

Coherency-maximizing exploration in the supermarket

Peter S. Riefer^{1,2*}, Rosie Prior², Nicholas Blair², Giles Pavey² and Bradley C. Love^{1,3}

In uncertain environments, effective decision makers balance exploiting options that are currently preferred against exploring alternative options that may prove superior^{1,2}. For example, a honeybee foraging for nectar must decide whether to continue exploiting the current patch or move to a new location^{3–6}. 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^{7–11}. However, rewards in these laboratory studies were objective (for example, monetary payoff), whereas many real-world decision environments involve subjective evaluations of reward (for example, satisfaction with food choice). In such cases, rather than choices following preferences, preferences may follow choices with subjective reward (that is, value) to maximize coherency between preferences and behaviour^{12,13}. 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^{7–9,14}—the longer the exploitation streak for a product, the less likely people were to explore an alternative. Furthermore, customers preferred coupons to explore alternative products when they had recently started an exploitation streak. These findings suggest interventions to promote healthy lifestyle choices.

Effective decision-making requires balancing exploratory and exploitative behaviour^{1,2,15}. For example, finding a restaurant that is better than 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 increases⁸. 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 (for example, money) behave in a manner consistent with the ideal actor^{7–9,14}, exploring more often when uncertainty is high. This efficient, systematic exploration appears to demand capacity-limited cognitive resources⁹ and rely on frontal dopamine brain circuitry^{14,15}. However, as in the restaurant example, rewards can be subjective rather than objective. Although it is clear that higher monetary rewards are better, comparing the reward associated with two dining experiences is more subjective and multidimensional (for example, 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)^{16–18}. 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'¹². In such studies, people altered their preferences to align with their previous behaviour, which can affect future choice^{17–20}. Such coherence-seeking behaviour is in line with people's preference for information that is consistent with their current views and behaviour^{21–23}.

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 to that found with objective monetary rewards. With objective rewards, the likelihood of exploring increases the longer it has been since exploring (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 (Fig. 1b), which we refer to as coherency-maximizing. In coherency maximization, the longer people repetitively exploit

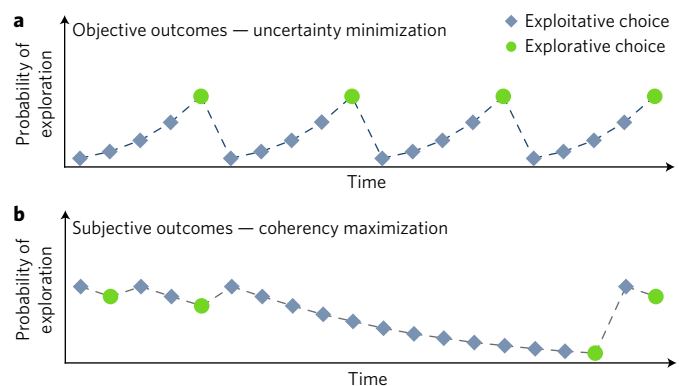


Figure 1 | Predicted exploration patterns for uncertainty minimization and coherency maximization. a, In changing environments, uncertainty-minimizing decision makers tend to explore more as the time since the last exploration increases. This normatively correct pattern of non-random exploratory behaviour, where people explore more as uncertainty about the relative goodness of competing options increases, has been found in humans with objective rewards (for example, 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 (for example, tasting food), decision makers change their preferences to match their recent choices to increase coherency. We predict that this self-reinforcing pattern will hold for supermarket shoppers.

¹Department of Experimental Psychology, University College London (UCL), 26 Bedford Way, London WC1H 0AP, UK. ²dunnhumby Ltd, 184 Shepherd's Bush Road, London W6 7NL, UK. ³The Alan Turing Institute, 96 Euston Road, London NW1 2DB, UK. *e-mail: peter@peterriefer.net

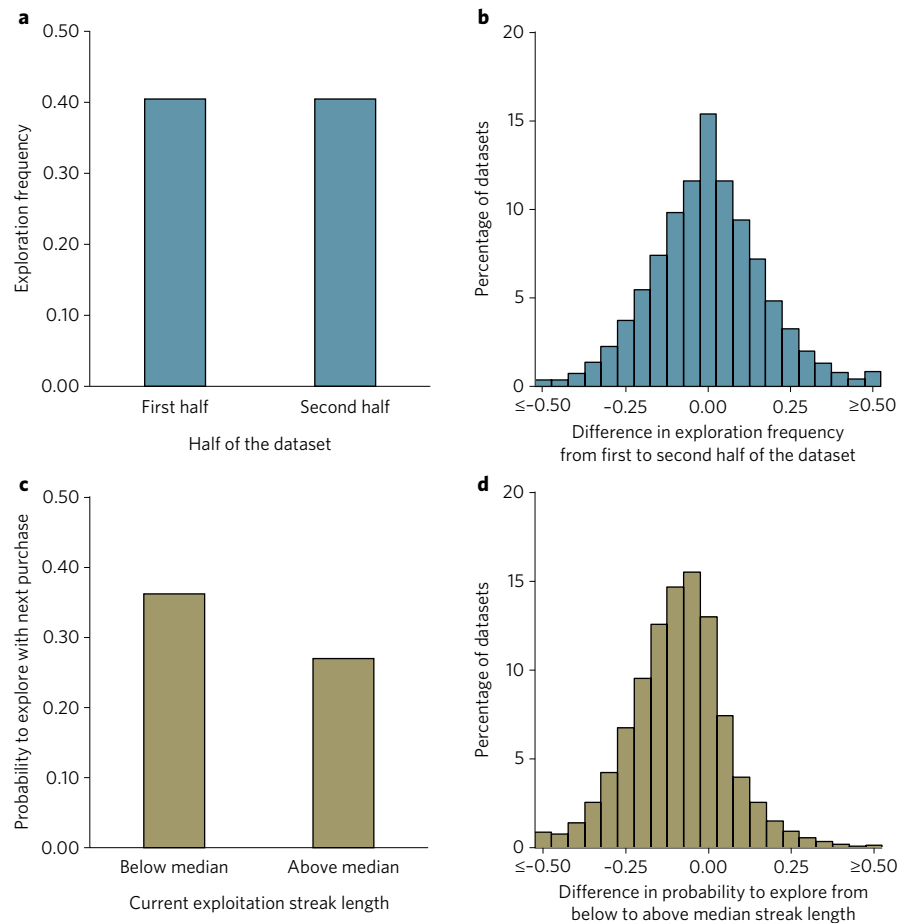


Figure 2 | Exploration changes locally but not globally. **a**, A median split of each shopper's purchases revealed that the overall rate of exploration was stable over time. **b**, Likewise, the distribution of differences between the first- and second-half exploration rates for each shopper showed 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 **c**, most individuals showed a decline in probability to explore from exploitation streaks shorter to those longer than their median streak.

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 learning²⁴, 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 (s.d.=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 (that is, purchases) were coded as exploitations, whereas explorations involved non-repetitive (that is, switching) choice (see Methods for further details).

On average, people explored with a relative frequency of 0.404 and this global tendency to explore was stable over time (Fig. 2a,b), mirroring the results in laboratory studies using objective rewards^{8,9}. Both uncertainty minimization and coherency maximization (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 (mean (M)=8.56 purchases, s.d.=18.33 purchases) than expected in 92.8% of cases by a permutation test (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 the coherency-maximizing view, shoppers were less likely to explore the longer they had been exploiting a product (Fig. 2c,d). This result is in stark contrast to studies with objective rewards that find uncertainty-minimizing exploration.

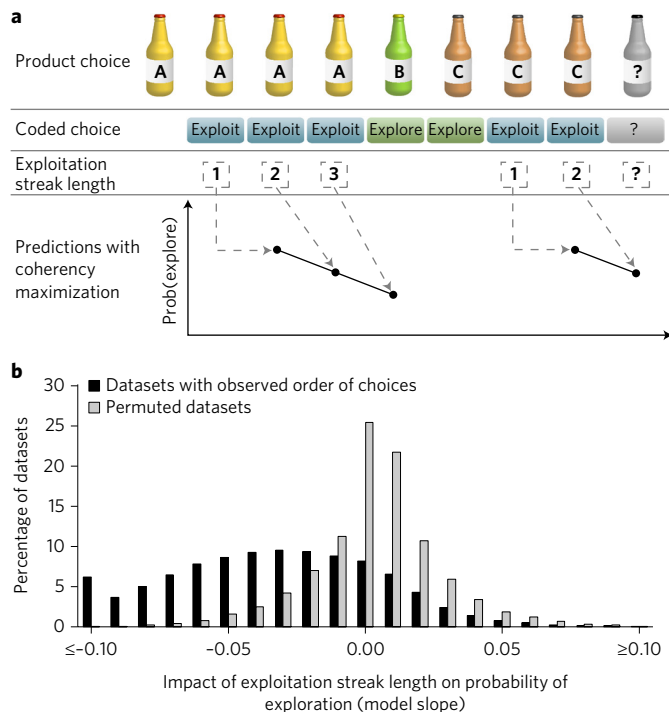


Figure 3 | Predicting exploration from exploitation streak lengths.

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 (Supplementary Information). Slopes were more negative in the actual than in the permuted data, providing further support that people are coherency-maximizing.

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 (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 (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 (for example, memory retrieval) and external (for example, foraging) search tasks²⁵, and exploratory behaviour has been found to systematically vary with factors such as impulsivity, genotype, depressive symptoms and age^{7,9,14,26}. 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 (Supplementary Information), which is remarkable given the diversity of the product categories considered.

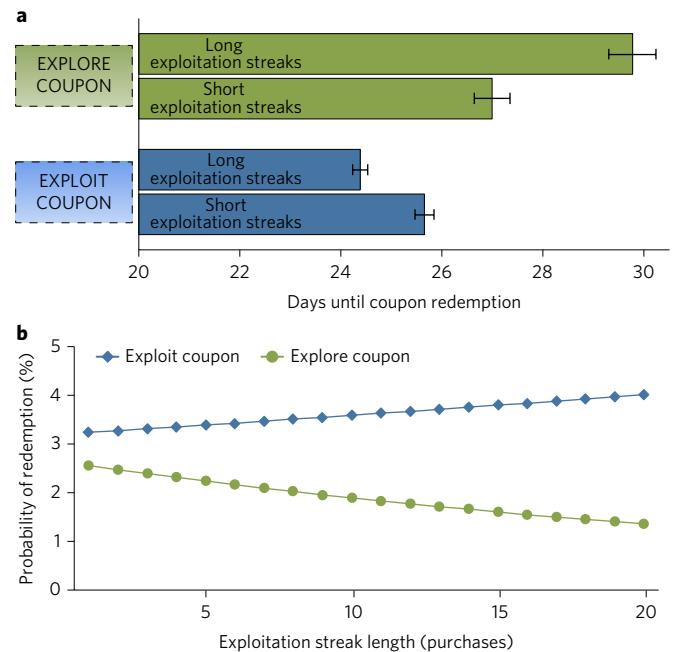


Figure 4 | Coupon redemption depending on current exploitation streak length.

a, Consistent with coherency maximization, customers redeemed coupons to exploit products more quickly the longer they had been exploiting the product (that is, on long exploitation streaks). Coupons to explore alternatives were used more quickly when customers were only beginning to exploit (that is, on short exploitation streaks). Error bars represent standard errors. **b**, Model fits of the observed relationship between exploitation streak length and the probability of coupon redemption are shown. A shopper was more likely to redeem a coupon to exploit the longer the exploitation streak, whereas a coupon to explore was more likely to be redeemed the more recently a shopper had explored.

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 towards 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 (that is, lengths) of their exploitation streaks. Based on 69,664 coupon redemptions in our choice datasets (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 with 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 of the logistic regression revealed a significant interaction term of coupon type (that is, whether the coupon meant exploration or exploitation to the customer) and current exploitation streak length, $|z| = 3.623$, $P < 0.001$ (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 (that is, 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 (for example, money) finds the opposite—uncertainty-minimizing exploration^{7–11,14}. One explanation is that subjective rewards involve an evaluative process (for example, satisfaction with food choice) in which the individual constructs value to justify the choice and maximize coherency^{18,27}. Indeed, our ability to *a priori* predict who would redeem a coupon relied on people's preferences being shaped by recent behaviours. Effectively, preferences may follow choices, which might appear irrational²⁸ but could be an effective strategy in some environments. For example, preferring food sources that have been frequently and recently sampled could be an effective means for avoiding foodborne illnesses. The same approach we utilized, linking big data with psychological theory, could be leveraged to properly time interventions aimed at improving diet and exercise regime. Given that we found individuals' patterns of exploration were consistent across diverse product categories, it may be possible to predict who would benefit most from such interventions. One basic lesson from our research is that people periodically enter periods of exploration with a predictable likelihood, creating a window of opportunity to modify behaviour for better or worse.

Methods

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

Code availability. Computer code to replicate the findings of the present study can be found online at the Open Science Framework: <https://osf.io/e76wy>.

Data availability. The datasets generated and analysed for the present study are available from dunhumby, a customer science company and subsidiary of Tesco, upon request: data_questions@dunhumby.com. Further information about the data and analyses is available online at the Open Science Framework: <https://osf.io/e76wy>.

Received 3 June 2016; accepted 11 November 2016;
published 9 January 2017

References

- Cohen, J. D., McClure, S. M. & Yu, A. J. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Phil. Trans. R. Soc. B* **362**, 933–942 (2007).
- Hills, T. T. *et al.* Exploration versus exploitation in space, mind, and society. *Trends Cogn. Sci.* **19**, 46–54 (2015).
- Hutchinson, J. M., Wilke, A. & Todd, P. M. Patch leaving in humans: can a generalist adapt its rules to dispersal of items across patches? *Anim. Behav.* **75**, 1331–1349 (2008).
- Kramer, D. L. & Weary, D. M. Exploration versus exploitation: a field study of time allocation to environmental tracking by foraging chipmunks. *Anim. Behav.* **41**, 443–449 (1991).
- Krebs, J. R., Kacelnik, A. & Taylor, P. Test of optimal sampling by foraging great tits. *Nature* **275**, 27–31 (1978).
- Mobbs, D. *et al.* Foraging under competition: the neural basis of input-matching in humans. *J. Neurosci.* **33**, 9866–9872 (2013).
- Blanco, N. J., Otto, A. R., Maddox, W. T., Beevers, C. G. & Love, B. C. The influence of depression symptoms on exploratory decision-making. *Cognition* **129**, 563–568 (2013).
- Knox, W. B., Otto, A. R., Stone, P. & Love, B. The nature of belief-directed exploratory choice in human decision-making. *Front. Psychol.* **2**, 398 (2012).
- Otto, A. R., Knox, W. B., Markman, A. B. & Love, B. C. Physiological and behavioral signatures of reflective exploratory choice. *Cogn. Affect. Behav. Neurosci.* **14**, 1167–1183 (2014).
- Riefer, P. S. & Love, B. C. Unfazed by both the bull and bear: strategic exploration in dynamic environments. *Games* **6**, 251–261 (2015).
- Speekenbrink, M. & Konstantinidis, E. Uncertainty and exploration in a restless bandit problem. *Top. Cogn. Sci.* **7**, 351–367 (2015).
- Hall, L., Johansson, P., Tärning, B., Sikström, S. & Deutgen, T. Magic at the marketplace: choice blindness for the taste of jam and the smell of tea. *Cognition* **117**, 54–61 (2010).
- Johansson, P., Hall, L. & Chater, N. In *Neuroscience of Preference and Choice: Cognitive and Neural Mechanisms* (eds Dolan, R. J. & Sharot, T.) 121–138 (Academic Press, 2012).
- Blanco, N. J. *et al.* A frontal dopamine system for reflective exploratory behavior. *Neurobiol. Learn. Mem.* **123**, 84–91 (2015).
- Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876–879 (2006).
- Egan, L. C., Bloom, P. & Santos, L. R. Choice-induced preferences in the absence of choice: evidence from a blind two choice paradigm with young children and capuchin monkeys. *J. Exp. Soc. Psychol.* **46**, 204–207 (2010).
- Sharot, T., Fleming, S. M., Yu, X., Koster, R. & Dolan, R. J. Is choice-induced preference change long lasting? *Psychol. Sci.* **23**, 1123–1129 (2012).
- Sharot, T., Velasquez, C. M. & Dolan, R. J. Do decisions shape preference? Evidence from blind choice. *Psychol. Sci.* **21**, 1231–1235 (2010).
- Hall, L., Johansson, P. & Strandberg, T. Lifting the veil of morality: choice blindness and attitude reversals on a self-transforming survey. *PLoS ONE* **7**, e45457 (2012).
- Johansson, P., Hall, L., Sikström, S. & Olsson, A. Failure to detect mismatches between intention and outcome in a simple decision task. *Science* **310**, 116–119 (2005).
- Festinger, L. A. *Theory of Cognitive Dissonance* (Stanford Univ. Press, 1957).
- Frey, D. Recent research on selective exposure to information. *Adv. Exp. Soc. Psychol.* **19**, 41–80 (1986).
- Jonas, E., Schulz-Hardt, S., Frey, D. & Thelen, N. Confirmation bias in sequential information search after preliminary decisions: an expansion of dissonance theoretical research on selective exposure to information. *J. Pers. Soc. Psychol.* **80**, 557–571 (2001).
- Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, 1998).
- Hills, T. T., Todd, P. M. & Goldstone, R. L. Search in external and internal spaces evidence for generalized cognitive search processes. *Psychol. Sci.* **19**, 802–808 (2008).
- Frank, M. J., Doll, B. B., Oas-Terpstra, J. & Moreno, F. Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nat. Neurosci.* **12**, 1062–1068 (2009).
- Johansson, P., Hall, L., Tärning, B., Sikström, S. & Chater, N. Choice blindness and preference change: you will like this paper better if you (believe you) chose to read it! *J. Behav. Decis. Making* **27**, 281–289 (2014).
- Ariely, D. & Norton, M. I. How actions create—not just reveal—preferences. *Trends Cogn. Sci.* **12**, 13–16 (2008).

Acknowledgements

We thank P. Todd for comments. This work was supported by the Leverhulme Trust grant RPG-2014-075, National Institutes of Health (grant 1P01HD080679) and Wellcome Trust Senior Investigator Award WT106931MA to B.C.L. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. At the time of submission, R.P., N.B. and G.P. were employed by dunhumby Ltd. This work was carried out as part of P.S.R.'s PhD thesis, which was co-sponsored by dunhumby Ltd. and UCL. dunhumby Ltd. did not place any restrictions on the design, data collection and analysis, decision to publish or preparation of the manuscript, beyond the requirement that this work was to be done in compliance with its data policy.

Author contributions

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

Additional information

Supplementary information is available for this paper.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to P.S.R.

How to cite this article: Riefer, P. S., Prior, R., Blair, N., Pavey, G. & Love, B. C. Coherency-maximizing exploration in the supermarket. *Nat. Hum. Behav.* **1**, 0017 (2017).

Competing interests

The authors declare no competing interests.