Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 179 (2019) 476–487

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

A note on optimal experimentation under risk aversion ☆

Godfrey Keller [a], Vladimír Novák [b], Tim Willems [c,*]

[a] *Department of Economics, University of Oxford, Manor Road Building, Oxford OX1 3UQ, UK*

[b] *CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Politických vězňů 7, 111 21 Prague 1, Czech Republic*

[c] *International Monetary Fund, 700 19th Street NW, 20431, Washington DC, United States*

Received 13 April 2017; final version received 20 November 2018; accepted 26 November 2018
Available online 3 December 2018

Abstract

In a standard two-armed bandit setup, this paper 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.

JEL classification: D81; D83

Keywords: Experimentation; Learning; Risk aversion

☆ We 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 Zápal, and participants of the 2016 Applied Bayesian 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, 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.

* Corresponding author.
E-mail addresses: godfrey.keller@economics.ox.ac.uk (G. Keller), vladimir.novak@cerge-ei.cz (V. Novák), twillems@imf.org (T. Willems).

https://doi.org/10.1016/j.jet.2018.11.006
0022-0531/

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 et al., 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 et al. (2005). Following Roberts and Weitzman (1981) and Bolton and Harris (1999), who in turn built upon the seminal work of Rothschild (1974a, 1974b), 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 et al., 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. Studies which do not take this into account may produce biased estimates.

2. A bandit model with non-linear utility

In this section, we employ the exponential bandit model of Keller et al. (2005) but extend it by allowing for a decision maker (henceforth 'DM') that need not be risk-neutral. For ease of exposition, we focus on the one-player case.[1]

Time $t \in [0, \infty)$ is continuous and the discount rate is $r > 0$. The player is facing a two-armed bandit problem, and at time t can allocate an amount $k_t \in \{0, 1\}$ of his 'informational' resource to the risky arm R, and thus $1 - k_t$ to the safe arm S.

The safe arm provides lump-sum payoffs of $s > 0$ (with the value of s fixed and known to the player, hence why this arm is called 'safe') according to a Poisson process with parameter 1

[1] The results in this section and the next carry over straightforwardly to an N-agent cooperative setup.

(which is also known). So if the player uses the safe arm over an interval $[t, t+dt)$, he receives $s\,dt$ in expectation.

The source of risk in the other arm lies in the fact that its type θ, and hence the size of its payoff, is unknown to the agent at $t=0$. He knows that the arm is either 'good' ($\theta=1$) or 'bad' ($\theta=0$). At time t, the player holds a belief p_t that the risky arm is good. 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 good, it yields lump-sum payoffs of h according to a Poisson process with parameter $\lambda>0$ (both h and λ are fixed and known by the player). When it is bad, it never pays off. Consequently, if the player uses the risky arm over an interval $[t, t+dt)$, he receives $p_t\lambda h\,dt$ in expectation (where the expectation is taken over both the unknown state of the world θ and the probabilistic arrival of the lump-sums).

The player evaluates the lump-sums using a utility function u, and so the expected increase in his utility is $[(1-k_t)u(s)+k_t p_t\lambda u(h)]\,dt$. We will assume that $u(0)=0$ for the following reason. Consider a stream of zero payoffs that arrive according to a Poisson process with parameter ν. The total expected discounted utility from this stream, expressed in per-period terms is $\mathbb{E}\left[\int_0^\infty re^{-rt}\,\nu u(0)\,dt\right]=\nu u(0)$, and it is natural to require that this be independent of the rate ν – it should not matter at what frequency nothing is paid out; this translates into a requirement that $u(0)=0$.[2] Also, to make the problem meaningful, we require that the player strictly prefers R, if it is good, to S, and strictly prefers S to R, if it is bad. Consequently, we assume that:

Assumption 1. $0=u(0)<u(s)<\lambda u(h)$.

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 lump-sum arrives, his belief that the risky arm is good is revised downward:

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

On the other hand, if a lump-sum h does arrive, the belief p_t jumps to 1. The objective of the DM is to choose $\{k_t\}_{t\geq 0}$ so as to maximize the total expected discounted utility, expressed in per-period terms:

$$\mathbb{E}\left[\int_0^\infty re^{-rt}\left[(1-k_t)u(s)+k_t p_t\lambda u(h)\right]dt\right],$$

where the expectation is over the processes $\{k_t\}$ and $\{p_t\}$,[3] and with beliefs being the state variable. The solution procedure is analogous to that in Keller et al. (2005) – the only difference being the presence of the utility function. As in Keller et al. (2005), the Principle of Optimality implies that the value function V satisfies:

[2] We thank a referee for this justification of $u(0)$ being zero, and for noting that this requirement implies that two such utility functions represent the same preferences if and only if one is a monotone linear transformation of the other.

[3] The total expected discounted utility given in the main text is equivalent to:

$$\mathbb{E}\left[\int_0^\infty re^{-rt}\left[(1-k_t)u(s)\,dN_{1,t}+k_t p_t u(h)\,dN_{\lambda,t}\right]\right],$$

where $N_{\ell,t}$ is a standard Poisson process with intensity ℓ, since $N_{\ell,t}-\ell t$ is a martingale.

$$V(p) = \max_{k\in\{0,1\}} \left\{ r\left[(1-k)u(s) + kp\lambda u(h)\right] dt + e^{-r\,dt}\mathbb{E}\left[V(p+dp)|p,k\right] \right\}.$$

To eliminate the expectations operator, observe that with subjective probability $pk\lambda\,dt$ a lump-sum h arrives, revealing to the DM that the risky arm is of the good type. In that case, the value function jumps to $V(1) = \lambda u(h)$. With complementary probability $1 - pk\lambda\,dt$, no lump-sum arrives; then, application of Bayes' rule (1) enables us to write $V(p+dp) \approx V(p) + V'(p)\,dp = V(p) - k\lambda p(1-p)V'(p)\,dt$. Combining this with $1 - r\,dt$, the approximation to $e^{-r\,dt}$, leads to the following Bellman equation:

$$V(p) = \max_{k\in\{0,1\}} \left\{ (1-k)u(s) + kp\lambda u(h) + kp\lambda\left[\lambda u(h) - V(p) - (1-p)V'(p)\right]/r \right\}.$$

As the maximand is linear in k, the DM would never optimally choose an interior allocation even were it allowed. If the DM chooses $k = 0$, then $V(p) = u(s)$. If he chooses $k = 1$, then V satisfies the following first-order ordinary differential equation:

$$\lambda p(1-p)V'(p) + (r + \lambda p)V(p) = (r+\lambda)\lambda u(h)p,$$

whose solution is given by:

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

where C is the constant of integration.

This solution has the exact same structure as that in Keller et al. (2005). It therefore inherits the feature that there exists a cut-off belief p^* above which it is optimal for the DM to play the risky arm R, while playing the safe arm S becomes optimal when the DM's belief $p \leq p^*$. By imposing value matching ($V^*(p^*) = u(s)$) and smooth pasting ($(V^*)'(p^*) = 0$), we can derive the cut-off belief[4] as:

$$p^* = \frac{(r/\lambda)u(s)}{(r/\lambda+1)\left[\lambda u(h) - u(s)\right] + (r/\lambda)u(s)}. \tag{2}$$

Now consider two DMs (indexed by i), with the utility function of DM_i being u_i. The difference between their cut-off beliefs $p_2^* - p_1^*$ satisfies:

$$\begin{aligned}
\operatorname{sgn}(p_2^* - p_1^*) &= \operatorname{sgn}\left(u_2(s)\left[\lambda u_1(h) - u_1(s)\right] - u_1(s)\left[\lambda u_2(h) - u_2(s)\right]\right) \\
&= \operatorname{sgn}\left(u_2(s)u_1(h) - u_1(s)u_2(h)\right) \\
&= \operatorname{sgn}\left(\frac{u_2(s)}{u_1(s)} - \frac{u_2(h)}{u_1(h)}\right).
\end{aligned} \tag{3}$$

Note that if DM_1 and DM_2 are both risk-neutral, then the right-hand side of the above equation is zero and hence $p_2^* = p_1^*$.

3. The effects of risk aversion

For the remainder of this article, we assume that the DM's utility function over payoffs u is increasing and concave, and recall that $u(0) = 0$. This assumption captures the notion of risk aversion in the sense that when our DM compares two streams of lump-sum payoffs, in each

[4] Note that this belief is invariant to a monotone linear transformation of u.

interval $[t, t+dt)$ he is facing a lottery over payoff increments, and when one of these lotteries second-order stochastically dominates the other, he prefers the less risky of the two.

Consider one stream $\mathcal{P}_{h,\lambda}$ that delivers lump-sum payoffs h according to a Poisson process with parameter λ, and another one $\mathcal{P}_{c,1}$ that delivers lump-sum payoffs c according to a Poisson process with parameter 1, with $c = \lambda h$. In $[t, t+dt)$, the probability of no payoff from $\mathcal{P}_{h,\lambda}$ equals $1 - \lambda\, dt$, while the probability of no payoff from $\mathcal{P}_{c,1}$ equals $1 - dt$. Also note that the lottery with the larger probability of no payoff is a mean-preserving spread of the other. Consequently, the lottery with the smaller probability of no payoff second-order stochastically dominates the other, and $\mathcal{P}_{h,\lambda}$ is preferred to $\mathcal{P}_{c,1}$ iff $\lambda \geq 1$. In terms of the total expected discounted utility from the two streams, which are $\lambda u(h)$ and $u(c) = u(\lambda h)$, we note that $\lambda u(h) \geq u(\lambda h)$ iff $\lambda \geq 1$.

This ordering of 'lumpy' payoff streams manifests itself as follows: other things being equal, a DM prefers a stream of modest payoffs that arrive with a high expected frequency to a stream of larger payoffs that are expected to arrive infrequently. This favoring of 'less risky' payoff streams and of 'smaller payoffs at higher expected frequency' are simply two manifestations of the same preference. Consequently, we couch most of our discussion below in terms of higher (expected) frequency rather than in terms of second-order stochastic dominance or lower risk.

To analyze the effects of risk aversion, let DM_2 be more risk averse than DM_1. In particular, the more risk-averse DM_2 has an increasing, concave utility function u_2 which is a concave transformation of u_1, the utility function of the less risk averse DM_1.

In Appendix A, we show that $u_2(x)/u_1(x)$ is strictly decreasing in x; this, together with reference back to equation (3), leads to our main result:

Proposition 1. *The ordering of the cut-off beliefs for DM_1 and DM_2 is as follows: (a) when $h > s$, $p_2^* > p_1^*$; (b) when $h < s$, $p_2^* < p_1^*$; (c) finally, when $h = s$, $p_2^* = p_1^*$.*[5,6]

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, 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 cut-off formula (2). Both collapse to:

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

From (4), one can see that pushing λ up (making the risky arm more attractive), lowers the cut-off belief (thus increasing the DM's willingness to try the risky arm). Crucially, however, when $h = s$, the cut-off 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 cut-off location), thereby keeping $p_1^* = p_2^*$.

[5] Parts (b) and (c) of the proposition require λ 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 it was known to be of good quality).

[6] Following suggestions by a referee, Appendix B contains a generalization of this proposition by considering an infinitesimal increase in risk aversion, employing Pratt's (1964) representation.

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 (see equation (2)). 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 cut-off p_1^* falls by more than the more risk-averse DM's cut-off 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 λ has to be rather high by Assumption 1 (otherwise the DM will always choose S and the problem is not meaningful). A high λ 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 cut-off p_i^*. So when λ 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.[7] In the case where $h < s$ (but λ 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 λ 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 λ is large, there is a high 'informational return' to pulling the risky arm, as there is a good 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

[7] Because of the difference in the concavity of the utility functions, the value of any increase in h is lower for the more risk-averse DM (due to decreasing marginal utility, which a risk-neutral DM for example does not experience).

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.[8,9]

In Appendix C we show that our result also arises in a simple two-period setup. It similarly carries over to an infinite-horizon, discrete-time version of the model.

4. Discussion

The counterintuitive part (b) of Proposition 1 is seemingly at odds with the result of Chancelier et al. (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 et al. (2009), however, reveals that the assumption made there restricts payoffs in such a way that it only covers case (a) of our Proposition 1.[10] 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 λ sufficiently high) in which the intuitive result does not arise.

Instances where λ is high (which means that the risky arm pays out frequently, conditional on it being 'good') are particularly likely to occur in online settings. There, information abounds and arrives at a high frequency. Reinforcement learning for example has the bandit problem at its core, there often referred to as the 'exploration vs. exploitation trade-off' (Sutton and Barto, 1998). Such algorithms are, among other things, used to customize webpage advertisements to user-preferences. At each page visit, the algorithm faces a choice between, say, showing a well-understood ad which is known to generate infrequent per-click payoffs of considerable size s (e.g. an ad for expensive watches), or show an ad for a new product (which comes with lower per-click payoffs, $h < s$). Suppose that the market for the associated product is very competitive and a priori it is not known whether the brand behind the advertisement will become popular (this is the source of risk in the problem). If the brand does take off, it is expected to generate frequent clicks (high λ) – improving upon the payoff generated by the safe ad (in this case, 'the risky arm is good'). If the new brand does not take off ('the risky arm is bad'), clicks on the ad will be infrequent (low λ) and the webpage would have been better off by sticking with the old ad. In such a setup, our result demonstrates that equipping the machine learning algorithm with a more risk averse objective function might lead to a greater appetite for the risky arm (the unknown ad).

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 pop-

[8] 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 λ_H, while this parameter equals λ_L for a bad arm (with $\lambda_H > \lambda_L$). Setting $\lambda_L = 0$ puts us back into the framework of Keller et al. (2005) and simplifies the algebra considerably.

[9] This can be rephrased in terms of entropy reduction: the entropy of a Poisson distribution is increasing in its parameter λ, as a result of which the expected entropy reduction (= uncertainty reduction = information production) is higher when the λ of the risky arm is higher. This makes it more attractive for a risk-averse DM to pull that arm.

[10] To see this, note that Theorem 1 of Chancelier et al. (2009) can be rewritten in our notation/model as: "Assume that there exists a concave increasing function $\varphi : \mathbb{R} \to \mathbb{R}$ such that $u_2(s) \geq \varphi(u_1(s))$ and $u_2(h) \leq \varphi(u_1(h)) \ldots$". Starting from a situation where $h > s$ (our case (a)), it is not possible to respect the concave increasing function φ 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 $u_2(h) \leq \varphi(u_1(h))$.

ularity of his product).[11] 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 (see 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.[12] 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.[13] In a principal-agent setup, it furthermore implies that appointing a more risk-averse agent is no guarantee that the principal will see more '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*[14]), 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).

Appendix A. $u_2(x)/u_1(x)$ is strictly decreasing in x

Lemma 1. *Let* $u_1 : [0,\infty) \to [0,\infty)$ *be a strictly increasing function with* $u_1(0) = 0$*; let* $\varphi : [0,\infty) \to [0,\infty)$ *be an increasing and concave function, strictly so on some interval that*

[11] See Kerr et al. (2014) and Manso (2015) for examples of this interpretation.

[12] I.e., if the information arrival rate λ is high enough. Bonatti and Hörner (2017) apply their bandit model to the labor market and argue that the information arrival rate λ 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).

[13] 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.

[14] In this game show, which is similar to Weitzman's (1979) Pandora's 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.

includes the origin, with $\varphi(0) = 0$; finally, let $u_2 = \varphi \circ u_1$. Then $u_2(x)/u_1(x)$ is strictly decreasing in x.

Proof. Measure u_1 on the horizontal axis and u_2 on the vertical axis. The graph of φ, the mapping from u_1 to u_2, is then a curve through the origin that is increasing and concave, strictly so on some interval that includes the origin, and the ratio $u_2(x)/u_1(x)$ is the slope of the chord from the origin to the point $(u_1(x), u_2(x))$. As we increase x, and thus $u_1(x)$, this slope strictly decreases and therefore $u_2(x)/u_1(x)$ is strictly decreasing in x. □

Appendix B. Generalizing Proposition 1

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 measure of absolute risk aversion associated with u. Let us consider an increase in the measure of absolute risk aversion by $\epsilon > 0$. This brings us to $u(x;\epsilon) = \int_0^x \exp\left(\int_0^y -(r(z)+\epsilon)\, dz\right) dy$.

From (2), we can rewrite the threshold belief associated with $u(x;\epsilon)$ as:

$$p^*(\epsilon) = \frac{ru(s;\epsilon)}{(r+\lambda)\lambda u(h;\epsilon) - \lambda u(s;\epsilon)}$$

and differentiate it with respect to ϵ to find that:

$$\begin{aligned}\operatorname{sgn}\left(\frac{\partial p^*}{\partial \epsilon}\right) &= \operatorname{sgn}\left(u(h;\epsilon)\,\partial u(s;\epsilon)/\partial\epsilon - u(s;\epsilon)\,\partial u(h;\epsilon)/\partial\epsilon\right)\\ &= \operatorname{sgn}\left(\frac{u(h;\epsilon)}{\partial u(h;\epsilon)/\partial\epsilon} - \frac{u(s;\epsilon)}{\partial u(s;\epsilon)/\partial\epsilon}\right)\end{aligned}$$

Writing $u_\epsilon(x;\epsilon)$ for the partial derivative with respect to the parameter ϵ, our aim is to show that $u(x;\epsilon)/u_\epsilon(x;\epsilon)$ is increasing $x > 0$, in which case 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 for an infinitesimal increase in risk aversion of any (twice-differentiable) Bernoulli utility function u.

To this end, define $f(y;\epsilon) = \int_0^y -(r(z)+\epsilon)\, dz$, giving

$$u(x;\epsilon) = \int_0^x \exp\left(f(y;\epsilon)\right) dy,$$

and $u_\epsilon(x;\epsilon) = \int_0^x f_\epsilon(y;\epsilon)\exp\left(f(y;\epsilon)\right) dy$. Noting that $f_\epsilon(y;\epsilon) = -y$, we see that

$$u_\epsilon(x;\epsilon) = -\int_0^x y\exp\left(f(y;\epsilon)\right) dy.$$

For future reference, we note that $u'(x;\epsilon) = \exp\left(f(x;\epsilon)\right)$, and $u'_\epsilon(x;\epsilon) = -x\exp\left(f(x;\epsilon)\right)$, where the prime denotes the derivative with respect to the variable x.

Lemma 2. *Given $u_\epsilon(x;\epsilon) < 0$ when $x > 0$, $u(x;\epsilon)/u_\epsilon(x;\epsilon)$ is increasing in $x \in [0,\infty)$.*

Proof.

$$\begin{aligned}&\operatorname{sgn}\left(u(x;\epsilon)/u_\epsilon(x;\epsilon)\right)' = \operatorname{sgn}\left(u_\epsilon(x;\epsilon)\,u'(x;\epsilon) - u(x;\epsilon)\,u'_\epsilon(x;\epsilon)\right)\\ &\quad= \operatorname{sgn}\left(-\exp\left(f(x;\epsilon)\right)\int_0^x y\exp\left(f(y;\epsilon)\right) dy + x\exp\left(f(x;\epsilon)\right)\int_0^x \exp\left(f(y;\epsilon)\right) dy\right)\end{aligned}$$

$$= \operatorname{sgn}\left(\exp\left(f(x;\epsilon)\right)\int_0^x (x-y)\exp\left(f(y;\epsilon)\right)dy\right)$$
$$\geq 0$$

with the inequality being strict when $x > 0$. □

Appendix C. A two-period model

Here, we will show that our result also holds in a discrete-time, two-period setup. Without loss of generality, we abstract from discounting.

In each period, the safe arm S pays out a lump-sum s with probability $\frac{1}{2}$, whereas the risky arm R pays out a lump-sum h with probability $\frac{1}{2}\gamma$ if it is of good quality, while a bad risky arm never pays off.[15] As in the main text, we use p to denote the DM's belief that R is of good quality.

At this stage, we rephrase Assumption 1 as follows:

Assumption C1. $0 = u(0) < \frac{1}{2}u(s) < \frac{1}{2}\gamma u(h)$.

In this simple setup, one can analyze the expected utilities resulting from the four possible strategies:

1. Playing the safe arm in both periods (SS) yields a total expected utility equal to $u(s)$.
2. Playing RR yields a total subjective expected utility equal to $\gamma u(h)p$.
3. Playing SR yields a total subjective expected utility equal to $\frac{1}{2}u(s) + \frac{1}{2}\gamma u(h)p$. This is dominated by SS if p is low, and dominated by RR if p is high.
4. Playing RS conditionally, i.e. only sticking with R after a success in period 1, yields a total subjective expected utility equal to $\frac{1}{2}\gamma u(h)p + \left[\frac{1}{2}\gamma u(h)\left(\frac{1}{2}\gamma p\right) + \frac{1}{2}u(s)\left(1 - \frac{1}{2}\gamma p\right)\right] = \frac{1}{2}u(s) + \frac{1}{2}\gamma\left[u(h) + \frac{1}{2}\gamma u(h) - \frac{1}{2}u(s)\right]p$.

Equating expected utility from SS and RS gives a lower cut-off belief:

$$p^{\ell} = \frac{\frac{1}{2}u(s)}{\frac{1}{2}\gamma\left[\left(1 + \frac{1}{2}\gamma\right)u(h) - \frac{1}{2}u(s)\right]},$$

while equating expected utility from RR and RS gives an upper cut-off belief:

$$p^{u} = \frac{\frac{1}{2}u(s)}{\frac{1}{2}\gamma\left[\left(1 - \frac{1}{2}\gamma\right)u(h) + \frac{1}{2}u(s)\right]}.$$

When the DM's belief $p < p^{\ell}$, it is optimal for him to play S in both periods. Similarly, when $p > p^{u}$ it is optimal to play R in both periods (even if no lump-sum arrived in period 1). For intermediate beliefs, i.e. when $p^{\ell} < p < p^{u}$, it is optimal to play R in the first period and switch to S in period 2 if no lump-sum arrived.

[15] Our maintained assumption is that a good risky arm is preferred to the safe arm. A necessary condition for the counterintuitive part of Proposition 1 from the main text is that the lump-sum h can nevertheless be smaller than the lump-sum s. Consequently we need to allow for the possibility that the probability of a payoff from R is greater than the probability of a payoff from S, and hence the requirement that the probability of a payoff from S is < 1.

Defining utility functions u_1 and u_2 as in the main body of the paper (with u_2 exhibiting greater risk aversion), it is straightforward to show that the sign of the difference in cut-offs again satisfies:

$$\operatorname{sgn}(p_2^\ell - p_1^\ell) = \operatorname{sgn}(p_2^u - p_1^u) = \operatorname{sgn}\left(\frac{u_2(s)}{u_1(s)} - \frac{u_2(h)}{u_1(h)}\right).$$

In this case, the counterintuitive part of the parameter space opens up when $\gamma > 1$ (but notice that because probabilities cannot be greater than 1, we also need that $\gamma \leq 2$). When $\gamma > 1$, the risky arm is expected to pay out more frequently than the safe arm and R second-order stochastically dominates S. This makes the risky arm more attractive to the more risk averse DM_2.

By following similar steps in Heidhues et al.'s (2015) discrete-time formulation of the infinite-horizon Poisson bandit model, one can verify that the same logic continues to apply in that setup.

References

Andersen, S., Di Girolamo, A., Harrison, G.W., Lau, M.I., 2014. Risk and time preferences of entrepreneurs: evidence from a Danish field experiment. Theory Decis. 77 (3), 341–357.

Andersen, S., Harrison, G.W., Lau, M.I., Rutström, E.E., 2008. Risk aversion in game shows. In: Cox, J.C., Harrison, G.W. (Eds.), Risk Aversion in Experiments. In: Res. Exp. Econ., vol. 12. Emerald, Bingley, UK.

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

Barsky, R., Juster, F., Kimball, M., Shapiro, M., 1997. Preference parameters and behavioral heterogeneity: an experimental approach in the health and retirement survey. Q. J. Econ. 112 (2), 537–579.

Bolton, P., Harris, C., 1999. Strategic experimentation. Econometrica 67 (2), 349–374.

Bonatti, A., Hörner, J., 2017. Career concerns with exponential learning. Theor. Econ. 12 (1), 425–475.

Cantillon, R., 1755. Essai sur la nature du commerce en general. Gyles, London.

Chancelier, J.P., De Lara, M., De Palma, A., 2009. Risk aversion in expected intertemporal discounted utilities bandit problems. Theory Decis. 67 (4), 433–440.

Cramer, J., Hartog, J., Jonker, N., Van Praag, C., 2002. Low risk aversion encourages the choice for entrepreneurship: an empirical test of a truism. J. Econ. Behav. Organ. 48 (1), 29–36.

De Roos, N., Sarafidis, Y., 2009. Decision making under risk in Deal or No Deal. J. Appl. Econom. 25 (6), 987–1027.

Hamilton, B.H., 2000. Does entrepreneurship pay? An empirical analysis of the returns to self-employment. J. Polit. Econ. 108 (3), 604–631.

Heidhues, P., Rady, S., Strack, P., 2015. Strategic experimentation with private payoffs. J. Econ. Theory 159, 531–551.

Herranz, N., Krasa, S., Villamil, A.P., 2015. Entrepreneurs, risk aversion, and dynamic firms. J. Polit. Econ. 123 (5), 1133–1176.

Keller, G., Rady, S., Cripps, M., 2005. Strategic experimentation with exponential bandits. Econometrica 73 (1), 39–68.

Keller, G., Rady, S., 2010. Strategic experimentation with Poisson bandits. Theor. Econ. 5 (2), 275–311.

Kerr, W.R., Nanda, R., Rhodes-Kropf, M., 2014. Entrepreneurship as experimentation. J. Econ. Perspect. 28 (3), 25–48.

Kihlstrom, R., Laffont, J.J., 1979. A general equilibrium entrepreneurial theory of firm formation based on risk aversion. J. Polit. Econ. 87 (4), 719–749.

Knight, F., 1921. Risk, Uncertainty and Profit. Hart, Schaffner & Marx, Boston.

Lilienfeld, S.O., et al., 2012. Fearless dominance and the U.S. Presidency: implications of psychopathic personality traits for successful and unsuccessful political leadership. J. Pers. Soc. Psychol. 103 (3), 489–505.

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

Moskowitz, T., Vissing-Jorgensen, A., 2002. The returns to entrepreneurial investment: a private equity premium puzzle? Am. Econ. Rev. 92 (4), 745–778.

Post, T., van den Assem, M.J., Baltussen, G., Thaler, R.H., 2008. Deal or no deal? Decision making under risk in a large-payoff game show. Am. Econ. Rev. 98 (1), 38–71.

Pratt, John W., 1964. Risk aversion in the small and the large. Econometrica 32 (1–2), 122–136.

Roberts, K., Weitzman, M.L., 1981. Funding criteria for research, development, and exploration projects. Econometrica 49 (5), 1261–1288.

Rothschild, M., 1974a. A two-armed bandit theory of market pricing. J. Econ. Theory 9 (2), 185–202.
Rothschild, M., 1974b. Searching for the lowest price when the distribution of prices is unknown. J. Polit. Econ. 82 (4), 689–711.
Schiller, B., Crewson, P., 1997. Entrepreneurial origins: a longitudinal inquiry. Econ. Inq. 35 (3), 523–531.
Sutton, R.S., Barto, A.G., 1998. Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA.
Weitzman, M.L., 1979. Optimal search for the best alternative. Econometrica 47 (3), 641–654.