Games **2015**, *6*, 251-261; doi:10.3390/g6030251

OPEN ACCESS

games

ISSN 2073-4336

www.mdpi.com/journal/games

Article

Unfazed by Both the Bull and Bear: Strategic Exploration in Dynamic Environments

Peter S. Riefer * and Bradley C. Love

Department of Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, UK; E-Mail: b.love@ucl.ac.uk

* Author to whom correspondence should be addressed; E-Mail: peter@peterriefer.net; Tel.: +44-7871-655-465; Fax: +44-2074-364-276.

Academic Editors: Andrew M. Colman and Briony D. Pulford

Received: 8 July 2015 / Accepted: 12 August 2015 / Published: 18 August 2015

Abstract: People in a changing environment must decide between exploiting options they currently favor and exploring alternative options that provide additional information about the state of the environment. For example, drivers must decide between purchasing gas at their currently favored station (*i.e.*, exploit) or risk a fruitless trip to another station to evaluate whether the price has been lowered since the last visit. Previous laboratory studies on exploratory choice have found that people choose strategically and explore alternative options when it is more likely that the relative value of competing options has changed. Our study extends this work by considering how global trends (which affect all options equally) influence exploratory choice. For example, during an economic crisis, global gas prices may increase or decrease at all stations, yet consumers should still explore strategically to find the best option. Our research question is whether people can maintain effective exploration strategies in the presence of global trends that are irrelevant in that they do not affect the relative value of choice options. We find that people explore effectively irrespective of global trends.

Keywords: decision-making; strategic choice; exploration and exploitation; dynamic environments; global trends

---

1. Introduction

Let us imagine the following scenario: A driver needs to refuel her car and knows of two gas stations in the area. As these stations change their prices frequently, it is difficult for the driver to predict which one is currently the cheapest. Let us further assume that the person has been to both stations and remembers the prices from the most recent trips to each of them. One possibility for the driver is to *exploit* this information and go to the station that would be cheaper according to current knowledge. Alternatively, the person could decide to *explore* and drive to the station believed to be more expensive, in the hope of finding that this station has now lowered the price. Exploration and exploitation are both associated with different costs which makes it crucial to balance the two optimally [1,2]. The exploitation of current beliefs incurs the cost of missing changes in dynamic environments. For example, only visiting one gas station makes it impossible to discover prices have been lowered at another station. Exploration of the environment entails potential costs of choosing inferiorly to update beliefs, such as driving to another gas station and observing that the prices are still higher. The optimal balance of exploration and exploitation is particularly important as beliefs about outcomes in changing environments can become outdated over time [3–6]. Therefore, optimal decision-making requires exploration and exploitation at the right point of time, whenever their respective costs are assumed to be minimal.

The exploration *vs.* exploitation dilemma has been studied as one-armed bandit problems in reinforcement learning e.g., [7–10], but gained high relevance in various fields, such as foraging theory [11–16], information search [1,17–19] or organizational planning and networking [20,21]. Research has investigated human exploratory decision-making e.g., [2,22,23] and found that people are able to make systematic exploratory decisions with regard to how changes occur in their environment [3–6]. Knox and colleagues [4] found that people explore their environment whenever it is more likely to gain new information. On the other hand, people exploit their current beliefs whenever it is more likely that these beliefs are still up-to-date. In order to examine how people balance exploration and exploitation in changing environments, Knox *et al.* designed the leapfrog task (see Figure 1). In this experiment, people choose repeatedly between two options, one of which is always better than the other. However, on each trial, the inferior option might improve and leapfrog the other option with a constant probability throughout the experiment. Participants are not informed about these jumps, but have to observe them directly by choosing the options. Since subjects can only see the current values of options by choosing them, they have to decide between exploiting the option they currently believe to be better and exploring the other option in order to check whether it has leapfrogged. Results from several studies [3–6] show that people update their beliefs through exploration while considering how the environment changes, meaning that they take into account when exploration is more likely to be informative.

Due to the fact that only the inferior option can change in the leapfrog task, it is possible to describe an optimal exploration strategy for it. Exploring effectively in an environment such as in the leapfrog task means exploring with respect to how recently one has previously explored. If subjects have just recently explored, chances are low that there has been an unobserved change with the other choice. In this case, the subjects should continue to exploit the outcome that they believe to be better. On the other hand, with longer exploitation streaks, chances for unobserved changes increase with every

additional exploitation. For example, drivers should explore different gas stations for price changes if they haven't checked their prices for a while. Optimal exploration requires one to become more likely to explore, the further back the last exploration dates or, in other words, the more likely it is that the previous exploration does not reflect the current state of the environment any more. Knox and colleagues [4] call this positive relation of subjects' probability to explore and the length of their exploitation streak the hazard rate of exploration. In the leapfrog experiments, the majority managed to explore systematically with respect to the length of their current exploitation streak. This means that subjects did not explore randomly, but in accordance with the prediction of a model that considers how the environment changes. There remains, however, the question of whether these findings still hold when people encounter noisy choice environments with both relevant and irrelevant information.

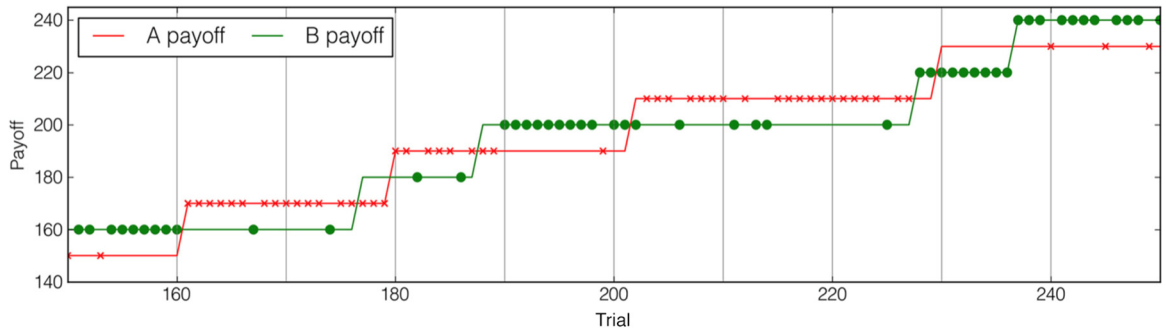

Figure 1. Example from [4] for the choice payoffs in the leapfrog task from trial 150 to trial 250. One of the options always features lower payoff, but also a random chance of becoming the higher payoff choice in the next trial (leapfrogging). Participants could only see the payoff of the options they chose and therefore had to decide whether to exploit the option that they believed to be better or explore the payoff of the other option to see whether it leapfrogged. The crosses and dots represent example choices of a participant.

In this paper, we examine whether global trends affecting all options equally have an impact on people's exploratory choices, although they should be ignored. Since global trends, such as with inflation or deflation of currencies, affect the whole decision space, they are irrelevant to discriminate between choices within this decision space. For example, if the currency is inflating, prices at all gas stations underlie the same inflation and therefore inflation shouldn't matter to choose between these stations. This applies to simultaneous exploratory decisions, such as making an instant decision for one gas station out of a set of known stations. Though global trends are mostly irrelevant when choosing simultaneously amongst a set of choices, they can have implications for exploratory decision-making with sequential choice (e.g., driving from one station to the other and checking prices before deciding). This is due to the fact that sequential search entails costs or benefits from a dynamic environment, depending on how the environment changes [24]. For example, if the currency is inflating, continued exploration might be costly, as the individual's current wealth will lose in value, while with deflating prices, continuing to explore can be advantageous as the prices change for the better. However, in this paper we focus on exploratory decisions that are made simultaneously and therefore without such an effect of global trends. People should ignore global trends with simultaneous choices since they affect

all options in the same way and therefore, these options will not differ from each other due to the impact of global trends.

The leapfrog studies have also shown that systematic exploration is difficult for people with depression symptoms [3] or for healthy subjects under cognitive load [5]. Therefore, we aim to study people's ability to make effective exploratory choices despite distractors that they typically encounter in the world. Research on exploratory decision-making in dynamic environments is still scarce [2], however, environments usually evolve and require people to update their beliefs from time to time. In these cases, some encountered information is relevant to discriminate between available options and some is not. If we think about inflation or deflation once again, prices for all goods and services will change simultaneously, and therefore these price changes aren't informative to discriminate between offers within the affected economy. But inflation and deflation have surprising effects on people's behavior [25–27]. For example, with inflation one might choose differently simply because the increasing prices make them feel anxious about their current situation. Research has shown that it is difficult for people to ignore irrelevant inputs and focus on information that is relevant for a task [28–31]. Effective exploration is resource-intensive and requires people to estimate when exploration is more likely to be informative [5,32,33]. Therefore, we assume that global trends could distract decision-makers from the actual exploration-exploitation task, making it harder for them to explore effectively. In other words, we would like to examine whether people can ignore global distractors that are irrelevant in a choice between two options in an exploration-exploitation task. Given the findings of previous studies, it is possible that subjects explore more randomly with distracting global trends. It is furthermore possible that they generally over- or under-explore due to the misleading effects of global trends. For example, an increasing global trend could suggest that the currently exploited option is doing better with every trial, making exploration less necessary. We examine these possible effects with the leapfrog task that has been successfully used to study exploratory behavior in the past and add global trends to both options in order to observe whether people can still explore effectively as in the standard experiment without global trends.

2. Experimental Section

2.1. Participants

One hundred ninety-nine participants (110 male, 89 female) were recruited via Amazon Mechanical Turk (MTurk). MTurk has been shown to be an inexpensive, yet reliable and efficient way of collecting data from a demographically diverse sample [34,35]. All participants were US citizens and required to have at least a 95% approve rate on their previous experiments on MTurk. The mean age of the sample was 31.0 years (SD = 9.2), with an age range of 20 to 63 years. Participants received $1 (US Dollars) for their participation, and 0.5¢ was rewarded for every correct choice in the experiment. In 200 trials of the experiment, participants scored on average 132 correct choices (66%), which means an average participant received $1.66 for about 17 minutes to finish the experiment. Participants were randomly assigned to three environmental conditions: Constantly increasing, constantly decreasing, or stable, which is the control condition and corresponds standard leapfrog task. There were 66, 65 and 68 participants in each condition respectively.

2.2. Design

The leapfrog task featured two options, buttons A and B from which the participant had to choose in each of 200 trials. Participants had to choose the respective button to reveal its value. The value of the other button remained covered. The number associated with button A and B started at 500 and 510 respectively. Throughout the experiment, the lower button number could randomly, with a probability of 0.10, increase by 20 in the next trial, therefore jumping over the value of the other button. The three experimental conditions differed in the way global trends were integrated into the leapfrog task (see Figure 2). In addition to the random chance of increasing by 20, the increasing and decreasing conditions also featured global trends. In the increasing trend condition, button A's and B's number steadily and simultaneously increased their values by 1 point per trial. In the decreasing trend condition, both numbers steadily decreased their values by 1 point per trial. There was no such global trend in the control condition.

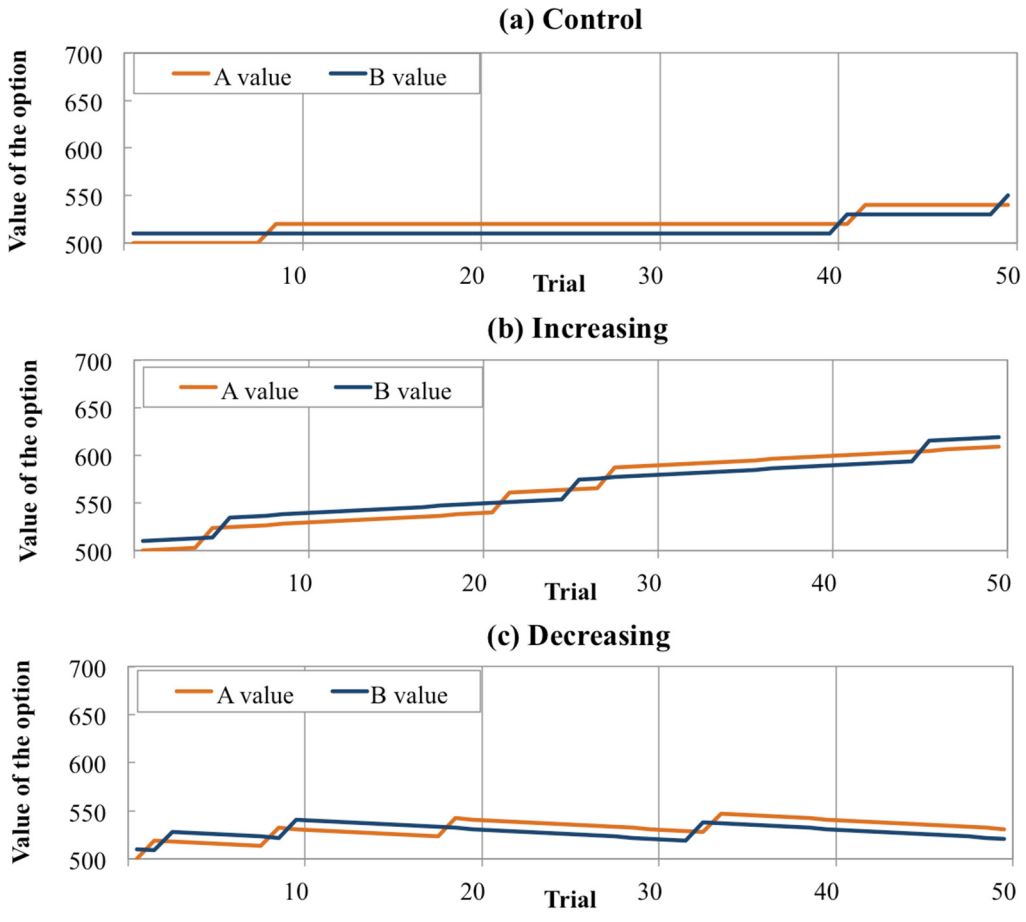

Figure 2. Examples for the number values of the two options in the experiment over the first 50 of 200 trials. In all experimental conditions, the lower value had a random chance of .10 to increase by 20 every trial. On top of this, there were global trends affecting both options in the increasing and decreasing condition. (**a**) There was no global trend in the control condition that corresponded to the standard leapfrog task; (**b**) In the increasing condition, both options increased their value by 1 every trial; (**c**) In the decreasing condition, both options decreased their value by 1 every trial.

2.3. Procedure

At the beginning of the experiment, participants were told that the task involves them choosing between two buttons, A and B (see Figure 3a), in each of the 200 trials with a 2 s time limit. Participants were informed that each button had a number associated that could change over time. Subjects were furthermore instructed to choose the button with the higher number in order to receive an additional ¢0.5 bonus in the respective trial. The numbers of the buttons were hidden and revealed for 1.5 s for the chosen button after each trial (see Figure 3b). If no choice was made within 2 s, the phrase "TOO SLOW" would appear below the buttons and that particular trial would be skipped. The timings of choice and feedback in our experiment were identical to the ones used by Knox and colleagues [4]. Immediately after choosing, participants only saw the number of the button they chose, not whether their choice was correct (*i.e.*, whether they chose the higher button number), as this would have informed them about the complete current state of the environment. Instead they saw a total count of their correct choices at the end of the experiment.

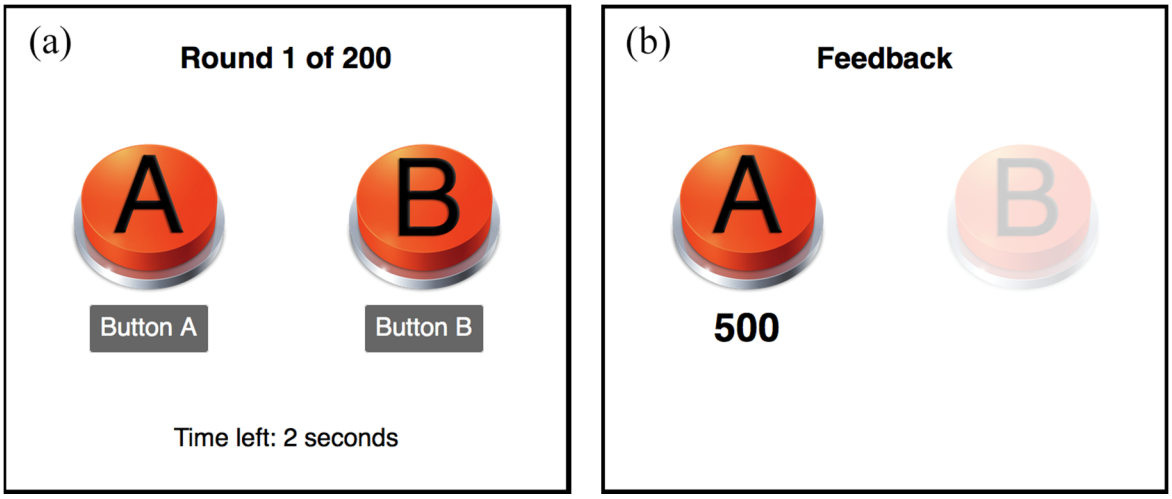

Figure 3. (**a**) Participants had 2 s to choose either button A or button B; (**b**) After their choice (button A had been chosen in the picture), they could see the number of the chosen button for 1.5 s and then returned to the choice screen again.

Before the experiment began, participants were asked three questions about the task to make sure they understood the instructions. They could only proceed if they answered all three questions correctly. After this, participants watched 100 demonstration trials, where unlike in the actual experiment, they did not make a choice but saw the numbers of both buttons in each round. The button numbers changed exactly as in actual choice trials and also featured global trends in the respective conditions. This was to ensure that subjects understood our descriptions of the task and how the numbers change in general. Participants then proceeded to the main part of the experiment and completed 200 trials of the task.

3. Results and Discussion

3.1. Overall Effectiveness and Frequency of Exploration

Subjects missed 1.6% of choices on average due to time out. There was no significant difference regarding missed choices between the three conditions, $F(2, 196) = 0.46$, $p = 0.63$. Choices were coded as explorations and exploitations according to the subject's observations. Hence, subjects decided either to exploit the button that had the higher number according to their observations or explore whether the other button with the lower number has leapfrogged. First, we examine whether global trends had an impact on exploration frequency. On average, subjects explored about every fifth trial with a relative frequency of 0.201 (SD = 0.073). The relative exploration frequency was not significantly different between conditions, $F(2, 196) = 0.001$, $p > 0.99$. To further test whether the experimental treatments made any difference with respect to people's exploration frequency, we conduct a Bayes factor analysis. Here, we always compare how the obtained data changes the beliefs in the null hypothesis of no differences between experimental treatments and the competing hypothesis that the treatments differ. A Cauchy-distribution is used for the priors to allow for non-uniform prior probabilities of different effect sizes [36]. Our analysis indicates that the data strongly supports the null hypothesis of no differences. It is 19.30 times more likely that the exploration frequencies were of equal value in the three conditions than assuming that they were different. Furthermore, participants accurately chose the option with the higher number in 67.5% (SD = 6.3) of the cases (omissions excluded). This accuracy also did not differ across conditions, $F(2, 196) = 1.80$, $p = 0.17$. Regarding the Bayes factors, the hypothesis of no differences in accuracy is 4.01 times more likely than the hypothesis of differences between conditions. Figure 4 depicts the similarity of the experimental conditions regarding exploration frequency and accuracy. In conclusion, exploration frequency and task performance appear unaffected by global trends.

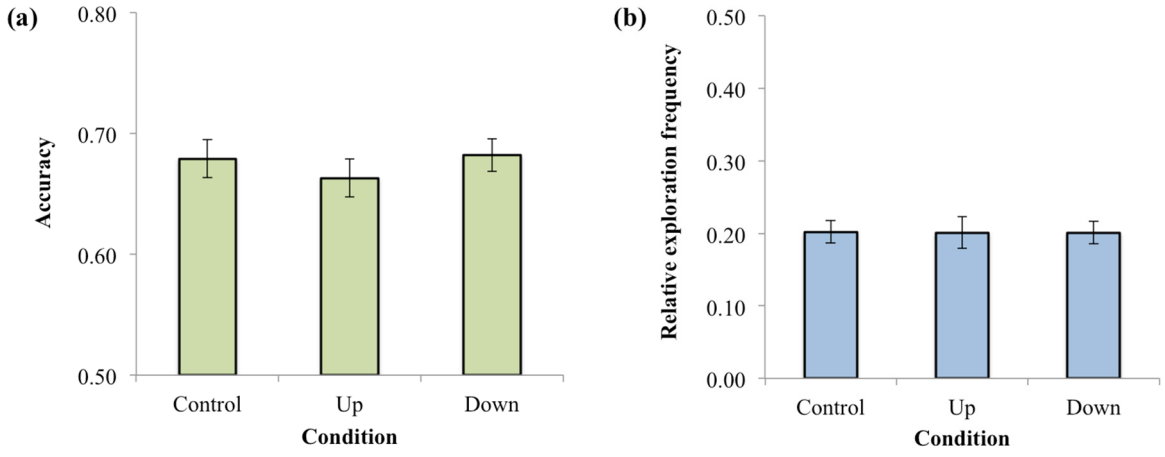

Figure 4. Group means of performance and behavior with 95% confidence interval for each value. (**a**) Accuracy of choosing the button with the highest value across conditions; (**b**) Relative exploration frequency in three experimental conditions.

3.2. Individual Timing of Exploration

Previous studies found that subjects become more likely to explore the longer they have been exploiting without interruption see e.g., [4]. Here, individuals seem to realize that the longer they have been exploiting an option, the more likely it gets for the other, unobserved option to have changed. Hence, exploration is more likely to provide new information and benefits after longer streaks of exploitation. As illustrated in Figure 5, we found that overall, subjects in each experimental condition become more likely to explore the longer they have already been exploiting. To examine this on the subject level, we calculate individual logistic regressions to estimate subjects' probability to explore in every trial of the experiment, depending on their current exploitation streak length in that trial. A positive relationship of exploitation streak length and probability of exploration would indicate that people become more likely to explore the higher the chances are to find that the alternative option has actually changed.

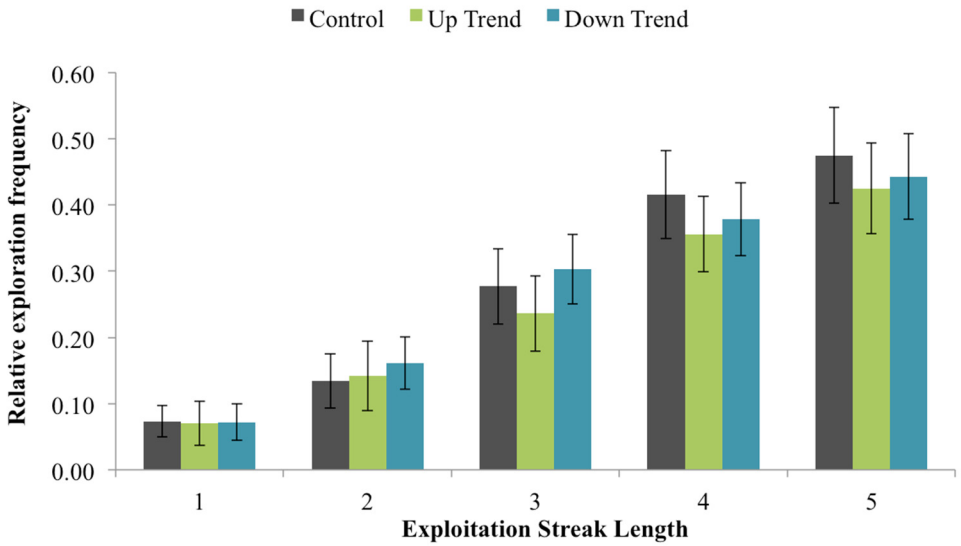

Figure 5. Subjects' relative exploration frequency with different exploitation streak lengths. Error bars represent 95% confidence intervals for the mean values between subjects. All conditions show similar patterns and subjects explore more frequently after longer exploitation streaks.

For 83.4% of the subjects, exploitation streak length plays a significant role to explain probability of exploration. In 162 of the 164 significant cases, this relationship is significantly positive so that exploration is more likely after longer exploitation streaks. This means we are able to reproduce the previous findings of Knox and colleagues. The parameter estimates for the relationship of exploitation streak length and probability of exploration do not differ across conditions, $F(2, 196) = 0.935$, $p = 0.39$. Using Bayes factors, we find that it is 8.52 more likely for the parameter estimates to not differ across conditions than assuming differences. We can deduce that despite global trend variables, subjects had similar timing of explorations that corresponded to the need to explore the environment whenever new information was more likely to be available. Summing up, it appears as if both, the frequency and timing of subjects' explorations, were unaffected by global trends.

4. Conclusions

In this study, we examined whether individuals can identify relevant information to explore changing environments effectively despite irrelevant global trends. Our results show that global trends neither affect people's exploration frequency nor the timing of their exploratory choices. This is important since people are usually exposed to both relevant and irrelevant information when they make exploratory decisions in their lives. For example, currency inflation affects all prices in an economy equally and is therefore irrelevant to distinguish between offers within this economy. In such a context, difficulties for people to explore effectively might arise for two different reasons. First, relevant information has to be selected over irrelevant information to evaluate choices adequately. Previous studies showed that in some situations, it is difficult for people to ignore irrelevant stimuli and only select information that is task-relevant [28–31]. Second, due to this additional selection step, fewer resources can be spent on actually using selected information to make exploratory decisions. Effective exploration is a demanding task that asks for people's full attention and cognitive resources [5,32,33]. For example, depression symptoms or cognitive load have negative effects on people's effectiveness to explore environments. Here, we showed that healthy subjects are able to handle distracting global trends and both filter and interpret information properly to make well-timed exploratory choices. But why might we still observe varying behavior with respect to global trends in other situations [25,27]? The answer could be related to the social context of decisions, where, for example, media or social environment influence people's perception [37] and use of information, as well as their general attitudes and preferences [38]. In these cases, social pressure could elicit panic decision-making where people ignore their ability to actually use information optimally when these factors are absent. Assuming this could explain some of the unexpected behavior during recessions, it would be worth studying to what extent individuals influence each other's information perception and exploratory decision-making in consequence.

Acknowledgments

This work was supported by the Leverhulme Trust grant RPG-2014-075 and Wellcome Trust Senior Investigator Award WT106931MA to B.L.

Author Contributions

P.R. was involved in all parts of this project and received input from B.L. for the design, analysis and write-up of the article.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Hills, T.T.; Todd, P.M.; Lazer, D.; Redish, A.D.; Couzin, I.D.; Group, C.S.R. Exploration *vs.* exploitation in space, mind, and society. *Trends Cognit. Sci.* **2015**, *19*, 46–54.

2. 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. *Philos. Trans. R. Soc. B Biol. Sci.* **2007**, *362*, 933–942.
3. 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* **2013**, *129*, 563–568.
4. Knox, W.B.; Otto, A.R.; Stone, P.; Love, B. The nature of belief-directed exploratory choice in human decision-making. *Front. Psychol.* **2012**, *2*, 398, doi:10.3389/fpsyg.2011.00398.
5. Otto, A.R.; Knox, W.B.; Markman, A.B.; Love, B.C. Physiological and behavioral signatures of reflective exploratory choice. *Cognit. Affect. Behav. Neurosci.* **2014**, *14*, 1167–1183.
6. Otto, A.R.; Markman, A.B.; Gureckis, T.M.; Love, B.C. Regulatory fit and systematic exploration in a dynamic decision-making environment. *J. Exp. Psychol. Learn. Mem. Cognit.* **2010**, *36*, 797–804.
7. Gittins, J.; Jones, D. *Progress in Statistics*; North-Holland: Amsterdam, The Netherland, 1974; Volume 2, p. 9.
8. Kaelbling, L.P.; Littman, M.L.; Moore, A.W. Reinforcement learning: A survey. *J. Artif. Intell. Res.* **1996**, *4*, 237–285.
9. Sutton, R.S.; Barto, A.G. *Introduction to Reinforcement Learning*; MIT Press: Cambridge, MA, USA, 1998.
10. Thrun, S.B. *Efficient Exploration in Reinforcement Learning*; Carnegie Mellon University: Pittsburgh, PA, USA, 1992.
11. Chen, H.; Zhu, Y.; Hu, K. *Adaptive Bacterial Foraging Optimization*; Abstract and Applied Analysis; Hindawi Publishing Corporation: New York, NY, USA, 2011.
12. Eliassen, S.; Jørgensen, C.; Mangel, M.; Giske, J. Exploration or exploitation: Life expectancy changes the value of learning in foraging strategies. *Oikos* **2007**, *116*, 513–523.
13. Hills, T.T. Animal foraging and the evolution of goal-directed cognition. *Cognit. Sci.* **2006**, *30*, 3–41.
14. Kramer, D.L.; Weary, D.M. Exploration *vs.* exploitation: A field study of time allocation to environmental tracking by foraging chipmunks. *Anim. Behav.* **1991**, *41*, 443–449.
15. Krebs, J.R.; Kacelnik, A.; Taylor, P. Test of optimal sampling by foraging great tits. *Nature* **1978**, *275*, 27–31.
16. Mobbs, D.; Hassabis, D.; Yu, R.; Chu, C.; Rushworth, M.; Boorman, E.; Dalgleish, T. Foraging under competition: The neural basis of input-matching in humans. *J. Neurosci.* **2013**, *33*, 9866–9872.
17. Betsch, T.; Haberstroh, S.; Glöckner, A.; Haar, T.; Fiedler, K. The effects of routine strength on adaptation and information search in recurrent decision making. *Organ. Behav. Hum. Decis. Process.* **2001**, *84*, 23–53.
18. 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.* **2008**, *75*, 1331–1349.
19. Pirolli, P. Rational analyses of information foraging on the web. *Cognit. Sci.* **2005**, *29*, 343–373.
20. Lazer, D.; Friedman, A. The network structure of exploration and exploitation. *Adm. Sci. Q.* **2007**, *52*, 667–694.
21. March, J.G. Exploration and exploitation in organizational learning. *Organ. Sci.* **1991**, *2*, 71–87.

22. Daw, N.D.; O'Doherty, J.P.; Dayan, P.; Seymour, B.; Dolan, R.J. Cortical substrates for exploratory decisions in humans. *Nature* **2006**, *441*, 876–879.
23. McClure, S.M.; Gilzenrat, M.S.; Cohen, J.D. An exploration-exploitation model based on norepinephrine and dopamine activity. *Adv. Neural Inf. Process. Syst.* **2006**, *18*, 867-874.
24. Dudey, T.; Todd, P.M. Making good decisions with minimal information: Simultaneous and sequential choice. *J. Bioecon.* **2001**, *3*, 195–215.
25. Kamakura, W.A.; Du, R.Y. How economic contractions and expansions affect expenditure patterns. *Kamakura Wagner A. Econ. Contract. Expans. Affect Expend. Patterns J. Consum. Res.* **2012**, *39*, 229–247.
26. Katona, G. Psychology and consumer economics. *J. Consum. Res.* **1974**, *1*, 1–8.
27. Sharma, V.; Sonwalkar, J. Does consumer buying behavior change during economic crisis? *Int. J. Econ. Bus. Adm.* (*IJEBA*) **2013**, *1*, 33–48.
28. Lavie, N. Perceptual load as a necessary condition for selective attention. *J. Exp. Psychol. Hum. Percept. Perform.* **1995**, *21*, 451–468.
29. Lavie, N. Distracted and confused? Selective attention under load. *Trends Cognit. Sci.* **2005**, *9*, 75–82.
30. Lavie, N.; Cox, S. On the efficiency of visual selective attention: Efficient visual search leads to inefficient distractor rejection. *Psychol. Sci.* **1997**, *8*, 395–396.
31. Lavie, N.; Hirst, A.; de Fockert, J.W.; Viding, E. Load theory of selective attention and cognitive control. *J. Exp. Psychol. Gen.* **2004**, *133*, 339–354.
32. Blanco, N.J.; Love, B.C.; Cooper, J.A.; McGeary, J.E.; Knopic, V.S.; Maddox, W.T. A frontal dopamine system for reflective exploratory behavior. *Neurobiol. Learn. Mem.* **2015**, *123*, 84–91.
33. Laureiro-Martínez, D.; Brusoni, S.; Canessa, N.; Zollo, M. Understanding the exploration-exploitation dilemma: An mri study of attention control and decision-making performance. *Strateg. Manag. J.* **2014**, *36*, 319–338.
34. Buhrmester, M.; Kwang, T.; Gosling, S.D. Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? *Perspect. Psychol. Sci.* **2011**, *6*, 3–5.
35. Crump, M.J.; McDonnell, J.V.; Gureckis, T.M. Evaluating amazon's mechanical turk as a tool for experimental behavioral research. *PLoS ONE* **2013**, *8*, e57410.
36. Rouder, J.N.; Morey, R.D.; Speckman, P.L.; Province, J.M. Default bayes factors for anova designs. *J. Math. Psychol.* **2012**, *56*, 356–374.
37. Richardson, D.C.; Street, C.N.; Tan, J.Y.; Kirkham, N.Z.; Hoover, M.A.; Cavanaugh, A.G. Joint perception: Gaze and social context. *Front. Hum. Neurosci.* **2012**, *6*, doi:10.3389/fnhum.2012.00194.
38. Gardner, M.; Steinberg, L. Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: An experimental study. *Dev. Psychol.* **2005**, *41*, 625–635.