoning in the prefrontal  
619 cortex. *Science* **344**, 1481–1486 (2014) <http://doi.org/10.1126/science.1252254>
- 620 47. Lee, S. W., Shimojo, S., & O'Doherty, J. P. Neural Computations Underlying Arbitration  
621 between Model-Based and Model-free Learning. *Neuron* **81**, 687–699 (2014)  
622 <http://doi.org/10.1016/j.neuron.2013.11.028>
- 623 48. Payzan-LeNestour, E., Dunne, S., Bossaerts, P., & O'Doherty, J. P. The Neural Representation  
624 of Unexpected Uncertainty during Value-Based Decision Making. *Neuron* **79**, 191–201 (2013)  
625 <http://doi.org/10.1016/j.neuron.2013.04.037>
- 626 49. Hájek, A. Dutch book arguments. *The Oxford Handbook of Rational and Social Choice*, 173–  
627 195. (Oxford University Press, 2008)
- 628 50. Boldt, A., & Yeung, N. Shared Neural Markers of Decision Confidence and Error Detection. *The*  
629 *Journal of Neuroscience* **35**, 3478–3484 (2015) [http://doi.org/10.1523/JNEUROSCI.0797-](http://doi.org/10.1523/JNEUROSCI.0797-14.2015)  
630 [14.2015](http://doi.org/10.1523/JNEUROSCI.0797-14.2015)
- 631 51. Koizumi, A., Maniscalco, B., & Lau, H. Does perceptual confidence facilitate cognitive control?.  
632 *Attention, Perception, & Psychophysics* **77**, 1295–1306 (2015) [http://doi.org/10.3758/s13414-](http://doi.org/10.3758/s13414-015-0843-3)  
633 [015-0843-3](http://doi.org/10.3758/s13414-015-0843-3)
- 634 52. Summerfield, C. Building bridges between perceptual and economic decision-making: neural  
635 and computational mechanisms. *Frontiers in neuroscience* **6**, 70 (2012)  
636 <http://doi.org/10.3389/fnins.2012.00070/abstract>

- 637 53. Brainard, D. H. The psychophysics toolbox. *Spatial Vision* **10**, 433–436 (1997)
- 638 54. Cornelissen, F. W., Peters, E. M., & Palmer, J. The Eyelink Toolbox: eye tracking with  
639 MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, &*  
640 *Computers* **34**, 613–617 (2002)
- 641 55. Rapp, K., et al. Fasting blood glucose and cancer risk in a cohort of more than 140,000 adults  
642 in Austria. *Diabetologia* **49**, 945-952 (2006)
- 643 56. Bates, D., Maechler, M., Bolker, B., & Walker, S. lme4: Linear mixed-effects models using  
644 Eigen and S4. *R Package Version*, **1** (2014)
- 645 57. Halekoh, U., & Højsgaard, S. A kenward-roger approximation and parametric bootstrap  
646 methods for tests in linear mixed models—the R package pbkrtest. *Journal of Statistical*  
647 *Software* **59**, 1–30 (2014)
- 648 58. Gelman, A., & Hill, J. Data analysis using regression and multilevel/hierarchical models.  
649 (Cambridge University Press, 2006)
- 650 59. Bolker, B. How trustworthy are the confidence intervals for lmer objects through the effects  
651 package? [http://stats.stackexchange.com/questions/117641/how-trustworthy-are-the-](http://stats.stackexchange.com/questions/117641/how-trustworthy-are-the-confidence-intervals-for-lmer-objects-through-effects-pa)  
652 [confidence-intervals-for-lmer-objects-through-effects-pa](http://stats.stackexchange.com/questions/117641/how-trustworthy-are-the-confidence-intervals-for-lmer-objects-through-effects-pa) (2014, accessed 10th December  
653 2015)
- 654 60. Gelman, A., & Pardoe, I. Bayesian measures of explained variance and pooling in multilevel  
655 (hierarchical) models. *Technometrics* **48**, 241-251 (2006)

656  
657  
658  
659  
660  
661  
662

663

## 664 Acknowledgements

665 *This work has been supported by the Wellcome Trust and Royal Society (Henry Dale Fellowship to*  
666 *BDM – 102612/Z/13/Z) and the Economics and Social Research Council (PhD Scholarship for TF).*  
667 *The funders had no role in study design, data collection and analysis, decision to publish, or*  
668 *preparation of the manuscript. We would like to specially thank Yoshitsugu Yamamoto for sharing*  
669 *the methods he developed to rank choice in Experiment 1 and Chris Street and Sebastian*  
670 *Bobadilla Suarez for help in collecting the data and pre-processing the eye-tracking raw data used*  
671 *in Experiment 1.*

672

## 673 Author Contribution Statement

674 *BDM, CJ and SF designed the first experiment reported in this paper. The data for the first*  
675 *experiment was collected by CJ. The second experiment was designed by TF and BDM. The data*  
676 *for the second experiment was collected by TF, the data from both experiments was analysed by*  
677 *TF. The article was written by BDM and TF. All the authors revised the manuscript*

678

## 679 Competing Interests Statement

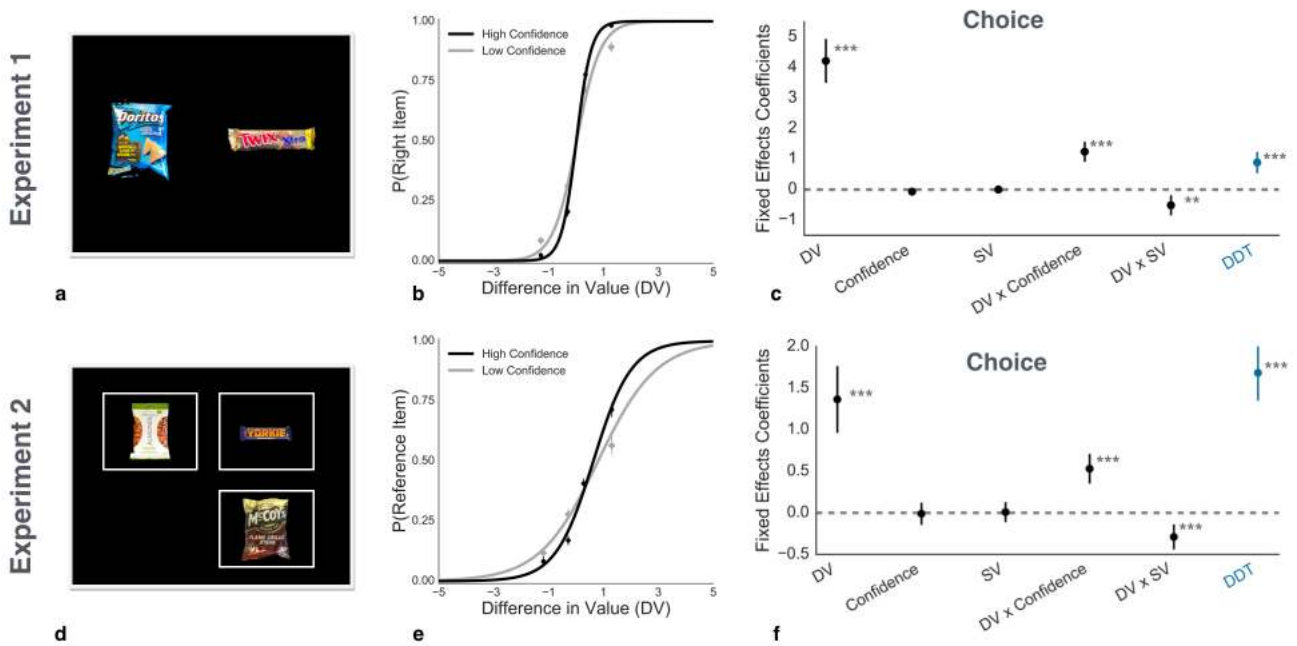
680 *The authors declare no competing interests.*

681

## 682 Figure legends

683

684



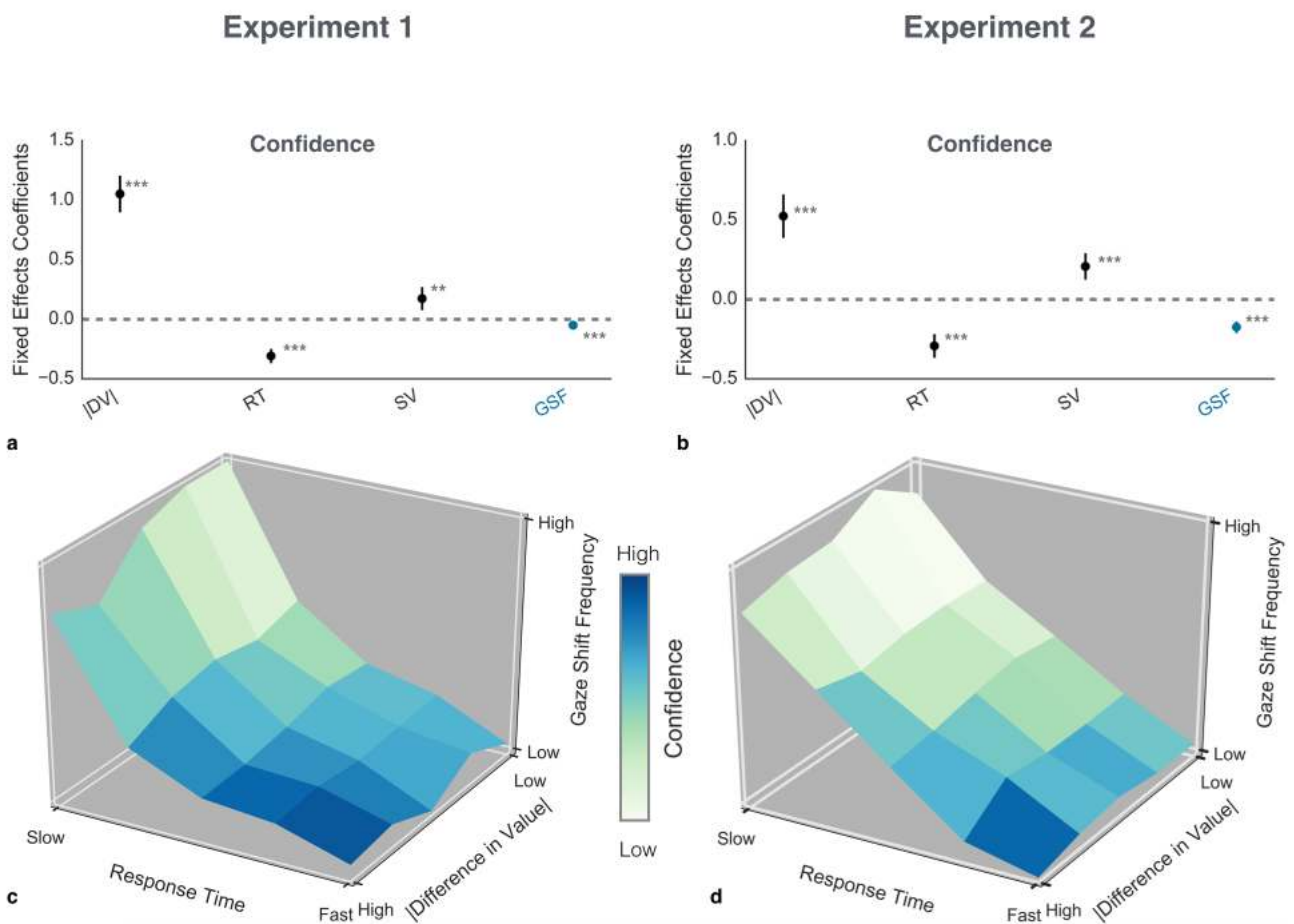
686

687

### 688 **Figure 1: Relation between confidence and choice**

689 *Eye-tracking tasks: (a) In Experiment 1, participants were presented with two snack items and*  
 690 *were then required to choose one item to consume at the end of the experiment. (d) In Experiment*  
 691 *2, participants chose between three options, and the presentation of the stimuli was contingent on*  
 692 *which box participants looked at. In both experiments, participants indicated their confidence that*  
 693 *they had made a correct decision on a visual analogue scale after each choice they made. (b)*  
 694 *Probability of choosing the item on the right as a function of the difference in value between the*  
 695 *options, data from. (e) Probability of choosing the reference item (see methods), as a function of*  
 696 *the value difference between the reference item and the mean value of the alternatives. Black line,*  
 697 *high confidence trials, grey line, low confidence trials (as determined by a median split). Each*  
 698 *graph shows the z-scored data pooled across participants. Points represent quartiles of DV. Error*  
 699 *bars show standard errors. (c and f) Fixed effects coefficients from hierarchical logistic regression*

700 models predicting choice (DV= difference in value; SV= summed value; DDT = difference in dwell  
 701 time, DV x Confidence= Interaction of difference in value and confidence; DV x SV= Interaction of  
 702 difference in value and summed value). The graph for experiment 1 (C) shows the coefficients  
 703 predicting the probability of choosing the right-hand option; the graph for experiment 2 (F) shows  
 704 the coefficients predicting the probability of choosing the reference option (see Methods). Error  
 705 bars show 95% CIs. The sample size for experiment 1 was 28 participant (each completing 240  
 706 trials), the sample size for experiment 2 was 24 participants (each completing 144 trials) \*\*\* =  $p <$   
 707  $.001$ ; \*\* =  $p < .01$ ; \* =  $p < .05$  (two-sided).  
 708

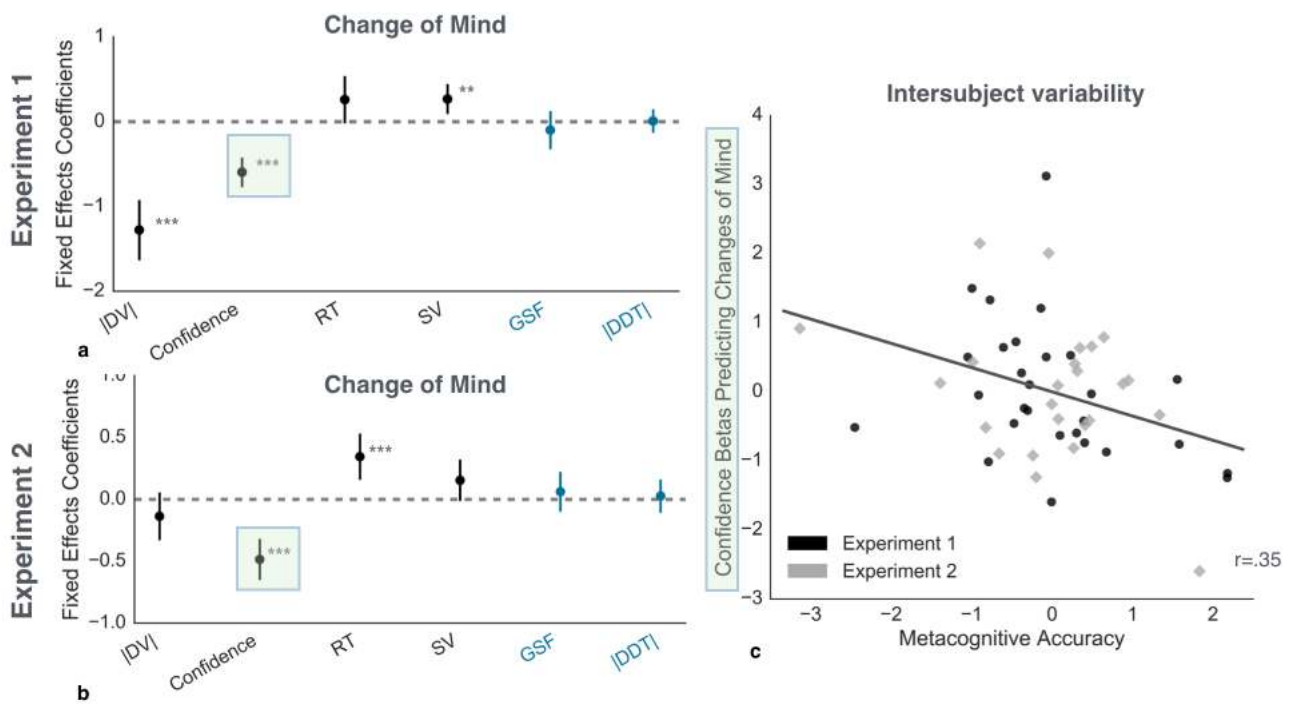


709

710

711 **Figure 2. Factors that contribute to confidence**

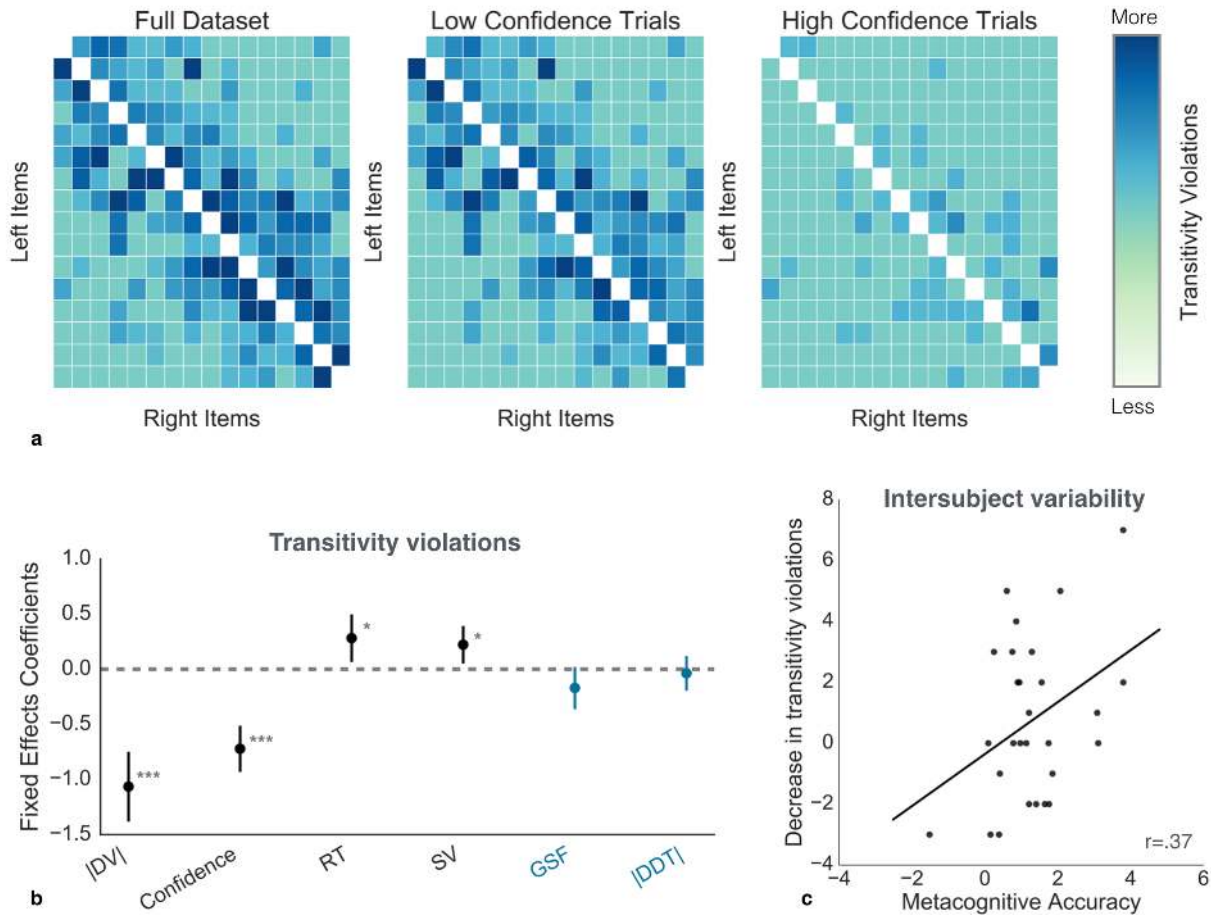
712 **(a-b)** Fixed-effect coefficients in hierarchical regression models predicting confidence for  
 713 experiment 1 and 2, respectively. Error bars show 95% CIs. \*\*\* =  $p < .001$ ; \*\* =  $p < .01$ ; \* =  $p < .05$   
 714 (two-sided). (IDVI= absolute difference in value; RT = reaction time; SV= summed value; GSF =  
 715 Gaze Shift Frequency). **(c-d)** 4-D heat maps showing mean z-scored confidence as a function of  
 716 subject specific quantiles of response time, absolute difference in value and gaze shift frequency.  
 717 The sample size for experiment 1 was 28 participants (each completing 240 trials); the sample size  
 718 for experiment 2 was 24 participants (each completing 144 trials)



719  
 720  
 721  
 722  
 723  
 724

**Figure 3. Confidence predicts change of mind.**

725 **(a-b)** Fixed effects coefficients from hierarchical logistic regression models predicting future  
726 changes of mind. Error bars show 95% CIs. \*\*\* =  $p < .001$ ; \*\* =  $p < .01$ ; \* =  $p < .05$  (two-sided).  
727 (IDVI= absolute difference in value; RT = reaction time; SV= summed value; GSF = gaze shift  
728 frequency; IDDTI= absolute difference in dwell time) **(c)** Correlation between metacognitive  
729 accuracy and the coefficients for confidence ratings predicting future changes of mind (highlighted  
730 in pale green). Participants with greater metacognitive accuracy are more likely to change their  
731 mind following a low-confidence judgment; note that the correlation is negative because the  
732 relationship between confidence and changes of mind is itself negative (lower confidence  
733 increases the probability of subsequent changes of mind). Participants from experiment 1 are  
734 represented by black dots; participants from experiment 2 are represented by grey diamonds. Both  
735 axes (x and y) are z-scored for each experiment separately. The sample size for experiment 1 was  
736 28 participants (each completing 240 trials); the sample size for experiment 2 was 24 participants  
737 (each completing 144 trials)



738

739 **Figure 4. Link between confidence and transitivity**

740 **(a)** Heat maps showing the number transitivity violations for the full sample and for high and low  
 741 confidence trials (median split). The middle diagonal line is empty because no item was ever  
 742 paired with itself. Note most transitivity violations took place on low-confidence trials. **(b)** Fixed  
 743 effects coefficients from a hierarchical logistic regression model predicting transitivity violations  
 744 Error bars show 95% CIs. \*\*\* =  $p < .001$ ; \*\* =  $p < .01$ ; \* =  $p < .05$  (two-sided). (|DVI|= absolute  
 745 difference in value; RT = reaction time; SV= summed value; GSF = gaze shift frequency; |DDTI|=  
 746 absolute difference in dwell time). **(c)** Decreases in transitivity violations between the first and  
 747 second presentation for each participant, as a function of metacognitive accuracy. The graph  
 748 shows that participants who are more metacognitively accurate tend to become more transitive  
 749 over time. The sample size for experiment 1 was 28 participants (each completing 240 trials).

750