

**Working Paper Series**  
(ISSN 1211-3298)

**618**

**A Note on Optimal Experimentation  
under Risk Aversion**

Vladimír Novák  
Tim Willems

CERGE-EI  
Prague, June 2018

**ISBN 978-80-7343-425-0 (Univerzita Karlova, Centrum pro ekonomický výzkum a doktorské studium)**  
**ISBN 978-80-7344-461-7 (Národohospodářský ústav AV ČR, v. v. i.)**

# A Note on Optimal Experimentation under Risk Aversion <sup>\*</sup>

Vladimír Novák<sup>†</sup> and Tim Willems<sup>‡</sup>

June 2018

## Abstract

This paper solves the two-armed bandit problem when decision makers are risk averse. It shows, counterintuitively, that a more risk-averse decision maker might be *more* willing to take risky actions. The reason relates to the fact that pulling the risky arm in bandit models produces information on the environment – thereby reducing the risk that a decision maker will face in the future. This finding gives reason for caution when inferring risk preferences from observed actions: in a bandit setup, observing a greater appetite for risky actions can actually be indicative of more risk aversion, not less. Studies which do not take this into account may produce biased estimates.

JEL-codes: D81, D83.

Key words: Experimentation, Learning, Risk aversion.

---

<sup>\*</sup>Many thanks to Godfrey Keller who was of crucial help in setting up the framework employed in this paper. We furthermore thank two anonymous referees, the Editor (Laura Veldkamp), an Associate Editor, Yeon-Koo Che, Mark Dean, Andrew Ellis, Sebastian Foster, Filip Matějka, Pietro Ortoleva, Kevin Roberts, Jakub Steiner, Felix Vardy, Jan Zápál, and participants of the 2016 Applied Bayesian Statistics Summer School (Como), the 2017 Summer School of the Econometric Society (Seoul), and SEAM 2017 for useful comments and discussions. Vladimír Novák was supported by Charles University in Prague, project GA UK No. 197216 and received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 678081). The views expressed in this paper are those of the authors and should not be attributed to the International Monetary Fund, its Executive Board, or its management.

<sup>†</sup>CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Politických veznu 7, 111 21 Prague, Czech Republic. E-mail: vladimir.novak@cerge-ei.cz.

<sup>‡</sup>Corresponding author. International Monetary Fund, 700 19th Street NW, 20431, Washington DC, United States. E-mail: twillems@imf.org

# 1 Introduction

This paper analyzes how a rational, risk-averse decision maker solves the two-armed bandit problem of having to choose between a safe alternative that yields a known reward, and a risky one that generates an unknown payoff.

At first sight, it seems intuitive that decision makers who are more risk averse will be less willing to take the risky action. Indeed, an earlier paper (Chancelier, De Lara, and De Palma, 2009) arrives at such a conclusion. Below, we however show that there exists a previously overlooked part of the parameter space where this result is overturned.

Our model is based upon the exponential bandit model of Keller, Rady, and Cripps (2005). Following Roberts and Weitzman (1981) and Bolton and Harris (1999), who in turn built upon the seminal work of Rothschild (1974a,b), it uses a continuous-time framework. We extend the standard model by allowing for risk aversion on behalf of the decision maker. In doing so, we uncover the previously overlooked result that a more risk-averse decision maker might be *more* willing to pull the risky arm than a less risk-averse colleague.

The reason for this counterintuitive result relates to the notion that risk in bandit models can be reduced through experimentation with the risky arm. It is most likely to arise in settings where information arrives at a high frequency, which makes our finding of specific relevance to the machine learning literature (where reinforcement learning algorithms have the bandit problem at their core; see Sutton and Barto (1998)).

It is furthermore important to understand how risk aversion of decision makers affects the decisions they make. Willingness to take risks has been linked to the success of entrepreneurs (Cantillon, 1755; Knight, 1921; Kihlstrom and Laffont, 1979; Herranz, Krasa, and Villamil, 2015), while it has also been studied in a principal-agent setup – for example analyzing decision making by CEOs (Bandiera *et al.*, 2011) and politicians (Lilienfeld *et al.*, 2012). A common narrative that can be found in this literature is that more risk-averse individuals can be expected to take less risky actions. The point of this paper is to show that the introduction of learning and experimentation can overturn this wisdom: appointing a more risk-averse decision maker is no guarantee for the implementation of

less risky actions.

Finally, our findings imply that it is not obvious to infer risk preferences from observed actions: when there is scope for experimentation, observing a greater appetite for risky actions might actually be indicative of *more* risk aversion, not less.

## 2 A Keller-Rady-Cripps model with risk aversion

In this section, we employ the exponential bandit model of Keller, Rady, and Cripps (2005) but extend it by allowing for risk-averse decision makers (henceforth “DMs”).

Time  $t \in [0, \infty)$  is continuous and there are  $N \geq 1$  players (who discount the future at rate  $r > 0$ ), all facing a two-armed bandit problem. Each player  $n$  can allocate a fraction  $k_{n,t} \in [0, 1]$  of his “informational” resource to the risky arm  $R$ , leaving  $(1 - k_{n,t})$  for the safe arm  $S$ . The aggregate intensity of experimentation is denoted by  $K_t \equiv \sum_{n=1}^N k_{n,t}$ .

The safe arm provides lump-sum payoffs  $s > 0$  (with the value of  $s$  being known to all agents) according to a Poisson process with parameter 1 (which is known as well, hence why this arm is safe). So if player  $n$  uses the safe arm  $S$  for a fraction  $(1 - k_{n,t})$  over an interval  $[t, t + dt)$ , he receives  $(1 - k_{n,t}) s dt$ .

The source of risk in the risky arm  $R$  lies in the fact that its type (indicated by  $\theta$ ) is unknown to all agents at  $t = 0$ . They know that there are two possibilities: the arm is either “good” ( $\theta = 1$ ) or “bad” ( $\theta = 0$ ). At time  $t$ , each player  $n$  holds a common belief  $p_t$  that the risky arm is of good quality. The DM’s learning process on the risky arm’s type is obstructed by the presence of noise in the associated payoff stream. When the arm is of good quality, it yields a lump-sum payoff equal to  $h$  with Poisson parameter  $\lambda > 0$  (both  $h$  and  $\lambda$  are known to all players). When its quality is bad, it will never pay off. As more information arrives over time, the belief  $p_t$  is revised according to Bayes’ rule. When the DM plays the risky arm but no payment is observed, the DM’s belief that the risky arm is “good” is revised downward by:

$$dp_t = -K_t \lambda p_t (1 - p_t) dt. \tag{1}$$

Upon arrival of a payoff  $h$ ,  $p_t$  jumps to 1 (as a bad arm would never pay off).

Crucially, however, DMs are risk averse: they have an increasing, concave utility function  $u(x)$ , with  $u'(x) > 0$ ,  $u''(x) \leq 0$  (for  $x \geq 0$ ), normalized such that  $u(0) = 0$ . Consequently, the utility that a DM derives from the aforementioned payoffs is equal to  $u(s)$  (for the safe arm) and  $u(h)$  (in case the risky arm pays off).

To analyze the impact of risk aversion, we distinguish between two DM-types (indexed by  $i$ ), where the DM with  $i = 1$  is less risk averse than the DM with  $i = 2$ . In particular, the more risk averse DM “2” has an increasing, concave utility function  $u_2(\cdot)$  which is a concave transformation of  $u_1(\cdot)$  (the utility function of the less risk averse DM “1”).<sup>1</sup>

We make the problem meaningful for both DM-types by assuming that:

ASSUMPTION 1.  $0 < u_i(s) < \lambda u_i(h)$  for  $i = 1, 2$ .

This assumption ensures that a DM pulls the risky arm if he knows that it is of the good type. (If  $\lambda u_i(h)$  were smaller than  $u_i(s)$ , a DM would never play the risky arm.)

At this stage, we are able to define the expected current (myopic) payoff from  $R$ , where the expectation is formed given current belief  $p_t$ , as:

$$m_i(p_t) = p_t \lambda u_i(h). \quad (2)$$

### 3 The single-agent case

We proceed by focusing on the case in which there is only a single agent of type  $i = 1, 2$  who takes decisions. This simplest-possible case illustrates the essence of our argument about why higher risk aversion might increase a DM’s willingness to pull the risky arm. Section 4 will subsequently demonstrate that our core result can be generalized to a setup in which  $N > 1$  agents cooperate in a team.

In the single-agent case, the objective of the DM is to choose  $\{k_t\}_{t \geq 0}$  so as to maximize

---

<sup>1</sup>I.e.:  $\exists \phi(\cdot)$  with  $\phi' > 0$ ,  $\phi'' \leq 0$  such that  $u_2(x) = \phi(u_1(x)) \forall x$ . As shown in Mass-Colell, Whinston, and Green (1995: 191), this is equivalent to measuring risk preferences via the absolute measure of risk aversion  $r_i(x) = -u_i''(x)/u_i'(x)$ . In particular, they show that  $u_2(x) = \phi(u_1(x)) \forall x \Leftrightarrow r_2(x) \geq r_1(x) \forall x$ .

the total expected discounted payoff:

$$\mathbb{E} \left[ \int_0^\infty r e^{-rt} [(1 - k_t) u_i(s) + k_t m_i(p_t)] dt \right].$$

From this point onward, the solution procedure is very much analogous to that in Keller, Rady, and Cripps (2005) – the only difference being the presence of risk aversion. As in Keller, Rady, and Cripps (2005), the Principle of Optimality implies that the value function  $V_i$  for DM-type  $i$  satisfies:

$$V_i(p) = \max_{k \in [0,1]} \left\{ r [(1 - k) u_i(s) + k m_i(p)] dt + e^{-rdt} \mathbb{E} [V_i(p + dp) | p, k] \right\}.$$

To eliminate the expectations operator, observe that with subjective probability  $pk\lambda dt$ , a payoff is observed – revealing to the DM that the risky arm is of the good type. In that case, the value function jumps to  $V_i(1) = \lambda u_i(h)$ . With complementary probability  $1 - pk\lambda dt$ , no payoff is observed. Then, application of Bayes' rule (1) enables us to write  $V_i(p) + V_i'(p) dp = V_i(p) - k\lambda p(1 - p)V_i'(p) dt$ . Combining this with equation (2), while approximating  $e^{-rdt}$  by  $(1 - rdt)$ , leads to the following Bellman equation:

$$V_i(p) = \max_{k \in [0,1]} \left\{ (1 - k) u_i(s) + pk\lambda u_i(h) + pk\lambda \left[ \lambda u_i(h) - V_i(p) - (1 - p) \frac{V_i'(p)}{r} \right] \right\}.$$

At this stage, it is interesting to note that the maximand in the Bellman equation continues to be linear in  $k$ , which implies that the solution will be of a bang-bang nature.

As a result, and exactly as in Keller, Rady, and Cripps (2005), it is always optimal for the DM to play either  $k = 0$  or  $k = 1$ . In the former case,  $V_i(p) = u_i(s)$ . In the latter case,  $V_i$  satisfies the following first-order ordinary differential equation:

$$\lambda p(1 - p)V_i'(p) + (r + \lambda p)V_i(p) = (r + \lambda)\lambda u_i(h)p,$$

whose solution is given by:

$$V_i(p) = \lambda u_i(h)p + C(1-p) \left( \frac{1-p}{p} \right)^{r/\lambda},$$

where  $C$  is the constant of integration.

This solution has the exact same structure as that in Keller, Rady, and Cripps (2005). It therefore inherits the feature that there exists a cutoff belief  $p_i^*$  below which it is optimal for the DM to play the safe arm  $S$ , while playing the risky arm  $R$  becomes optimal when the DM's belief  $p > p_i^*$ . By imposing value matching ( $V_i^*(p_i^*) = u_i(s)$ ) and smooth pasting ( $(V_i^*)'(p_i^*) = 0$ ), we can derive the cutoff belief  $p_i^*$  as:

$$p_i^* = \frac{(r/\lambda) u_i(s)}{(r/\lambda + 1) (\lambda u_i(h) - u_i(s)) + (r/\lambda) u_i(s)}. \quad (3)$$

After some algebra, it then emerges that the difference  $p_2^* - p_1^*$  is equal to:

$$p_2^* - p_1^* = \frac{(r/\lambda + 1) (r/\lambda) \lambda [u_1(h)u_2(s) - u_1(s)u_2(h)]}{[(r/\lambda + 1) (\lambda u_2(h) - u_2(s)) + (r/\lambda) u_2(s)][(r/\lambda + 1) (\lambda u_1(h) - u_1(s)) + (r/\lambda) u_1(s)]}.$$

Hence:

$$\text{sgn}(p_2^* - p_1^*) = \text{sgn}(u_1(h)u_2(s) - u_1(s)u_2(h)) = \text{sgn}(u_2(s)/u_1(s) - u_2(h)/u_1(h)).$$

Noting that  $u_2(x)/u_1(x)$  is decreasing in  $x$ ,<sup>2</sup> we can establish our main result:

**PROPOSITION 1.** *In the single-agent problem, the ordering of the cutoff beliefs for DM-types 1 and 2 is as follows: (a) When  $h > s$ ,  $p_2^* > p_1^*$ . (b) When  $h < s$ ,<sup>3</sup>  $p_2^* < p_1^*$ . (c) Finally, when  $h = s$ ,  $p_2^* = p_1^*$ .<sup>4</sup>*

Part (a) of Proposition 1 implies that the more risk-averse DM needs a more optimistic belief on the quality of the risky arm to become willing to play  $R$ . In case (b) however,

---

<sup>2</sup>See Appendix A for a proof.

<sup>3</sup>This, as well as part (c) of the proposition, requires  $\lambda$  to be high enough so that Assumption 1 is not violated. If it were violated, the DM would never pull the risky arm (even if the latter was known to be of good quality).

<sup>4</sup>Following suggestions by a referee, we show in Appendix B that one can generalize this proposition to any utility function for which we can measure risk attitudes by the absolute measure of risk aversion by studying an infinitesimal increase in risk aversion using Pratt's (1964) representation.

the more risk-averse DM has the **lower** threshold  $p^*$  – implying that he will play  $R$  at more pessimistic beliefs relative to the less risk-averse DM.

To gain intuition for Proposition 1, start with part (c). When  $h = s$ , the safe arm gives rise to the exact same payoff as a good-quality risky arm (only at a different frequency given that  $\lambda > 1$ ). As a result,  $h/s = u_i(h)/u_i(s) = 1$  for  $i = 1, 2$  and all payoff-related terms disappear from the cutoff formulae (3). Both collapse to:

$$p_1^* = p_2^* = \frac{r/\lambda}{r + \lambda - 1} \quad (4)$$

From (4), one can see that pushing  $\lambda$  up (making the risky arm more attractive), lowers the cutoff belief (thus increasing the DM’s willingness to try the risky arm). Crucially, however, when  $h = s$ , the cutoff belief falls at the same rate for all DMs irrespective of their degree of risk aversion (because  $u_i(h)/u_i(s) = 1$  for  $i = 1, 2$  and transformations of payoffs no longer affect the cutoff location), thereby keeping  $p_1^* = p_2^*$ .

This no longer holds true when we increase  $h$  slightly to  $h' > s$ . Again, we have made the risky arm more attractive, in response to which both  $p_1^*$  and  $p_2^*$  fall (*cf.* equation (3)). But since marginal utility of a more risk-averse DM decreases at a faster rate when the payoff rises, he gains fewer utils from the increase in  $h$  than his less risk-averse counterpart. As a result, the less risk-averse DM’s cutoff  $p_1^*$  falls by more than the more risk-averse DM’s cutoff  $p_2^*$  – putting us in case (a) of Proposition 1, where the less risk-averse DM is more willing to pull the risky arm.

Further understanding of the difference between parts (a) and (b) of Proposition 1 can be gained by taking learning incentives into account and by realizing that our DM solves two fundamental trade-offs:

1. Choosing between an arm of known quality (the safe one,  $S$ ) and an arm of unknown quality (the risky one,  $R$ ), where information on the latter’s quality can be gathered through experimentation. (This is the learning-dimension of the problem.)
2. Choosing between an arm that provides a relatively frequent stream of modest payoffs, and an arm that provides a less frequent stream of larger payoffs. (This is

the dimension along which curvature in the utility function plays a role.)

When  $h < s$ , the risky arm's payoff frequency  $\lambda$  has to be rather high by Assumption 1 (otherwise the DM will always choose  $S$  and the problem is not meaningful). A high  $\lambda$  implies that pulling the risky arm is relatively informative: if the arm is of the good type, a payoff  $h$  should be observed soon; if not, the belief about the nature of the risky arm will quickly be revised downward by equation (1) – ending the experimentation process once the belief  $p$  falls below the cutoff  $p_i^*$ . So when  $\lambda$  is high, uncertainty about the quality of the risky arm (captured under point 1) is likely to be short-lived. Consequently, the consideration under point 2 becomes more important. Along this dimension, a more risk-averse DM prefers a frequent stream of modest payoffs to an infrequent stream of larger payoffs.<sup>5</sup> In the case where  $h < s$  (but  $\lambda$  is high enough to meet Assumption 1), arm  $R$  is the one that offers a relatively frequent stream of modest payoffs (provided the arm is of good quality). The more risk-averse DM does not like the fact that this arm is risky (its quality is initially unknown and may turn out to be low, in which case it will never pay) but when  $\lambda$  is high, this risk is likely to be resolved soon and therefore of subordinate importance.

It is thus the trade-off between these two forces that determines a DM's decision to pull  $R$  or  $S$ . On the one hand, a more risk-averse DM is drawn towards the safe arm (the fact that the risky arm is of unknown quality introduces extra uncertainty in its payoff stream, which he dislikes). But, on the other hand, the DM realizes that pulling the risky arm enables him to reduce risk (which a risk-averse DM particularly likes). When  $\lambda$  is large, there is a high “informational return” to pulling the risky arm, as there is a good

---

<sup>5</sup>Because of the concavity in his utility function, the more risk-averse DM values increases in  $h$  less and less (due to decreasing marginal utility, which a risk-neutral DM for example does not experience). To see this mathematically, consider two cases, indexed A and B. Assume that  $\lambda_A > \lambda_B$  and  $h_B > h_A = 1$  (a normalization), but  $\lambda_A h_A = \lambda_A = \lambda_B h_B$  – implying that  $\lambda_B = \lambda_A/h_B$ . Hence, option A represents a rather frequent stream of relatively minor payments, while option B represents a rarer stream of larger payments. Given that the product  $\lambda h$  is equal for both options, a risk-neutral DM is indifferent between them. A DM whose utility follows a concave transformation  $u(\cdot)$  of any payout  $h$ , prefers the frequent, modest flow provided by option A: his expected flow utility for option A is given by  $U_A = p\lambda_A u(1)$ , while for option B we have  $U_B = p\lambda_B u(h_B) = p\lambda_A [u(h_B)/h_B]$ . Clearly  $U_A = U_B$  for  $h_B = 1$  and since  $u(h_B)/h_B$  is decreasing in  $h_B$  when  $u(\cdot)$  is concave, we have that  $U_A > U_B$  for our case where  $h_B > 1$ . More generally, this implies that a more risk-averse DM prefers a relatively frequent flow of modest payoffs to rarer arrivals of larger amounts. We thank a referee for pointing this out to us.

chance that pulling  $R$  eliminates risk (which happens when a payoff  $h$  is observed, no matter how small  $h$  is). Subsequently, the DM is able to enjoy the (higher) utility stream provided by  $R$  in a world that no longer exhibits uncertainty on the nature of  $R$ . This explains the counterintuitive part of our result that a more risk-averse DM might be more willing to pull the risky arm than a less risk-averse DM.<sup>6,7</sup>

The counterintuitive part (b) of Proposition 1 is seemingly at odds with the result of Chancelier, De Lara, and De Palma (2009), who conclude that more risk-averse DMs are always more likely to pull the safe arm in bandit problems. Closer inspection of Theorem 1 in Chancelier, De Lara, and De Palma (2009), however, reveals that the assumption made there restricts payoffs in such a way that it only covers case (a) of our Proposition 1.<sup>8</sup> There, we obtain the same result. Since the continuous-time framework employed in this paper makes the existence of different regimes more transparent, it becomes apparent that there is a part of the parameter space (with  $h < s$  and  $\lambda$  sufficiently high) in which the intuitive result does not arise.

Instances where  $\lambda$  is high are particularly likely to occur in online settings, where information abounds and arrives at a high frequency. Websites often use machine learning algorithms (such as reinforcement learning), for example to customize the display of advertisements to user-preferences. Such algorithms have the bandit problem at their core (*cf.* Sutton and Barto (1998)) and our result points out that increasing the degree of risk aversion in the objective function might actually give rise to a greater appetite for the risky alternative.

---

<sup>6</sup>Proposition 1 continues to hold in the more general framework of Keller and Rady (2010): in their setup, even bad arms generate occasional payoffs equal to  $h$  – only at a lower frequency than good arms. More specifically, a good arm pays off according to a Poisson process with parameter  $\lambda_H$ , while this parameter equals  $\lambda_L$  for a bad arm (with  $\lambda_H > \lambda_L$ ). Setting  $\lambda_L = 0$  puts us back into the framework of Keller, Rady, and Cripps (2005) and simplifies the algebra considerably.

<sup>7</sup>This can be rephrased in terms of entropy reduction: the entropy of a Poisson distribution is increasing in its parameter  $\lambda$ , as a result of which the expected entropy reduction (= uncertainty reduction = information production) is higher when the  $\lambda$  of the risky arm is higher. This makes it more attractive for a risk-averse DM to pull that arm.

<sup>8</sup>To see this, note that Theorem 1 of Chancelier, De Lara, and De Palma (2009) can be rewritten in our notation/model as: “Assume that there exists a concave increasing function  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  such that  $u_2(s) \geq \varphi(u_1(s))$  and  $u_2(h) \leq \varphi(u_1(h))$ ...” Starting from a situation where  $h > s$  (our case (a)), it is not possible to respect the concave increasing function  $\varphi(\cdot)$  and move to a situation in which  $h < s$  (our case (b)) while satisfying the assumption that  $u_2(s) \geq \varphi(u_1(s))$  and  $u_2(h) \leq \varphi(u_1(h))$ .

Alternatively, our counterintuitive result can arise in a labor market setup. Consider an interpretation of the two-armed bandit model which captures the career choice between becoming a worker or becoming an entrepreneur. When going down the latter route, the DM will face greater uncertainty about his long-run payoffs – at least initially (less so after he has learned the popularity of his product).<sup>9</sup> Our DM is uncertain on – say – his organizational talent, which can be either high or low. If it is high, he will obtain a higher utility level as an entrepreneur; if it is low, his firm will never take off and he is better off as a worker. Following the seminal paper by Kihlstrom and Laffont (1979), most earlier papers featuring this choice have started from the (widely accepted) premise that more risk-averse individuals will choose to become workers (the safer option, which immediately gives greater clarity on long-run payoffs), while the less risk-averse ones will choose to start a business. In this literature, the narrative of the “risk tolerant entrepreneur” has been proposed as a parsimonious and plausible fix to the puzzling observation that entrepreneurs tend to earn less and bear more risk than salaried workers (*cf.* Hamilton (2000) and Moskowitz and Vissing-Jorgensen (2002)).

But by taking learning and experimentation dynamics into account, this paper demonstrates that this popular narrative does not necessarily hold true. It opens up the possibility that more risk-averse individuals might be *more* willing to start with the riskier action (setting up a business), if taking such a risk produces a sufficient amount of information on their organizational talent.<sup>10</sup> This theoretical ambiguity could explain why previous empirical studies have reported mixed results regarding the effect of risk aversion on the decision to become an entrepreneur.<sup>11</sup> In a principal-agent setup, it furthermore implies that appointing a more risk-averse agent is no guarantee that the principal will see more

---

<sup>9</sup>See Kerr, Nanda, and Rhodes-Kropf (2014) and Manso (2015) for examples of this interpretation.

<sup>10</sup>I.e.: if the information arrival rate  $\lambda$  is high enough. Bonatti and Hörner (2017) apply their bandit model to the labor market and argue that the information arrival rate  $\lambda$  is increasing in the amount of effort exerted by the DM, which seems intuitive. This suggests that high-effort exerting, relatively risk-averse DMs are more likely to display a greater preference for the risky arm than their less risk-averse counterparts (especially those who are not inclined to exert much effort).

<sup>11</sup>Compare Schiller and Crewson (1997, who report mixed results themselves), Barsky *et al.* (1997, who find no significant effect; Andersen *et al.* 2014 also falls in this category) and Cramer *et al.* (2002, who do find a significant negative effect of risk aversion on the probability of becoming self-employed, but conclude that they are not able to make statements on causality).

“safe” actions implemented.

More generally, our findings illustrate that it is not straightforward to infer risk preferences from observed actions when the setup is dynamic and offers scope for experimentation. Risk-aversion estimates obtained from game shows (such as *Deal or No Deal*<sup>12</sup>), which neglect the point made by this paper, might suffer from a serious bias (not only along the quantitative dimension, but even along the qualitative one).

## 4 Generalization to a team of $N > 1$ agents

In this section, we will show that Proposition 1 generalizes to a setup in which  $N > 1$  agents cooperate in a team (with all team members having the same utility function). The problem now consists of a social planner setting  $\{(k_{1,t}, \dots, k_{N,t})\}_{t \geq 0}$  so as to maximize:

$$\mathbb{E} \left[ \int_0^\infty r e^{-rt} \left[ \left(1 - \frac{K_t}{N}\right) u_i(s) + \frac{K_t}{N} m_i(p_t) \right] dt \right].$$

Taking the same steps as before (and as in Keller, Rady, and Cripps (2005)) leads to the following expression for the cutoff belief:

$$p_{i,N}^* = \frac{(r/\lambda) u_i(s)}{(r/\lambda + N) (\lambda u_i(h) - u_i(s)) + (r/\lambda) u_i(s)}.$$

Consequently, we have:

$$p_{2,N}^* - p_{1,N}^* = \frac{(r/\lambda + N) (r/\lambda) \lambda [u_1(h)u_2(s) - u_1(s)u_2(h)]}{[(r/\lambda + N) (\lambda u_2(h) - u_2(s)) + (r/\lambda u_2(s))] [(r/\lambda + N) (\lambda u_1(h) - u_1(s)) + (r/\lambda) u_1(s)]},$$

---

<sup>12</sup>In this game show, which is very similar to Weitzman’s (1979) Pandora Box problem, a participant is typically presented with 26 suitcases – one of which becomes “his” at the start of the game. Each case contains a monetary prize, the value of which is hidden to the participant (but he knows that the distribution of prizes is uniform over 26 pre-specified amounts). Subsequently, a game unfolds in which the participant has to open all remaining suitcases in a sequential manner. After each round, the show makes a cash offer to the participant, which he has to accept or reject. If he rejects (“no deal”), the game proceeds to the next round – until the participant makes a deal or all suitcases are opened (and the participant is left with the prize in “his” case). Studies that try to elicit risk preferences from this game show include Post *et al.* (2008) and De Roos and Sarafidis (2009). See Andersen *et al.* (2008) for an extensive overview of studies inferring risk aversion from behavior in (dynamic) game shows.

and it still is the case that:

$$\text{sgn}(p_2^* - p_1^*) = \text{sgn}(u_1(h)u_2(s) - u_1(s)u_2(h)) = \text{sgn}(u_2(s)/u_1(s) - u_2(h)/u_1(h)).$$

Consequently, Proposition 1 continues to apply to the  $N$ -agent cooperative problem.

## 5 Conclusion

We have shown that increasing a decision maker's degree of risk aversion in a two-armed bandit setup may make him more willing to experiment with the risky arm. The reason relates to the fact that experimenting with the risky action produces information on the environment, which reduces the risk that a decision maker will face going forward.

This insight has important implications. In a principal-agent setup it for example suggests that appointing a more risk-averse decision maker is no guarantee of less risky decisions, while it also implies that it is not straightforward to infer risk preferences from observed actions: when there is scope for experimentation, observing a greater appetite for risky actions might be indicative of more risk aversion, not less. This suggests that risk-aversion estimates stemming from dynamic game shows might suffer from serious biases – not only along the quantitative dimension, but even along the qualitative one.

## 6 References

Andersen, S., A. Di Girolamo, G.W. Harrison, and M.I. Lau (2014), “Risk and Time Preferences of Entrepreneurs: Evidence from a Danish Field Experiment”, *Theory and Decision*, 77 (3): 341-357.

Andersen, S., G.W. Harrison, M.I. Lau, and E.E. Rutström (2008), “Risk Aversion in Game Shows”, in: J. C. Cox and G. W. Harrison (eds.), *Risk Aversion in Experiments* (Vol. 12), Bingley, UK: Emerald, Research in Experimental Economics.

Bandiera, O., A. Prat, L. Guiso, and R. Sadun (2011), “Matching Firms, Managers and Incentives”, NBER Working Paper No. 16691.

Barham, B.L., J.P. Chavas, D. Fitz, V. Rios Salas, and L. Schechter (2014), “The Roles of Risk and Ambiguity in Technology Adoption”, *Journal of Economic Behavior & Organization*, 97: 204-218.

Barsky, R., F. Juster, M. Kimball, and M. Shapiro (1997), “Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Survey”, *Quarterly Journal of Economics*, 112 (2): 537-579.

Bolton, P. and C. Harris (1999), “Strategic Experimentation”, *Econometrica*, 67 (2), pp. 349-374.

Bonatti, A. and J. Hörner (2017), “Career Concerns with Exponential Learning”, *Theoretical Economics*, 12 (1): 425-475.

Cantillon, R. (1755), *Essai sur la nature du commerce en general*, London: Gyles.

Chancelier, J.P., M. De Lara, and A. De Palma (2009), “Risk Aversion in Expected Intertemporal Discounted Utilities Bandit Problems”, *Theory and Decision*, 67 (4): 433-440.

Cramer, J., J. Hartog, N. Jonker, and C. Van Praag (2002), “Low Risk Aversion Encourages the Choice for Entrepreneurship: An Empirical Test of a Truism”, *Journal of Economic Behavior and Organization*, 48 (1): 29-36.

De Roos, N. and Y. Sarafidis (2009), “Decision making under risk in Deal or No Deal”, *Journal of Applied Econometrics*, 25 (6): 987-1027.

Hamilton, B.H. (2000), “Does Entrepreneurship Pay? An Empirical Analysis of the

Returns to Self-Employment”, *Journal of Political Economy*, 108 (3): 604-631.

Herranz, N., S. Krasa, and A.P. Villamil (2015), “Entrepreneurs, Risk Aversion, and Dynamic Firms”, *Journal of Political Economy*, 123 (5): 1133-1176.

Keller, G., S. Rady, and M. Cripps (2005), “Strategic Experimentation with Exponential Bandits”, *Econometrica*, 73 (1): 39-68.

Keller, G. and S. Rady (2010), “Strategic Experimentation with Poisson Bandits”, *Theoretical Economics*, 5 (2): 275-311.

Kerr, W.R., R. Nanda, and M. Rhodes-Kropf (2014), “Entrepreneurship as Experimentation”, *Journal of Economic Perspectives*, 28 (3): 25-48.

Kihlstrom, R. and J.J. Laffont (1979), “A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion”, *Journal of Political Economy*, 87 (4): 719-749.

Knight, F. (1921), *Risk, Uncertainty and Profit*, Boston: Hart, Schaffner & Marx.

Lilienfeld, S.O. *et al.* (2012), “Fearless Dominance and the U.S. Presidency: Implications of Psychopathic Personality Traits for Successful and Unsuccessful Political Leadership”, *Journal of Personality and Social Psychology*, 103 (3): 489-505.

Manso, G. (2015), “Experimentation and the Returns to Entrepreneurship”, Mimeo, Haas School of Business, UC Berkeley.

Mass-Colell, A., M.D. Whinston, and J.R. Green (1995), *Microeconomic Theory*, Oxford: Oxford University Press.

Moskowitz, T. and A. Vissing-Jorgensen (2002), “The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?”, *American Economic Review*, 92 (4): 745-778.

Post, T., M.J. van den Assem, G. Baltussen, and R.H. Thaler (2008), “Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show”, *American Economic Review*, 98 (1): 38-71.

Pratt, John W. (1964), “Risk Aversion in the Small and the Large”, *Econometrica*, 32 (1-2): 122-136.

Roberts, K. and M.L. Weitzman (1981), “Funding Criteria for Research, Development,

and Exploration Projects ”, *Econometrica*, 49 (5), pp. 1261-1288.

Rothschild, M. (1974a), “A Two-armed Bandit Theory of Market Pricing”, *Journal of Economic Theory*, 9 (2): 185-202.

Rothschild, M. (1974b), “Searching for the Lowest Price When the Distribution of Prices Is Unknown”, *Journal of Political Economy*, 82 (4): 689-711.

Schiller, B. and P. Crewson (1997), “Entrepreneurial Origins: A Longitudinal Inquiry”, *Economic Inquiry*, 35 (3): 523-531.

Sutton, R.S. and A.G. Barto (1998), *Reinforcement Learning: An Introduction*, Cambridge, MA: MIT Press.

Weitzman, M.L. (1979), “Optimal Search for the Best Alternative, *Econometrica*, 47 (3): 641-654.

## 7 Appendix A: $u_2(x)/u_1(x)$ is decreasing in $x$

LEMMA 1.  $u_2(x)/u_1(x)$  is decreasing in  $x$ , where  $u_2(x)$  is a concave transformation of  $u_1(x)$  for all  $x \in [0, \infty)$  and  $u_1(x)$  is a positive, increasing, concave function which is normalized such that  $u_1(0) = 0$ .

PROOF. The first derivative of the fraction is given by:

$$\left(\frac{u_2(x)}{u_1(x)}\right)' = \left(\frac{\phi(u_1(x))}{u_1(x)}\right)' = \frac{u_1'(x)}{(u_1'(x))^2} \left(\phi'(u_1(x)) u_1(x) - \phi(u_1(x))\right)$$

Since  $u_1(x)$  is an increasing function, the first term of this expression is positive. Consequently, the sign of this first derivative is equal to the sign of  $F(x) \equiv \phi'(u_1(x)) u_1(x) - \phi(u_1(x))$ . At this stage, it is easily verified that  $F(0) = 0$ , while:

$$F'(x) \equiv \phi''(u_1(x)) u_1'(x) u_1(x) < 0.$$

## 8 Appendix B: Generalizing Proposition 1 using Pratt's representation

Using Pratt's (1964) representation, we can write  $u(x) = \int_0^x \exp\left(\int_0^y -r(z) dz\right) dy$ , where  $r(x) = -u''(x)/u'(x)$  is the absolute measure of risk aversion associated with  $u(\cdot)$ . Let us consider an increase in the absolute measure of risk aversion by  $\epsilon > 0$ . This brings us to  $u(x; \epsilon) \approx \int_0^x \exp\left(\int_0^y - (r(z) + \epsilon) dz\right) dy$ .

Using Pratt's representation, the threshold belief associated with  $u(x; \epsilon)$  equals:

$$p^* = \frac{(r/\lambda) \int_0^s \exp\left(\int_0^y - (r(z) + \epsilon) dz\right) dy}{(r/\lambda + 1) \left(\lambda \int_0^h \exp\left(\int_0^y - (r(z) + \epsilon) dz\right) dy - \int_0^s \exp\left(\int_0^y - (r(z) + \epsilon) dz\right) dy\right) + (r/\lambda) \int_0^s \exp\left(\int_0^y - (r(z) + \epsilon) dz\right) dy}$$

Differentiating this expression with respect to  $\epsilon$  shows how an increase in the absolute measure of risk aversion affects the location of the threshold. We have that:

$$\begin{aligned} \frac{\partial p^*}{\partial \epsilon} = & \frac{r(\lambda + r) \left( -\int_0^s \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) \int_0^h -\exp \left( \int_0^y - (r(z) + \epsilon) dz \right) y dy \right)}{\lambda \left( -(\lambda + r) \int_0^h \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) dy + \int_0^s \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) dy \right)^2} + \\ & + \frac{r(\lambda + r) \left( \int_0^h \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) dy \right) \int_0^h -\exp \left( \int_0^y - (r(z) + \epsilon) dz \right) y dy}{\lambda \left( -(\lambda + r) \int_0^h \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) dy + \int_0^s \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) dy \right)^2} \end{aligned}$$

From this, we can see that:

$$\operatorname{sgn} \left( \frac{\partial p^*}{\partial \epsilon} \right) = \operatorname{sgn} \left( \frac{\int_0^s \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) dy}{\int_0^s \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) y dy} - \frac{\int_0^h \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) dy}{\int_0^h \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) y dy} \right)$$

If the fraction  $\frac{\int_0^x \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) dy}{\int_0^x \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) y dy}$  is decreasing in  $x > 0$ , we are back to Proposition 1 formulated in the main text. The following Lemma (which is a generalization of Lemma 1 in Appendix A) establishes this result and thereby generalizes the main result of our paper to any concave and increasing utility function for which we can measure risk attitudes by the absolute measure of risk aversion.

**LEMMA 2.** Assuming that  $\int_0^x \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) y dy \neq 0$ , we have that  $\frac{\int_0^x \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) dy}{\int_0^x \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) y dy}$  is decreasing in  $x \in [0, \infty)$ .

**PROOF.** The first derivative of the fraction is given by:

$$\frac{\int_0^x \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) dy \left( \int_0^x \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) y dy - x \int_0^x \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) dy \right)}{\left( \int_0^x \exp \left( \int_0^y - (r(z) + \epsilon) dz \right) y dy \right)^2}$$

Note that the sign of this first derivative is equal to the sign of

$$F(x) \equiv \int_0^x \exp\left(\int_0^y -(r(z) + \epsilon) dz\right) y dy - x \int_0^x \exp\left(\int_0^y -(r(z) + \epsilon) dz\right) dy.$$

At this stage, it is easily verified that  $F(0) = 0$ , while:

$$F'(x) = - \int_0^x \exp\left(\int_0^y -(r(z) + \epsilon) dz\right) dy < 0.$$

## Abstrakt

V tomto článku zkoumáme model optimálního experimentování s hráči, kteří mají averzi k riziku. Ukazujeme, překvapivě, že hráč s větší averzí k riziku může optimálně volit riskantnější experiment. K tomuto efektu dochází, protože riskantnější experiment produkuje informace, které snižují riziko, kterému hráči čelí v budoucnosti. Větší apetit k rizikovějšímu experimentování tedy může indikovat větší, nikoliv nižší, averzi vůči riziku. Empirické a experimentální studie, které by ignorovaly tento efekt, mohou produkovat vychýlené odhady míry averze vůči riziku.

Working Paper Series  
ISSN 1211-3298  
Registration No. (Ministry of Culture): E 19443

Individual researchers, as well as the on-line and printed versions of the CERGE-EI Working Papers (including their dissemination) were supported from institutional support RVO 67985998 from Economics Institute of the CAS, v. v. i.

Specific research support and/or other grants the researchers/publications benefited from are acknowledged at the beginning of the Paper.

(c) Vladimír Novák and Tim Willems, 2018

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical or photocopying, recording, or otherwise without the prior permission of the publisher.

Published by  
Charles University, Center for Economic Research and Graduate Education (CERGE)  
and  
Economics Institute of the CAS, v. v. i. (EI)  
CERGE-EI, Politických vězňů 7, 111 21 Prague 1, tel.: +420 224 005 153, Czech Republic.  
Printed by CERGE-EI, Prague  
Subscription: CERGE-EI homepage: <http://www.cerge-ei.cz>

Phone: + 420 224 005 153  
Email: [office@cerge-ei.cz](mailto:office@cerge-ei.cz)  
Web: <http://www.cerge-ei.cz>

Editor: Byeongju Jeong

The paper is available online at [http://www.cerge-ei.cz/publications/working\\_papers/](http://www.cerge-ei.cz/publications/working_papers/).

ISBN 978-80-7343-425-0 (Univerzita Karlova, Centrum pro ekonomický výzkum a doktorské studium)  
ISBN 978-80-7344-461-7 (Národohospodářský ústav AV ČR, v. v. i.)