Reward, Risk and Ambiguity in Human Exploration: A Wheel of Fortune Task

Andra Geana
Department of Psychology
Princeton University
ageana@princeton.edu

Robert C. Wilson
Department of Psychology
Princeton University
rcw2@princeton.edu

Jonathan D. Cohen
Department of Psychology
Princeton University
jdc@princeton.edu

Abstract

In realistic environments, organisms are frequently faced with multiple resource alternatives, and must balance the tradeoff between pursuing the known options (exploitation), and searching the environment for unknown opportunities (exploration). Exploration can be most beneficial in the presence of environmental uncertainty - when the range and benefits of all reward options are not fully known, exploration can lead to the discovery of new, better resources and an ultimately higher overall reward. However, uncertainty can take many forms, and it is unclear how different types of uncertainty impact people's exploratory behaviour. We used a 'wheel of fortune' task to separate two well-established sources of uncertainty: risk (when outcomes are stochastic, but the probabilities of outcomes are known) and ambiguity (when the probabilities and/or the outcomes are unknown), and examine how they impact exploration. The results suggest that the presence of ambiguity in the environment drives people to explore in order to acquire more information and reduce the ambiguity. Conversely, a higher risk level in the environment increases exploration by increasing decision noise and making people less sensitive to the reward values of the available options. We examined these effects under two different decision horizons, and found that ambiguity-, and not risk-related exploration increases with decision horizon. These findings imply that different sources of uncertainty impact exploration differently, and may shed light on the mechanisms behind two distinguishable types of exploration that have been previously identified: random (characterized by an increase in decision noise) and directed (information-seeking) exploration.

Keywords: explore-exploit, decisions under uncertainty, risk, ambiguity

Acknowledgements

1   Introduction

Imagine you are driving home from work during evening rush hour. Just as you are about to turn onto the freeway, you see a side road that looks like it might avoid the rush hour traffic jam, and potentially get you home faster. Do you continue onto your usual freeway route, or do you take the unknown side road which might reduce your commute time, or might cause you to get lost? This is an example of the "exploration - exploitation dilemma": the trade-off between choosing a certain resource alternative (in our example, the freeway) and searching the environment for other options (the side road), that may or may not be better.

This kind of scenario is not restricted to humans. A foraging animal, such as a bee for instance, might choose between searching for nectar in a nearby food patch, or searching farther away for a potentially better patch [1]. Whether we are discussing animals foraging for food, or people driving home from work, the question is the same. With limited time and energy to spend, how should an organism best choose between exploiting what is known and exploring less known options? This open question is encountered across fields, from cognitive neuroscience [2], to animal cognition [3], to ecology [4], economics[5], and reinforcement learning [6]. Nevertheless, due to the complex nature of realistic environments, the precise cognitive and neural mechanisms that drive exploration have not yet been fully discovered.

Psychological and economic theories identify two different types of exploration: random - an increase in decision noise which makes people less sensitive to the values of the available options -, and directed exploration - a result of intentionally sampling unknown options to acquire information. The exact mechanisms behind these two types of exploration, as well as the factors that might lead a decision-maker to choose one over the other, are not known. However, evidence suggests that the decision to explore is related to environmental uncertainty - numerous studies found that uncertainty regulates exploratory behaviour [2,7], and relates to the neural underpinnings of the exploration - exploitation dilemma [8,9].

Uncertainty is a broad term, and the literature has differentiated at least two separable concepts contained within it. Expected uncertainty, or risk, is associated with choices where the decision-maker knows the possible outcomes, and their probabilities. The uncertainty stems from the stochastic nature of the choice, which makes it impossible to predict the exact outcome. Unexpected uncertainty, or ambiguity, appears when the decision-maker does not know the probabilities, and may not even know all of the multiple possible outcomes [10, 11]. Ambiguity makes it impossible to directly calculate the expect reward associated with a choice, because the outcome probabilities are unknown.

It is likely that people incorporate these two sources of uncertainty differently into their decision processes, which may bias them either toward exploration or away from it. Whether the current choice is risky or ambiguous is a crucial difference for deciding a strategy. Several studies argue that this distinction is important enough for the brain to maintain separate representations of risk and ambiguity [9, 12], and that different neural substrates respond to risk (orbito-frontal cortex, striatum) and ambiguity (lateral prefrontal cortex), and that different neurotransmitters are released in the presence of risk (acetylcholine) and ambiguity (norepinephrine).

Despite this important distinction, there have not been many attempts to establish how risk and ambiguity independently influence the degree to which people explore, or whether they impact random and directed exploration differently. Many studies that examine the neural and behavioural responses to risk and ambiguity pay little attention to the exploration framework, and instead search for evidence that people have separate risk- and ambiguity preferences [10, 12]. Studies investigating people's exploration strategies, on the other hand, often focus heavily on the effect of relative reward magnitude, and either disregard the ambiguity framework, or pool risk and ambiguity together under the broader uncertainty heading [7, 13]. To bridge this gap, we designed a task that allowed for a clear separation and empirical manipulation of risk and ambiguity, and examined how these two factors impact exploratory behaviour.

2   Methods

2.1. Behavioural Task

Participants played multiple games of a sequential decision task that required them to choose between two virtual wheels of fortune (figure 1A). The rewards (number of points) that each wheel could pay out were written on the slices. One wheel, referred to as the ambiguous wheel, always had some of the reward values covered with a question mark. This design made it possible to clearly operationalize reward, risk and ambiguity, and allowed each of the three parameters to be manipulated individually, while the other two were kept fixed.

The mean reward of each wheel was given by the mean of the numbers on the slices. Risk in this task was operationalized as the variance of the numbers on the wheel, and could take five different values that ranged from no risk (all numbers were the same) to high risk (the variance of the numbers was high). Ambiguity was operationalized as the number of question marks on the wheel, and could be either zero (all the wheel slices remained uncovered) or four (four slices were covered by the question mark).

The decision horizon was manipulated as the length of a game: each game consisted of either one choice between two wheels (horizon 1), or five sequential choices (horizon 5). The number of choices left in a game was always displayed on the screen (figure 1A). The wheels remained the same within a game, but they changed between games. The changes were always signalled.
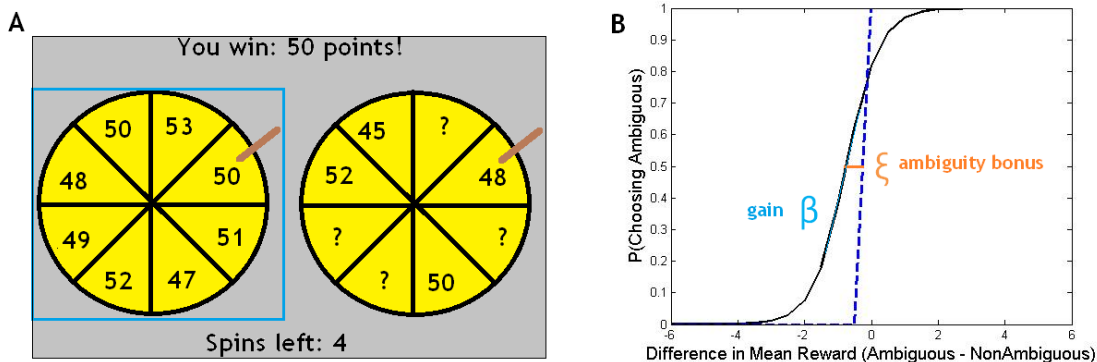

Figure 1:A. Wheel of fortune task. The blue rectangle around the left wheel show that it was the chosen wheel. The pin stopped on the slice with the number '50', earning the participant 50 points for that spin. B. Example choice curve generated using model. The black curve is described by the parameters β (slope) and ξ (centre), which in the choice model correspond to gain (inverse decision noise) and ambiguity bonus. The dotted blue line shows what the choice curve would be if the participant chose exclusively based on mean reward, and had no decision noise.

Each trial began with the two wheels being displayed on the screen. When participants made their choice, the chosen wheel spun a random number of times before landing on a slice, and the computer displayed the number on the slice as the reward earned for that trial. If that slice had been covered, the computer generated a value for it based on a Gaussian distribution that maintained the variance of the rewards on the wheel. The reward was displayed on the corresponding slice, and stayed visible for the remaining trials in that game.

2.2. Model and Analysis

We modelled participants' behaviour by assuming that they assign a value $V(W_i)$ to each wheel W_i , and then make their choices using a softmax function that assigns a choice probability $P(W_i)$ to each option, according to:

$$P(W_i) = \frac{1}{e^{\beta \Delta V(W_i)}} \tag{1}$$

where β is the gain parameter which determines how sensitive the choice probabilities are to the values. The value of an option was modelled as the expected value of the rewards (both the known - the uncovered slices in the wheel of fortune task - and the unknown, or the covered slices). In addition to mean reward, an ambiguity bonus parameter, ξ, was added to the value computation, to quantify how much reducing the ambiguity about that particular option is worth in units of value.

$$V(W_i) = EV[uncovered\ slices] + EV[covered\ slices] + ξ \tag{2}$$

The β and ξ parameters define a choice curve for each participant (figure 1B), which shows, for any given difference in mean reward between the two wheels, how frequently the participant chose the ambiguous wheel. The gain parameter in the choice model was used to quantify decision noise - a measure of random exploration - in the participants' behaviour. The ambiguity bonus parameter was used as a measure of information-seeking. Exploration was defined as choosing the ambiguous wheel.

3   Results

Participants' choice curves (see figure 2A) show that they were generally sensitive to the mean reward values of the two wheels, with their probability of choosing the ambiguous wheel increasing as the difference in means between the two wheels increased. Overall average exploration on the first trial of a game was significantly higher in the horizon 5 condition than in the horizon 1 condition ($P_{exp,horizon1} = 0.38 \pm 0.15$, $P_{exp,horizon5} = 0.56 \pm 0.19$).

There was a significant main effect of risk level ($F(5,24) = 5.62$, $p<0.01$) on the fit decision noise parameter, with the participants' decision noise higher in high-risk conditions (fig 2C). This effect is also visible as the choice curves becoming flatter at increased risk level (fig. 2A). There was also a main effect of decision horizon ($F(2,24) = 3.12$, $p<0.05$), with decision noise higher in the horizon 5 condition; the interaction was not significant.
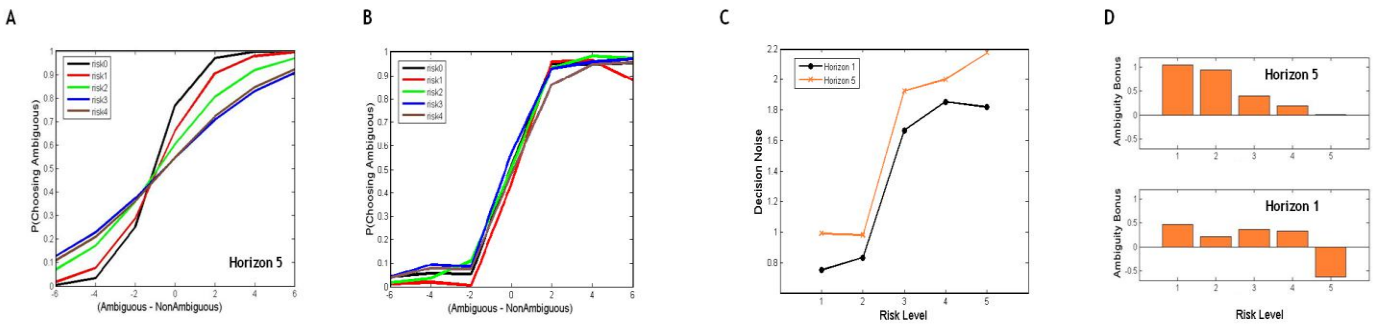

Figure 2: A. Choice curves showing the probability of choosing the ambiguous wheel as a function of the difference in mean rewards, across risk levels, for decision horizon 5. B. Choice curves from a control task in which both wheels were non-ambiguous superpose. C. Decision noise increases as a function of risk, in both horizon 1 (black) and horizon 5 (orange). D. Ambiguity bonus decreases (curves move further to the right) as a function of risk level in horizon 5 (top), not in horizon 1 (bottom panel).

For the fit ambiguity bonus, there was a main effect of horizon, with ambiguity-seeking significantly higher for the longer horizon, a main effect of risk level, with ambiguity-seeking significantly higher in lower risk conditions, and an interaction: the decreasing trend in ambiguity bonus values was present only in the horizon 5 condition (figure 2D).

4   Discussion

These results offer evidence for the separate roles of risk and ambiguity in exploratory decision-making. Higher risk levels were associated with a lower gain parameter in the choice function (figure 2A,C), suggesting that risk might influence the strength of people's preference for the more rewarding option by modulating decision noise. A control task run with no ambiguous options showed no effect of risk level, suggesting that it is not simply a computational load that causes increased decision noise in higher-risk conditions - rather, the presence of ambiguity (and thus the potential benefit of exploration) seems to be necessary for risk to affect exploration (figure 2B).

A longer decision horizon was associated with a higher ambiguity bonus parameter, suggesting that the degree to which ambiguity biases value increases when there is more opportunity to use the acquired information to make better choices. These results are consistent with the idea that the degree to which ambiguity drives exploration is strongly correlated with the value of information [2, 7].

Interestingly, the ambiguity bonus decreased with risk level - suggesting that people become less ambiguity seeking when the environment becomes more variable. This effect, however, is only present in the long-horizon condition, when information acquired by exploring can be usefully employed in future choices. This, along with the increasing decision noise in high risk conditions, indicates a potential trade-off between random and directed exploration, with more risky environments driving increased randomness and decreased information-seeking, while more ambiguous environments bias people toward information-seeking. The current findings suggest that both risk and ambiguity should be taken into account when modelling exploratory behaviour, and future work is required to describe a more precise mechanism of how these two factors interact to affect exploration.

References

[1] C.R. Gallistel. *The Organization of Learning*. Cambridge, MA: Bradford Books/MIT Press, 1990

[2] J.D. Cohen, S.M. McClure & A.J. Yu. 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: Biological Sciences. 362*, 933–942, 2007

[3] C.R. Gallistel, A.P. King, D. Gottlieb, F. Balci. Is matching innate? *J. Exp. Anal. Beh.* 87: 161–199, 2007

[4] T. Caraco. On foraging time allocation in a stochastic environment. *Ecology , 61*, 119-128, 1980.

[5] J.S. Banks, & R.K. Sundaram. Switching costs and the Gittins index. *Econometrica: Journal of the Economical Society, 62*, 687–694, 1994

[6] L.P. Kaelbling, M.L. Littman, and A. W. Moore. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4, 1996.

[7] T.E.J. Behrens, M.W. Woolrich, M.E. Walton, M.F.S. Rushworth. Learning the value of information in an uncertain world. *Nature Neuroscience* 10: 1214 - 1221, 2007

[8] G. Aston-Jones & J.D. Cohen. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu Rev Neurosci*, 28:403–50, 2005.

[9] A.J. Yu & P. Dayan. Uncertainty, neuromodulation and attention. *Neuron* 46, 681–692, 2005

[10] M. Hsu, M. Bhatt, R. Adolphs, D. Tranel, C.F. Camerer. Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310, 1680–1683, 2005

[11] Bach, D.R., Hulme, O., Penny, W.D. & Dolan, R.J. (2011) The known unknowns: neural representation of second-order uncertainty, and ambiguity. *J Neurosci*. 31, 4811 - 4820

[12] Huettel, S.A., Stowe, C.J., Gordon, E.M., Warner, B.T. & Platt, M. L. (2006) Neural signatures of economics preferences for risk and ambiguity. *Neuron* 49, 765 - 775

[13] N.D. Daw, J.P. O'Doherty, P. Dayan, B. Seymour & R.J. Dolan. Cortical substrates for exploratory decisions in humans. *Nature*, 441(7095):876–9, 2006.