

# Coherency Maximizing Exploration in the Supermarket

Peter S. Riefer<sup>1,2\*</sup>, Rosie Prior<sup>2</sup>, Nicholas Blair<sup>2</sup>, Giles Pavey<sup>2</sup> & Bradley C. Love<sup>3,1</sup>

<sup>1</sup> Department of Experimental Psychology, University College London (UCL), 26 Bedford Way, London WC1H 0AP, UK.

<sup>2</sup> dunnhumby Ltd., 184 Shepherd's Bush Road, London W6 7NL, UK.

<sup>3</sup> The Alan Turing Institute, 96 Euston Road, London NW1 2DB, UK.

\* Corresponding author. E-mail: peter@peterriefer.net

In uncertain environments, effective decision makers balance exploiting options that are currently preferred against exploring alternative options that may prove superior<sup>1,2</sup>. For example, a honeybee foraging for nectar must decide whether to continue exploiting the current patch or move to a new location<sup>3-6</sup>. When the relative reward of options changes over time, humans explore in a normatively correct fashion, exploring more often when they are uncertain about the relative value of competing options<sup>7-11</sup>. However, rewards in these laboratory studies were objective (e.g., monetary payoff), whereas many real-world decision environments involve subjective evaluations of reward (e.g., satisfaction with food choice). In such cases, rather than choices following preferences, preferences may follow choices with subjective reward (i.e., value) to maximize coherency between preferences and behaviour<sup>12,13</sup>. If so, increasing coherency would lessen the tendency to explore while uncertainty increases, contrary to previous findings. To evaluate this possibility, we examined the exploratory choices of more than 280,000 anonymized individuals in supermarkets over several years. Consumers' patterns of exploratory choice ran counter to normative models for objective rewards<sup>7-9,14</sup> – the longer the exploitation streak for a product, the less likely were

people to explore an alternative. Furthermore, customers preferred coupons to explore alternative products when they have recently started an exploitation streak. These findings suggest interventions to promote healthy lifestyle choices.

Effective decision making requires balancing exploratory and exploitative behaviour<sup>1,2,15</sup>. For example, finding a restaurant that bests one's current favourite requires some exploration. The timing of exploration is also critical. Normatively, the rate of exploration should increase as uncertainty about the relative goodness of options increases<sup>8</sup>. For example, one may give a restaurant a second chance after a year has passed because the service could have improved in the interim. People in laboratory studies with objective rewards (e.g., money) behave in a manner consistent with the ideal actor<sup>7-9,14</sup>, exploring more often when uncertainty is high. This efficient, systematic exploration appears to demand capacity-limited cognitive resources<sup>9</sup> and rely on frontal dopamine brain circuitry<sup>14,15</sup>. However, as in the restaurant example, rewards can be subjective rather than objective. Although it's clear that higher monetary rewards are better, comparing the reward associated with two dining experiences is more subjective and multidimensional (e.g., atmosphere, service, food quality). In such cases, determining value becomes an interpretive exercise. This interpretive process can be self-reinforcing such that people come to prefer what they happen to choose (or believe they chose)<sup>16-18</sup>. For example, in a jam tasting task, the jam people initially disfavoured was deceptively presented as the favoured option for a re-taste. Not only did people frequently fail to detect the switch, but they also provided rich justifications for their "choice"<sup>12</sup>. In such studies, people altered their preferences to align with their previous behaviour, which can affect future choice<sup>17-20</sup>. Such coherence seeking behaviour is in line with people's preference for information that is consistent with their current views and behaviour<sup>21-23</sup>.

These coherency seeking tendencies in subjective choice have implications for exploratory behaviour. Most choices, like those in a supermarket, involve subjective interpretation of reward. If people alter their preferences to match their choices, then patterns of exploration should be opposite of that found with objective monetary rewards. With objective rewards, the likelihood of exploring increases the longer it has been since exploring (see Fig. 1a). We refer to this manner of exploration as *uncertainty minimizing*, as it responds to the possibility of missing changes in the choice environment while exploiting preferred options. If instead preferences conform to choices, then people should become less likely to explore the more they exploit (see Fig. 1b), which we refer to as *coherency maximizing*. In coherency maximization, the longer people repetitively exploit an option, the more entrenched their preference becomes. Such increased liking for chosen options strengthens coherence between preference and past behaviour, while also promoting coherent future behaviour based on this preference. Unlike common approaches to balancing exploration and exploitation in machine learning<sup>24</sup>, both views predict that exploration is structured and non-random in that the likelihood of exploring varies with recent choice history. Although these two views of exploration differ in their predictions for local timing of exploratory choice, they both predict the global exploration frequency should be stable over longer timescales. For example, under coherency maximization, once one eventually explores and discovers a new choice to exploit, the entrenchment process starts anew. In effect, the exploratory choice reduces the burden of continuing to choose coherently to justify past choices, thereby resetting people's preferences to the level before the entrenchment started. This makes it possible to settle on a new choice once the exploitation streak of a former choice has ended.

Whereas laboratory studies with objective rewards find uncertainty minimizing exploration, we predict that coherency maximizing exploration will dominate with subjective rewards. To

test this hypothesis, we evaluated how people explore with subjective rewards by examining shoppers' behaviour in the supermarket. Tesco, a major UK supermarket chain, provided approximately 283,000 fully anonymized datasets, each representing the purchases of a shopper within a specific product category over a period of 250 weeks, involving 152.2 (SD = 89.9) store visits on average. We examined how individual shoppers explored product options within six different product categories: beers, breads, coffees, toilet papers, washing detergents and yogurts. For example, a shopper may prefer and exploit beer brand A for a number of store visits before exploring brand B. Exploration and exploitation coding was based on repetition – repeated choices (i.e., purchases) were coded as exploitations whereas explorations involved non-repetitive (i.e., switching) choice (see Methods for further details).

On average, people explored with a relative frequency of .404 and this global tendency to explore was stable over time (see Fig. 2a and 2b), mirroring the results in laboratory studies using objective rewards<sup>8,9</sup>. Both uncertainty minimization and coherency maximization (see Fig. 1) anticipate this result while also predicting that local patterns of exploration should be non-random. Indeed, people's patterns of exploratory purchases were non-random, as evidenced by exploitation streaks that were longer ( $M = 8.56$ ,  $SD = 18.33$ ) than expected in 92.8% of cases by a permutation test (see Supplementary Information). This result indicates that people systematically explore when shopping. The key question is whether people's local exploration patterns are more akin to those predicted by uncertainty minimization (Fig. 1a) or coherency maximization (Fig. 1b). As predicted and consistent with coherency maximizing view, shoppers were less likely to explore the longer they had been exploiting a product (see Fig. 2c and 2d). This result is in stark contrast to studies with objective rewards that find uncertainty minimizing exploration.

Model-based analyses, which treat exploitation streak length as a continuous predictor of exploration rate, corroborated the conclusion that people are coherency maximizers. Choices were modelled with logistic regression to predict the probability of exploration given the current exploitation streak length (see Fig. 3a). The results showed that the impact of exploitation streak length on probability to explore was negative for 79.3% of the shopper datasets, implying that people explored less the longer they have been exploiting. A permutation test for all regression slopes revealed that 82.6% were lower than expected (see Fig. 3b). The findings suggest non-random exploration in line with the predictions for subjective outcomes and coherency maximization.

In other domains, exploratory behaviour is viewed as a stable characteristic of individuals and groups. For example, individuals' strategies tend to agree across internal (e.g., memory retrieval) and external (e.g., foraging) search tasks<sup>25</sup> and exploratory behaviour has been found to systematically vary with factors such as impulsivity, genotype, depressive symptoms, and age<sup>7,9,14,26</sup>. Analogously, we consider whether people's pattern of coherency maximizing exploration is consistent at the individual level across different products. Using the model-based estimates, we found that individuals' patterns of coherency maximization were consistent across the product categories considered. For example, for 20.3% more shoppers than expected by chance, either all or none of five product category datasets were associated with strong coherency maximizing behaviour (see Supplementary Information), which is remarkable given the diversity of the product categories considered.

One controversial aspect of the coherency maximizing view is that preferences may follow from choices. We assessed this possibility by examining consumer's choices with coupon offers. First, we analysed how customers reacted toward product coupons, where they

received points on a bonus card or price discounts for buying a promoted product. If people's preferences change with exploitation streak length, they should prefer coupons to exploit or explore products differently at different stages (i.e. lengths) of their exploitation streaks. Based on 69,664 coupon redemptions in our choice datasets (see Fig. 4a), we observed that customers redeemed coupons to explore products more quickly when they were on short exploitation streaks ( $M = 27.0$  days) compared to long ones ( $M = 29.8$  days). Conversely, customers redeemed coupons to exploit more quickly on long exploitation streaks ( $M = 24.4$  days) and slower on short streaks ( $M = 25.7$  days). This strong interaction is predicted by the coherency maximization view.

Second, rather than relying on existing data, we conducted a follow-up coupon study in which we issued coupons for instant coffee to 8,623 randomly selected households who regularly buy instant coffees. A logistic regression model was fit to the group to predict coupon redemption probability based on exploitation streak length. Consistent with the previous coupon analysis and coherency maximizing exploration, the results revealed a significant interaction of coupon type (i.e. whether the coupon meant exploration or exploitation to the customer) and current exploitation streak length,  $|z| = 3.623$ ,  $p < .001$  (see Fig. 4b, details in Supplementary Information). Hence, we find support for the idea that people's choices induced preference changes, as their interest in coupon rewards depended on how well the coupon matched their recent choices (i.e. their exploitation streak length).

The overall pattern of results strongly indicates that shoppers are coherency maximizing explorers, which is striking given that research with objective rewards (e.g., money) finds the opposite, uncertainty minimizing exploration<sup>7-11,14</sup>. One explanation is that subjective rewards involve an evaluative process (e.g., satisfaction with food choice) in which the

148 individual constructs value to justify the choice and maximize coherency<sup>18,27</sup>. Indeed, our  
149 ability to a priori predict who would redeem a coupon relied on people's preferences being  
150 shaped by recent behaviours. Effectively, preferences may follow choices, which might  
151 appear irrational<sup>28</sup> but could be an effective strategy in some environments. For example,  
152 preferring food sources that have been frequently and recently sampled could be an effective  
153 means for avoiding foodborne illnesses. The same approach we utilized, linking big data with  
154 psychological theory, could be leveraged to properly time interventions aimed at improving  
155 diet and exercise regime. Given that we found individuals' patterns of exploration were  
156 consistent across diverse product categories, it may be possible to predict who would benefit  
157 most from such interventions. One basic lesson from our research is that people periodically  
158 enter periods of exploration with a predictable likelihood, creating a window of opportunity  
159 to modify behaviour for better or worse.

- 161     1     Cohen, J. D., McClure, S. M. & Angela, J. Y. Should I stay or should I go? How the  
162         human brain manages the trade-off between exploitation and exploration.  
163         *Philosophical Transactions of the Royal Society of London B: Biological Sciences*  
164         **362**, 933-942 (2007).
- 165     2     Hills, T. T. *et al.* Exploration versus exploitation in space, mind, and society. *Trends*  
166         *in cognitive sciences* **19**, 46-54 (2015).
- 167     3     Hutchinson, J. M., Wilke, A. & Todd, P. M. Patch leaving in humans: can a generalist  
168         adapt its rules to dispersal of items across patches? *Animal Behaviour* **75**, 1331-1349  
169         (2008).
- 170     4     Kramer, D. L. & Weary, D. M. Exploration versus exploitation: a field study of time  
171         allocation to environmental tracking by foraging chipmunks. *Animal Behaviour* **41**,  
172         443-449 (1991).
- 173     5     Krebs, J. R., Kacelnik, A. & Taylor, P. Test of optimal sampling by foraging great  
174         tits. *Nature* **275**, 27-31 (1978).
- 175     6     Mobbs, D. *et al.* Foraging under competition: the neural basis of input-matching in  
176         humans. *The Journal of Neuroscience* **33**, 9866-9872 (2013).
- 177     7     Blanco, N. J., Otto, A. R., Maddox, W. T., Beevers, C. G. & Love, B. C. The  
178         influence of depression symptoms on exploratory decision-making. *Cognition* **129**,  
179         563-568 (2013).
- 180     8     Knox, W. B., Otto, A. R., Stone, P. & Love, B. The Nature of Belief-Directed  
181         Exploratory Choice in Human Decision-Making. *Frontiers in Psychology* **2**, 398  
182         (2012).
- 183     9     Otto, A. R., Knox, W. B., Markman, A. B. & Love, B. C. Physiological and  
184         behavioral signatures of reflective exploratory choice. *Cognitive, Affective, &*  
185         *Behavioral Neuroscience* **14**, 1167-1183 (2014).
- 186     10    Riefer, P. S. & Love, B. C. Unfazed by Both the Bull and Bear: Strategic Exploration  
187         in Dynamic Environments. *Games* **6**, 251-261 (2015).
- 188     11    Speekenbrink, M. & Konstantinidis, E. Uncertainty and exploration in a restless  
189         bandit problem. *Topics in cognitive science* **7**, 351-367 (2015).
- 190     12    Hall, L., Johansson, P., Tärning, B., Sikström, S. & Deutgen, T. Magic at the  
191         marketplace: Choice blindness for the taste of jam and the smell of tea. *Cognition*  
192         **117**, 54-61 (2010).
- 193     13    Johansson, P., Hall, L. & Chater, N. in *Neuroscience of preference and choice:*  
194         *Cognitive and neural mechanisms* (eds Raymond J Dolan & Tali Sharot) 121-138  
195         (Academic Press, 2012).
- 196     14    Blanco, N. J. *et al.* A frontal dopamine system for reflective exploratory behavior.  
197         *Neurobiology of learning and memory* **123**, 84-91 (2015).
- 198     15    Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. Cortical  
199         substrates for exploratory decisions in humans. *Nature* **441**, 876-879 (2006).
- 200     16    Egan, L. C., Bloom, P. & Santos, L. R. Choice-induced preferences in the absence of  
201         choice: Evidence from a blind two choice paradigm with young children and capuchin  
202         monkeys. *Journal of Experimental Social Psychology* **46**, 204-207 (2010).
- 203     17    Sharot, T., Fleming, S. M., Yu, X., Koster, R. & Dolan, R. J. Is choice-induced  
204         preference change long lasting? *Psychological science* **23**, 1123-1129 (2012).
- 205     18    Sharot, T., Velasquez, C. M. & Dolan, R. J. Do decisions shape preference? Evidence  
206         from blind choice. *Psychological science* **21**, 1231-1235 (2010).



207 19 Hall, L., Johansson, P. & Strandberg, T. Lifting the veil of morality: Choice blindness  
208 and attitude reversals on a self-transforming survey. *PloS one* **7**, e45457 (2012).  
209 20 Johansson, P., Hall, L., Sikström, S. & Olsson, A. Failure to detect mismatches  
210 between intention and outcome in a simple decision task. *Science* **310**, 116-119  
211 (2005).  
212 21 Festinger, L. *A theory of cognitive dissonance*. (Stanford University Press, 1957).  
213 22 Frey, D. Recent research on selective exposure to information. *Advances in*  
214 *experimental social psychology* **19**, 41-80 (1986).  
215 23 Jonas, E., Schulz-Hardt, S., Frey, D. & Thelen, N. Confirmation bias in sequential  
216 information search after preliminary decisions: an expansion of dissonance theoretical  
217 research on selective exposure to information. *Journal of personality and social*  
218 *psychology* **80**, 557 (2001).  
219 24 Sutton, R. S. & Barto, A. G. *Reinforcement learning: An introduction*. (MIT press,  
220 1998).  
221 25 Hills, T. T., Todd, P. M. & Goldstone, R. L. Search in external and internal spaces  
222 evidence for generalized cognitive search processes. *Psychological Science* **19**, 802-  
223 808 (2008).  
224 26 Frank, M. J., Doll, B. B., Oas-Terpstra, J. & Moreno, F. Prefrontal and striatal  
225 dopaminergic genes predict individual differences in exploration and exploitation.  
226 *Nature neuroscience* **12**, 1062-1068 (2009).  
227 27 Johansson, P., Hall, L., Tärning, B., Sikström, S. & Chater, N. Choice blindness and  
228 preference change: You will like this paper better if you (believe you) chose to read  
229 it! *Journal of Behavioral Decision Making* **27**, 281-289 (2014).  
230 28 Ariely, D. & Norton, M. I. How actions create—not just reveal—preferences. *Trends in*  
231 *cognitive sciences* **12**, 13-16 (2008).

232    **Acknowledgements**

233    We thank Peter Todd for his helpful comments.

234    This work was supported by the Leverhulme Trust grant RPG-2014-075, NIH (Grant  
235    1P01HD080679), and Wellcome Trust Senior Investigator Award WT106931MA to B.L.

236    The funders had no role in study design, data collection and analysis, decision to publish, or  
237    preparation of the manuscript.

238    **Author Contributions**

239    P.R. was involved in all parts of this project, supported by R.P., N.B. and G.P. regarding the  
240    data analysis and with input from B.L. for the design, analysis and write-up of the article.

241    **Competing Interests statement**

242    The authors declare no competing interests.

## Figures

**Figure 1**



**Figure 1: Predicted exploration patterns for uncertainty minimization and coherency maximization.**

**a,**

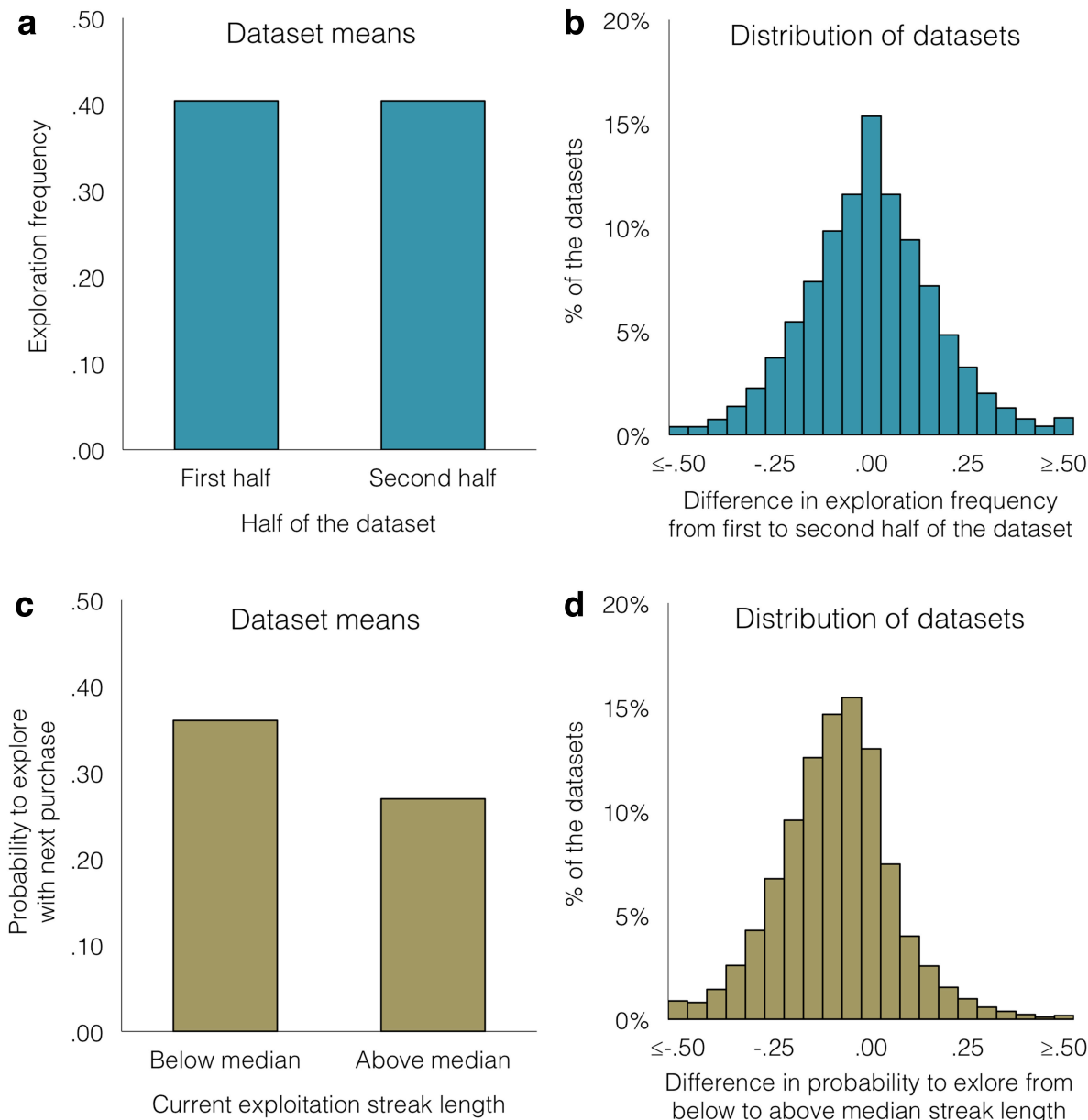
In changing environments, uncertainty minimizing decision makers tend to explore more as the time since the last exploration increases. This normatively correct pattern of non-random exploratory behaviour, where people explore more as uncertainty about the relative goodness of competing options increases, has been found in humans with objective rewards (e.g. monetary values).

**b,**

In contrast, coherency maximizing decision makers tend to explore less as the interval since the last exploration increases. One possibility is that when outcomes require subjective

interpretation (e.g. tasting food), decision makers change their preferences to match their recent choices in order to increase coherency. We predict that this self-reinforcing pattern will hold for supermarket shoppers.

**Figure 2**



**Figure 2: Exploration changes locally but not globally.**

a,

A median split of each shopper's purchases revealed that the overall rate of exploration was stable over time.

**b,**

Likewise, the distribution of differences between the first and second half exploration rates for each shopper showed no systematic variations over time.

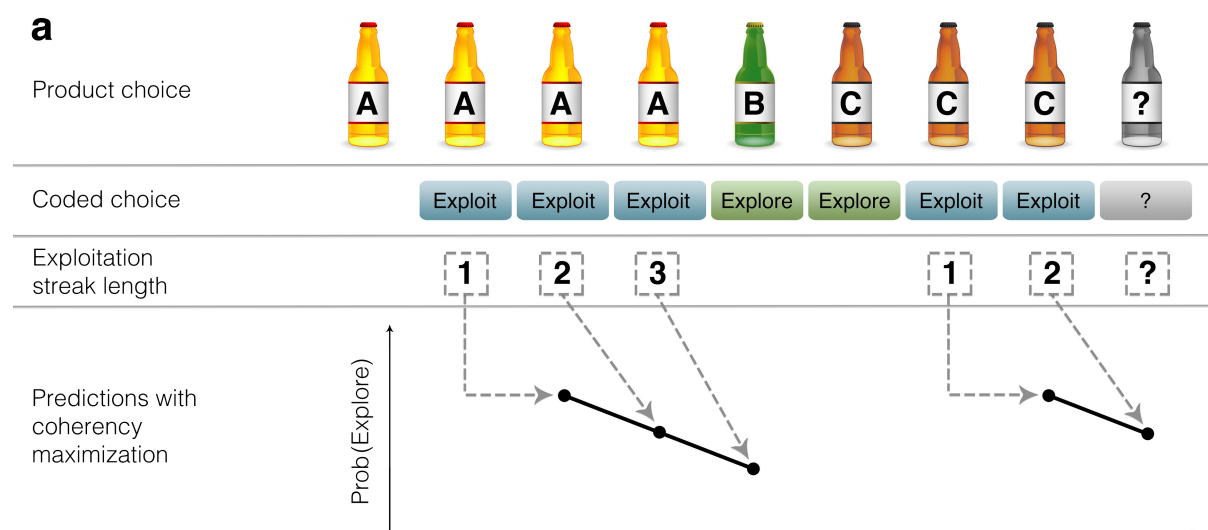
**c,**

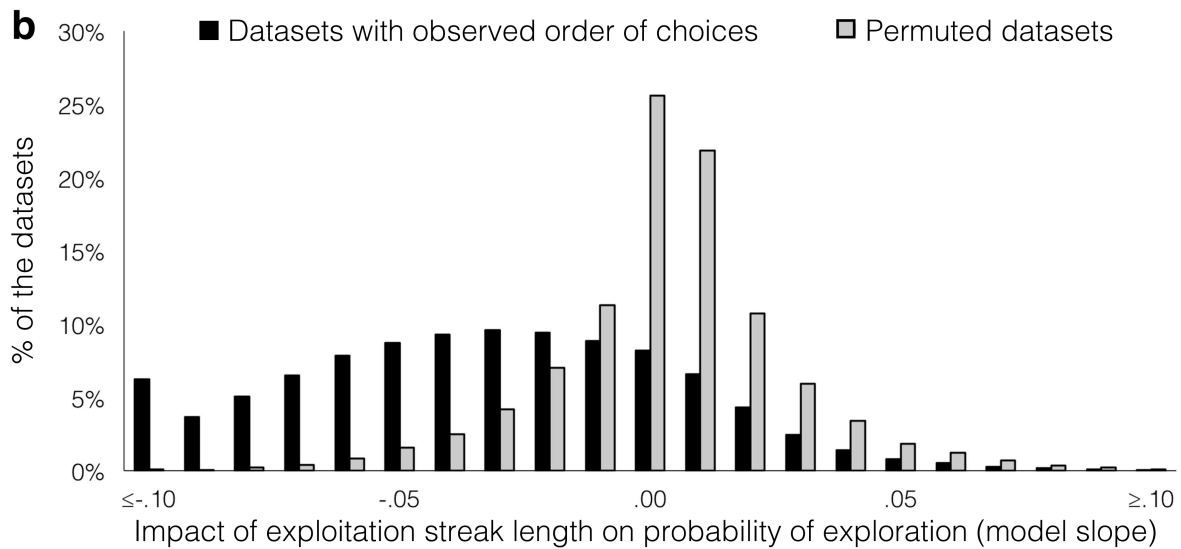
A median split of exploitation streaks by shopper revealed that, in line with the predictions for coherency maximization, people were overall less likely to explore on their next purchase when currently on a long run of exploitative choices.

**d,**

In line with panel c, most individuals showed a decline in probability to explore from exploitation streaks shorter to those longer than their median streak.

**Figure 3**





**Figure 3: Predicting exploration from exploitation streak lengths.**

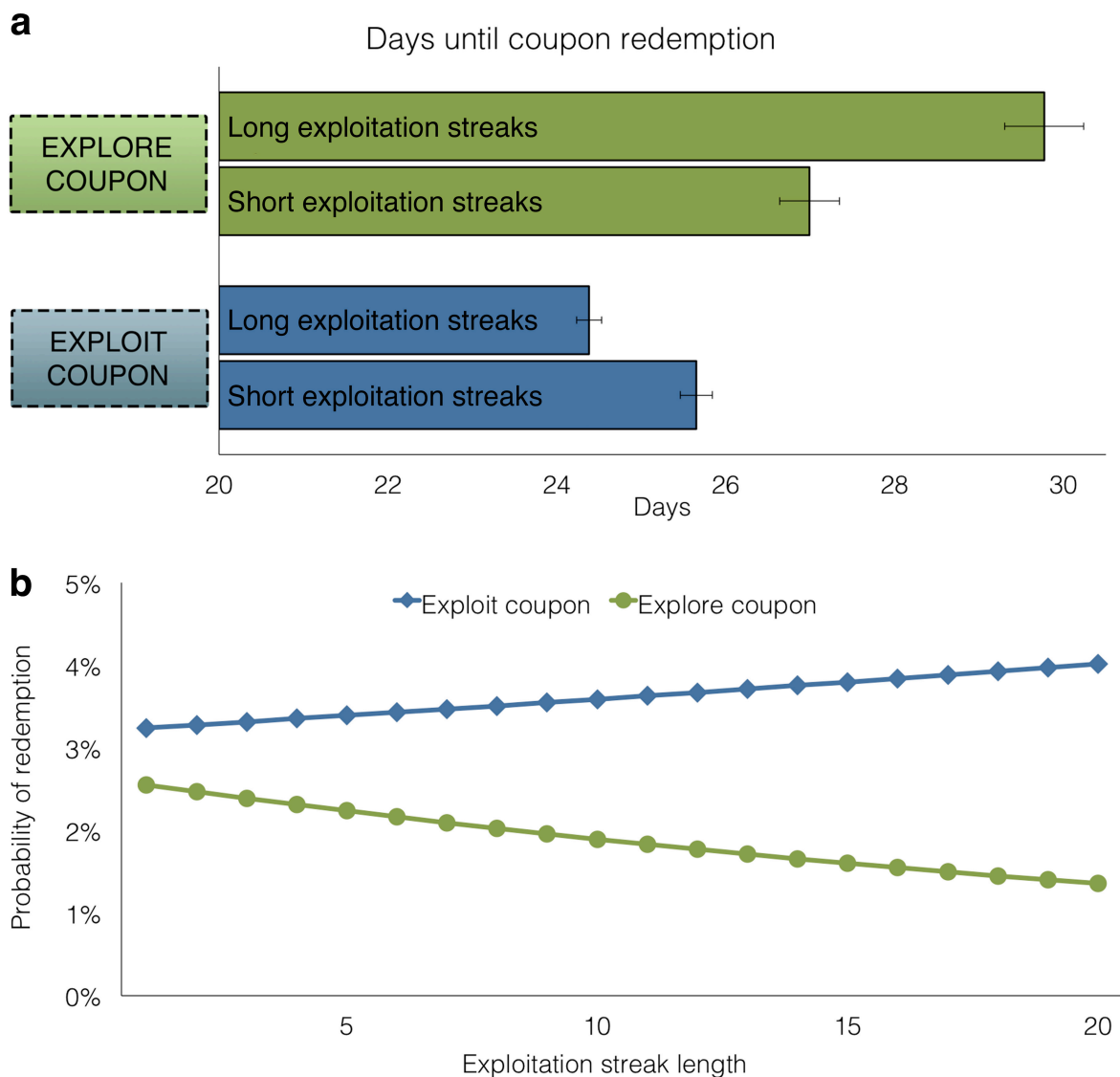
**a,**

As shown, a coherency maximizing shopper is less likely to explore alternatives the longer the exploitation streak, which is characterized by a negative slope in the logistic regression model. In contrast, the slope would be positive under uncertainty minimization and flat for random exploration.

**b,**

Consistent with coherency maximization, the slope of the fitted logistic regression model was negative for the majority of individuals. For comparison, we permuted the order of each individual's purchases and fitted the model (see Supplementary Information). Slopes were more negative in the actual than in the permuted data, providing further support that people are coherency maximizing.

293 **Figure 4**



296 **Figure 4: Coupon redemption depending on current exploitation streak length.**

297 **a,**

298 Consistent with coherency maximization, customers redeemed coupons to exploit products  
 299 more quickly the longer they have been exploiting the product (i.e. on long exploitation  
 300 streaks). Coupons to explore alternatives were used more quickly when customers were only  
 301 beginning to exploit (i.e. on short exploitation streaks). Error bars represent standard errors.

302 **b,**

303 Model fits of the observed relationship between exploitation streak length and the probability  
304 of coupon redemption are shown. A shopper was more likely to redeem a coupon to exploit  
305 the longer the exploitation streak, whereas a coupon to explore was more likely to be  
306 redeemed the more recently a shopper has explored.



## 307    **Method**

308    We analysed 282,972 anonymous datasets containing the chronologically ordered purchases  
309    of individual supermarket customers at Tesco, a UK supermarket chain, regarding one of six  
310    different product categories. Tesco provided access to these datasets in collaboration with  
311    dunnhumby, a customer science company and subsidiary of Tesco (see Data Availability  
312    statement for requests). Customers' product choices were recorded in a database every time  
313    they checked out using a personalized bonus card. This data use was in accord with the card  
314    agreement, which stipulated that anonymized shopping data would be used and shared  
315    outside of Tesco. Individuals in this database can only be identified by an anonymous dataset  
316    number, but not by any personal information. Thus, the analysed datasets only contained  
317    purchase-related information (e.g. quantities, prices, discounts, etc.), but no personal  
318    information about the shopper. Our sample was restricted to people with at least 50 purchases  
319    within a specific product category, which was necessary to model individual behaviour and to  
320    select from loyal customers (i.e., for which we have good coverage of their purchases). We  
321    did not select customers who never explored or who did over 75% of the time. We coded  
322    exploration and exploitation based on repetition, where repeated choices were coded as  
323    exploitations and non-repetitions as explorations (see Fig. 3a).

## 324    **Data Availability**

325    The datasets generated and analysed for the present study are available from dunnhumby, a  
326    customer science company and subsidiary of Tesco, upon request:  
327    [data\\_questions@dunnhumby.com](mailto:data_questions@dunnhumby.com).  
328    Further information about the data and analyses is available online at the Open Science  
329    Framework: [osf.io/e76wy](https://osf.io/e76wy).

330    **Code Availability**

331    Computer code to replicate the findings of the present study can be found online at the Open

332    Science Framework: [osf.io/e76wy](https://osf.io/e76wy).