When to Walk Away: The Effect of Variability on Keeping Options Viable

Anastasia Ejova (anastasia.ejova@adelaide.edu.au)
School of Psychology, University of Adelaide
Adelaide, SA 5005 Australia

Daniel J. Navarro (daniel.navarro@adelaide.edu.au)
School of Psychology, University of Adelaide
Adelaide, SA 5005 Australia

Amy F. Perfors (amy.perfors@adelaide.edu.au)
School of Psychology, University of Adelaide
Adelaide, SA 5005 Australia

Abstract

In this work we explore a variant of a classic explore/exploit dilemma based on a recent study by Shin and Ariely (2004). We presented participants with a scenario in which they were forced to explore among 9 different options of unknown value; for some, options left unexplored long enough disappeared. Contrary to expectations, the mere presence of additional options did not substantially increase people's ability to walk away from them. However, when options differed in value, people were able to locate the more valuable options and became much more willing to walk away from at least a few. Finally, our analysis of individual differences revealed that while people differ greatly in the strategies they approach the task with, almost every participant showed an "explore to exploit" shift over time.

Keywords: decision making; information search; explore-exploit dilemmas

"You got to know when to hold 'em, know when to fold 'em, know when to walk away – and know when to run."
– The Gambler, by Kenny Rogers

Introduction

At almost any point in time, the set of possible actions available to a person is extremely large. When ordering lunch at many sandwich delicatessens, for instance, several different types of bread, meat, salads and sauces lead to a combinatorial explosion in the number of potential sandwiches. When applying to a University, high school students have hundreds of possible institutions and degree structures to consider. In such situations, people need to select the best course of action from a very large number of possibilities, despite the fact that they lack the resources (time, effort, and money) to thoroughly investigate all aspects of the domain. As a result, the learner is faced with a classic "explore-exploit" dilemma: the value of some of the options is known to them from previous experience, but the value of other possibilities remain unknown. Thus, a higher-level decision is required: what proportion of resources should be allocated to the problem of searching for good options (exploration), and what proportion to taking advantage of them (exploitation)?

In a recent study, Shin and Ariely (2004) presented people with a limited version of this dilemma, which took the

following form. Participants were presented with a computerized display showing three doors, which could be opened by clicking the mouse. Once a door was opened, further clicks yielded rewards sampled randomly from a probability distribution not known to the participants. Each door had a different distribution of rewards (though, unbeknownst to participants, the distributions had identical expected values), so participants needed to investigate the three doors. However, since participants were only allowed a fixed number of clicks in the experiment, and the opening of one door caused all other doors to close, engaging in search was a costly behavior. Formally, this is an example of a three-armed bandit problem with switching costs, a problem that is well studied in statistics (e.g., Robbins, 1952) and has begun to be studied in the decision-making literature (e.g., Steyvers, Lee, & Wagenmakers, in press).

A novel innovation in the Shin and Ariely (2004) study was the fact that, in some conditions, the doors could shrink and vanish if left unopened for a sufficiently long time, removing that option for good. This matches the structure of many real-world problems: if one does not study mathematics in high school, for instance, the option of becoming a mathematician later in life eventually disappears. The results showed that when an option threatens to vanish, people were willing to expend resources to keep it viable, even though this is a mathematically suboptimal strategy for this task.

While switching costs and the threat of vanishing is an interesting extension to the standard multi-armed bandit problem, the experiment in Shin and Ariely (2004) has some shortcomings in terms of how well it maps onto the situations faced in real life.

First, in many (probably most) situations in life, people are faced with more than the three options presented to subjects in the experiment; as the number of options increases, it becomes increasingly difficult to preserve all possible future actions. However, given that reward-generating processes in real life also tend to be nonstationary (e.g., profits in farming alter as a function of climate), it would be unwise to follow the optimal "pick one and stick with it" strategy from the original Shin and Ariely (2004) experiment. Are people sensitive to the total number of options possible, and are they more
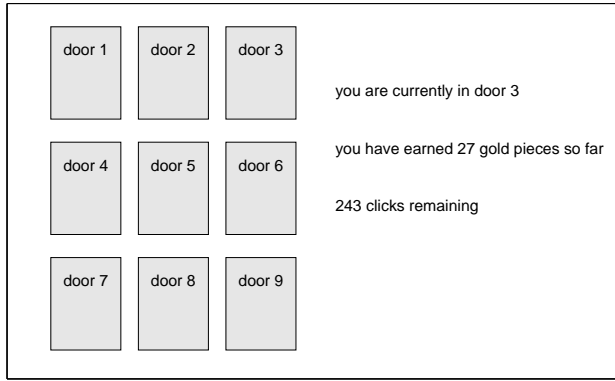
Figure 1: Participants encountered doors to nine rooms, each of which they could click on to open the door. Once open, clicking again yielded a variable amount of treasure (sampled from that room's payoff distribution). Participants were allocated a total of 300 clicks, and moving to new rooms cost them a click. Their goal was to maximize the treasure gained over the course of the experiment.

willing to "know when to walk away" when there are more than three available?

Second, in the Shin and Ariely (2004) experiment, participants were presented with doors (options) that were equal in expected value. In real life, options differ widely in their value to us: in choosing among future careers, "doctor" is an option that more people want to preserve than, say, "garbage collector." When options differ in value, the mathematically optimal strategy is no longer to "pick one and stick with it": rather, there must be sufficient exploration of all of the options to be able to identify which are the most valuable. Furthermore, it may also be easier to walk away from some of options when some are clearly better than others.

In this paper we present and analyse an experiment, modelled after the Shin and Ariely (2004) work, but with two major differences. First, participants were presented with 9 doors (options) rather than 3; and in some conditions, the doors differed in expected value. Our results suggest that, contrary to expectations, the mere existence of additional options did not make people more able to walk away from some of them. However, when options differed in expected value, they were much more willing to abandon options, and performed quite similar to the conditions in which options did not vanish with disuse. In general, these results indicate that people have trouble knowing when to walk away primarily when the options all appear quite similar. We discuss the implications of this finding.

Experimental Method

The Task

Following the method outlined by Shin and Ariely (2004), the overall structure of the experiment involved a sequential search task. In the original experiment, the alternatives were represented as three doors on a computer screen; in our ex-

Table 1: Distributions of payment in the nine doors by expected value condition. Following the approach used by Shin and Ariely (2004), the distributions have different variances and shape. In two conditions, the doors had equivalent expected value, while in the other two conditions some doors were objectively better choices than others.

# doors	distribution	mean	variance	range
Constant Expected Value:				
3	normal	3	2.25	0 to 7
3	normal	3	0.64	1 to 5
3	chi-square	3	10.00	−2 to 10
Variable Expected Value:				
1	normal	3	2.25	0 to 7
1	normal	3	0.64	1 to 5
1	chi-square	3	10.00	−2 to 10
1	normal	1	2.25	−2 to 5
1	normal	1	0.64	−1 to 3
1	chi-square	1	10.00	−4 to 8
1	normal	5	2.25	2 to 9
1	normal	5	0.64	3 to 7
1	chi-square	5	10.00	0 to 12

periment, there were nine. Each door was associated with a certain payoff distribution, which participants could claim by clicking on the door. As in many real-life search tasks, the search itself was costly; the task was to maximize earnings by spending their limited clicks wisely. Figure 1 shows a schematic depiction of the experiment as it appeared to our participants.

In the experiment (implemented in Visual Basic) each participant was presented with nine doors, each of a different, random, color. Clicking with the mouse on a door (door-click) allowed participants to open that door and enter the room; once inside, each click (room-click) would yield a different amount of payoff sampled randomly from that room's payoff distribution. It was not possible to claim treasure with a room-click before the door was open, and upon leaving a room to click on another, the first door closed automatically. This was explained to participants by telling them that each door hid a pile of treasure, which they could claim by first opening the door, and then clicking on it to grab handfuls of treasure.

Participants were informed that each door might contain different amounts of treasure, and that the size of each handful would vary; because some of the payoff distributions occasionally resulted in negative points, we mentioned that some doors might occasionally "steal" a participant's treasure when they reached in. Each person was allocated a total of 300 clicks – three times as many as in the Shin and Ariely (2004) experiment, since there were three times as many doors – and told that they would be paid in proportion to the amount of treasure they gathered. The number of clicks they had remaining, as well as their total treasure earned, was displayed prominently over the course of the experiment.
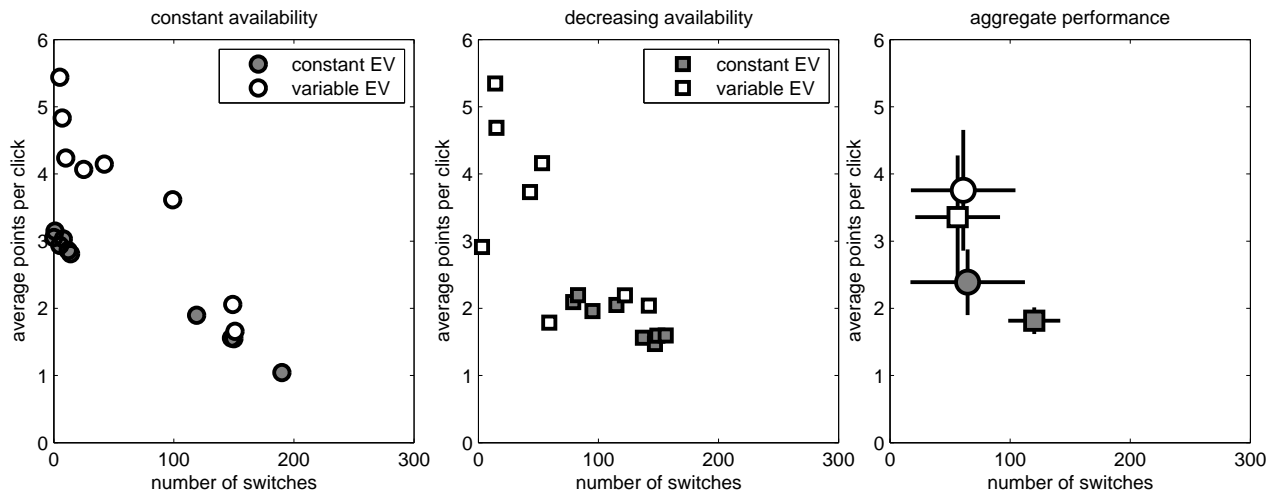
Figure 2: Total number of switches made during the experiment by each participant (x-axis), plotted against the average number of points won per click (y-axis). The panel on the left plots the data from participants in the constant availability conditions (circles), while the middle panel plots data from the decreasing availability conditions (squares). White markers denote participants in a variable expected value condition, while grey markers plot data from the constant expected value conditions. The panel on the right summarizes the data from all four conditions, plotting means and 95% confidence intervals.

The Manipulations

There were two main manipulations to this experiment. First, as in Shin and Ariely (2004), we varied the *availability* of the doors. In the CONSTANT AVAILABILITY condition, the doors did not disappear if they were not clicked on occasionally. In the DECREASING AVAILABILITY condition, the doors decreased in size with every click that was not on them, and disappeared altogether after 15 clicks; once a door disappeared, it was forevermore inaccessible. If a door was clicked on before disappearing entirely, it was immediately restored to its full size, and the process began again.

The second manipulation was to vary the distribution of the expected value of the payoffs assigned to individual doors. In three of the four original Shin and Ariely (2004) experiments, all of the doors always had the same expected payoff, although they varied in the variance and nature of the distribution (some were Normal distributions, and some were Chi-Squares).[1] In our experiment, in the CONSTANT EXPECTED VALUE condition, all of the doors had an expected value of 3 points per click, although (as before) they varied in the variance and nature of the distribution). By contrast, in the VARIABLE EXPECTED VALUE condition, three of the doors had an expected value of 1 point per click, three had an expected value of 3 points per click, and three had an expected value of 5 points per click. Participants were given the same instructions regardless of condition, and received no indication of which doors had higher expected value. Table 1 shows the distribution of payments in the nine doors across the different conditions. Conditions were crossed in a 2x2 factorial design: we denote the conditions as CA-CE (constant availability, constant expected value), CA-VE (constant availability, constant expected value), CA-VE (constant availability,

ity, variable expected value), DA-CE (decreasing availability, constant expected value) and DA-VE (decreasing availability, variable expected value).

The Participants

Thirty-two participants (7 male) aged 19-54, 8 in each condition, were recruited from a paid participant pool largely consisting of undergraduate psychology students and their acquaintances. The experiment took place as part of a series of three unrelated studies, which took approximately 1 hour to complete. Payment varied as a function of performance on this task, but averaged $12 each across the hour. This experiment took approximately 15-20 minutes.

Results

Overview

To give a broad overview of the data, Figure 2 plots the total number of door switches made by each participant against the average reward received per click. In the panels on the left and middle, each marker corresponds to a participant, and the type of marker indicates which experimental condition was involved. As one might expect, participants who switched frequently tended to have much poorer returns-per-click. However, as is also clear from the aggregate plots shown in the right-hand panel, there are some significant differences between conditions. In agreement with the general pattern of results presented by Shin and Ariely (2004), the decreasing availability conditions (squares) tended to produce more door-switching behavior than their corresponding constant availability conditions (circles). However, as predicted, the variable expected-value conditions tended to *decrease* the door switching behavior by a similar amount. Finally, as predicted, people were able to explore the rooms in such a way as

[1]In the fourth, the payoffs were different, but only slightly - ranging from 2.5 points to 3.5 points per door.
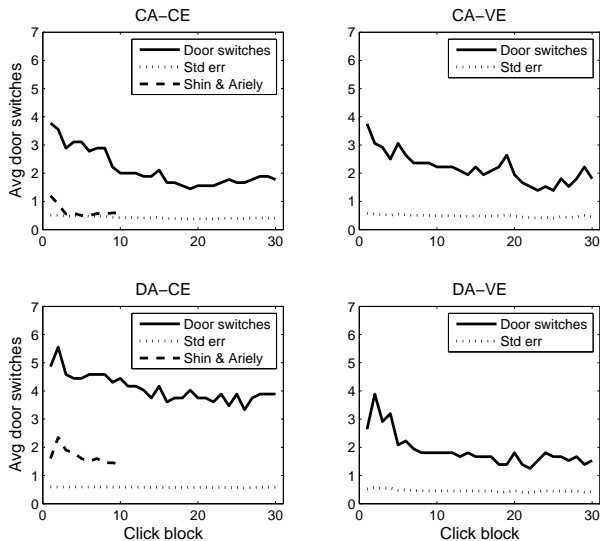
Figure 3: Average number of door switches within each block of 10 clicks in each condition. Average number of door switches is indicated with the solid line; the dotted line indicates standard error at each block. The *y* axis reflects the average number of door switches, and the *x* axis indicates the block number (30 in total). In the constant availability conditions, the Shin and Ariely (2004) data is plotted as a comparison (only extending 10 blocks because their participants were only given 100 clicks, not 300).
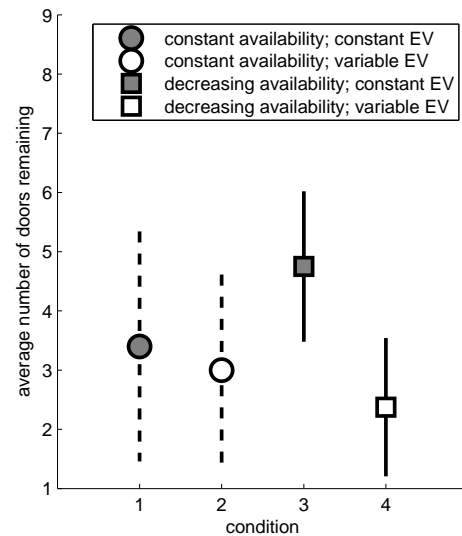

Figure 4: Average number of doors that remain viable at the end of the experiment, for the decreasing availability conditions (squares). For comparative purposes, an analogous measure is shown for the constant availability conditions (circles), which counts the number of unique doors clicked in the final 15 trials. Error bars depict 95% confidence intervals. Note that the variable EV conditions result in fewer doors being kept viable, since participants are generally able to detect bad options.

to find good doors and ignore bad ones: the variable expected value conditions involve higher rates of return.

Learning & Option Trimming

How did people's behavior change as the experiment progressed? We calculate the average number of door switches during each block of 10 clicks over the course of the 300-click experiment, as shown in Figure 3. Several patterns emerge.

In all of the conditions, the average number of door switches decreases over the course of the experiment. This is a replication of a result by Shin and Ariely (2004), and is quite sensible, since it reflects the change in the nature of the explore/exploit tradeoff as the experiment progresses. At the beginning of the experiment, participants have little to no information about the nature of the payoff distributions of each door, and thus it is more in their interest to explore each of the doors, rather than to exploit the ones they know. As the experiment progresses, the expected benefit to exploring decreases (since the experiment will end after 300 clicks), and the need for exploring also decreases (since the participants already have some information about the payment distributions of the doors).

Despite this general trend, participants in the DA-CE condition showed little decrease in the amount of door switching per block, especially relative to participants in the other conditions. Even in the final block, when keeping options open is no longer even remotely sensible, participants switched on average four times – nearly three times as many as the participants in the Shin and Ariely (2004) experiment, and approx-

imately twice as many times as the participants in our other conditions. This result came as a surprise because we had hypothesized that if the number of options was increased significantly beyond three, they would be more willing to walk away from at least a few of them. However, switching nearly three times as often is what one would expect if there were three times as many options and one's willingness to walk away was almost the same. Thus, contrary to expectation, the number of options *per se* did not substantially affect participants' willingness to let some options go.

The variable expected-value environments pose an explore-exploit dilemma for participants because discovering good options requires engaging in some door-switching behavior. By contrast, within the constant expected-value environments – as in the Shin and Ariely (2004) experiments – all options are equally good. However, the participants have no way of knowing which kind of environment they are in *a priori*. Decreasing availability sharpens this problem considerably, since participants must engage in a lot of effort to preserve the ability to explore. Figure 3 shows that differences in the expected value of the distributions of the doors makes a difference in behavior. As we have seen, participants in the DA-CE condition continued to switch relatively often, but those in the DA-VE condition switched much more. In fact, even when availability is constant, variable expected value conditions switched less often than when all of the doors were equally valuable. Figure 4 shows this as well: people tended to keep 1 to 3 options in the variable expected value condi-

tions and 4 to 6 in the constant expected value condition. To facilitate comparison, we also plot the analogous measure for the constant availability conditions: the number of different doors clicked on in the last 15 trials of the experiment.

Individual Differences

At this stage, we have evidence to suggest that the basic effects discussed by Shin and Ariely (2004) replicate: people do work to preserve their exploration options, and are willing to take financial losses to do so. That being said, it is clear that in general people do not attempt to preserve all options, with few people keeping more than half of the doors viable. Moreover, we generalize the findings to show that people use this exploration ability to improve their choices in inhomogeneous environments, and show strong signs of learning during the process. However, one extremely important issue that remains unexplored is the matter of individual differences, since it is well-known (e.g. Estes, 1956) that aggregate measures can sometimes be highly misleading in the presence of individual variation.

With that in mind, Figures 5–8 attempt to provide an explicit representation of the entire raw data set. Each of these figures displays the data from one of the four experimental conditions, with each participant's data corresponding to one panel. The way data are displayed in each panel is as follows: the horizontal axis corresponds to trial number (running from 1 to 300), and the vertical axis corresponds to door number (running from 1 to 9). Doors with higher expected value are plotted at the top of the figure; if two doors have the same expected value, the higher variance one is plotted above; and doors with identical distributions are plotted in a random order. Each black segment corresponds to a door click – so if the first door clicked on was the best one, there is a slim black line in the top left corner of the plot. If a participant clicked on the same door on all 300 trials, the panel would show a single black strip. In contrast, if they switched between all 9 doors very frequently, the display would look like a chaotic scatterplot.

Inspection of these figures illustrates that there are very strong individual differences in this task, and moreover that averaged performance measures act to hide the fundamental strategies that people follow in solving the task. Broadly speaking, almost every participant shows an "explore to exploit" shift over time regardless of experimental conditions, with a gradual winnowing of the doors chosen. Given that a large part of the task involves an information search process – learning which doors are good, which are bad, and which ones are equivalent, this general pattern is anything but irrational: how else can a learner discover good options except by exploring (switching doors)?

The interesting variation across people regards the rate of the switch from exploration to exploitation patterns. In the constant availability/constant EV condition (Figure 5) there is a stark difference between two groups of people. Six of the eight participants hone in on an option easily, while two continue to explore for the entire 300 trials. For the constant availability/variable EV condition (Figure 6) seven participants find one good option, though in one case a low-level exploration continues through the experiment.

Turning to the two decreasing availability conditions, the same general strategies seem to be in play, but with the added complication that people are required to switch more frequently in order to preserve the ability to explore. In the decreasing availability/constant EV condition, this is particularly problematic for people, since there is in fact no useful structure to uncover, but it is not easily searchable due to the enforced switching during search. As a result, the various response patterns displayed in Figure 7 do show a slow trimming of options, but most participants are not able to complete the process during the allotted time. In contrast, in the decreasing availability/variable EV condition (Figure 8), there is a genuine structure to be discovered, and so even in spite of the enforced switching, all participants are able to sensibly trim the options; in doing so, there is a strong bias to trim the bad ones. Most of the black marks migrate to the top of the panels across trials.

Conclusion

As in Shin and Ariely (2004), we found that people paid certain costs in order to keep options open, even when there were many more options (nine rather than three). However, our results support a more subtle and interesting explanation of this behavior, because they show that people were far less willing to keep many options open when some were clearly better than others. This suggests that the "keeping options open" problem is acute only when the situation is structured so that it is difficult to explore sufficiently thoroughly to tell them apart. Finally, an analysis of individual differences indicates substantial variety in the strategies people use; it is not the case that everyone opts for irrational strategies, even in the most irrationality-inducing of situations.

References

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, *53*, 134-140.

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, *55*, 527-535.

Shin, J., & Ariely, D. (2004). Keeping doors open: The effect of unavailability on incentives to keep options viable. *Management Science*, *50*, 575-586.

Steyvers, M., Lee, M., & Wagenmakers, E. (in press). A Bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology*.
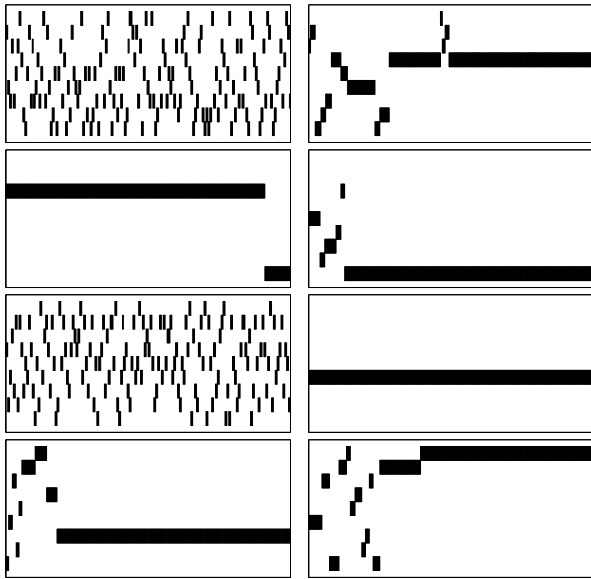
Figure 5: Raw data for the participants in the constant availability–constant expected value (CA-CE) condition. There are two qualitatively different strategies: six participants engaged in a brief exploration to find a good option but spent the majority of their clicks on one door. The remaining two participants appeared to remain in "explore" mode for most of the experiment. (See main text for a detailed description of the plots)
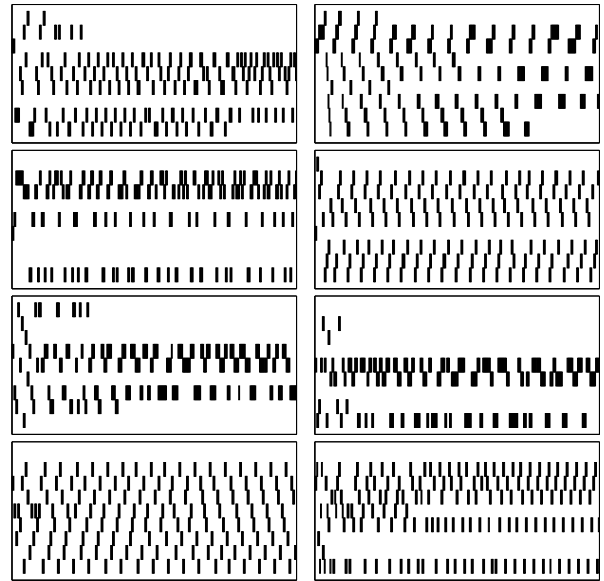

Figure 7: Raw data for the participants in the decreasing availability–constant expected value (DA-CE) condition. As expected given the Shin and Ariely (2004) data, many more participants engage in effort to keep more options viable.
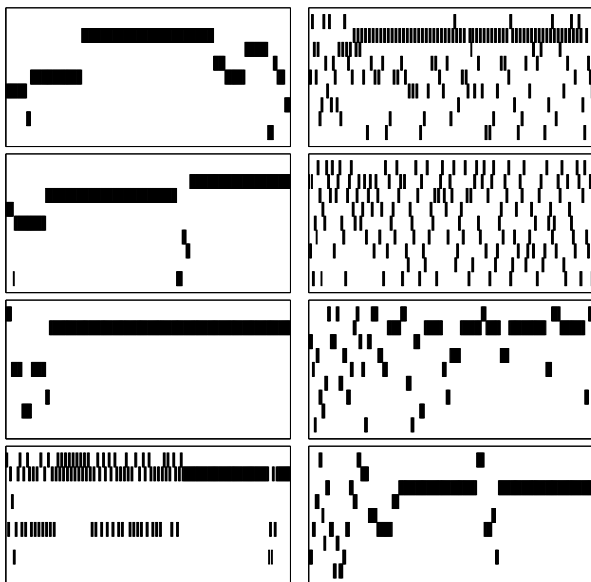

Figure 6: Raw data for the participants in the constant availability–variable expected value (CA-VE) condition. While one participant stays in pure explore mode for the whole experiment, most use their exploration to find good options (toward the top of the figure).
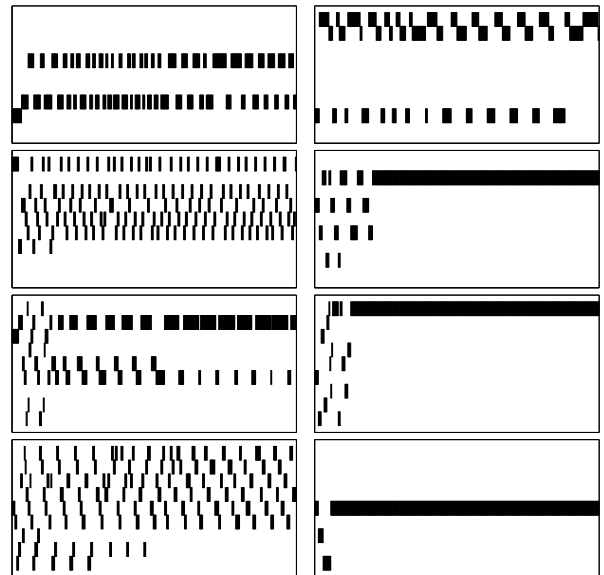

Figure 8: Raw data for the participants in the decreasing availability–variable expected value (DA-VE) condition.