Coherency Maximizing Exploration in the Supermarket

Peter S. Riefer[1,2]*, Rosie Prior[2], Nicholas Blair[2], Giles Pavey[2] & Bradley C. Love[3,1]

[1] Department of Experimental Psychology, University College London (UCL), 26 Bedford Way, London WC1H 0AP, UK.

[2] dunnhumby Ltd., 184 Shepherd's Bush Road, London W6 7NL, UK.

[3] 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 [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 (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 [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 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 [1,2,15]. 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 [8]. 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 [7-9,14], exploring more often when uncertainty is high. This efficient, systematic exploration appears to demand capacity-limited cognitive resources [9] and rely on frontal dopamine brain circuitry [14,15]. 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) [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 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 [24], 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 [8,9]. 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 [25] 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 (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 [7-11,14]. One explanation is that subjective rewards involve an evaluative process (e.g., 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 irrational [28] 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.

References

1 Cohen, J. D., McClure, S. M. & Angela, J. Y. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **362**, 933-942 (2007).

2 Hills, T. T. *et al.* Exploration versus exploitation in space, mind, and society. *Trends in cognitive sciences* **19**, 46-54 (2015).

3 Hutchinson, J. M., Wilke, A. & Todd, P. M. Patch leaving in humans: can a generalist adapt its rules to dispersal of items across patches? *Animal Behaviour* **75**, 1331-1349 (2008).

4 Kramer, D. L. & Weary, D. M. Exploration versus exploitation: a field study of time allocation to environmental tracking by foraging chipmunks. *Animal Behaviour* **41**, 443-449 (1991).

5 Krebs, J. R., Kacelnik, A. & Taylor, P. Test of optimal sampling by foraging great tits. *Nature* **275**, 27-31 (1978).

6 Mobbs, D. *et al.* Foraging under competition: the neural basis of input-matching in humans. *The Journal of Neuroscience* **33**, 9866-9872 (2013).

7 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).

8 Knox, W. B., Otto, A. R., Stone, P. & Love, B. The Nature of Belief-Directed Exploratory Choice in Human Decision-Making. *Frontiers in Psychology* **2**, 398 (2012).

9 Otto, A. R., Knox, W. B., Markman, A. B. & Love, B. C. Physiological and behavioral signatures of reflective exploratory choice. *Cognitive, Affective, & Behavioral Neuroscience* **14**, 1167-1183 (2014).

10 Riefer, P. S. & Love, B. C. Unfazed by Both the Bull and Bear: Strategic Exploration in Dynamic Environments. *Games* **6**, 251-261 (2015).

11 Speekenbrink, M. & Konstantinidis, E. Uncertainty and exploration in a restless bandit problem. *Topics in cognitive science* **7**, 351-367 (2015).

12 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).

13 Johansson, P., Hall, L. & Chater, N. in *Neuroscience of preference and choice: Cognitive and neural mechanisms* (eds Raymond J Dolan & Tali Sharot) 121-138 (Academic Press, 2012).

14 Blanco, N. J. *et al.* A frontal dopamine system for reflective exploratory behavior. *Neurobiology of learning and memory* **123**, 84-91 (2015).

15 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).

16 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. *Journal of Experimental Social Psychology* **46**, 204-207 (2010).

17 Sharot, T., Fleming, S. M., Yu, X., Koster, R. & Dolan, R. J. Is choice-induced preference change long lasting? *Psychological science* **23**, 1123-1129 (2012).

18 Sharot, T., Velasquez, C. M. & Dolan, R. J. Do decisions shape preference? Evidence from blind choice. *Psychological science* **21**, 1231-1235 (2010).

19 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).

20 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).

21 Festinger, L. *A theory of cognitive dissonance*. (Stanford University Press, 1957).

22 Frey, D. Recent research on selective exposure to information. *Advances in experimental social psychology* **19**, 41-80 (1986).

23 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. *Journal of personality and social psychology* **80**, 557 (2001).

24 Sutton, R. S. & Barto, A. G. *Reinforcement learning: An introduction*. (MIT press, 1998).

25 Hills, T. T., Todd, P. M. & Goldstone, R. L. Search in external and internal spaces evidence for generalized cognitive search processes. *Psychological Science* **19**, 802-808 (2008).

26 Frank, M. J., Doll, B. B., Oas-Terpstra, J. & Moreno, F. Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature neuroscience* **12**, 1062-1068 (2009).

27 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! *Journal of Behavioral Decision Making* **27**, 281-289 (2014).

28 Ariely, D. & Norton, M. I. How actions create–not just reveal–preferences. *Trends in cognitive sciences* **12**, 13-16 (2008).

Acknowledgements

We thank Peter Todd for his helpful comments.

This work was supported by the Leverhulme Trust grant RPG-2014-075, NIH (Grant 1P01HD080679), and Wellcome Trust Senior Investigator Award WT106931MA to B.L. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

P.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.L. for the design, analysis and write-up of the article.

Competing Interests statement

The authors declare no competing interests.

Figures

Figure 1

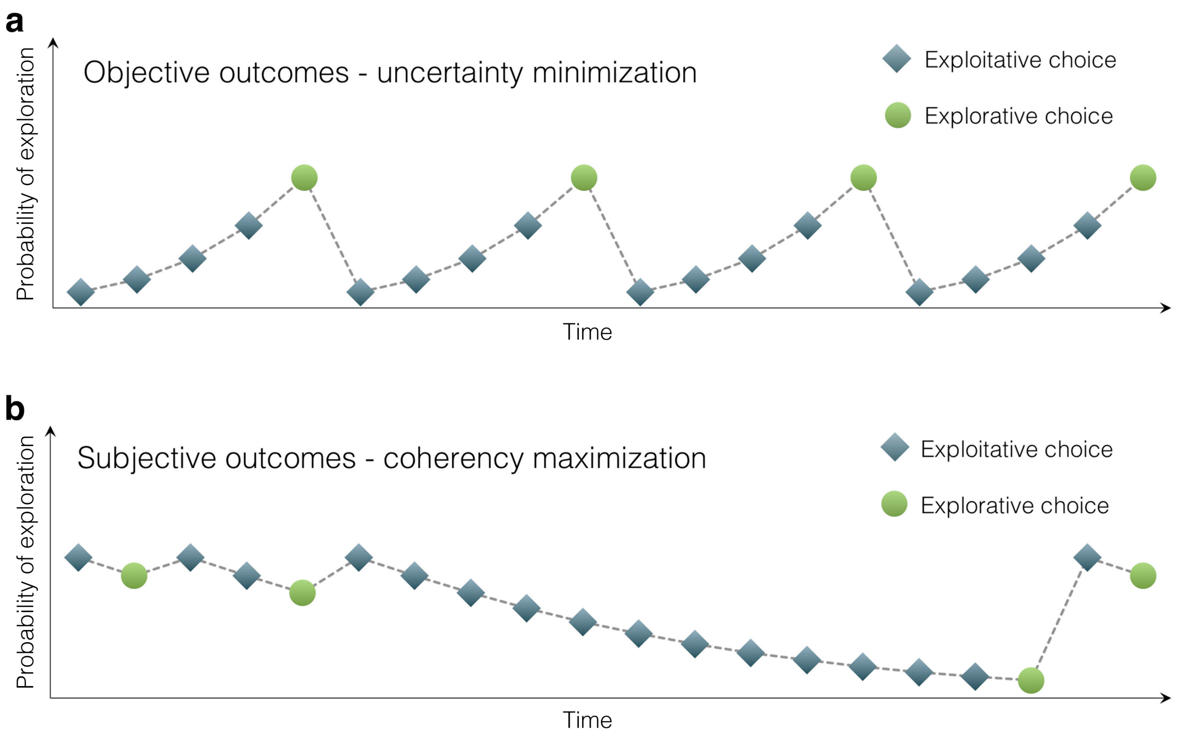


Figure 1: Predicted exploration patterns for uncertainty minimization and coherency maximization.

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

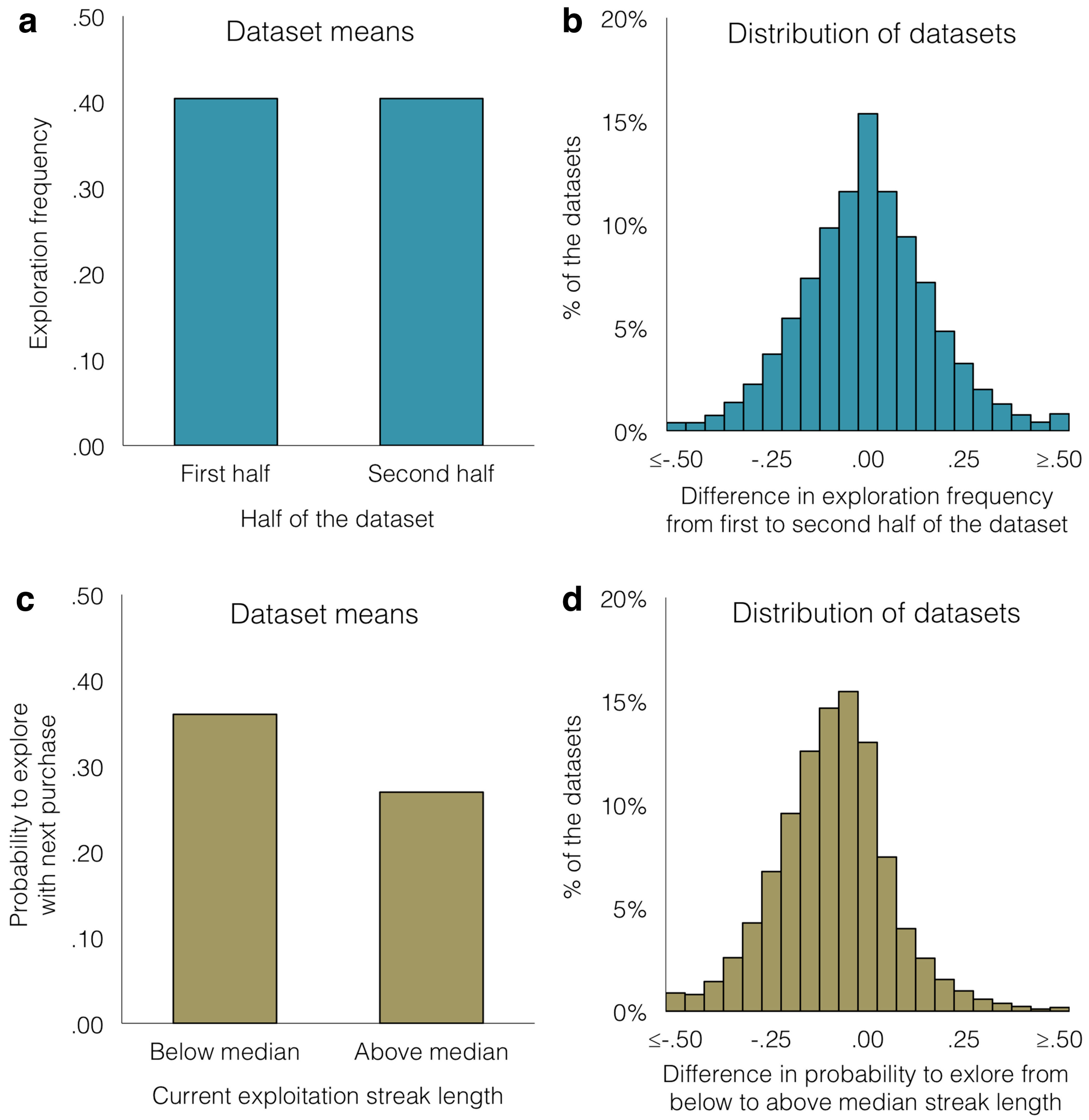


Figure 2: Exploration changes locally but not globally.

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.

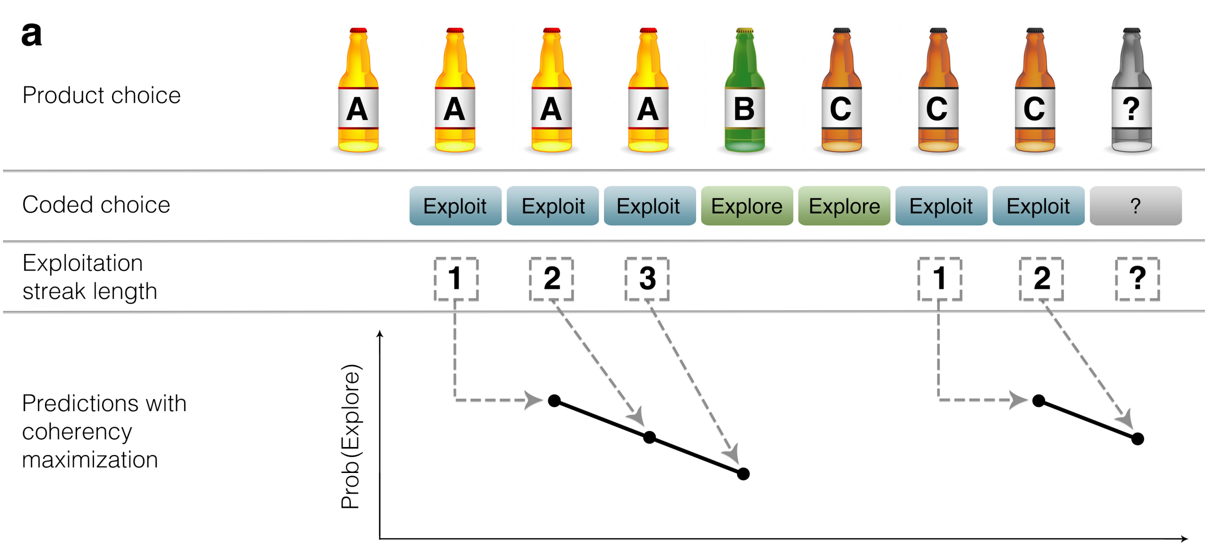
c,

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



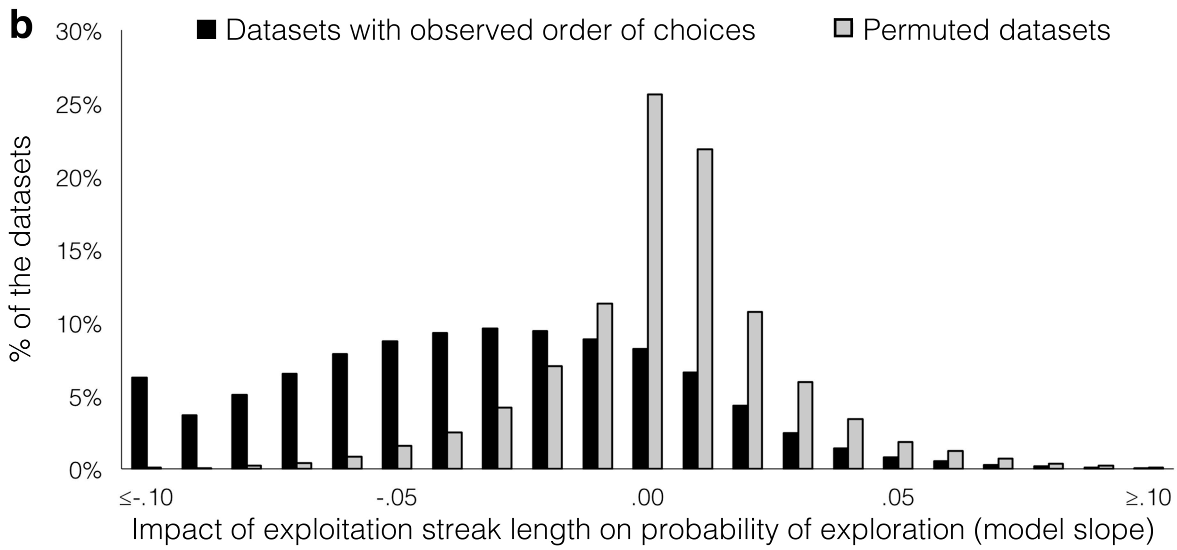


Figure 3: Predicting exploration from exploitation streak lengths.

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.

Figure 4

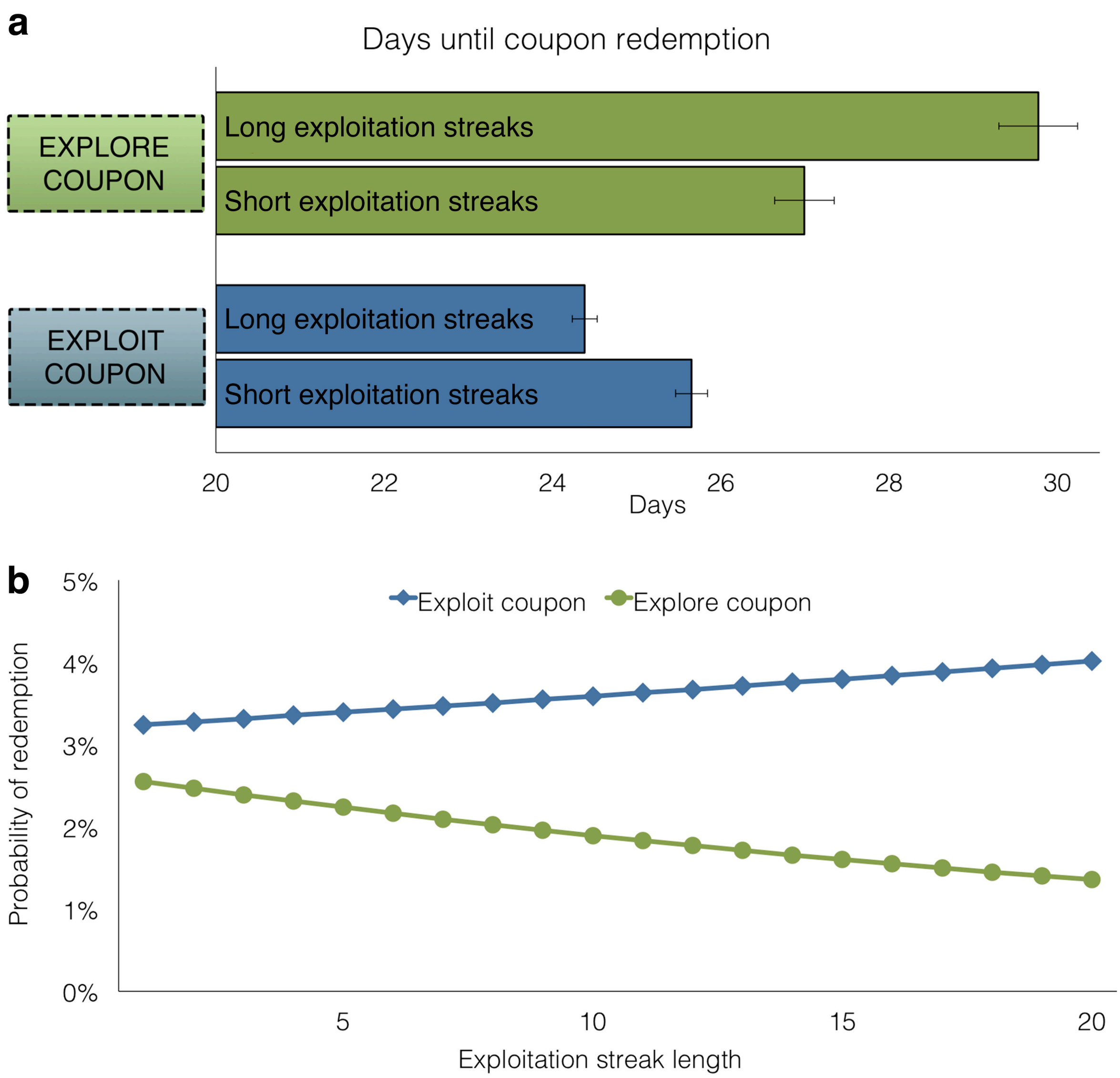


Figure 4: Coupon redemption depending on current exploitation streak length.

a,

Consistent with coherency maximization, customers redeemed coupons to exploit products more quickly the longer they have been exploiting the product (i.e. on long exploitation streaks). Coupons to explore alternatives were used more quickly when customers were only beginning to exploit (i.e. 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 has explored.

Method

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 dunnhumby, a customer science company and subsidiary of Tesco (see Data Availability statement 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 (e.g. quantities, prices, discounts, etc.), 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 (i.e., 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 (see Fig. 3a).

Data Availability

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

Coherency Maximizing Exploration in the Supermarket

Peter S. Riefer[1,2]*, Rosie Prior[2], Nicholas Blair[2], Giles Pavey[2] & Bradley C. Love[3,1]

[1] Department of Experimental Psychology, University College London (UCL), 26 Bedford Way, London WC1H 0AP, UK.

[2] dunnhumby Ltd., 184 Shepherd's Bush Road, London W6 7NL, UK.

[3] The Alan Turing Institute, 96 Euston Road, London NW1 2DB, UK.

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

Supplementary Information

Customer datasets

We gained access to the Clubcard bonus program of Tesco, a major UK supermarket chain. As Clubcard members, people receive points for their purchases at Tesco that can be transferred into store discounts and non-monetary rewards. Every time a customer shops at Tesco and uses the personalized Clubcard at the checkout, all purchased items and additional information, such as given discounts or redeemed coupons are recorded in a database. Individuals in this database can only be identified by an anonymous dataset number, but not by any personal information. We analysed anonymized datasets that only contained purchase-related information (e.g. quantities, prices, discounts, etc.), but no personal information about the shopper. Thus, it was impossible to draw any connection between the analysed data and the shopper who carried out these purchases. The Clubcard database provides 250 weeks of purchase histories for approximately 1.55 million customers, which is a ten per-cent random sample of all Clubcard users. In order to capture a set of typical choices in the supermarket, we examined purchases within the product categories of beers, breads, coffees, toilet papers, washing detergents and yogurts. The range of every category was defined according to the retailer's descriptions, which corresponds to the segmentation that people encounter in their stores. In order to assure a clearer differentiation between the choice options (e.g. ignore package sizes and special editions), we differed products on the brand level. Our datasets included chronological individual purchases of customers within one of the six above-mentioned product categories. For some customers, we had multiple datasets from different categories available.

We selected people with at least 50 purchases within a specific product category to gain sufficient data for individual analyses and also to retrieve customers that likely use their Clubcard with every store visit. Each product category contributed between 39,105 and 79,988 datasets, summing up to 318,294 in total. We coded exploration and exploitation based on repetition, where repeated choices were coded as exploitations and non-repetitions as explorations (see Fig. 3a). After the coding, we removed about 3.5% of the datasets, as subjects chose the same item on every single occasion. Furthermore, extreme cases of exploration proportions beyond .75 were also excluded from all analyses (7.6% of the datasets). Eventually, a total of 282,972 datasets remained for our analyses. The mean time between purchases was 14.57 days (SD=7.46) with breads on the low end at 8.57 days (SD=3.35) and washing detergent on the high end at 23.38 (SD=6.46) days. Of the available options in a product category, customers tried 8.22 (SD=5.84) different brands on average. This ranged from 4.59 (SD=2.05) options for washing detergents to 15.83 (SD=9.82) for beers.

Coding of explorative and exploitative choices

We identified choice options at the brand level, which subsumes minor differences in unit sizes (e.g. the same product in a 0.5kg and 1.5kg package). In previous research, the classification of explorative and exploitative choices is often based on the numeric value of experienced rewards [1-3]. As there is no numeric choice outcome in our brand choice dataset, we focus on the individual's decision between either staying or going – repeating a brand choice or breaking out of a repetitive habit [4]. This definition assumes that the brand people stay with at the moment is the one they currently perceive as the best choice. Our results proved robust to various definitions of exploratory choice.

Thus, we coded customers' repeated choice of a brand as exploitation. Non-repetitions were coded as explorations. For example, a choice sequence of A-A-B-A-A would be coded Exploit-Exploit-Explore-Explore-Exploit (assuming the first choice is a repetition). In some cases (less than 10%), people chose multiple items from the same category within the same store visit. We coded these cases as explorations when at least one of the items was not purchased on the previous store visit.

Permutation tests

We used permutation tests to determine when results differed from chance. We created 100 permutations of the originally observed choice sequences for each choice dataset, which served as a null distribution. As a next step, we applied the coding of exploration and exploitation as described above to the original and the permuted datasets. All statistical analyses could hence be conducted with original and permuted datasets. Comparing the results of observed and permuted datasets could therefore be used to identify whether and how people's actual choices differed from randomly ordered choices. The statistical values of every observed dataset (e.g. exploration frequency) were ranked in comparison with their permuted control (i.e., null) datasets. Here, we captured whether the dataset showed a relatively low or high value compared to the control datasets. Table 1 displays the mean values of exploration frequency, exploitation streak length and exploration streak length for observed and permuted data. Further, the relative number of observed datasets with low values compared to their control is shown. For example, 7.2% of shoppers have unusually short exploitation streaks compared to a permuted control. On the other hand, this means a vast majority of people has unusually long exploitation streaks. This suggests that the observed data is not randomly ordered and further, people go on longer exploitation streaks than one would expect if their choices were ordered randomly.

Supplementary Table 1
Mean values of observed and control data

	M_{Obs} (SD)	$M_{Control}$ (SD)	% with low values
f(explore)	.404 (.196)	.477 (.240)	83.4
Exploitation streak length	8.56 (18.33)	6.51 (17.95)	7.2
Exploration streak length	2.36 (.72)	2.56 (1.06)	76.7

Note. M=Mean. SD=Standard Deviation. Obs=Observed.

Individual logistic regression models

For each dataset, we modelled the probability of exploration for all choices in observed and permuted control data (see previous paragraph) using logistic regressions. A baseline model with intercept only (one parameter) attributed constant probability of exploration independent of recent choices. This baseline model was compared to a streak model that included an intercept and a slope linked to the current exploitation streak length (two parameters). For each dataset, both models were fitted to all choices (M=114.0 choices, SD=66.3 choices). The slope parameter in the streak model for the randomized control data shows a slight positive bias, due to the limited amount of data and the fact that the longest exploitation streak will definitely end, therefore causing f(explore | exploitation streak length) to inherently rise with streak length. In the observed data, 79.3% of the slopes in the streak model are negative across all product categories. Compared to the permuted control data, 82.6% of the datasets have unusually small or negative slope estimates. Table 2 lists these slope parameter estimates per category.

Supplementary Table 2
Slope estimates for the streak model in all product categories

	Observed datasets			Permuted control datasets	
Product category	Mean slope (SD)	% negative slope values	% values in lower 50th percentile of control data	Mean slope (SD)	% negative slope values
Beers	-.019 (.034)	66.8	71.9	.002 (.008)	27.6
Breads	-.040 (.037)	85.5	86.5	-.001 (.009)	55.6
Coffees	-.013 (.036)	59.2	68.3	.005 (.009)	24.2
Toilet tissues	-.026 (.039)	71.7	77.4	.004 (.010)	36.3
Washing detergents	-.016 (.035)	61.4	69.2	.003 (.007)	23.4
Yogurts	-.025 (.036)	71.0	74.8	.002 (.006)	29.3

Note. Means and standard deviations were calculated across all observed customer datasets in the respective categories. For the control datasets, the 100 slopes per customer dataset were first averaged and the mean and standard deviation of these averages across customer datasets are reported.

Cross-category analysis

Some customers were represented in multiple datasets that contained their shopping decisions in different product categories. In these cases, we tested whether customers showed similar behaviour across these categories. We yielded slope estimates from a logistic regression (see previous paragraph) for all datasets. Next, we examined how many individuals had unusually negative slopes, i.e. ranked in the lower fifth percentile (slope ranked ≤ .05) of the permuted control in either all or none of their choice datasets. Such individuals were labelled consistent in their exploratory behaviour. We focused on the lower fifth percentile of slopes as these individuals showed clear and strong coherency maximizing behaviour. In each category, this still applied to almost half of the people. In order to check for unexpected behavioural consistency across product categories we compared the number of observed consistent individuals with the expected frequencies from chance. We included individuals who were represented in three to five categories in our analysis to assure an appropriate number of constellations cases to analyse.

From the observed frequencies, we obtained a maximum-likelihood estimate for the probability that a dataset yielded a slope estimate in the lower fifth percentile rank amongst its permuted versions. This probability was estimated separately for cases where three, four or five datasets were available. We then created an expected frequency distribution for the number of datasets with slopes ranked in the lower fifth percentile given the total number of datasets available (i.e. either three, four or five) and compared this to the observed distribution. Table 3 gives the probability estimates for datasets to yield slopes in the lower fifth percentile, given the total number of datasets and also the observed and expected distributions for datasets with such slopes.

Supplementary Table 3

Observed and expected frequencies of slopes in lower 5th percentile when multiple datasets available per person

	Datasets available		
	3 datasets	4 datasets	5 datasets
No slope ranked ≤.05	4,946 (4,748)	1,317 (1,314)	163 (136)
1 slope ranked ≤.05	11,793 (12,062)	4,383 (4,285)	544 (561)
2 slopes ranked ≤.05	10,159 (10,213)	5,100 (5,239)	911 (923)
3 slopes ranked ≤.05	3,007 (2,882)	2,817 (2,846)	736 (759)
4 slopes ranked ≤.05	---	647 (580)	325 (311)
5 slopes ranked ≤.05	---	---	62 (51)
Total	29,905	14,264	2,741

Note. Expected frequencies in brackets.

Coupon analysis

First, we analysed how quickly customers redeemed product coupons, which provided points on their bonus card or price discounts for buying a promoted product. Customers received these product coupons in their mail and could choose to redeem them when buying the promoted product within 60 days. In all our datasets, we found a total of 69,664 purchases that were associated with the use of a product coupon. From the view of coherency maximizing exploration, coupons to explore fit customers' current behaviour better on short than on long exploitation streaks. Similarly, coupons to exploit a favourite fit better on long exploitation streaks than on short ones. Therefore, having exploited a product for a long time rather than just having started to exploit a new favourite, customers should be quicker to redeem a coupon for the exploited product. On the other hand, customers should be slower to redeem a coupon to explore a different product when they are currently on a long exploitation streak. This means on the other hand, that shoppers should redeem such an exploration coupon quicker when their exploitation streak is still short. We divided the exploitation streaks in each individual's dataset into short and long streaks by a median split. For each observed coupon redemption, we noted whether at the time of reception the customer was on a short or long exploitation streak and whether the coupon product would mean exploration or exploitation for the next purchase. 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).

Based on the findings that we reported for the redemption times of coupons in the database, we mailed product coupons to randomly selected households in order to examine coupon redemption with respect to recent exploratory choices. We obtained 8,623 anonymized and randomly selected datasets from the instant coffees category where the customer made at least 50 purchases, independent of what was chosen. In the instant coffees category, people could choose between a maximum of 27 options of which five had a market share of at least 5% (including the target brand on our coupon). We made a priori predictions about how likely individuals would redeem a coupon for a certain instant coffee brand based on their recent choices. Then, we sent the 8,623 anonymous customer IDs to a mailing house so that they could be mailed a coupon offering of 100 bonus points if they bought a specific instant coffee brand, which was the same for all customers. Customers could redeem the coupon within three weeks after reception.

We defined the coupon type for each customer individually. Coupons served customers to exploit if their last purchase before receiving the coupon was the coupon product. On the other hand, the coupon was identified as exploration coupon if the customer bought a different instant coffee brand with the most recent purchase. Table 4 summarizes a logistic regression with probability of coupon redemption as dependent variable. We used coupon type (exploration or exploitation coupon) and exploitation streak length at coupon reception, as well as the interaction of the two variables as predictors. The significant negative interaction term implies that redemption was less likely for exploration coupons on long exploitation streaks. Conversely, redemption was more likely for exploitation coupons the longer customers have already been exploiting. Regarding the odds ratios, we can see that for exploitation coupons, the chance of coupon redemption increases 1.012 times with every additional exploitative choice or 1.127 times every ten exploitations. On the other hand, redemption decreases by a factor of 1.035 for exploration coupons with every additional exploitative product purchase or, in other words, by 1.411 with every ten additional exploitations.

Supplementary Table 4

Logistic regression to predict coupon redemption from exploitation streak length

Model parameter	Estimate	Std. Error	z-value	p-value
Intercept	-3.4077	.1079	-31.584	<.001
Coupon type (0=Exploit, 1=Explore)	-.2003	.1628	-1.231	.219
Exploitation streak length	.01168	.0037	3.185	.001
Coupon type * Exploitation streak length	-.0456	.0126	-3.623	<.001

References

1 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).

2 Knox, W. B., Otto, A. R., Stone, P. & Love, B. The Nature of Belief-Directed Exploratory Choice in Human Decision-Making. *Frontiers in Psychology* **2**, 398 (2012).

3 Otto, A. R., Knox, W. B., Markman, A. B. & Love, B. C. Physiological and behavioral signatures of reflective exploratory choice. *Cognitive, Affective, & Behavioral Neuroscience* **14**, 1167-1183 (2014).

4 Cohen, J. D., McClure, S. M. & Angela, J. Y. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences* **362**, 933-942 (2007).