Whom to Observe?

Bøg, Martin

Erasmus University Rotterdam

1 December 2006

Whom to Observe?*

Martin Bøg†
Erasmus University Rotterdam

May 14, 2008

*This article is based on Chapter 2 of my dissertation. I am grateful to Tilman Börgers for supervision. The paper has benefitted from inputs from Erik Eyster, Sanjeev Goyal, Cloda Jenkins, Nicola Persico, Pedro Rey Biel and Peyton Young. Financial support from ESRC (grant R42200134547), ELSE, Carl Christiansen og Hustrus Legat and Konsul Axel Nielsens Mindelegat is gratefully acknowledged.

†Martin Bøg, Department of Economics, Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. Tel.: +31 (0)104081479. Email: mbog@few.eur.nl

Abstract

This paper considers the problem of a decision maker who is faced with a dynamic decision problem with several alternatives, and additionally can engage in prior consultation on one of the alternatives. Information received from others is coarse. When consulting on an alternative that the decision maker is pre-disposed to, she either consults someone that shares precisely her convictions, or she consults someone who is more "picky" than herself. Optimality depends on the attractiveness of alternatives; when another alternative becomes sufficiently attractive the decision maker prefers a picky contact. When the decision maker consults on a lower ranked alternative, optimal consulting depends non-monotonically on the value of the alternative she is pre-disposed to. For high and low values of the pre-disposed alternative she prefers to consult someone with her own convictions, but for medium values she prefers to consult a picky contact. Finally a decision maker may prefer to consult on a lower ranked alternative.

JEL: C73, D83, D85

1 Introduction

Imagine having to choose between two different products, say A(jax) and B(otox), whose characteristics are unknown to you. The products are such that the short-term expected gain from each of the products are negative, but possibly, given suitable experiences, are worthwhile in the longer-term vis-a-vis the product, S(afe), you are currently using. Your initial assessment is that A is more likely to generate long run joy than product B, so that if you had to choose between A and B you would choose A. Suppose now that prior to making your choice you have the opportunity to consult someone who has prior experience with one of the new products (A or B). Whom should you consult? The answer to this question has two dimensions. First, should you consult someone who has experience with the product you already are pre-disposed to, or should you consult on the less likely alternative? Second, given that you consult on an alternative should you consult someone who likes the same characteristics of the product as you do, or someone who likes different characteristics?

An important assumption for the main results of this paper is that the information we receive from others is coarse. If consulting others perfectly reveals the characteristics of the product then the recommender's preferences over product characteristics are immaterial, since his description of the characteristics reveals to you how much pleasure you will get from the product. The paper therefore fits in the literature on how decision makers deal with the particular problem of information coarsening. Meyer (1991) shows how organizations that have to deal with fundamental coarsening in communication in promotion decisions optimally biases a sequential promotion contest to maximize informational content. Calvert (1985) and Suen (2004) show that a decision maker faced with a binary choice between two alternative finds it optimal to seek information from contacts that are pre-disposed to the same alternative that the decision maker is already pre-disposed to.

The paper contains three main results. First, suppose that the decision maker must consult on the alternative she is predisposed to (say alternative A for concreteness). We show that unless the alternative B is relatively attractive compared to alternative A, then the decision maker prefers to consult someone like herself in the sense that she and the contact likes and dislikes the same product characteristics. That is preferential attachment is given to contacts that are like

the decision maker herself. When alternative B becomes sufficiently attractive, in the sense that even "good" news about product characteristics, relative to the status-quo product S, is not enough to fully convince the decision maker about option A, then she prefers to consult a biased contact. An optimal contact is biased in a specific way. In particular she is more "picky" than the decision maker herself, in the sense that there are outcomes that the decision maker likes, that the contact does not like, but not vice-versa. The intuition for why contacts are biased in this specific way is that it allows the decision maker to distinguish between the truly good news and the not so good news. This comes at a cost: the decision maker may miss out on outcomes that improves on the status quo. But this cost is outweighed by the potential benefits of trying out alternative B on her own.

Second, suppose that the decision maker is restricted to consulting on alternative B, rather than alternative A that she is predisposed to. In this case whether it is optimal to consult a biased contact depends on the value of the outside option (alternative A) in a non-monotonic way. In particular whenever alternative A is either ex-ante relatively unattractive or very attractive then it is optimal to consult a contact who likes the same outcomes as the decision maker. For intermediate values of A it becomes optimal to consult a biased contact. The intuition for the result is as follows. When A is of low value any good news from alternative B is sufficient to convince the decision maker to stick with B. In this case information on B has a first-order effect on payoffs, even entertaining the possibility that the news on B is not so good. However as the value of alternative A increases, by not distinguishing between the good and the not so good news on B the decision maker bears a cost of trying arm B only to realize that the generated outcome was worse than what she hoped for. Alternatively she may choose to ignore the information for the moment and try out A, knowing that if A turns out to be less than pleasing she can use the information on B in a future period. In this case information on B has a second order effect on payoffs. If the decision maker instead consults a contact with an optimal bias she can avoid this (at the cost of missing out on the not so good news on B), in order to learn only when B is truly attractive.

The third main result concern the choice of which alternative to consult on. Should the decision maker consult on the alternative which she is pre-disposed to or should she consult on the less attractive alternative? We show that a decision

maker prefers to consult a contact that tries out alternative B provided that alternative B is sufficiently "risky". That is the reason why the decision maker prefers A over B is that B contains some very bad outcomes. By consulting on B the decision maker can avoid experiencing such outcomes.

Many if not most decisions that we make in daily economic life involve an element of learning. We learn what consumer goods we like by trying out new products, what projects to invest in, which jobs we like, where we like to spend our holidays, etc. Rarely is this learning based only on private experiences, rather it typically takes place in the context of a social network where we observe the experiences of and/or receive advice from friends, neighbors, colleagues, and family members (Arndt 1967, Bandiera and Rasul 2006, Foster and Rosenzweig 1995). At least since the 1920's sociologist have been studying the composition of social networks (McPherson, Smith-Lovin, and Cook 2001). One of the most robust key findings is that links in social networks, whether these are networks for social support, friends or advice networks, tend to be organized by the "homophily principle". This principle states that people who share similar attributes such as race, education, age, line of work, etc. are relatively more likely to form links with one another than they are to form links with people who are dis-similar[1]. Currarini, Jackson, and Pin (2007) construct a model where agents form multiple friendships within a setting of distinct groups of agents. Agents value both similar and dis-similar friends. Their model displays homophily[2]. The present paper contributes to the literature on social networks, by providing a micro-foundation for preferential attachment in networks. In particular the idea of this paper is that we form relationships with other people for informational reasons, namely for the advice they can potentially give us, or for the experiences they will expose us to. By providing such a micro-foundation we are able to more precisely pin down when preferential attachment are given to others who are similar to oneself. It also provides novel insights into when preferential attachment is given to people different from oneself, and we are able to say precisely what dissimilarities attract.

[1]Lazarsfeld and Merton (1954) introduced the term "homophily" to describe this robust phenomenon.

[2]In fact it displays a stronger form of homophily known as "inbreeding homophily" (McPherson, Smith-Lovin, and Cook 2001), which means that the equilibrium network displays more homophily than a network formed randomly where the probability of a link is proportional to the size of the group.

Perhaps the most straightforward application of our model is to consumer choice, which we used to introduce the basic question of the paper and the main results. Most products available in the market place are experience goods, and therefore a consumer will try out goods in order to determine the attractiveness of its characteristics. Talking to others or the mere observation that others have already adopted a product may make it more or less likely that a given consumer tries out the product as well, depending on the characteristics and tastes of the consumers who have adopted the product. If some inference is possible about the tastes of the adopter such information is valuable. Another application is to mentoring. The mentor tutors her mentee by relating advice on choice of action in important decisions, usually based on the lived experience of the mentor. A successful mentorship requires that actions taken in the past are informative about what sort of consequences actions taken today will likely lead to. The mentoring case has a natural separation of time where first the mentor makes her decisions and thereafter the mentee takes her decisions.

The remainder of the paper is organized as follows. Section 2 introduces a dynamic learning environment for a single decision maker, and presents some preliminary results. In section 3 we introduce the full model which contains the possibility of learning from contacts. Section 4 presents our main results. In section 5 we relate the model and results to literature on this topic. Section 6 concludes.

2 Learning in Autarky

This section considers the simple problem of decision making in autarky. The section is a building block towards the next section where we introduce learning by observation on top of the individual decision problem analyzed here. We first describe the decision problem and then characterize optimal strategies.

We model the decision problem as a bandit problem. Bandit problems have been used previously in the economics literature to study both individual learning (Rothschild 1974), and learning in groups (Bolton and Harris 1999, Cripps, Keller, and Rady 2002). A bandit problem is an interactive decision problem. The decision maker interacts with the environment (by taking an action), receives payoff relevant feedback (an outcome) about the desirability of the action and then interacts with the environment again.

The specification of the bandit that we choose is the simplest such environment we can think of that allow us to capture the effects highlighted in the introductory section. In particular, the decision horizon is two periods. Relative to the "standard" bandit problem the decision problem is simple since one period of experimentation with an arm reveals all payoff relevant information about that arm[3].

The bandit problem studied in this paper has three independent arms[4] $\mathcal{A} = \{a, b, s\}$. Each arm has a finite set of states, Θ^i, $i = a, b, s$. Nature chooses a state for each arm i from the commonly known prior distribution $p^i : \Theta^i \to (0, 1]$, where $\sum_\theta p^i(\theta) = 1$.

Time is discrete. At each time $t = 1, 2$ the decision maker (DM) must take an action. If arm i is chosen at time t and the state of the arm is θ then the DM receives the (deterministic) outcome $x_i(\theta)$. In order to make the characterization as simple as possible, we assume that arms a and b have three possible states, whereas arm c has a single (known) state. Thus from an ex-ante perspective arms a and b are *risky*, whereas arm c is *safe*. The set of possible outcomes X are given by: $X = \{x_a(1), x_a(2), x_a(3), x_b(1), x_b(2), x_b(3), x_s\}$. The outcome function $f : \mathcal{A} \times \Theta \to X$ associates to each arm and state the outcome received.

The DM is Bayesian rational and has a von Neumann-Morgenstern utility function: $u : X \to \mathbb{R}$. We normalize the outcome on the *safe* arm, s, to 0. Except where explicitly stated we also restrict attention to preferences over outcomes such that ex-ante the expected per period payoff from actions a and b are strictly negative, but there is at least one outcome that is preferred to the *safe* outcome (Call this assumption [A1]). This underlines our initial assertion that a and b are *risky*, but potentially attractive to the decision maker. These restrictions are not essential and only serve to rule out trivial cases.

The objective of the decision maker is to maximize the expected sum of utility.

2.1 Optimal Strategies

In this section we characterize optimal strategies for decision makers in autarky. We also present results on learning in autarky, which will be useful when characterizing environments with social learning. None of the results presented in

[3] In technical terms conditional on the state of arm i the outcome distribution on arm i is degenerate.

[4] We use the terms arms, actions and alternatives interchangeably.

this section are new. However they show that the insights from more general specifications of the decision problems carries over to this simpler setting.

The fundamental trade-off for the decision maker is that between exploration and exploitation. By taking a risky action the decision maker faces a cost of acquiring information today, but the information gathered can be used to exploit the arm tomorrow. More generally the question is how much information to gather before starting to exploit the information which has been gathered.

In order to solve for optimal strategies it is useful to define the time t *history*. A time $t = 1, 2$ history lists actions taken and outcomes experienced up until time $t-1$. A *terminal* history is the history listing actions and outcomes experienced up until and including $t = 2$. Optimal strategies can be solved for by backwards induction. In particular at time $t = 2$ for any possible history an optimal strategy involves playing the action that maximizes her myopic payoff. This follows since $t = 2$ is the final period of play, hence at this point the decision maker should exploit the information she has gathered. It then follows from our assumption that acquiring information is costly, that a candidate for an optimal strategy takes the following forms: (i) either play the safe action s in both periods, (ii) play either a or b at $t = 1$ and play sequentially rational at $t = 2$. In case (i) prior beliefs are such that the value of acquiring information is outweighed by the costs. Case (ii) is the complementary case where in particular an exploration phase is followed by an exploitation phase.

To see this in more detail, suppose that the DM played s at $t = 1$. Then the expected payoff from taking actions a or b are strictly negative. Thus the optimal action after history $\{x_s\}$ is to play s.

Now suppose the DM played risky action $i = a, b$ at $t = 1$. At $t = 2$ playing risky action $-i$ is clearly suboptimal since the expected payoff is strictly negative. Thus the optimal action at $t = 2$ is either action i again or the safe action s.

Given the preferences of the DM we let Θ^i_+ be the list of states on arm i that yield outcomes that are strictly preferred to the safe outcome, more formally:

$$\Theta^i_+ \quad = \quad \{\theta \in \Theta |\ u\left(f(i,\theta)\right) > u(x_s)\}$$

Following this convention we will let X^i_+ be the set of outcomes on arm i which are preferred to the safe outcome.

The *worth* of the strategy in which the DM plays arm c, W_c, in both periods

is 0. The worth of the strategy that starts by pulling risky arm i, $i = a, b$ is:

$$W_i \;=\; \mathbb{E}\left[u\left(f\left(i,\theta\right)\right)|p\right] + \sum_{\theta\in\Theta^i_+} p^i(\theta)u(f(i,\theta))$$

where the first term may be interpreted as the cost of exploration and the second term as the benefit of exploitation. An optimal strategy is a strategy that maximizes the worth among all possible candidate strategies.

We summarize our findings in the following lemma:

Lemma 1. *The candidate optimal strategies lead to the following behavior on the path:*

1. *Play the safe action at $t = 1$ and $t = 2$.*
2. *Play risky action i at $t = 1$. At $t = 2$ play risky action i if $x \in X^i_+$ and s otherwise.*

For what follows it will be helpful to know whether learning in this setting is *complete* in the sense that the DM eventually (with probability one) adopts the action that would be optimal had she observed the state of the world. Let this action (which depend on both the state and preferences) be denoted by $\alpha \in \mathcal{A}$.

Since the DM will experiment with at most one risky alternative, say r, when using an optimal strategy, it follows that with positive probability $r \neq \alpha$. Therefore the DM will not learn the ex-post optimal action with probability one. This result is summarized in the following lemma:

Lemma 2. *With positive probability the DM does not learn the ex-post optimal action.*

The result can be stated in much more general settings, see e.g. Rothschild (1974). The basic message, however, is the same: it is generally not in the DM's interest to learn the ex post optimal action with probability one. In the present model a DM could, if she wished to do so, learn the ex-post optimal action with probability one by experimenting with both risky alternatives. In our setting this is not optimal, because the decision horizon ends after $t = 2$ (or equivalently the DM uses a discount sequence $(1, 1, 0, \dots)$). Thus there would be no time to exploit the information gathered.

3 Learning from Others

This section contains the description of the main model. The following section details our main results. We introduce a model where a DM can learn from observing the experiences of others prior to having to choose actions herself. Such a setup is meant to capture a variety of examples, e.g. the recommendations in consumer-to-consumer networks of experience goods; in this example the separation in the timing of decision between agents captures the information available to early and late adopters. Formally, we augment the single person decision problem considered in the last section with a prior observational stage.

3.1 Description of the Game

Suppose we augment the decision problem analyzed in the previous section in the following way. There is a finite set of players, which consists of a set of players called *contacts*, and a single player called the decision maker (DM). The set of contacts is denoted $\mathcal{C}$. Contacts may differ in their preferences over outcomes, which are represented by a VNM-utility function. At the beginning of play a state is drawn from the commonly known probability distribution p on Θ. Then each contact solves in autarky a decision problem identical to the problem analyzed in the previous section. The player DM on the other hand can engage in observational learning. In particular at the beginning of play she chooses which contact in $\mathcal{C}$ to observe. Let the contact that DM observes be denoted $\hat{c}$. Whereas the starting history of all contacts is the empty history and their time t history is their private experiences up until time $t-1$, the starting history of DM is the terminal "history" of $\hat{c}$ (exactly what we mean by "history" will depend on the informational setting specified below). It is in this sense that DM engages in observational learning by sharing the "history" of $\hat{c}$.

We assume that DM has perfect information about the preferences of her contacts. Technically, by observing a particular contact, DM chooses a particular distribution of terminal histories as her starting history. In the terminology of the experience good example, it may be that there are specific characteristics about the product (e.g. the color of the product) which the consumer does not care about, but she may care about other characteristics of the product (e.g. how much electricity it consumes). Intuitively in this case if the consumer is restricted to asking questions of the form: "Did you like this product or not?" then she

will receive a different distribution of answers to the question depending on the preferences of the person she asks.

We consider two informational settings. In the most informative setting both the sequence of outcomes and actions experienced by the contact ($\hat{c}$) is observable to DM. In the least informative setting only the sequence of actions taken by $\hat{c}$ is observable to DM. Let the informational setting where also outcomes are observable be denoted by $\mathcal{I}_O$, and the setting where only actions are observable be denoted $\mathcal{I}_A$.

Our main interest is a characterization of the equilibrium mapping (we will look for sequential equilibria) between the preferences of DM and the preferences of $\hat{c}$. As will become clearer below the behavior of a contact is independent of the preferences of DM due to the exogenously imposed time separation. Therefore in this paper the characterization will be a mapping from the set of preferences of DM to the set of preferences of $\hat{c}$. In the example of consumer adoption of an experience good, the characterization answers the question: Who would you like to consult? Note that we require that DM chooses whom to consult at the beginning of play, that is the characterization we are interested in is the optimal *ex-ante* correspondence between the preference of DM and the preferences of $\hat{c}$.

The philosophy of the modeling is to pick the simplest learning environment that allow us to display the effects we are interested in. Thus e.g. the choice of three outcomes on a risky arm is necessary and sufficient to show that a decision maker may prefer to pick a contact more "picky" than herself. Likewise the choice of two risky arms is precisely enough to fix the interesting trade-off that we identify between learning about one alternative vs. another. Finally we have assumed that learning is particularly easy, in that one period of experimentation perfectly reveals all the outcome relevant information about an arm. The qualitative insights that we develop should be robust to a relaxing of this assumption. We comment more on this below.

An obvious reservation about the model, in the context of the consumer example, is that people may consult more than one contact before deciding whether or not to buy a particular product. However there is evidence that the majority of consumers only relies on a small set of acquaintances before deciding whether to adopt a specific product (Brown and Reingen 1987). This may be due to time constraints, or it may be that unless the decision maker knows the preferences

of the other decision makers well, the informational content of such recommendations may be low. From a modeling perspective the reason is simply that this modeling choice allow us to focus more precisely on how the incentives to learn from particular decision makers depend on preferences over outcomes. In particular by observing one contact rather than the other, entails an opportunity cots. The decision maker will learn something but not all about the environment she has been placed in.

The assumption that DM knows perfectly the preferences of all her contacts is strong. The model applies to situations where decision makers know each other well, such as long term relationships. From a modeling perspective the assumption focuses the analysis on the value of diverse information.

4 Main Results

This section presents our main results. The analysis proceeds from the perspective of DM, and asks the question: Which preferences would I like the recommender to have? The answer naturally depends on the informational setting and the preferences of contacts. First we shall answer the following question: Are there informational settings and preferences of contacts such that with probability one observing such a decision maker allows DM to infer her ex-post optimal action? We show that this question has a positive answer, and we characterize the informational setting and the set of preferences that allow such inference. The most interesting insight is that $\hat{c}$ will be more "picky" than DM. In this set-up the relation "more picky than" refers to liking a fewer number of outcomes of an action. Why is it good to observe someone who is picky? Suppose that only actions can be observed, but not actual outcomes. Then an observer may value precise information about when an action is attractive higher than the knowledge that it is preferred to a safe choice. In this cardinal information may be more valuable than ordinal information. This is the real distinction between first best contacts and second best contacts. In the first best case contacts must be picky so that they are willing to switch action. In the second best case pickiness is good because it reduces the uncertainty about the value of the prize on an arm. The up shot of pickiness is that the observer may miss outcomes that are also preferred to the safe outcome.

Our primary interest is in situations where DM cannot learn the ex-post op-

timal action with probability one. However also in this *second best* setting do we find an informationally based intuition about "pickiness" identified in the *first best* setting. In addition we identify decision problems such that a decision maker prefers to learn from a contact willing to experiment with alternatives that appear ex-ante non-attractive, rather than a risky alternative which is ex-ante relatively more attractive.

4.1 First Best Observation

The purpose of this section is to characterize behavior of contacts and environments that allow the observer to infer her ex-post optimal action with probability one. We show that if information transmission is precise, in the sense that outcomes can be observed, then such contacts exists, provided [A1] does not hold. Generally and unsurprisingly such contacts are characterized by a high willingness to experiment. An important insight that we will re-discover when we turn to second best observation is that the DM who is being observed will in general be "more picky than" the observer.

We will say that a pair $(c^{FB}, \mathcal{I}^{FB})$ is *first best* if there exists a contact, c^{FB}, and an informational setting, $\mathcal{I}^{FB}$, such that DM can infer her ex-post optimal action with probability one after having observed c^{FB} in environment $\mathcal{I}^{FB}$.

As anticipated the existence of c^{FB} relies heavily on the willingness of c^{FB} to experiment. In order to learn the ex-post optimal action with probability one c^{FB} must be willing to experiment with both alternatives a and b, even if she finds herself in period 2 having experienced an unattractive outcome on one of the risky arms in period 1. In other words the ex-ante expected per period payoff must be strictly positive for both risky actions (Assumption [A1] rules out preferences of this form). However a willingness to experiment with both alternatives is not sufficient for first best observation. A second condition must be satisfied. This condition relates to the order in which experimentation takes place and the sensitivity of preferences to outcomes.

Some notation will be needed to state the result. Let $\bar{u}_i = \max_\theta u(f(i,\theta))$ be the maximum period payoff on arm $i = a, b$, and let $\hat{u}_i = \min_{\theta \in \Theta^i_+} u(f(i,\theta))$. Without loss of generality assume that $\bar{u}_a > \bar{u}_b > 0$. Note that we may have $\bar{u}_i = \hat{u}_i$ in which case there is a unique outcome on i that is preferred to the safe outcome.

Proposition 1. *Let the preferences of DM be represented by u. A pair $(\hat{c}, \mathcal{I})$ is* first best *if and only if:*

1. *Actions and Outcomes are observable ($\mathcal{I} = \mathcal{I}_O$).*
2. *$\hat{c}$ starts experimenting with arm a, and switches to arm b whenever $u(f(a,\theta)) < \bar{u}_b$.*

Note that it follows from the proposition that a contact who is not willing to experiment with both risky actions can not be *first best* independent of the informational setting.

The result is easily generalized to any finite number of states and an arbitrary number of risky alternatives. The result of complete learning from contacts *is* special to our setting, and does not generalize. E.g. in the environment of (Rothschild 1974) it is not true that if a firm was able to observe arbitrarily many periods of experimentation by another firm then it would learn the ex-post optimal action with probability one. We present the result here partly for completeness but also because it provides important insights into the case of second best learning from contacts.

Earlier we alluded to a decision maker being "more picky than" another decision maker. We can now say more precisely what this means, and why it benefits DM to observe a decision maker who is more picky than she is. First note that the characterization imply that whenever $\hat{c}$ likes an outcome[5] on a then DM likes it as well. And whenever DM dislikes an outcome on a then $\hat{c}$ dislikes it as well. To a large extent then their ordinal rankings on a relative to the safe outcome are aligned. However it is necessary that they disagree about outcomes that DM values positively but does not care about very passionately. Here their ordinal rankings must be opposed, since this is precisely the point where DM has a positive valuation for information on arm b. To satisfy incentive compatibility it must be that $\hat{c}$ dislikes such outcomes. It is in this sense that $\hat{c}$ is "more picky than" DM[6].

[5] These statements are relative to the safe outcome.

[6] $\hat{c}$'s preferences over outcomes on arm b does not play a role as long they satisfy that ex-ante the expected per period payoff of b i strictly positive. However if we were to add a third risky arm (and extend the decision horizon) then a similar result of "more picky than" on arm b could be stated.

4.2 Second Best Observation

The remainder of the paper focuses on the case where assumption [A1] is satisfied for all decision makers. Under this assumption it follows from Proposition 1 that there is no first best contact informational environment pair. Consequently we will focus on the second best. This section considers the informational setting where only actions are observable, i.e. $\mathcal{I}_A$. The characterization in environment $\mathcal{I}_O$ is trivial given the simple setup.

The analysis proceeds as follows. First we characterize second best observation under the assumption that contacts only experiment with the risky alternative that is ex-ante most attractive to DM. We show that in this case DM may prefer to observe contacts that have preferences different from herself on the arm. We then turn to the question of whom to observe when it is only possible to learn about the risky alternative which is the least favored ex-ante among the two risky alternatives. Here we re-discover that it may be optimal to learn from observing picky contacts. Finally we turn to the question of which alternative to learn about. The most interesting finding is that a DM may prefer to learn about a risky alternative that she does not find attractive ex-ante.

To proceed some notation is needed. It is well known that optimal behavior in bandit problems with independent arms can be characterized via Gittins Indices. Consider the following simplified bandit with only two arms: a risky arm i and a safe arm s. Suppose the DM starts by using arm i. At any time thereafter she can switch to arm s (without being able to return to i). What is the period expected discounted payoff if the DM optimally abandons arm i? This is the value of the Gittins index of arm i. In particular the Gittins index of arm i is the per period discounted value of the solution to an optimal stopping problem. In our context the solution to the optimal stopping problem is particularly simple: At time $t = 2$ arm i should be abandoned if and only if $x \notin X^i_+$. We let Λ_i denote the value of the Gittins index of arm i.

Once the Gittins index have been calculated the optimal strategies are simply: in each period play the alternative with the largest Gittins index. At every time t a DM then has a ranking over the alternatives available to her, which may evolve as history accumulates. For our purposes here we will say that a risky arm i is *ex-ante* preferred to risky arm j if $\Lambda_i \geq \Lambda_j$.

4.2.1 Learning about a Favored Alternative

Suppose we restrict the contacts in $\mathcal{C}$ to only contain contacts whose preferences are such that they are willing to experiment with DM's ex-ante preferred alternative, which wlog we will assume is alternative a. We place no further restrictions on the preferences of contacts over outcomes on arm a.

The following dichotomy which identifies whether two decision makers have the "same" ordinal rankings over alternatives will be useful:

Definition 1 (Aligned)**.** *A pair of preference relations restricted to outcomes on arm i (X^i), and represented by utility functions, u, $u' : X^i \to \mathbb{R}$, are* aligned on *i if either:*

1. *For all θ: $u(f(i,\theta)) > 0$ if and only if $u'(f(i,\theta)) > 0$,* or
2. *For all θ: $u(f(i,\theta)) > 0$ if and only if $u'(f(i,\theta)) < 0$.*

If preferences are not aligned in the sense defined above then we shall say that preferences are *biased*.

Our first result shows that if all that matters to DM is that she learns whether or not she prefers the outcome on arm a to the safe outcome, then it is optimal to observe a contact whose preferences are aligned with hers.

Proposition 2 (Aligned Contacts)**.** *Suppose that (i) actions are observable, and (ii) contacts are willing to experiment with arm a only. DM finds it optimal to observe a contact with aligned preferences when:*

1. *DM is not willing to provide experimentation with arm b ($\Lambda_b < 0 = u(x_s)$)*
2. *DM is willing to provide experimentation with arm b but any preferred outcome on a discourages experimentation with b ($0 \leq \Lambda_b < \frac{2\hat{u}_a}{1+\sum_{\theta\in\Theta^b_+} p^b(\theta)}$).*

The intuition is that if any outcome on a preferred to the safe is enough to discourage experimentation on b then DM does not need to discriminate between preferred outcomes; she only requires ordinal information to determine her optimal action. The benefit of observing a biased contact is that she can discriminate between preferred outcomes. In other words by observing a biased contact she can get cardinal information. As we shall see in the next result when such cardinal information is important then DM optimally observes a biased contact.

The reader may have noticed that the same inference can be made from a contact whose opinions are diametrically opposed to DM's. While this is an interesting insight, we consider it to be non-robust. In particular the insight is not robust in decision problems where an additional risky arm is added and the decision horizon extended suitably. In this case observing a contact with diametrically opposed preferences is less useful, since such a contact will be staying with the same action, precisely when the observer values additional information on other alternatives. Thus the reader should emphasize likeness rather than opposition in interpreting the result.

The next result shows that when DM may be willing to experiment with b even after learning that a has a preferred outcome, then it is optimal to observe a biased contact. The nature of the bias is also characterized.

Corollary 1 (Biased Contacts)**.** *Suppose (i) that actions are observable, (ii) contacts are willing to experiment with arm a only. DM finds it optimal to observe a* biased *contact when DM is willing to experiment with b and is not discouraged from experimentation with b by learning that a has a preferred outcome* ($\Lambda_b \geq \frac{2\hat{u}_a}{1+\sum_{\theta\in\Theta^b_+} p^b(\theta)}$)*. Moreover, the contact is biased on DM's lower ranked outcome* ($\hat{u}_a$)*.*

In the present context when a decision maker wants to discriminate between outcomes that are preferred to the status quo she can do this by getting recommendations from someone who has a biased view (relative to her own preferences). Her demand for information is not purely ordinal, i.e. knowing whether the outcomes are preferred to a safe alternative; for optimal decision making she also requires cardinal information, i.e. how much better is the outcome relative to the safe outcome. To distinguish between outcomes the decision maker consults contacts that are more picky than she is[7]. Such contacts are picky in a particular way. In particular they dislike the outcome which is the least preferred among outcomes that are preferred to the status quo. In this sense the bias between the preferences of the decision maker and the contact is minimal.

The corollary shows that in a setting where only actions are observable DM may want to consult a contact who is biased in a particular way relative to her own preferences. Such an effect is only possible when arm b (the lower ranked of

[7]We emphasize pickiness here since, as argued above, we believe the insight of diametrically opposed preferences is less robust

the risky alternatives) is ex-ante more attractive than the safe alternative. The distinction between the demand for ordinal and cardinal information on outcomes would not appear if the bandit only had one risky arm. These insights clearly generalize to any finite number of outcomes on the arms.

4.2.2 Learning about a non-favored Alternative

We now turn our attention to contacts that are different from the decision maker in that their ex-ante ranking of the risky alternatives are different. In particular DM herself ex-ante favors arm a but she only have contacts who favor arm b. As was the case in the preceding section observation in the informational setting where outcomes are observable proves "too" informative to give interesting results in the context of the simplicity of the decision problem. In this section we therefore restrict attention to the informational setting where only actions are observable ($\mathcal{I}_A$). The goal of this section is to understand if there are situations where the decision maker prefers to learn from contacts who are biased, and characterize the nature of the bias. In order to keep notation at a minimum without losing important insights we specialize the model somewhat.

We start out be treating the case where both X_+^a and X_+^b are singleton. In this case DM does not need to distinguish between preferred outcomes on b[8]. By observing an aligned contact DM can become fully informed about the desirability of arm b.

Proposition 3 (Aligned Contacts)**.** *Suppose (i) that actions are observable, (ii) contacts are willing to experiment with arm b only, (iii) there is a unique preferred outcome on a and b (X_+^a and X_+^b are singleton for all players). DM finds it optimal to observe an* aligned *contact.*

We now turn to the case where X_+^b is not singleton, but for simplicity X_+^a is. We show that DM may want to observe a biased contact. In particular the relationship between when it is optimal to observe an aligned contact is non-monotonic in the ex-ante attractiveness of the favored risky alternative.

Proposition 4. *Suppose (i) that actions are observable, (ii) that DM can observe experimentation with risky arm b, (iii) arm a has a unique preferred outcome (X_+^a*

[8]Moreover we do not have to treat the case where DM might consider going back to b after having learned that a has an outcome that she likes somewhat, i.e. she gets utility $\hat{u}_a$.

is singleton), whereas arm b has several preferred outcomes (X^b is not singleton). For $\bar{u}_b - \hat{u}_b > 0$ sufficiently large, there exists scalars $0 < \underline{\Lambda} < \bar{\Lambda}$ such that

1. *If arm a is either ex-ante unattractive ($\Lambda_a < \underline{\Lambda}$) or ex-ante attractive ($\Lambda_a > \bar{\Lambda}$) then DM observes an* aligned *contact. Otherwise she observes a* biased *contact ($\underline{\Lambda} < \Lambda_a < \bar{\Lambda}$).*

2. *An optimal biased contact, is biased on DM's lower ranked outcome ($\hat{u}_b$).*

The intuition is relatively straightforward. When any "good" news on arm b is enough to convince DM to try arm b then observing an aligned contact is optimal. Likewise when even "good" news on arm b is not enough to convince DM to try out arm b it is again optimal to consult an aligned contact. In the second case there is an added benefit to consulting an aligned contact. Since no "good" news on b is enough to convince DM to use arm b herself, the only way such information will be valuable is if it turns out that arm a is "bad". In this case the information gathered from an aligned contact can be used in the second period. In other words information has only a second order effect on payoffs, since the information does not convince DM to switch from her ex-ante favored action.

When DM prefers to consult biased contacts then information has a first order effect on behavior. That is there is sufficiently "good" news about b that convinces DM to try out arm b. A biased contact is used to discriminate between the truly "good" news and plain "good" news (which an aligned contact cannot). This comes at a cost. In particular the not so "good" news is bundled with the "bad" news, and this discourages use of arm b. However if the utility difference between the truly good news and the relatively good news is large then it pays to consult a biased contact to be able to discriminate between outcomes. Discriminating the truly "good" news from the "good" news, allows DM to pursue the relatively attractive arm a rather than ending up with a "mediocre" outcome on b.

A more general insight follows about the desirability of observing a biased contact. Information about a non-favored alternative can either confirm the initial pre-disposition against the alternative or reject it. Learning from a biased contact tends to have a first-order effect on payoffs, because the revelation of good outcomes are sufficient to convince the decision to switch her ordering of the risky alternatives. In contrast aligned information tends to confirm the pre-disposition and information only has a second-order effect on payoff, that is the

information will only be used upon learning that alternative a was unattractive. In other words aligned information tends not overturn the ex-ante ordering of alternatives, but information is still useful in that it can be exploited if the decision maker's conviction turns out to be false.

4.2.3 When Do Opposites Attract?

Do opposites attract? Will the decision maker ever prefer to learn from a contact with an ex-ante ordering of the risky alternatives different from hers? Is it ever preferable to get information about alternatives you would not consider trying out yourself, to getting more information about alternatives you deem to be more enticing ex-ante?

To study this question in its purest form we abstract from the issue of whether it is optimal to consult an aligned or biased contact, that is we specialize to the case where X_+^a and X_+^b are singleton. Furthermore, to ease notation we assume that all outcomes on arm j, $j = a, b$ that are not preferred to the safe outcome, yields the same level of utility, $\underline{u}_j < 0$. Essentially we specialize to two outcomes to drive home the point as cleanly as possible. The following result shows that DM may prefer to learn about an alternative that is not favored ex-ante, rather than her ex-ante favored alternative.

Proposition 5 (Opposite Attracts)**.** *Suppose (i) actions are observable, (ii) DM can observe contacts who experiments with either arm a or b, (iii)* X_+^a *and* X_+^b *are singleton. Given* $\Lambda_a > \Lambda_b$*, there exist* $\bar{u} > 0$ *and* $\underline{u} < 0$*, such that DM prefers to learn about arm b iff* $\bar{u}_b > \bar{u} > 0$ *and* $\underline{u}_b < \underline{u} < 0$*.*

Notice that in both cases the proof works by driving a wedge of sufficient size between "good" outcomes and "bad" outcomes on arm b. Eventually the outcome on b becomes sufficiently attractive that the observer prefers to learn about arm b although arm b is not ex-ante the most attractive arm. The DM piggybacks on her contact, which helps her avoid "bad" outcomes, and identify "good" outcomes.

The result can be interpreted as follows. For a fixed ex-ante ranking of alternatives, the more "passionate" a DM reacts to outcomes "good" and "bad" on b the more likely that she will want to consult a contact with experience of alternative b. This insight is related to the fact that a mean preserving spread (i.e. more variance) makes an alternative more attractive ex-ante. Here we use this

insight with a vengeance; we increase the spread while keeping the arm equally attractive from the perspective of providing own experimentation, it becomes more attractive from the perspective of learning through observation.

5 Related Literature

This paper is not the first to consider modeling learning as a bandit problem. Rothschild (1974) studies the problem faced by a monopolist who is uncertain about how demand is affected by price. He models the monopolists choice between setting different prices in order to learn about demand as a two-armed bandit problem. Bolton and Harris (1999) and Cripps, Keller, and Rady (2002) studies a social learning problem in the context of a two armed bandit problem where several agents simultaneously choose between two actions. Payoff realizations are publically observable. This leads to the usual incentive problems and under provision of the public good (here experimentation with a single risky alternative). Our model contains no such free-riding problem, nor does our model display the encouragement effect identified in Bolton and Harris. The encouragement effect is a dynamic effect: a higher rate of experimentation by a player today, induces a higher rate of experimentation by others tomorrow. This effect does not appear in our model for the following reasons. First, the player being observed does not move again after the observing player has moved, she has no incentives to induce more experimentation. Second, one unit of experimentation fully reveals the state, once this has happened the demand for information on that arm drops to zero, so there is no mechanism through which the player providing the experimentation can induce the other player to experiment more. More generally there is a literature on social learning that studies how efficient adoption depends on simple rule of thumbs (Ellison and Fudenberg 1993, Ellison and Fudenberg 1995) and properties of the sampling rules that players use when learning from others (Banerjee and Fudenberg 2004). In a network context Bala and Goyal (1998) studies social learning where players learn from a only their neighbours. Their focus is on the question of which networks structures facilitate optimal adoption.

The subject of the paper is most closely related to the literature on coarsening of information. Apart from information coarsening, our paper shares with this literature the assumption that communication is non-strategic, that is the player

providing advice reports truthfully her information, given the informational constraints imposed on her[9]. Calvert (1985) studies the problem of a decision maker faced with a once-and-for-all choice between two alternatives A and B. Before deciding the decision maker can consult on the alternatives. A consultation leads to either a recommendation or a rejection of the alternative. Thus information is coarse. Calvert shows that the decision maker may benefit from getting advice from an advisor who is biased towards the alternative the decision maker is himself pre-disposed to. The reason is that evidence from such an advisor against the alternative is strong evidence that alternative B is indeed the better choice. Suen (2004) studies a related model, but his main errand is to study how in the presence of heterogeneous beliefs, the optimal demand for biased information among agents, slows the process of convergence of beliefs. Meyer (1991) studies how a principal should organize sequential promotion contests when she only receives coarse information of the form: "Employee A performed better than employee B". Meyer shows that quite generally it is optimal to bias subsequent contests in favor of the current leader. Suppose the principal wants to promotes the most able of two employees (A and B) whom she cannot distinguish between ex-ante. Suppose A wins the first unbiased contest. The informational gain from organizing another unbiased contest is zero. Either A wins, or B wins. If A wins she should promote A and if B wins she is indifferent between A and B, so she might as well promote A. However organizing a biased contest in favor of A provides information. If B wins this contest then the principal optimally promotes B since she was able to win in a relative hostile environment.

6 Conclusion

This paper has studied the problem of a decision maker who must make sequential decisions, and can rely on prior "advice" from a suitably chosen contact out of a pool of contacts. Many economic applications fit such a basic scenario.

When the decision maker consults on an alternative that she is already predisposed to, then she gives preferential attachment to people like that a similar to herself. As the attractiveness of the second best alternative increases she shifts preferential attachment towards a more "picky" contact. This contact helps

[9]Suen (2004) argues that information coarsening can also be given a strategic foundation in the vein of Crawford and Sobel (1982).

her to distinguish between favorable outcomes. If the decision maker instead must consult on a lower ranked alternative then the relation between optimal contacts depends non-monotonically on the value of the best alternative. When the best alternative is either of low or high value then it is optimal to consult an aligned contact. When the best alternative is of medium value it becomes attractive to consult a picky contact. For low values information of the best alternative information has a first order impact on payoffs, whereas for high values information has a second order impact. A picky contact helps the decision maker to distinguish between when her pre-disposition was faulty and when it was (partially) correct.

In a wider sense our study highlights the potential benefits of diverse encounters. In the language of Granovetter (1995 [1975]) diverse contacts can be thought of as "weak" ties because they are likely to give access to information which is not normally present within the "strong" tie network characterized by homophily.

References

Arndt, J. (1967): "Role of Product-Related Conversations in the Diffusion of a New Product," *Journal of Marketing Research*, 4(3), 291–295.

Bala, V., and S. Goyal (1998): "Learning from Neighbours," *Review of Economic Studies*, 65(3), 595–621.

Bandiera, O., and I. Rasul (2006): "Social Networks, and Technology Adoption in Northern Mozambique," *Economic Journal*, 116(514), 869–902.

Banerjee, A., and D. Fudenberg (2004): "Word-of-mouth learning," *Games and Economic Behavior*, 46(1), 1–22.

Bolton, P., and C. Harris (1999): "Strategic Experimentation," *Econometrica*, 67(2), 349–374.

Brown, J. J., and P. H. Reingen (1987): "Social Ties and Word-of-Mouth Referral Behavior," *Journal of Consumer Research*, 14, 350–362.

Calvert, R. L. (1985): "The Value of Biased Information: A Rational Choice Model of Political Advice," *The Journal of Politics*, 47(2), 530–555.

Crawford, V. P., and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50(6), 1431–1451.

Cripps, M., G. Keller, and S. Rady (2002): "Strategic Experimentation: The Case of Poisson Bandits," CESifo Working Paper No. 737.

Currarini, S., M. O. Jackson, and P. Pin (2007): "An Economic Model of Friendship: Homophily, Minorities and Segregation," `http://ssrn.com/abstract=1021650`.

Ellison, G., and D. Fudenberg (1993): "Rules of Thumb for Social Learning," *The Journal of Political Economy*, 101(4), 612–643.

——— (1995): "Word-of-Mouth Communication and Social Learning," *Quarterly Journal of Economics*, 110(1), 93–125.

Foster, A. D., and M. R. Rosenzweig (1995): "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy*, 103(6), 1176–1209.

Granovetter, M. (1995 [1975]): *Getting a Job: A study of Contacts and Careers*. The University of Chicago Press, 2nd edn.

Lazarsfeld, P., and R. Merton (1954): "Friendship as a social process: a substantive and methodological analysis," in *Freedom and Control in Modern Society*, ed. by M. Berger, pp. 18–66. Van Nostrand, New York.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001): "BIRDS OF A FEATHER: Homophily in Social Networks," *Annual Review of Sociology*, 27(1), 415–444.

Meyer, M. A. (1991): "Learning from Coarse Information: Biased Contests and Career Profiles," *Review of Economic Studies*, 58(1), 15–41.

Rothschild, M. (1974): "A two-armed bandit theory of market pricing," *Journal of Economic Theory*, 9(2), 185–202.

Suen, W. (2004): "The Self-Perpetuation of Biased Beliefs," *Economic Journal*, 114(495), 377–396.

A Omitted Proofs

A.1 Proof of Proposition 1

Proof.
$\Leftarrow$: Suppose that we are in informational setting $\mathcal{I}_O$. Given the prescribed behavior of $\hat{c}$, it follows that if the state is such that $u(f(a,\theta)) > \bar{u}_b > 0$ then after the first period of play DM1 knows that the ex-post optimal action is a. In words the value of further information about the state is 0.

Suppose now that the state of arm a is such that $u(f(a,\theta)) < \bar{u}_b$. In this case with positive probability the ex-post optimal action is b. $\hat{c}$ switches to arm b exactly if this condition is met. Thus DM learns whether the ex-post optimal action is b or s in period 2. In other words in this case DM has a positive valuation for additional information about the state.

$\Rightarrow$: We now show that if either of the stated conditions fails then the pair $(\hat{c}, \mathcal{I})$ is not first best. Suppose first that the informational setting is $\mathcal{I}_A$, while we keep behavior of $\hat{c}$ fixed as is in the proposition. Suppose $\hat{c}$ does not switch arm in period 2. Then DM can deduce that the ex-post optimal action is a. However suppose $\hat{c}$ switches to b in period 2. Since the informational setting is $\mathcal{I}_A$ DM only knows that $\alpha \in \{b, s\}$.

Finally, suppose we are in informational setting $\mathcal{I}_O$ but $\hat{c}$'s switching behavior does not follow the rule switch to b if: $u(f(a,\theta)) < \bar{u}_b$. Suppose $\hat{c}$ stays with a for some θ such that $u(f(a,\theta)) < \bar{u}_b$ then with positive probability the ex-post optimal action is on b but this is not revealed to DM. □

A.2 Proof of Proposition 2

Proof. First consider the trivial case where DM is not willing to provide experimentation with b herself, i.e. $\Lambda_b < 0$. In this case observing a contact who is aligned reveals to DM when arm a is preferred to the safe arm, and when its not with probability one. Since arms are independent such information does not affect the index of arm b. If DM observes a contact who is not aligned then with positive probability she will not know whether arm a or s is preferred.

Next, consider the case where $\Lambda_b > 0$. Note that if X^a_+ is singleton then it necessarily follows that $\Lambda_b < \bar{u}_a \equiv \hat{u}_a$.

To see this note that by assumption we have: $\Lambda_a \geq \Lambda_b$. Wlog assume that $u(x_a(1)) = \bar{u}_a$. We have that: $\Lambda_a < \bar{u}_a$, since

$$\Lambda_a = \frac{p^a(1)2\bar{u}_a + \sum_{\theta \notin \Theta^a_+} p^a(\theta)u(f(a,\theta))}{1 + p^a(1)}$$

the claim follows as $\sum_{\theta \notin \Theta^a_+} p^a(\theta)u(f(a,\theta))$ is strictly negative.

Therefore we need only consider the case where there are two outcomes on arm a preferred to the safe outcome. Without loss of generality assume that $\bar{u}_a = u(f(a,1)) > u(f(a,2)) = \hat{u}_a$.

If an aligned contact is observed and good news is received then this information should be exploited immediately. This follows from the observation that:

$$\Lambda_a(\{1,2\}) = \frac{\frac{p^a(1)}{p^a(1)+p^a(2)}2\bar{u}_a + \frac{p^a(2)}{p^a(1)+p^a(2)}2\hat{u}_a}{1+1}$$

As a matter of fact: $\Lambda_a(\{1\}) > \Lambda_a(\{1,2\}) > \Lambda_a$ If an aligned contact is observed, then states on arm a are partitioned: $\{\{1,2\},\{3\}\}$, leading to expected payoff:

$$(1-p^a(3))\left(\frac{p^a(1)}{1-p^a(3)}2\bar{u}_a + \frac{p^a(2)}{1-p^a(3)}2\hat{u}_a\right) + p^a(3)(1+\sum_{\theta\in\Theta^b_+} p^b(\theta))\Lambda_b$$

Suppose now that DM chooses to observe a biased contact. There are two types of biased contacts to consider: (i) a player who only likes $x_a(2)$ (or alternatively likes both $x_a(1)$ and $x_a(3)$), (ii) a player who only likes $x_a(1)$ (or alternatively likes both $x_a(2)$ and $x_a(3)$). In the former case states on arm a are partitioned: $\{\{1,3\},\{2\}\}$, and expected payoff is:

$$V_1 = p^a(2)\max\left(2\hat{u}_a, (1+\sum_{\theta\in\Theta^b_+} p^b(\theta))\Lambda_b + (1-\sum_{\theta\in\Theta^b_+} p^b(\theta))\hat{u}_a\right) + (1-p^a(2))\max\left((1+\sum_{\theta\in\Theta^b_+} p^b(\theta))\Lambda_b, \tilde{V}_a(\{1,3\})\right)$$

In the latter case states on arm a are partitioned: $\{\{2,3\},\{1\}\}$. It then follows that if "good" news is received about a then DM should immediately exploit this news. If "mixed" news is received then DM either tries out b (and does not go back to a in period 2, since the myopic expected payoff is negative), or tries a. This gives expected payoff:

$$V_2 = \max\left((1+\sum_{\theta\in\Theta^a_+} p^a(\theta))\Lambda_a, p^a(1)2\bar{u}_a + (1-p^a(1))(1+\sum_{\theta\in\Theta^b_+} p^b(\theta))\Lambda_b\right)$$

Now compare the value of observing a biased contact with that of an aligned contact. First observe that when $V_j = (1 + \sum_{\theta \in \Theta^a_+} p^a(\theta))\Lambda_a$, $j = 1, 2$ then observing an aligned contact is optimal. When this is not the case, there are many cases to treat. When the contact is biased on outcome 1 the relevant case is where DM changes her behavior after getting the information. Aligned contact is preferred to a biased iff:

$$p^a(1)(\bar{u}_a - (1 + \sum_{\theta \in \Theta^b_+} p^b(\theta)))\Lambda_b) > 0$$

and an aligned contact is preferred to a contact biased on outcome 2 iff:

$$p^a(2)(\hat{u}_a - (1 + \sum_{\theta \in \Theta^b_+} p^b(\theta))\Lambda_b) > 0$$

Since $\hat{u}_a < \bar{u}_a$ the statement follows. □

A.3 Proof of Corollary 1

Proof. Suppose $\Lambda_b \geq \frac{2\hat{u}_a}{1+\sum_{\theta \in \Theta^b_+} p^b(\theta)}$ then it follows from proposition 2 that it is not optimal to observed an aligned contact. Note that since $\Lambda_a > \Lambda_b$ it follows that $\Lambda_b < \bar{u}_a$. We now determine which form the bias takes, in particular we show that $V_1 > V_2$. Since $\Lambda_b \geq \frac{2\hat{u}_a}{1+\sum_{\theta \in \Theta^b_+} p^b(\theta)}$ it follows that:

$$V_1 = p^a(2)2\bar{u}_a + (1 - p^a(2))(1 + \sum_{\theta \in \Theta^b_+} p^b(\theta))\Lambda_b$$

and

$$V_2 = \max\left(p^a(1)2\hat{u}_a + (1 - p^a(1))(1 + \sum_{\theta \in \Theta^b_+} p^b(\theta))\Lambda_b, (1 + \sum_{\theta \in \Theta^a_+} p^a(\theta))\Lambda_a\right)$$

If $V_2 = 2\Lambda_a$ then the result is immediate. In the other case it follows that $V_1 > V_2$ since:

$$p^a(2)(2\bar{u}_a - (1 + \sum_{\theta \in \Theta^b_+} p^b(\theta))\Lambda_b) > 0 > p^a(1)(2\hat{u}_a - (1 + \sum_{\theta \in \Theta^b_+} p^b(\theta))\Lambda_b)$$

□

A.4 Proof of Proposition 3

Proof. The proof proceeds by dividing into two cases: (i) $\Lambda_a < 0$ and (ii) $\Lambda_a \geq 0$. Consider first case (i). DM is restricted to learning about arm b, $\Lambda_a < 0$ and the arms are independent. Thus DM never experiments with arm a. Consequently it is optimal to be able to distinguish perfectly when arm b is preferred to the safe action. This can only be achieved by observing an aligned contact. Next, consider case (ii). Suppose wlog that when arm i, $i = a, b$ is in state 1 then it generates the outcome preferred to the safe outcome. We have either $\bar{u}_b \geq \Lambda_a$ or $\bar{u}_b < \Lambda_a$. In the former case if an aligned contact is observed then the value is:

$$V^{align} = 2p^b(1)\bar{u}_b + (1 - p^b(1))(1 + p^a(1))\Lambda_a$$

Suppose now that a biased contact is consulted. First some notation is needed. Let $\tilde{V}_b(t, \Theta')$ be the *value* of a two armed bandit, with arms b and s, conditional on the information that state of arm b is in the set Θ', and when the decision horizon is t periods.

The value of consulting a biased contact, who dislikes outcome $x_b(j)$, $j = 2, 3$, may be written:

$$\begin{aligned} V^{bias} = \quad & p^b(j)(1 + p^a(1))\Lambda_a + (1 - p^b(j)) \max \Big(\tilde{V}_b(2, \{1, -j\}), \\ & (1 + p^a(1))\Lambda_a, (1 + p^a(1))\Lambda_a + (1 - p^a(1))\tilde{V}_b(1, \{1, -j\}) \Big) \end{aligned}$$

The first case pertain to a situation where knowing that the "good" state is bundled with one "bad" state is sufficient that DM now regards b the more attractive choice ($\tilde{V}_b(2, \{1, -j\}) > (1 + p^a(1))\Lambda_a > 0$). The second case pertains to a situation in which knowing that the "good" state is bundled with one "bad" state is not enough to convince DM to try arm b instead she will try arm a, and will not find it worthwhile to go back to b in period if a has a "bad" state ($(1 + p^a(1))\Lambda_a > \tilde{V}_b(2, \{1, -j\})$, $\tilde{V}_b(1, \{1, -j\}) < 0$). For the final case the last inequality is reversed, so that if a "fails" then it is worthwhile to return to b in the second period.

For the first and second case simple algebra shows that it is optimal to consult an aligned contact. Turning to the final case consulting an aligned contact is optimal provided that:

$$\bar{u}_b > \Lambda_a + \frac{1 - p^a(1)}{(1 + p^a(1))} \frac{p^b(-j)}{p^b(1)} u(x_b(-j))$$

Since the last term of the RHS is negative this inequality is satisfied.

Now turn to the case where: $\bar{u}_b < \Lambda_a$. Observing an aligned contact yields:

$$V^{align} = (1 + p^a(1))\Lambda_a + p^b(1)(1 - p^a(1))\bar{u}_b$$

For the biased case the value depends upon whether the expected value of the bundled states is greater than 0 or not.

$$V^{bias} = (1 + p^a(1))\Lambda_a + (1 - p^b(j))(1 - p^a(1)) \max\left(0, \tilde{V}_b(1, \{1, -j\})\right)$$

Clearly it is optimal to observe an aligned contact in this case. □

A.5 Proof of Proposition 4

Proof. Suppose DM consults a contact who experiments with b, and only actions are observable. Recall that $\Lambda_b < \Lambda_a$. DM can either observe a contact with aligned or biased preferences. Suppose that $u(x_b(1)) = \bar{u}_b > \hat{u}_b = u(x_b(2)) > 0$. If a biased contact is observed then it is clearly favorable to observe a player with preferences who is biased on the outcome corresponding to $\hat{u}_b$.

Suppose first that $\Lambda_a \leq 0$. When observing an aligned contact the value is:

$$V^{align} = (1 - p^b(3))\tilde{V}_b(2, \{1, 2\})$$

Whereas if a biased contact is observed:

$$V^{bias} = p^b(1)2\bar{u}_b$$

Clearly in this case an aligned contact is optimal.

Assume now that $\Lambda_a > 0$. Consider the value of observing an aligned contact. The value is composed of two elements. If DM learns that b holds an unattractive outcome, then she plays the bandit as if alternative b did not exist. On the other hand if she learns that b has a nice outcome (but not which), then she has two options. She can either exploit that information now (and forego alternative a) or she can try a and return to b if a does not give a nice outcome. Hence the value of observing an aligned contact is:

$$V^{align} = p^b(3)(1 + p^a(1))\Lambda_a + (1 - p^b(3)) \max\left(\tilde{V}_b(2, \{1, 2\}), (1 + p^a(1))\Lambda_a + (1 - p^a(1))\tilde{V}_b(1, \{1, 2\})\right)$$

When $\Lambda_a < \tilde{V}_b(1,\{1,2\})$ then the first argument applies, and if $\Lambda_a \geq \tilde{V}_b(1,\{1,2\})$ then the second argument applies.

If a biased contact is observed, then if DM learns that b does not hold the most attractive outcome, she must choose a and will not return to b. If she learns that b holds the most attractive outcome then she has two possibilities. Either "postpone" using the information till after she learns whether a is better, or start using the information immediately (ignoring alternative a). The value is:

$$V^{bias} = \quad p^b(1)\max\left(2\bar{u}_b, (1+p^a(1))\Lambda_a + (1-p^a(1))\bar{u}_b\right) + (1-p^b(1))(1+p^a(1))\Lambda_a$$

Similar to the aligned case the first argument of the max applies when $\Lambda_a < \bar{u}_b$ and the other argument if $\Lambda_a \geq \bar{u}_b$.

In the following table the interval where $\Lambda_a < 0$ is denoted by (I), the interval $0 < \Lambda_a < \tilde{V}_b(1,\{1,2\})\frac{1+p^a}{2}$ is denoted by (II), the interval $\tilde{V}_b(1,\{1,2\})\frac{1+p^a}{2} < \Lambda_a < \frac{1+p^a(1)}{2}\bar{u}_b$ is denoted by (III), and the interval where $\Lambda_a > \frac{1+p^a(1)}{2}\bar{u}_b$ is denoted by (IV). Denote the value functions associated with observing an aligned contact

Interval	Aligned	Biased
I	$(1-p^b(3))\tilde{V}_b(2,\{1,2\})$	$2p^b(1)\bar{u}_b$
II	$p^b(3)(1+p^a(1))\Lambda_a + (1-p^b(3))\tilde{V}_b(2,\{1,2\})$	$(1-p^b(1))(1+p^a(1))\Lambda_a + p^b(1)2\bar{u}_b$
III	$(1+p^a(1))\Lambda_a + (1-p^a(1))(1-p^b(3))\tilde{V}_b(1,\{1,2\})$	$(1-p^b(1))(1+p^a(1))\Lambda_a + p^b(1)2\bar{u}_b$
IV	$(1+p^a(1))\Lambda_a + (1-p^a(1))(1-p^b(3))\tilde{V}_b(1,\{1,2\})$	$(1+p^a(1))\Lambda_a + (1-p^a(1))p^b(1)\bar{u}_b$

and a biased contact V^{align} and V^{bias} respectively. Note that the value functions are weakly increasing in Λ_a everywhere.

Simple algebra shows that in interval I $V^{align} > V^{bias}$. In interval II we have $V^{bias} > V^{align}$ if $\Lambda_a > \hat{u}_b\frac{2}{1+p^a(1)}$. It is possible that $\bar{u}_b \notin II$. The precise condition that $\bar{u}_b \in II$ is $\bar{u}_b > \frac{\hat{u}_b}{p^b(1)}\frac{p^b(1)-p^b(2)p^a(1)}{1+p^a(1)}$. Which is satisfied for $\bar{u}_b - \hat{u}_b$ sufficiently large. Intuitively this makes it more profitable to be able to distinguish the two outcomes. It is always possible to construct such decision problems by adjusting the utility values on b where $u(x_b) < 0$ appropriately.

Turning now to interval III algebraic manipulations show that provided that $\Lambda_a < \bar{u}_b\frac{1+p^a(1)}{2} - \frac{p^b(2)}{p^b(1)}\frac{\hat{u}_b}{2}$ then it is optimal to observe a biased contact. The condition that the threshold belongs to interval III is: $\bar{u}_b > \hat{u}_b(1+\frac{1-p^b(3)}{p^b(1)(1+p^a(1))})$, which holds provided $\bar{u}_b - \hat{u}_b$ is sufficiently large. In interval IV we have that the aligned contact is better than a biased contact. $\square$

A.6 Proof of Proposition 5

Proof. Let $\mu_i \equiv \tilde{V}_i(1,\{1,2,3\}),\ i=a,b$ denote the ex-ante unconditional expected payoff from pulling arm i once. Recall that by assumption: $\mu_i < 0$. For convenience assume that when an arm is in state 1 then the outcome is preferred to the safe outcome. Let V^i denote the value of observing experimentation with arm i.

Consider first the case: $0 < \Lambda_a < \Lambda_b$. In this case we have:

$$\begin{aligned} V^a &= p^a(1)2\bar{u}_a \\ V^b &= p^b(1)2\bar{u}_b \end{aligned}$$

So the requirement that $V^a > V^b$ becomes:

$$\bar{u}_b > \frac{p^a(1)}{p^b(1)}\bar{u}_a$$

In order to see that for given Λ_a and Λ_b the inequality can always be satisfied, totally differentiate μ_b and Λ_b in order to obtain the rates of change needed to maintain the terms constant:

$$\begin{aligned} \mu_b &: \frac{d\underline{u}_b}{d\bar{u}_b} = \frac{-p^b(1)}{1-p^b(1)} \\ \Lambda_b &: \frac{d\underline{u}_b}{d\bar{u}_b} = \frac{-2p^b(1)}{1-p^b(1)} \end{aligned}$$

where $\underline{u}_b$ denotes the utility associated with outcomes $x_b(2)$ and $x_b(3)$.

Note that if we change $\underline{u}_b$ and $\bar{u}_b$ at a rate at least $\frac{-2p^b(1)}{1-p^b(1)}$ in absolute terms then both μ_b and Λ_b do not increase.

Next, we treat the case: $\Lambda_a > 0 > \Lambda_b$. In this case if DM learns about a her payoff is:

$$V^a = p^a(1)2\bar{u}_a$$

If she learns about arm b her payoff is:

$$V^b = p^b(1)\max\left(2\bar{u}_b, (1+p^a(1))\Lambda_a + (1-p^a(1))\bar{u}_b\right) + (1-p^b(1))(1+p^a(1))\Lambda_a$$

And so there are two subcases to consider depending on: $\Lambda_a \lessgtr \bar{u}_b$. Consider first the case: $\Lambda_a < \bar{u}_b$. Algebraic manipulations yields the following inequality for $V^b > V^a$:

$$\bar{u}_b > p^a(1)\bar{u}_a - \frac{1-p^b(1)}{p^b(1)}\frac{1-p^a(1)}{2}\underline{u}_a$$

As in the previous case we can adjust utilities over outcomes on b suitably, until the inequality holds, while keeping conditions $\Lambda_b < 0$ and $\mu_b < 0$ satisfied. Turning now to the second sub-case: $\Lambda_a \geq \bar{u}_b$, we get $V^b > V^a$, provided that:

$$\bar{u}_b \quad > \quad -\frac{\underline{u}_a}{p^b(1)}$$

Again we can use the same procedure as above.

Finally we turn to the case $\Lambda_a > \Lambda_b > 0$. We have:

$$\begin{aligned} V^a &= p^a(1)\max((1+p^b(1))\Lambda_b + (1-p^b(1))\bar{u}_a, 2\bar{u}_a) + (1-p^a(1))(1+p^b(1))\Lambda_b \\ V^b &= p^b(1)\max((1+p^a(1))\Lambda_a + (1-p^a(1))\bar{u}_b, 2\bar{u}_b) + (1-p^b(1))(1+p^a(1))\Lambda_a \end{aligned}$$

We show the case where: $V^a = p^a(1)2\bar{u}_a + (1-p^a(1))(1+p^b(1))\Lambda_b$ and $V^b = p^b(1)2\bar{u}_b + (1-p^b(1))(1+p^a(1))\Lambda_a$, the remaining cases are completely analogous and left out. The condition that $V^b > V^a$ becomes:

$$\bar{u}_b \quad > \quad \bar{u}_a + \frac{1-p^a(1)}{p^a(1)}\frac{1-p^b(1)}{p^b(1)}\frac{\underline{u}_b - \underline{u}_a}{2}$$

As in the previous case the idea is to show that we can always change utility levels on arm b to satisfy the inequality. From above we have the separate rates of change needed to keep μ_b and Λ_b constant. Hence if the rate of change is exactly $\frac{d\underline{u}_b}{d\bar{u}_b} = \frac{-2p^b(1)}{1-p^b(1)}$ then Λ_b remains constant. But $\frac{-2p^b(1)}{1-p^b(1)}$ is twice the rate of change required to keep μ_b constant so at this rate μ_b decreases. We continue this until the inequality is satisfied.

□