# Terminating Information Search

## Stefan Appelhoff

## November 9, 2016

## Contents

## List of Figures

## 1 Introduction

As agents in different environments, humans often find themselves in situations that require decisions based on limited knowledge about the available choice options. The topic of the present project is the process, how humans selectively sample choice options to arrive at a more refined knowledge about the properties of the options. Specifically, the interest lies in self-directed information search in sequential decision making problems. A short example illustrates the topic further:

Imagine that your laptop is broken and you want to acquire a new one. In the process, you may sequentially sample several available models and their respective properties such as prices, specifications, and reviews. To avoid indefinite browsing, you will need to terminate your search at some point and select a new laptop - most likely without having obtained full certainty about the best option.

For the project described here, the aforementioned example will be transformed into an experimental paradigm that allows for the examination of behavioral and neural aspects of choice in a controled manner. Similar to the real life example, the experimental paradigm exhibits the following characteristics:

- There is an environment in which several options exist that will yield outcomes upon selection.

- There exists limited knowledge about the options and the frequencies of their associated outcomes.

- There will be the opportunity to obtain information about the options through a sampling of the options.

- The overall goal is to maximize the positive outcomes obtained from the options in the long run.

In the paradigm, participants will make sequential decisions about which option to sample. In these trial to trial decisions, the question of interest to this project lies in how participants set their termination criteria. I.e., when do participants decide to terminate the information search in one option and switch to search another option? When do participants decide to stop sampling altogether and decide for one final option to be optimal based on current knowledge? The remainder of this document will be a concise formulation of the research questions followed by an outline of the experimental paradigm. Finally, the implementation of the experimental paradigm on a computer program will be described.

## 2 Research Questions

1. How are sampling efforts sequentially allocated to the different options?

2. How do we arrive at a decision to terminate the information search?

3. What is the quantitative link between neural signals obtained by EEG with the behavioral data of participants performing the experimental paradigm.

# 3 The Experimental Paradigm

The proposed experiment will be based on a combination of two different paradigms that are well known in the literature.

1. The n-armed Bandit Paradigm

2. The Sampling Paradigm

In the following, these two paradigms will be descibed in detail. Figure 1 provides an overview of the paradigms. Afterwards, the combination of both paradigms for the present experiment will be outlined.

## 3.1 The n-armed Bandit Paradigm

The n-armed bandit paradigm was first formalized by Robbins in 1952 [5]. It has since been the subject of interest in a variety of fields, see [1] for a review. Generally, participants are presented with n action that represent initially unknown outcome distributions with a set number of outcomes. Upon selection of an action, an outcome is yielded from the distribution and presented to the participant as a feedback. With this gained knowledge, the participant may continue with selecting the same or a different action.

The bandit paradigm offers a concise framework to study the exploration exploitation tradeoff common to the field of reinforcement learning. Here, exploration represents the goal to gather information about actions and their associated outcomes, while exploitation represents the goal to make use of the acquired information. A player who only explores will never exploit upon the accumulated knowledge, while a player who only exploits will most likely select suboptimal actions based on the inadequate knowledge about all actions. Hertwig & Erev label the bandit paradigm as Partial-Feedback-Paradigm [4].

## 3.2 The Sampling Paradigm

The sampling paradigm is similar to the bandit paradigm. The main difference is that exploration and exploitation are separated. This is achieved through permitting a participant to sample all actions without consequences for as long as the participant wishes. This stage represents purely the exploration, because the outcomes do not count towards a payoff. Once the participant decides to terminate the sampling process, there is one last opportunity to select any action. For this last choice, the outcome will affect the payoff and thus, this stage represents purely the exploitation. For more information, see [2–4]

## 3.3 Combination of the Two Paradigms
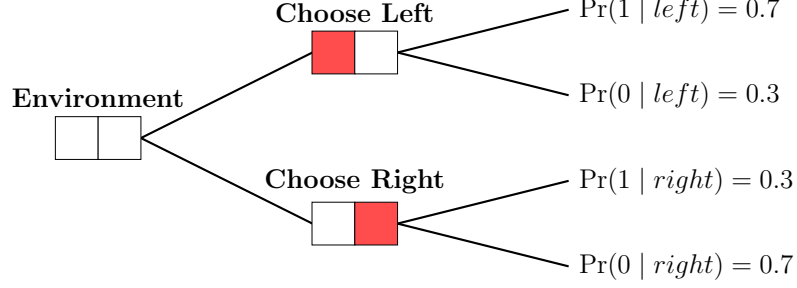
Need a figure here for the 2x2 design

# 4 Computerized Implementation of the Experiment
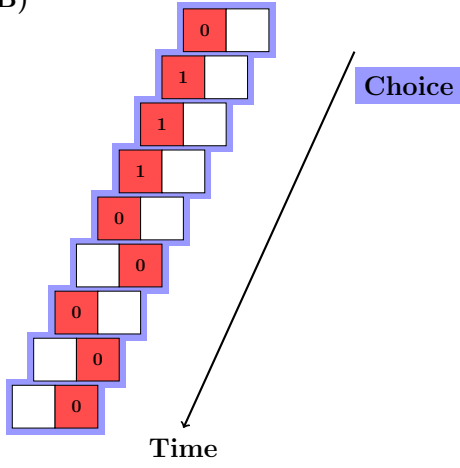
describe the computer code here

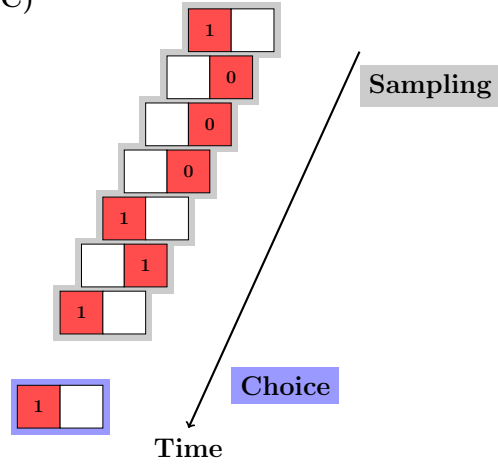need a figure for the flow ???

The stimuli used can be seen in Figure 3

**A)** Depiction of a two-armed bandit. Note that the number of options in the environment can be extended arbitrarily to form an n-armed bandit. Each option contains its own probability mass function (pmf) of outcomes. These pmfs are unknown to a subject performing the bandit task. **B)** A typical run of trials in a bandit task. A subject sequentially chooses among options and is provided with feedback. Each trial represents a exploration-exploitation problem. **C)** A typical run of trials in the sampling paradigm. A subject is allowed to sample the available options without an impact of the outcomes on the final payoff. At some point, the subject can decide to terminate sampling and choose one of the options, which will impact the final payoff. In the sampling paradigm, exploration and exploitation are thus distinct.

# Paradigm Condition

|  | Sampling Paradigm | Bandit Paradigm |
|---|---|---|
| **Active** | Active Sampling | Active Bandit |
| **Passive** | Passive Sampling | Passive Bandit |

**Interaction Condition**

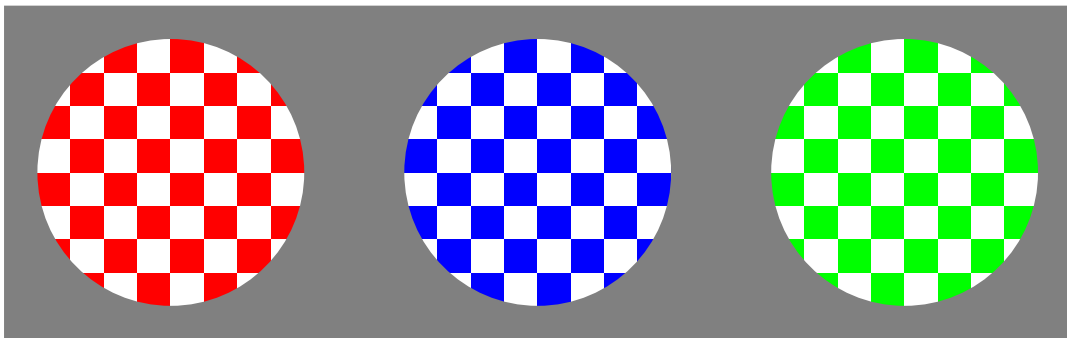**Figure 2:** The 2x2 factorial design of the experiment.



**Figure 3:** Three stimuli against a background as used in the experiment . The red and blue stimuli represent either a win or lose outcome. The green stimulus represents a distractor stimulus.

# 5 References

[1]  Donald A Berry and Bert Fristedt. *Bandit problems: sequential allocation of experiments (Monographs on statistics and applied probability)*. Springer, 1985.

[2]  Ido Erev and Alvin E. Roth. "Maximization, Learning and Economic Behavior". In: *Proceedings of the National Academy Sciences of the USA* 111.3 (2014), pp. 10818–10825. DOI: `10.1037/e573552014-045`. URL: `http://dx.doi.org/10.1037/e573552014-045`.

[3]  Ralph Hertwig. "The psychology and rationality of decisions from experience". In: *Synthese* 187.1 (Oct. 2011), pp. 269–292. DOI: `10.1007/s11229-011-0024-4`. URL: `http://dx.doi.org/10.1007/s11229-011-0024-4`.

[4]  Ralph Hertwig and Ido Erev. "The description–experience gap in risky choice". In: *Trends in Cognitive Sciences* 13.12 (Dec. 2009), pp. 517–523. DOI: `10.1016/j.tics.2009.09.004`. URL: `http://dx.doi.org/10.1016/j.tics.2009.09.004`.

[5]  Herbert Robbins. "Some Aspects Of The Sequential Design Of Experiments". In: *Bulletin of the American Mathematical Society* 58 (Sept. 1952), pp. 527–535.