



ISSN 2282-6483

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

**Learning and dynamic choices under
uncertainty: from weighted regret
and rejoice to expected utility**

Fabio Zagonari

Quaderni - Working Paper DSE N°1090



Learning and dynamic choices under uncertainty: from weighted regret and rejoice to expected utility

Fabio Zagonari

Dipartimento di Scienze Economiche, Università di Bologna, via Angherà 22, 47900 Rimini (Italy)
Phone: 0039 0541 434135 Fax: 0039 0541 434120 E-mail: fabio.zagonari@unibo.it

Abstract

This paper identifies the globally stable conditions under which an individual facing the same choice in many subsequent times learns to behave as prescribed by the expected-utility model. To do so, the analysis moves from the relevant behavioural models suggested by psychology (i.e., weighted probabilities applied to regret and rejoice theory), and by updating probability estimations and outcome preferences according to the learning models suggested by neuroscience (i.e., adaptive learning aimed at reducing surprises), and analogous to Bayesian updating. The search context is derived from experimental economics, whereas the learning framework is borrowed from theoretical economics. Analytical results show that obstinate and lucky individuals are better off in the short-run (i.e., a low density of events in the reference period), but they do not learn, and this is true to a greater extent in a simple context; in contrast, reactive and unlucky individuals are worse off in the short-run, but they learn and are better off in the long-run (i.e., all individuals are equally lucky or unlucky), and this is true to a greater extent in a complex context. The expected-utility model explains real behaviours in the long-run whenever unlucky events are more likely than lucky events.

Keywords

Learning; weighted probabilities; regret; rejoice; repeated choices; expected utility

JEL classification

D83

1. Introduction

Three main *econometric* analyses of dynamic choices under uncertainty in *real* or *experimental* contexts have been performed. Zagonari (1995) applied ordinary least-squares estimators to panel data (104 individuals) from a *real* rural daily-rated labour market in India, and showed that the expected-utility model performed about as well with linear and non-linear subjective probabilities for representing the decision-making process of women who were repeatedly dealing with the same kind of uncertainty in the same kind of framework. Moreover, an econometric analysis based on *experimental* data from two samples of subjects sequentially facing 90 similar pair-wise choice problems showed that the random preference model (which assumes uncertainty about preferences) is more successful than the Fechner model (which depicts uncertainty at the calculation stage) and the constant error model (which assumes uncertainty about actions), and the estimated random preference model tends towards the same results as in the expected-utility model as the number of choices tends to infinity (Loomes et al., 2002). Finally, Ben-Elia et al. (2013) applied a mixed multinomial logit model to *experimental* data (49 participants) from simple route choices involving different levels of time variability, and showed that experienced regret explains travel behaviour better than anticipated regret, and that learning seems to mitigate the amplitude of regret emotions, although to a smaller extent in more risky contexts.

Note that many *experimental* studies can be found that deal with repeated dynamic decision-making under uncertainty. For recent examples, see Hopfensitz & Van Winden (2008) in an investing context; Norman et al. (2012), Huang & Hutchinson (2013) in a shopping context; and Di Cagno et al. (2014) in a bidding context.

The purpose of the present study was to identify circumstances, in terms of observable conditions, in which the explanatory power of the expected-utility model for observed behaviours increases over time in repeated decision-making. Section 2 provides the theoretical economic background for such settings. To identify these circumstances, Section 3 borrows *the search context* from experimental economics. In particular, offers are independently drawn from a continuous distribution, which is either a uniform or an exponential distribution in $[0, n]$; the searcher knows these probability distributions and has a non-binding budget constraint; one offer can be presented each time and the cost of one search action is constant. Note that in this context, the searcher cannot accept a previously rejected offer.

Section 4 refers to psychological studies that attempted to identify *the initial biases* with respect to standard decision-making (i.e., unbiased expected-utility theory) in estimating probabilities and realisations. In particular, Loomes and Sugden (1982) provide a source for representing regret and rejoice, and Prelec (1998) depicts under-estimation of small probabilities and over-estimation of large probabilities. Note that these formulas were chosen for their simplicity (alternatively, the analysis could have used the two-parameter formulation by Lattimore et al., 1992) and their consistency with a dynamic context (alternatively, the analysis could have used the rank-dependent expected-utility theory of Quiggin, 1982 or the cumulative prospect theory of Tversky and Kahneman, 1992).

Section 5 borrows *the learning strategy* that has been described in neurological studies. In particular, at each point in time, individuals adapt their probability estimations to reduce their biases in estimating the probabilities of experienced events, and satisfaction evaluations to reduce their regret and rejoice perceived from experienced outcomes. The Appendix shows how this learning mechanism is related to the Bayesian updating rule. Note that the present search context differs from one-armed-bandit models, because individuals face a single stochastic machine with a known probability distribution, so no exploration is needed, whereas it is similar to reinforced learning, because individuals refer to optimal decision-making, although no long-run optimal behaviour is assumed, and thus, no long-term convergence is ensured.

Section 6 considers decision-makers who apply the reservation-value rule (i.e., they stop searching at a given time if the marginal cost of an additional search action exceeds its expected marginal benefit), consider the expected regret and rejoice, and potentially over-weight or under-weight

probabilities of the outcomes (i.e., an optimist over-estimates and a pessimist under-estimates the probability of lucky events).

Section 7 presents two main results of this analysis. Procedurally rational individuals (i.e., those who attempt to maximise their utility with imperfect information) could be better off in the short-run (i.e., few choices per reference time, or a low density of events in the reference period) by adopting *ex ante* a weighted regret and rejoice model (as opposed to the expected-utility model) if they turn out to be lucky *ex post*, this is true to a larger extent if they are obstinate (as opposed to reactive), although they will not learn, and to a greater extent in a simple world (here represented by a uniform distribution of events). In contrast, all individuals will learn to behave fully rationally (i.e., to maximise utility based on perfect information) according to the expected-utility model in the long-run (i.e., all individuals are equally lucky or unlucky, or the law of large numbers applies) in a complex world, and they will be better off than those who adopt a weighted regret and rejoice model, provided that unlucky events are more likely than lucky events. In other words, individuals learn over time to behave more rationally and optimally, on average, although this behaviour might not be optimal for a subsequent time. Thus, the model presented in this paper accounts for the econometric evidence obtained by Zagonari (1995), Loomes et al. (2002), and Ben-Elia et al. (2013).

2. Background

Three main *economic* analyses on convergence to the expected-utility model in a context of dynamic choices under uncertainty have been performed. The present study relates to these analyses as follows. First, like Agastya & Slinko (2015):

- Individuals are considered to be procedurally rational even if not fully rational.
- They do not care about the chronological order in which realisations have occurred (i.e., exchangeability).
- Realisations are not necessarily continuous.
- Preferences for all actions do not change if, in the current period, all actions receive the same (and expected) payoff (i.e., consistency).
- Path-dependence in preferences for actions across time is allowed (i.e., the history of decision-makers, or their past experiences, affect their current choices as they mould their preferences).
- Monetary rewards represent an arbitrary (but finite) set.
- The set of actions is finite.
- Preferences are based on rewards rather than on lotteries, as in Mengel & Rivas (2012).
- Preferences for actions are revised to account for the individual's history of outcomes.
- A preference relationship for any set of actions is a complete, transitive, and reflexive ordering of elements.
- The decision-maker observes the rewards of all actions, including those not chosen by the decision-maker, as in Easley & Rustichini (1999); this assumption is reasonable in a social context such as portfolio choice or labour supply, but not always reasonable in individual contexts such as route choice and consumption demand.

However, the decision-maker prefers one action over another if the former shows a larger average utility with respect to the empirical distribution of the historical rewards up to a given time; that is, the decision-maker remembers the historical frequencies of all rewards linked to all actions, and chooses actions that represent the best response to these frequencies. This assumption is reasonable in an analytical context such as portfolio choice, but not in intuitive contexts such as labour supply, route choice, or consumption demand. In addition, results by Agastya & Slinko (2015) suggest that a backward-looking (fictitious player) decision-maker will behave as a forward-looking (maximiser of expected utility) decision-maker if the stochastic process of events (rewards) is a (*restrictively*) exchangeable sequence (i.e., any permutation of events has the same probability distribution as the original sequence).

As in Oyarzun & Sarin (2013), individuals update their behaviour according to what has been observed. However, individuals show an increasing probability to choose the action suggested by the expected-utility model: this assumption is not suitable in intuitive contexts, in which an action is either chosen or not chosen. In addition, this research suggests that the learning rule converges with high probability on the set of actions that dominate all others according to a second-order stochastic criterion only if individuals are sensitive to second-order stochastic dominance in all states (*specific conditions*) and if an individual's learning is slow in response to their experience (*unobservable conditions*).

As in Gilboa & Schmeidler (1996), individuals learn by starting from a specific decision model other than expected utility. However, the set of conceivable cases is necessarily assumed to be infinite: this assumption is not suitable in intuitive contexts, in which infinity is an implausible concept to be applied. In addition, results by Gilboa & Schmeidler (1996) suggest that the decision-maker will choose actions that maximise their expected utility only if the aspiration level is (*bizarrely*) adjusted over time, both realistically (i.e., it relates to the average of its previous level and the best average performance so far) and ambitiously (i.e., it is higher than the maximum average performance sufficiently often to support this ambition).

In summary, this paper will combine in a single framework the same determinants identified by the economic literature (e.g., realised outcomes, learning rates) by disregarding the most implausible assumptions (i.e., that decision-makers remember the frequencies of all historical rewards linked to all actions, individuals choose probabilities to take an action rather than actions, and the set of conceivable cases is infinite) by adding observable insights, based on explicable assumptions, in realistic contexts (see Section 8 for details).

3. The search context

Many search contexts have been suggested by experimental economics; for recent examples, see Duffy and Puzzello (2014) and Kloosterman (2016). However, these contexts differ based on three main features. First, the *situation* that is referred to differs. For example, sometimes subjects are provided with a random offer, and choose to take that offer (Hey, 1987); sometimes searchers draw prices from a known distribution, choose the price at which they are willing to buy a fictitious commodity, and resell the commodity to the experimenter at a predetermined price (Kogut, 1990); and sometimes subjects receive an article, ask for one or more bids, and decide whether to sell the article to the experimenter for the highest bid so far (Sonnemans, 1998). Second, the *order of bids* that is given differs. Sometimes searchers face random sequences of offers, as in the case of Hey (1987) and Kogut (1990). Sometimes subjects face the same pre-selected sequences of offers; this is the case in the study by Sonnemans (1998), in both his experiment 1 (in which any search strategy was allowed) and his experiment 2 (in which the reservation-value rule was enforced). Third, the *information set* that is provided differs. In most experimental studies, subjects are sequentially presented with random offers from a known distribution. This was a normal distribution in Hey (1987), but a uniform distribution in Kogut (1990) and Sonnemans (1998).

In particular, the present study will refer to the following search context. Offers are independently drawn from a known continuous distribution, in which two alternative scenarios are considered: a uniform distribution in $[0, n]$ (i.e., $p = 1/n$) and an exponential distribution in $[0, n]$ (i.e., $p = \lambda \exp[-\lambda x]$). I define those individuals who correctly estimate probabilities as realists, and those who over-estimate and under-estimate the likelihood of bids with lower probability and higher pay-off as optimists and pessimists, respectively. The searcher has a non-binding budget constraint so that an unlimited number of searches can be assumed. One offer is presented at each point in time (t) and the cost of one search action is constant (c) so that at each time t , individuals must decide whether to accept the last bid $r(t)$ or to reject it; in the former case, they receive $r(t) - c$, but in the latter case, they pay c . The searcher cannot accept a previously rejected offer; thus, a time horizon of one period ahead is assumed.

4. Initial biases

Many biases with respect to the expected-utility decision-making have been suggested by psychological studies. For recent examples, see Glaser et al. (2012), Reyna (2012), Sherstyuk et al. (2013), Romm (2014), and Ben-Elia and Avineri (2015). However, the most relevant studies in our context refer to the biases in estimating probabilities of outcomes (section 4.1) and in predicting satisfaction from outcomes (section 4.2).

4.1. Weighted probabilities

Several *behavioural* probability-weighting functions have been suggested to account for observed choices that are inconsistent with the subjective expected-utility model. For recent examples, see Koster & Verhoef (2012), Riddell (2012), Capra et al. (2013), Charupat et al. (2013), Hensher et al. (2013), Petrova et al. (2014), Huang et al. (2015), Blavatskyy (2016), Boos et al. (2016), Keskin (2016), and Traczyk & Fulawka (2016). For the sake of simplicity, I will apply the probability-weighting function derived axiomatically by Prelec (1998), with weights (w) normalized over probabilities so that $w(0) = 0$ and $w(1) = 1$:

$$\begin{aligned} w_u[p(y)] &= \exp[-[-\ln(1/n)]^{\beta_u(t)}] \\ w_e[p(y)] &= \exp[-[-\ln(\lambda \exp^{-\lambda y})]^{\beta_e(t)}] \end{aligned}$$

where w_u represents the weight for a uniform distribution, w_e represents the weight for an exponential distribution, y represents a given outcome, n represents the largest outcome, λ represents the parameter characterising the exponential distribution, $\beta_u(t)$ and $\beta_e(t)$ represent the bias coefficients for uniform and exponential distributions (respectively) at time t , such that $\beta(t) > 1$ represents the degree of optimism, and $\beta(t) < 1$ represents the degree of pessimism. In other words, I will assume that probabilities are processed non-linearly and that the attention given to an outcome does not depend on its ranking with respect to other outcomes but only on its probability. Note that the assumption of the absence of two distinct domains, together with evidence by Cavagnaro et al. (2013) and Chechile & Barch (2013), which show that the (second) elevation parameter is domain-dependent, will justify the use of a one-parameter variant of the probability-weighting function in the case of multi-outcome (e.g., monetary) lotteries. Without loss of generality, I will fix n at 10 and approximate λ as 1, where $1/\lambda$ represents the mean of an exponential distribution, which is 0.9995 if n is 10. This lets me compare the uniform and exponential contexts by focusing on other crucial parameters.

4.2. Regret and rejoice

Several *behavioural* utility models expressed as a function of outcomes have been applied in the literature to explain observed choices that are inconsistent with the subjective expected-utility model. For recent examples, see Bracha & Brown (2012), Jouini & Napp (2012), Wolpert & Leslie (2012), Chorus & Bierlaire (2013), Krähmer & Stone (2013), Dekker (2014), Leong & Hensher (2014), Riella & Teper (2014), Qu (2015), and Charles-Cadogan (2016). For the sake of simplicity, I will assume a constant absolute risk aversion utility model as a function of outcomes:

$$u(y, t) = 1 - \exp^{-\alpha(t)y} - \int_y^n w[p(x)][1 - \exp^{-\gamma(t)x}]dx + \int_0^y w[p(x)][1 - \exp^{-\delta(t)x}]dx$$

where $\alpha(t)$, $\gamma(t)$, and $\delta(t)$ represent the degrees of risk aversion, regret, and rejoice at time t , respectively, and where $w[p(x)] = w_u[p(x)]$ and $w[p(x)] = w_e[p(x)]$ in the cases of uniform and exponential distributions, respectively. In other words, I will assume that individuals do not perceive outcomes as gains and losses, but perceive the regret from having taken a decision that *ex post* turns out to be the wrong decision and the rejoice from having taken a decision that *ex post* turns out to be the right decision. Note that the perceived regret for missing a better bid and the perceived rejoice for avoiding worse bids are weighted according to the (possibly biased) estimated probabilities of these events. For the sake of simplicity, I will assume that $\gamma(t) = \delta(t)$ for each t . This

lets me de-emphasize the learning impacts of unlucky events, without focusing on bad or good events.

5. The learning mechanism

Many alternative learning mechanisms have been suggested for estimating probabilities. For recent examples, see Fennell & Baddeley (2012), Hayashi (2012), Di Caprio et al. (2014), Balcombe & Frazer (2015), and Bisière et al. (2015). However, neuroscience (e.g., Schwartenbeck et al., 2013) suggests that individuals aim to reduce surprises. Here, individuals are assumed to compare the probability of what happened according to the perceived distribution of outcomes with the probability according to the known distribution, and to update the bias parameter β to reduce the difference between these two probabilities, in both the uniform and exponential contexts:

$$\begin{aligned}\beta_u(t+1) &= \beta_u(t) - \varepsilon_u\{1 - w_u[p(y)]/(1/n)\} \\ \beta_e(t+1) &= \beta_e(t) - \varepsilon_e\{1 - w_e[p(y)]/(\lambda \exp^{-\lambda y})\}\end{aligned}$$

Where ε_u and ε_e represent the learning rates (i.e., the inverse of the reliabilities attached to the perceived distributions of outcomes) in the uniform and exponential contexts, respectively, and y represents the experienced outcome. Note that the learning rule suggested here is *formally analogous* to Bayesian updating; that is, $p[H_i|E] = p[E|H_i] p[H_i]/p[E]$, where $p[H_i|E]$ is the posterior probability distribution over the space of hypothesis H_i for event E , $p[E|H_i]$ is the likelihood of observing event E according to generative distribution H_i , $p[H_i]$ is the prior probability distribution over the space of hypothesis H_i , and $p[E]$ is the *a priori* probability of the event E (Eberhardt and Danks, 2011). Indeed, the learning rule suggested here boils down to the Bayesian updating if $\varepsilon_u = \{\ln[1/n](2\beta_{u0} - 1) - \ln[p[H_i]]\} / \{\ln[1/n](1 - (1/n)^{\beta_{u0}-1})\}$ and $\varepsilon_e = \{y[2\beta_{e0} - 1] - \ln[p[H_i]]\} / \{y(1 - \exp[-y]^{\beta_{e0}-1})\}$, where $\beta_{u0} = \beta_u(0)$ and $\beta_{e0} = \beta_e(0)$ in the uniform and exponential contexts, respectively (see the Appendix for details). In other words, individuals are told how the world works in the long run (i.e., probability distributions on average), but do not think that the law of large numbers (i.e., the tendency of observed frequencies to approach the statistical expectation with large sample sizes) should be applied in calculating probabilities in the short-run (i.e., few choices) (Li et al., 2013). Actually, with few experiences, individuals could have events that differ from the average and thereby reinforce their bias in the short-run (Lu et al., 2016). See Van De Kuilen (2009) for findings based on *experimental* choices consistent with convergence towards linearity of the subjective probability-weighting function.

Many alternative learning mechanisms have been suggested for discovering preferences. For recent examples, see Friston et al. (2013), Hinvest et al. (2014), Shen et al. (2014), and Turi et al. (2015). However, neuroscience (e.g., Rutledge et al., 2015) suggests that individuals aim to reduce surprises. Individuals are here assumed to compare the *ex post* satisfaction obtained from experiencing an accepted bid, regret from missing better bids, and rejoice from avoiding worse bids, with the utility without regret and rejoice feelings, and to update the bias parameter γ to reduce this difference:

$$\gamma(t+1) = \gamma(t) - \zeta [1 - u(y, t)/U(y)]$$

with

$$U(y) = 1 - \exp^{-A y}$$

where ζ represents the learning rate (i.e., the inverse of the reliability attached to the assumed utility) at each time t and A represents the actual degree of risk aversion. Note that $u(t)$ and γ will be replaced by $u_u(t)$ and γ_u , $u_e(t)$ and γ_e , $u_{ub}(t)$ and γ_{ub} , and $u_{eb}(t)$ and γ_{eb} in the following sections to depict unbiased uniform, unbiased exponential, biased uniform, and biased exponential contexts, respectively. For the sake of simplicity, I will assume $\alpha(t) = \alpha$ for each t , as in Buckert et al.'s (2014) discussion of learning risk aversion. This lets me consider a potential gap between the perceived (α) and actual (A) degree of risk aversion by focusing on the dynamics of regret and rejoice. Note that individuals do *not* have meta-preferences (i.e., I would like to be an expected-utility decision-maker, but I am not, and it takes me effort and time to become this type of decision-maker). Otherwise, people would decide as an expected-utility decision-maker from the beginning.

Here, decision-makers are told how to behave in the long run (i.e., the expected-utility model, on average), but do not think that the same rules should be applied in the short run (i.e., they anticipate regret and rejoice from their choices; see Horvath and Sinha, 2015). Actually, if decision-makers are *ex ante* optimistic and *ex post* lucky, they could be better off by following a weighted regret and rejoice model so to reinforce their bias in the short-run (Chorus, 2014). See Matsumoto & Spence (2016) for findings based on reported expectations for *real* online textbook prices from a survey of 1224 college students, which is consistent with learning in beliefs about price distribution. Without loss of generality, I will assume $A = 1$, but see Analytis et al. (2014) for a discussion of multi-attribute alternatives. This amounts to a normalisation with respect to the utility obtained by an expected-utility decision-maker. Note that replacing these learning mechanisms based on comparing ratios (i.e., $u(y,t)/U(y)$) with learning mechanisms based on minimising squares (i.e., $[u(y,t) - U(y)]^2$) do not alter the results qualitatively.

6. The decision rule

Consider an individual who makes decisions at each time t by accounting for the level of satisfaction derived from experiencing a given outcome (y), together with the regret due to missing better bids (from y to n) and the rejoice due to avoiding worse bids (from 0 to y) (Bault et al., 2016). Individuals are *procedurally* rational (see Bordley and Uberti, 2015), by myopically applying the reservation rule at each time t (Shoji & Kanehiro, 2012). That is, they stop searching at a given time if the marginal cost of an additional search action exceeds its expected marginal benefit. In particular, individuals will accept a bid at time t if and only if it is larger than $r(t)$, where $r^*(t)$ solves the following implicit equation:

$$u[r(t), t] = \int_0^n w(p)[1 - \exp^{-\alpha[r(t)-c]}]dr - \int_{r(t)}^n w(p)[1 - \exp^{-\gamma(t)r(t)}]dx + \int_0^{r(t)} w(p)[1 - \exp^{-\gamma(t)r(t)}]dx$$

Where $w(p) = w_u(p)$ and $w(p) = w_e(p)$ in the cases of uniform and exponential distributions, respectively.

Individuals will *then* draw a bid y by accepting the bid if it is larger than or equal to the reservation bid [i.e., $y \geq r^*(t)$] and thereby perceiving $u(y, t)$, and by rejecting the bid if it is smaller than the reservation bid [i.e., $y < r^*(t)$] and thereby perceiving $u(-c, t)$.

Having experienced these outcomes, individuals *then* update their preference bias parameter (γ) in order to reduce the perceived regret or rejoice. *Contextually*, having observed an event y whose likelihood differs according to the estimated probability distribution and the known probability distribution, individuals update their probability bias parameter (β) in order to reduce this difference. Note that individuals are not fully *rational* in the short-run (i.e., they obtain something different from what they expected), unless $u(y,t) = U(y)$ at time t , although they could choose *optimally* in the short-run if they get more than they would obtain as suggested by the expected-utility model in the long-run.

Finally, individuals will apply the updated reservation rule at time $t+1$.

Note that the assumption that the searcher can never accept a previously rejected offer implies that the reservation bid for a risk-neutral individual ($\alpha = 0$) who is not affected by regret or rejoice ($\gamma = 0$) is the expected value, if there is no cost of search ($c = 0$): it will be 5 in the case of a uniform distribution and 1 in the case of an exponential distribution. Moreover, the reservation bid depends on how events affected the updating of the preferences. Cubitt et al. (2012) discuss the implications of a lack of evidence on the separability principle, in which behaviour is independent of history and of unreachable eventualities (i.e. events that are expected to be impossible to occur). Finally, uncertainty in the bid context might be (implicitly or explicitly) reversed with respect to labour or route contexts; this inversion is irrelevant, since individuals are procedural rational, and they do not change their reservation bid once the bid has been drawn.

7. Results

In this section, I will look for the existence of globally attractive fixed points for $\beta(t)$ and $\gamma(t)$ in alternative contexts, where individuals reach the equilibria $\beta^* = 1$ and $\gamma^* = 0$ through small steps from below if the initial values of $\beta_0 < 1$ and $\gamma_0 < 0$, and are thus smaller than the long-run optimal value, and if they are characterized by a large inertia (i.e., small learning rates ε and ζ); individuals learn through progressively smaller jumps above and below the long-run optimal value if they do not attach a sufficiently large reliability to the currently assumed degree of realism or of regret and rejoice (i.e., large learning rates ε and ζ); individuals reach the equilibrium through small steps from above if the initial values of $\beta_0 > 1$ and $\gamma_0 > 0$, and are thus larger than the long-run optimal value, and if they are characterized by a large inertia (i.e., small learning rates ε and ζ). Specifically, subsections 7.1 and 7.2 will show when individuals learn unbiased probabilities in the case of uniform and exponential probabilities, respectively, and then subsections 7.3 to 7.6 will compare learning conditions in four alternative scenarios (i.e., uniform probabilities without bias, exponential probabilities without bias, uniform probabilities with bias, and exponential probabilities with bias, respectively). This will reveal the most relevant variables in each learning process.

7.1. Learning unbiased probabilities with a uniform distribution

Consider an individual who over-estimates or under-estimates probabilities of outcomes (i.e., $\beta_u \neq 1$), but who updates their bias at separate time intervals (i.e., in a discrete time t) according to the following rule:

$$\beta_u(t+1) = \beta_u(t) - \varepsilon_u [1 - n(1/n)^{\beta_u(t)}]$$

where ε_u is the learning rate.

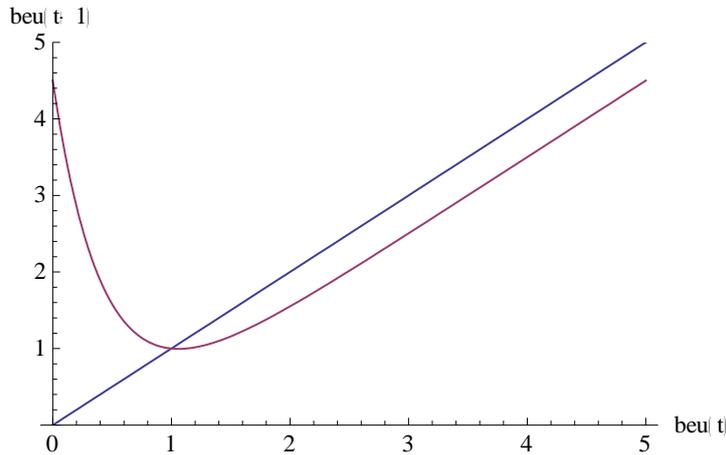
Result 1

In the case of a uniform distribution of bids, individuals learn unbiased probabilities if their learning rate is small enough: $\varepsilon \leq 0.868$.

Proof:

At $\beta_u^* = 1$, $-1 < \partial\beta_u(t+1)/\partial\beta_u(t) = 1 + \varepsilon_u \ln(1/10) < 1$.

Figure 1. Phase diagram for $\beta_u(t)$ with $\varepsilon_u = 0.5$. The increasing linear line represents $y = x$.



This result is illustrated graphically in Figure 1. Thus, $\beta_u^* = 1$ is globally stable, and the most commonly observed biases (i.e., $\beta_u > 1$) favour the learning process. If the previous learning rule is analysed in a continuous time t (i.e., individuals uninterruptedly update their bias), we obtain the following dynamics of $\beta_u(t)$:

$$\beta_u(t+1) = -\ln[10 + (1/10)^{t \varepsilon_u - \ln[-10 + 10^{\beta_u(0)}/\ln(10)]}/\ln(1/10)]$$

Note that time (t) is multiplied by the learning rate ε_u : in other words, the learning period (i.e., the learning rate multiplied by the time period) rather than the learning rate (ε_u) or time (t) taken

separately turns out to be crucial. The following dynamics of $\beta_u(t)$, which depend on the initial values β_{u0} and learning period η_u (i.e., $\eta_u = \varepsilon_u \times t$), will be used:

$$\beta_u(t) = 1 + (\beta_{u0} - 1) \exp^{-\eta_u}$$

Where β_{u0} depends on the individual's previous experiences.

7.2. Learning unbiased probabilities with an exponential distribution

Consider an individual who over-estimates or under-estimates probabilities of outcomes (i.e., $\beta_e \neq 1$), but who updates their bias according to the following rule in a discrete time t :

$$\beta_e(t+1) = \beta_e(t) - \varepsilon_e \left\{ 1 - \frac{\exp[-[-\ln(\lambda \exp^{-\lambda x})]^{\beta_e(t)}]}{(\lambda \exp^{-\lambda x})} \right\}$$

where ε_e is the learning rate.

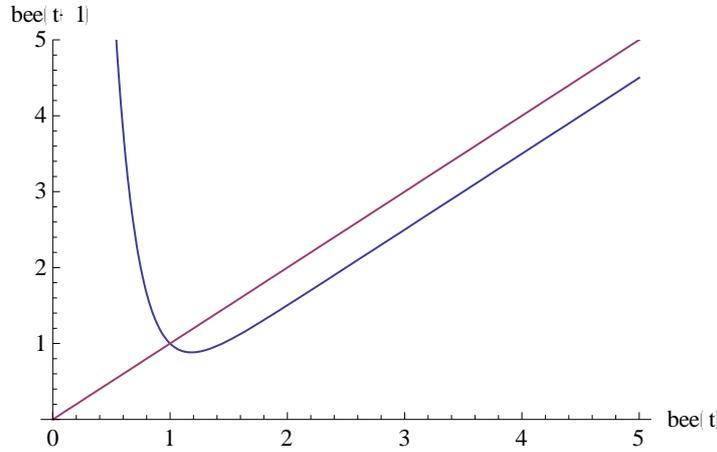
Result 2

In the case of an exponential distribution of bids, an individual learns unbiased probabilities if their learning rate is small enough: $\varepsilon_e \leq 2/y$.

Proof

At $\beta_e^* = 1$, $-1 < \partial\beta_e(t+1)/\partial\beta_e(t) = 1 - y \varepsilon_e < 1$.

Figure 2. Phase diagram for $\beta_e(t)$ with $\varepsilon_e = 0.5$ and $y = 5$. The increasing linear line represents $y = x$.



This result is illustrated graphically in Figure 2. Comparing Result 1 for a uniform distribution with Result 2 for an exponential distribution suggests that stability conditions for the exponential distribution are more demanding than those for the uniform distribution whenever $y > 2.30$. Thus, $\beta_e^* = 1$ is globally stable, and the most commonly observed biases (i.e., $\beta_e > 1$) *again* favour the learning process. If the previous learning process is analysed in a continuous time t , we obtain the following dynamics of $\beta_e(t)$:

$$\beta_e(t) = \ln[\exp^y + (\exp^{-y})^t \varepsilon_e + \beta_e^0] / y$$

Note that time t is *again* multiplied by the learning rate ε_e . The following dynamics of $\beta_e(t)$, which depend on the initial values of β_{e0} and the learning period η_e (i.e., $\eta_e = \varepsilon_e t$) will be used below:

$$\beta_e(t) = 1 + (\beta_{e0} - 1) \exp^{-\eta_e}$$

Where β_{e0} depends on the individual's previous experiences.

7.3. Learning expected utility without bias in estimating a uniform probability distribution

Consider an individual who perceives regret and rejoice (i.e., $\gamma_u \neq 0$) in accepting or rejecting bids that are uniformly distributed, but who updates their bias according to the following rule in a discrete time t (e.g., a daily labour supply without personal experiences so that $\beta_u = 1$):

$$\gamma_u(t+1) = \gamma_u(t) - \zeta_u [1 - u_u(y, t) / U(y)]$$

Where

$$u_u(y, t) = 1 - \exp^{-\alpha y} - \int_y^n (1/n)[1 - \exp^{-\gamma_u(t)x}]dx + \int_0^y (1/n)[1 - \exp^{-\gamma_u(t)x}]dx$$

$$U(y) = 1 - \exp^{-A y}$$

and ζ_u is the learning rate.

Result 3

Provided that an individual knows their absolute risk aversion (i.e., $\alpha = A$), they learn the expected utility (i.e., $\gamma_u = 0$) if their learning rate and experienced outcomes are small enough:

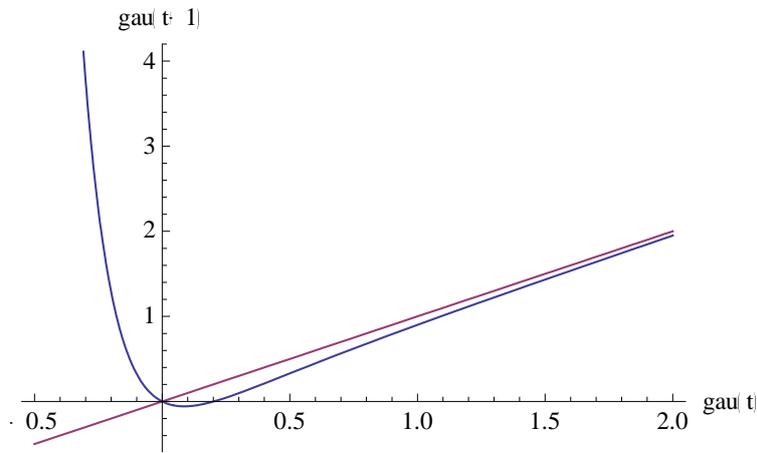
$$\zeta_u \leq 20(1 - \exp^{-y})/(50 - y^2)$$

$$y \leq 5\sqrt{2}$$

Proof

If and only if $\alpha = 1$, $\gamma_u^* = 0$ is an equilibrium for $\zeta_u \neq 0$. At $\gamma_u^* = 0$, $\partial\gamma_u(t+1)/\partial\gamma_u(t) = [-10 + \exp^y \zeta_u (10 + (-50 + y^2))]/[10(-1 + \exp^y)]$.

Figure 3. Phase diagram for γ_u with $\zeta_u = 1$ and $y = 5$. The increasing linear line represents $y = x$.



This result is illustrated graphically in Figure 3. Thus, even if an individual is not affected by biases in probability assessment, and probability distributions are simple (i.e., uniform), they learn the expected utility provided a single aspect of their psychological behaviour must be uncovered. Thus, $\gamma_u^* = 0$ is globally stable, and the most commonly observed biases (i.e., $\gamma_u > 0$) *again* favour the learning process.

Figure 4. Contour plot for $\gamma_u = 0$ stability conditions as a function of ζ_u and y . Numerical values on each line represent $\partial\gamma_u(t+1)/\partial\gamma_u(t)$ in $[-1,1]$.

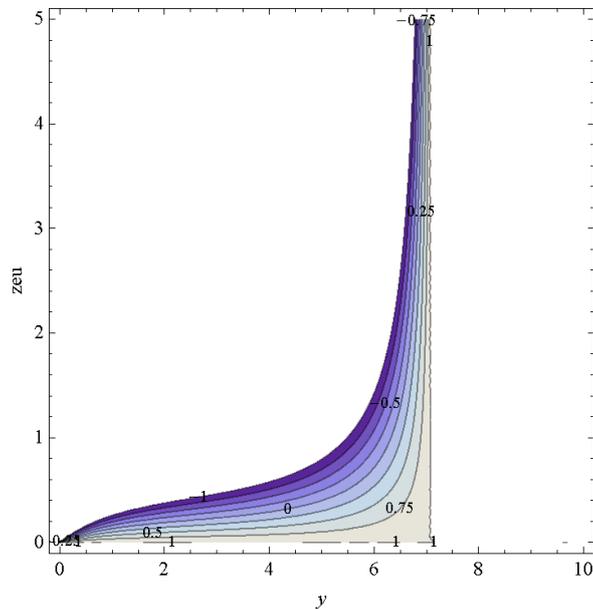


Figure 4 provides stability conditions as a function of the learning rate ζ_u and the observed outcome y , where these conditions must be met for each time t during the whole learning process.

Thus, everybody must be *moderately* unlucky (say, $y < 6/10$), but an individual who is less reactive (say, $\zeta_u < 1$) must be more unlucky. If the previous learning process is analysed in a continuous time t , we obtain the following dynamics of $\gamma_u(t)$:

$$\gamma_u(t) = \text{IF} \int_1^{\#} \frac{\exp^{10z+y} z}{2\exp^{2y+10z} [2\exp^{10z} - \exp^{yz} + 2\exp^{(y+10)z} (2y - 11)]} dz$$

Where IF means the inverse function, and

$$\# = \frac{\exp^{-y} t \zeta_u}{10(-1 + \exp^y)} + \gamma_{u0}$$

Where $\gamma_{u0} = \gamma_u(0)$. Thus, the learning period (here $\theta_u = \zeta_u \times t$) is *again* crucial. In case of convergence:

$$\gamma_u(t) = \gamma_{u0} \exp^{-\theta_u}$$

Where γ_{u0} depends on the individual's previous experiences. In particular, in the case of convergence, the difference between perceived regret and rejoice and perceived utility (i.e., $u_u - U$), which depend on the experienced outcome y and the learning time θ_u , is depicted in Figure 5.

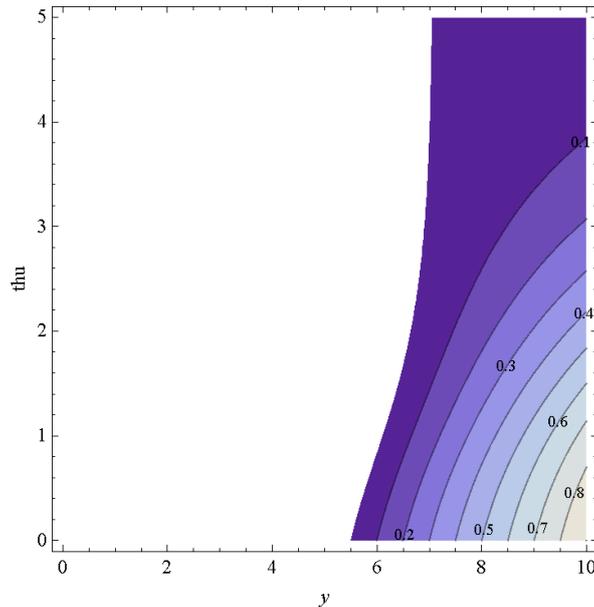
Thus, in the short-run (i.e., few experiences), individuals who are *pretty* lucky (say, $y > 6/10$) are better off if they adopt an expected regret and rejoice approach than if they adopt an expected-utility approach, and the improvement will be greater for individuals who are less reactive. Note that θ_u may increase either because individuals are more reactive (i.e., a larger ζ_u) or because they have additional experiences (i.e., a larger t). However, the mean of the difference between u_u and U over all y values (D_u) is given by:

$$D_u = \int_0^n \left(\frac{1}{n}\right) (u_u - U) dy$$

$$= \frac{1}{50} \exp^{-10\exp[-\theta_u]+\theta_u} (-5 - 5\exp^{10\exp[-\theta_u]} - \exp^{\theta_u} + \exp^{10\exp[-\theta_u]+\theta_u}) < 0$$

for each $\theta_u \geq 0$, where this difference tends to 0 if θ_u tends to infinity, the minimum D_u is -0.14 at $\theta_u = 1.25$ and its intercept is -0.08 at $\theta_u = 0$. Thus, in the long-run (i.e., many experiences), where there are no lucky or unlucky individuals (i.e., the law of large numbers applies), individuals are better off if they are reactive.

Figure 5. Gains from u_u (i.e., $u_u - U$) as a function of θ_u and y if $\gamma_{u0} = 1$. Numerical values on each line represent $u_u - U \geq 0$. The white area represents $u_u < U$.



Therefore, if individuals are reactive (say, $\zeta_u > 1$), they gain less in the short-run if they are lucky (say, $y > 6/10$), although they learn, and they lose less in the long-run once they have learned. In contrast, individuals who are obstinate (say, $\zeta_u < 1$) gain more in the short-run if they are lucky (say, $y > 6/10$), although they do not learn, and they lose more in the long-run. In other words, in the case of repeated unlucky events, reactive individuals will be better off than obstinate individuals, because the former will learn more quickly to behave according to the expected-utility approach.

7.4. Learning expected utility without bias in estimating an exponential probability distribution

Consider an individual who perceives regret and rejoice (i.e., $\gamma_e \neq 0$) in accepting or rejecting bids in the case with an exponential distribution, but who updates their bias according to the following rule in a discrete time t (e.g., a daily portfolio choice without personal experiences, so that $\beta_e = 1$):

$$\gamma_e(t+1) = \gamma_e(t) - \zeta_e[1 - u_e(y, t)/U(y)]$$

Where

$$u_e(y, t) = 1 - \exp^{-\alpha y} - \int_y^n (\lambda \exp^{-\lambda x}) [1 - \exp^{-\gamma_e(t) x}] dx + \int_0^y (\lambda \exp^{-\lambda x}) [1 - \exp^{-\gamma_e(t) x}] dx$$

$$U(y) = 1 - \exp^{-A y}$$

and ζ_e is the learning rate.

Result 4

Provided that the individual knows their absolute risk aversion (i.e., $\alpha = A$), they learn the expected utility (i.e., $\gamma_e = 0$) if their learning rate and experienced outcomes are small enough:

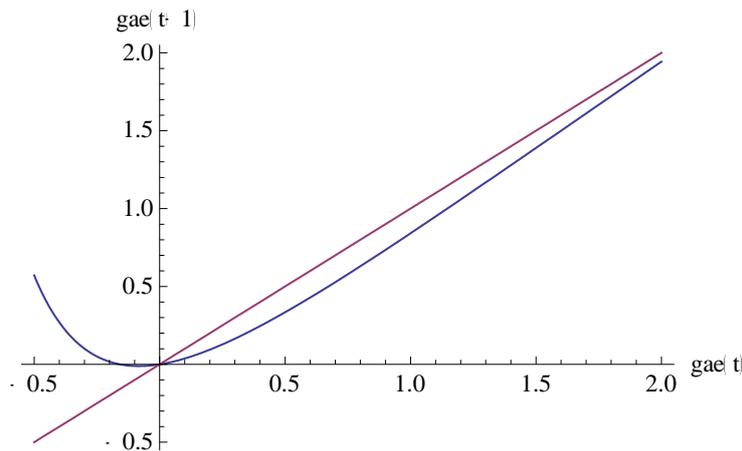
$$\zeta_e \leq - \frac{2 \exp^{10} (-1 + \exp^y)}{11 \exp^y + \exp^{10+y} - 2 \exp^{10} (1 + y)}$$

$$y \leq 1 - \text{ProductLog} \left[- \frac{11 + \exp^{10}}{2 \exp^{11}} \right]$$

Proof

If and only if $\alpha = 1$, $\gamma_e^* = 0$ is an equilibrium for $\zeta_e \neq 0$. At $\gamma_e^* = 0$, $\partial \gamma_e(t+1) / \partial \gamma_e(t) = [11 \exp^y \zeta_e + \exp^{10+y} (1 + \zeta_e) - \exp^{10} (1 + 2(1 + y) \zeta_e)] / [\exp^{10} (-1 + \exp^y)]$.

Figure 6. Phase diagram for γ_e with $\zeta_e = 1$ and $y = 5$. The increasing linear line represents $y = x$.



This result is illustrated graphically in Figure 6. Comparing Result 3 for a uniform distribution with Result 4 for an exponential distribution suggests that convergence conditions are more demanding in a more complex context: all individuals must be *pretty* unlucky (say, $y < 2/10$). Thus, $\gamma_e^* = 0$ is globally stable, and the most commonly observed biases (i.e., $\gamma_e > 0$) *again* favour the learning process.

Figure 7. Contour plot for $\gamma_e = 0$ stability conditions as a function of ζ_e and y . Numerical values on each line represent $\partial\gamma_e(t+1)/\partial\gamma_e(t)$ in $[-1,1]$.

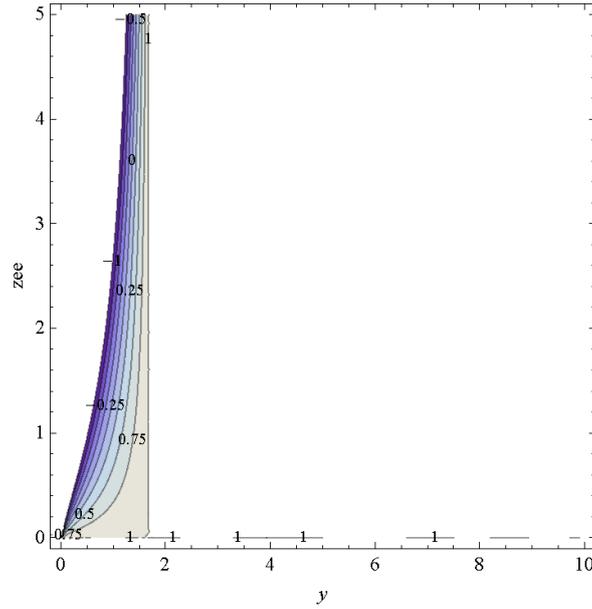


Figure 7 provides stability conditions as a function of the learning rate ζ_e and the observed outcome y , where these conditions must be met for each time t during the whole learning process.

Comparing Figure 4 and Figure 7 suggests that in a more complex world, there is less difference between reactive and obstinate people (say, ζ_e below and above 1, respectively). If the previous learning process is analysed in a continuous time t , we obtain the following dynamics of $\gamma_e(t)$:

$\gamma_e(t)$

$$= \text{IF} \int_1^{\#} \frac{\exp^{10z}(1+z)}{-\exp^y + 2\exp^{10+y+10z+y[-1-z]} + \exp^{10+y+10z}z - 2\exp^{10(1+z)}(1+z) + \exp^{y+10z}(1+z)} dz$$

Where IF means the inverse function, and

$$\# = \frac{t\zeta_e}{\exp^{10}(-1 + \exp^y)} + \gamma_{e0}$$

Where $\gamma_{e0} = \gamma_e(0)$. Thus, the learning period (here $\theta_e = \zeta_e \times t$) is *again* crucial. In the case of convergence,

$$\gamma_e(t) = \gamma_{e0} \exp^{-\theta_e}$$

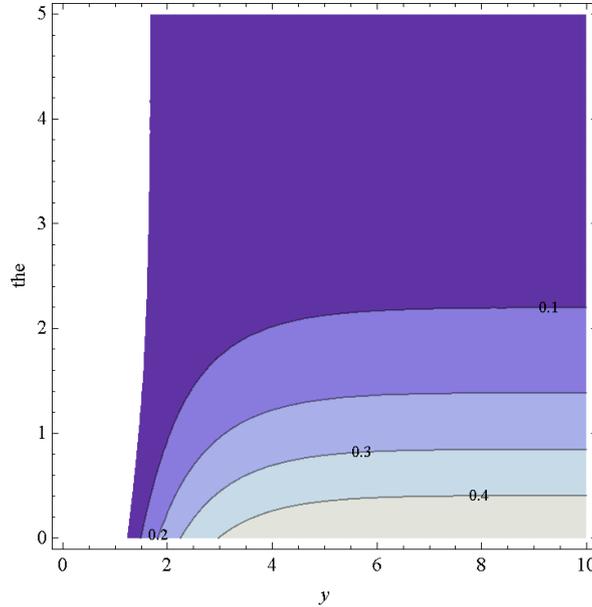
where γ_{e0} depends on the individual's previous experiences. In particular, in the case of convergence, the difference between perceived regret and rejoice and perceived utility (i.e., $u_e - U$), as a function of the experienced outcome y and the learning time θ_e , is depicted in Figure 8.

Comparing Figure 5 with Figure 8 suggests that obstinate people are better off in a simpler context. Thus, in the short-run (i.e., few experiences), individuals who are *moderately* lucky (say, $y > 1.6/10$) are *again* better off if they adopt an expected regret and rejoice approach than if they adopt an expected-utility approach, and the improvement is larger if they are less reactive. Note that θ_e may increase either because they are more reactive (i.e., a larger ζ_e) or because they have additional experiences (i.e., a larger t). However, the mean of the difference between u_e and U (D_e) over all values of y is given by:

$$\begin{aligned} D_e &= \int_0^n (\lambda y \exp^{-\lambda y})(u_e - U) dy \\ &= \frac{\exp^{-20-10\exp[-\theta_e]} + \theta_e (2\exp^{10+\theta_e}(-1 + \exp^{10\exp[-\theta_e]}) - (-1 + \exp^{10})(1 + \exp^{10(1+\exp[-\theta_e])}))}{(1 + \exp^{\theta_e})(1 + 2\exp^{\theta_e})} \\ &< 0 \end{aligned}$$

for each $\theta_e \geq 0$, where this difference tends to 0 if θ_e tends to infinity, with its minimum and its intercept (-0.15) at $\theta_e = 0$. Comparing D_u and D_e (i.e., $D_u < D_e$ for each θ) suggests that, in the long-run (i.e., many experiences), obstinate people are worse off than reactive people in a complex world, and that the magnitude of the difference is greater than in a simple world.

Figure 8. Gains from u_e (i.e., $u_e - U$) as a function of ζ_e and y . Numerical values on each line represent $u_e - U \geq 0$. The white area means $u_e < U$.



Therefore, in a more complex world, everybody must be more unlucky (i.e., $y < 16\%$ rather than $y < 66\%$) in order to learn in the long-run, whereas the difference between reactive and obstinate people is smaller in the short-run, both in terms of the relative unluckiness required to learn and in terms of gains with respect to the expected-utility choices. However, the opposite difference applies in the long-run, with reactive people being better off than obstinate people, and to a greater extent in a more complex world.

7.5. Learning expected utility with bias in estimating a uniform probability distribution

Consider an individual who misperceives uniform probabilities (i.e., $\beta_u \neq 1$) and who perceives regret and rejoice in accepting or rejecting bids in the case with a uniform distribution (i.e., $\gamma_{ub} \neq 0$), but who updates their bias according to the following rule in a discrete time t (e.g., a daily route choice with personal experiences, so $\beta_u \neq 1$):

$$\gamma_{ub}(t+1) = \gamma_{ub}(t) - \zeta_{ub}[1 - u_{ub}(y, t)/U(y)]$$

Where

$$\begin{aligned} u_{ub}(y, t) &= 1 - \exp^{-\alpha y} \\ &\quad - \int_y^n \exp[-[-\ln\left(\frac{1}{n}\right)]^{\beta u(t)}][1 - \exp^{-\gamma u(t) x}] dx \\ &\quad + \int_0^y \exp[-[-\ln\left(\frac{1}{n}\right)]^{\beta u(t)}][1 - \exp^{-\gamma u(t) x}] dx \\ U &= 1 - \exp^{-A y} \end{aligned}$$

Result 5

In the case of a biased uniform distribution of bids, an individual learns the expected utility (i.e., $\gamma_{ub} = 0$) if their learning rate and experienced outcomes are small enough:

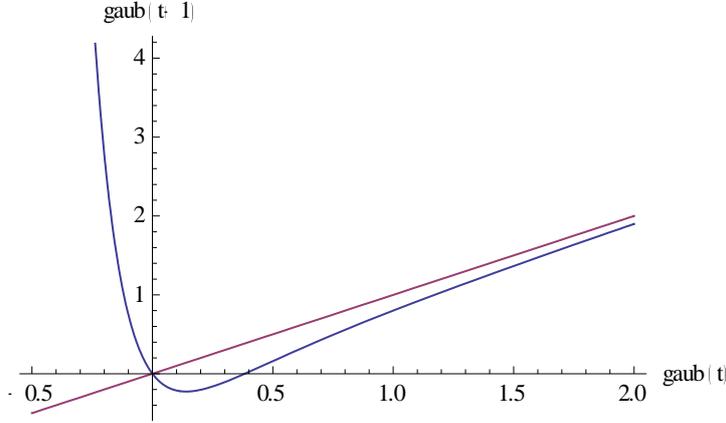
$$\zeta_{ub} \leq - \frac{2 \exp^{-y + (-1)^{(\beta u_0 - 1) \exp[-\eta u]}} \ln[10]^{1 - (\beta u_0 - 1) \exp[-\eta u]} (-1 + \exp^y)}{-50 + y^2}$$

$$y \leq 5\sqrt{2}$$

Proof

$$\text{At } \gamma_{ub}^* = 0, \partial\gamma_{ub}(t+1)/\partial\gamma_{ub}(t) = \left[-1 + \exp^y + \exp^{y-(-1)^{(-1+\beta_{u0})\exp[-\eta u] \ln[10] \exp[-\eta u] (-1+\beta_{u0}+\exp[\eta u])} \zeta_{ub}(-50 + y^2)} \right] / [-1 + \exp^y]$$

Figure 9. Phase diagram for γ_{ub} with $\zeta_{ub} = 0.001$, $\eta_u = 0$, $y = 5$, and $\beta_{u0} = 2$. The increasing linear line represents $y = x$.



This result is illustrated graphically in Figure 9. Comparing Result 5 for a biased uniform distribution with Result 3 for an unbiased uniform distribution suggests that the convergence conditions are extremely more demanding for Result 5 in terms of reactivity and equally demanding in terms of unluckiness: all individuals must be *moderately* unlucky (say, $y < 6/10$) and extremely obstinate (say, $\zeta_{ub} < 1/1000$).

Figure 10. Contour plot for γ_{ub} stability conditions as dependent on ζ_{ub} and y with $\eta_u = 1$ and $\beta_u(0) = 2$. Numerical values on each line represent $\partial\gamma_{ub}(t+1)/\partial\gamma_{ub}(t)$ in $[-1,1]$.

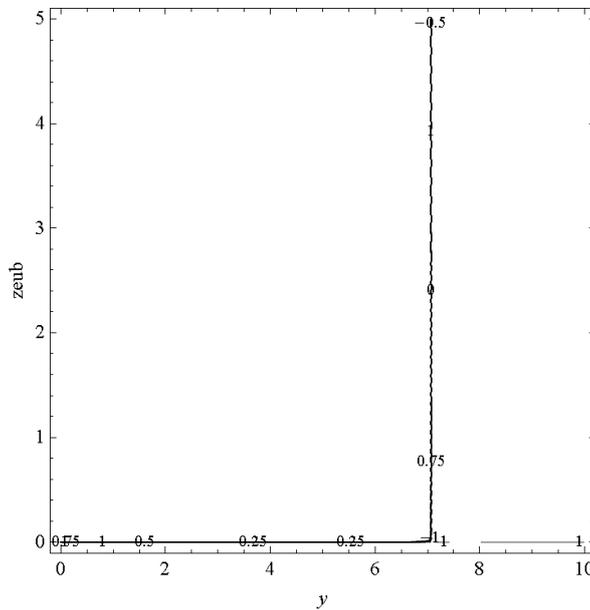


Figure 10 illustrates the stability conditions as a function of the learning rate ζ_{ub} and observed outcome y , where these conditions must be met for each time t during the whole learning process. Note that if η_u is large enough (e.g., $\eta_u > 50$) to depict previous learning of β_u , we obtain the same result obtained in section 7.4.

Therefore, in an *unrealistic* context in which people do not learn (uniform) probability distributions before learning to reduce regret and rejoice (although reducing biases in assessing the probabilities

might require a larger number of observations), only very obstinate individuals will learn to behave according to the expected-utility model, in which luck is not relevant.

7.6. Learning expected utility with bias in estimating an exponential probability distribution

Consider an individual who misperceives exponential probabilities (i.e., $\beta_e \neq 1$) and who perceives expected regret and rejoice in accepting or rejecting bids in the case with an exponential distribution (i.e., $\gamma_{eb} \neq 0$), but who updates their bias according to the following rule in a discrete time t (e.g., a daily consumption demand based on personal experiences, so $\beta_e \neq 1$):

$$\gamma_{eb}(t+1) = \gamma_{eb}(t) - \zeta_{eb}[1 - u_{eb}(y, t)/U(y)]$$

Where

$$\begin{aligned} u_{eb}(y, t) &= 1 - \exp^{-\alpha y} \\ &\quad - \int_y^n \exp[-[-\ln(\lambda \exp^{-\lambda x})]^{\beta_e(t)}][1 - \exp^{-\gamma_e(t) x}] dx \\ &\quad + \int_0^y \exp[-[-\ln(\lambda \exp^{-\lambda x})]^{\beta_e(t)}][1 - \exp^{-\gamma_e(t) x}] dx \\ U &= 1 - \exp^{-A y} \end{aligned}$$

Result 6

In the case of a biased exponential distribution of bids, an individual learns the expected utility (i.e., $\gamma_{eb} = 0$) if their learning rate and experienced outcomes are small enough:

ζ_{eb}

$$\leq \frac{2 \exp^{-\eta_e - y} (\#)^2 (\exp^y - 1)}{\exp^{\eta_e} + \exp^{-10 \# \exp[-\eta_e]} (10(-1 + \beta_{e0}) + 11 \exp^{\eta_e}) - 2(\exp^{-y})^{\# \exp[-\eta_e]} ((-1 + \beta_{e0})y + \exp^{\eta_e}(1 + y))}$$

$$\frac{1 + (1 + 10\beta_{e0})e^{-10\beta_{e0}} - 2(\exp^{-y})^{\beta_{e0}}(1 + y + (-1 + \beta_{e0})y) \leq 0}{\exp^{\eta_e} + \exp^{-10 \exp[-\eta_e](-1 + \exp[\eta_e])}(-10 + 11 \exp^{\eta_e}) - 2(\exp^{-y})^{\exp[-\eta_e](-1 + \exp[\eta_e])}(-y + \exp^{\eta_e}(1 + y)) \leq 0}$$

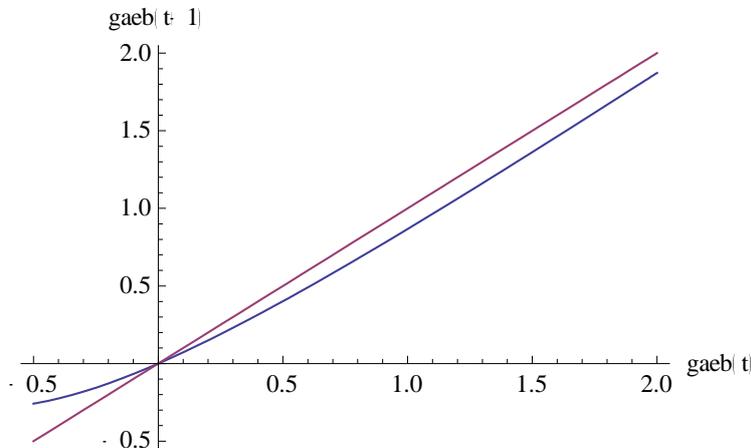
where

$$\# = -1 + \beta_{e0} + \exp^{\eta_e}$$

Proof

At $\gamma_{eb}^* = 0$, $\partial \gamma_{eb}(t+1)/\partial \gamma_{eb}(t) = 1 + [\exp^{\eta_e + y} \zeta_{eb} (\exp^{\eta_e} + \exp^{-10 \exp[-\eta_e](-1 + \beta_{e0} + \exp[\eta_e])}(-10 + 10\beta_{e0} + 11 \exp^{\eta_e}) - 2(\exp^{-y})^{\exp[-\eta_e](-1 + \beta_{e0} + \exp[\eta_e])}((-1 + \beta_{e0})y + \exp^{\eta_e}(1 + y)))]/(-1 + \beta_{e0} + \exp^{\eta_e})^2(-1 + \exp^y)$

Figure 11. Phase diagram for γ_{eb} with $\zeta_{eb} = 1$, $\eta_e = 1$, $y = 0.5$, and $\beta_{e0} = 2$. The increasing linear line represents $y = x$.



This result is illustrated graphically in Figure 11. Comparing Result 6 for a biased exponential distribution with Result 4 for an unbiased exponential distribution suggests that convergence

conditions are less demanding in terms of reactivity and more demanding in terms of unluckiness for Result 6: all individuals must be *definitely* unlucky (say, $y < 1/12$), regardless of their reactivity.

Figure 12. Contour plot for $\gamma_{eb} = 0$ stability conditions as a function of ζ_{eb} and y with $\eta_e = 1$ and $\beta_{e0} = 2$. Numerical values on each line represent $\partial\gamma_{eb}(t+1)/\partial\gamma_{eb}(t)$ in $[-1,1]$.

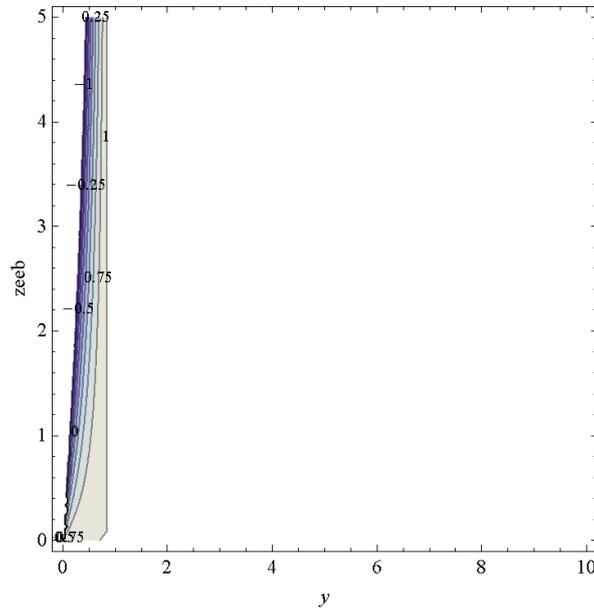


Figure 12 illustrates the stability conditions as a function of the learning rate ζ_{eb} and observed outcome y , where these conditions must be met for each time t during the whole learning process. Note that if η_e is large enough (e.g., $\eta_e > 50$) to depict previous learning of β_e , we obtain the same result as in section 7.5.

Therefore, in a *realistic* context where people do not learn (exponential) probability distributions before learning to reduce regret and rejoice (since reducing biases in assessing probabilities requires a larger number of observations), only very unlucky individuals will learn to behave according to the expected-utility model, in which their reactivity is slightly relevant.

8. Discussion

Generally speaking, by referring to keywords such as short-run vs. long-run, lucky vs. unlucky, learning vs. non-learning, reactive vs. obstinate, simple vs. complex world, better off vs. worse off., the insights obtained by this study can be summarised as follows.

In order to *learn* (in the long-run), individuals must be obstinate and unlucky in the short-run, although in a complex world, being unlucky is the most significant condition. However, individuals are better off if they are obstinate and lucky in the short-run, and the benefit is greater in a simple world, although they will be worse off in the long-run, and the drawbacks will be greater in a complex world. Note that obstinate individuals must be more unlucky than reactive individuals to learn in the long-run, and the magnitude of the effect is greater in a simple world.

Obstinate people are better off than reactive individuals in the short-run if they are lucky, although they do not learn and are worse off in the long-run, and the magnitude of the effect is greater in a complex world. However, unless individuals are repeatedly lucky, they must and will converge on the expected-utility model, and obstinate individuals must be even more unlucky than reactive people in a simple world.

In order to *be better off*, individuals must learn in the long-run and must be lucky in the short-run. However, if individuals are lucky (in the short-run), they do not need to learn and do not learn. Note that reactive people will learn more quickly (in the long-run) than obstinate people, and the magnitude of the effect will be greater in a complex world, whereas obstinate people will be better

off than reactive people, if lucky (in the short-run), and the magnitude of the effect will be greater in a simple world.

Lucky people (in the short-run) will not learn, although they will be better off in the short-run, and the magnitude of the effect will be greater in a simple world. However, if individuals are repeatedly lucky, they do not need to converge on the expected-utility model, and are even better off if they are obstinate.

The insights obtained by this paper can be summarised with respect to previous economic studies as follows. Like Gilboa & Schmeidler (1996), I show that individuals learn to behave according to the expected-utility model, but my findings do not depend on aspiration levels changing continuously. Like Oyarzun & Sarin (2013), I show that optimistic individuals choose a risky strategy, but my results do not depend on strictly preferences for second-order stochastic dominance. Like Agastya & Slinko (2015), I show that individuals will learn provided that certain events happen, but my insights do not depend on the stochastic process being an exchangeable sequence.

In summary, I combined the two main determinants of learning (i.e., reactions by individuals and the type of distribution of outcomes), but I obtained more realistic insights by starting from more realistic biases in more realistic contexts. Moreover, my results do not depend on individual characteristics that are difficult to observe (e.g., small learning rates) but on observable contextual characteristics (e.g., unlucky events that are more likely than lucky events). Finally, I develop insights by explaining how learning time is more relevant than the learning rate, where the former refers to an individual characteristic (i.e., large ε or ζ means high reactivity), whereas the latter refers to a contextual characteristic (i.e., small t means a relatively low level of repetition).

The main strengths of this approach can be summarised as follows:

1. The context is suitable to perform experiments on essential variables that lead to the expected-utility model.
2. The starting point is based on psychological studies.
3. The reason and approach to learning are based on neurological studies,
4. The endpoint is based on statistical studies.
5. The results are suitable for suggesting policies on the information that should be provided to lead to the expected-utility model.

The main weaknesses can be summarised as follows:

1. There is no exploration. *However*, exploitation is more likely to explain behaviours by the majority of individuals in real life, and exploration is meaningless here because individuals are assumed to know the probability distribution in their specific situation.
2. Specific utility functions are used. *However*, these are common to the literature and let me achieve analytical solutions, which in turn permitted a comparison of the results in alternative scenarios. This is not possible in the general learning framework used in the theoretical economics literature.
3. Regret and rejoice are weighted equally (i.e., $\delta = \gamma$). *However*, more realistic alternative assumptions (e.g., $\delta < \gamma$), which would differentiate a larger regret from being unlucky from a smaller rejoice from lucky events, would imply that individuals are expected to uncover one additional parameter, and individuals must be more unlucky.
4. A single-parameter, inverse-S-shaped probability-weighting function is used. *However*, this is well supported by empirical findings.
5. Individuals are assumed to know the probability distribution that exists for their specific situation. *However*, the comparison of scenarios in terms of their complexity suggests that in more complex worlds, where individuals are expected to learn additional parameters, convergence on the expected-utility model will require more unlucky individuals.

Note that the adopted simple search context allowed me to obtain the optimal decision at each time t in formal terms (i.e., the reservation bid). This let me explain the psychological determinants of the observed increasing adoption of the reservation rule over time obtained by experimental economics (e.g., Di Cagno et al., 2014) without relying on the context characteristics (e.g., recall vs. no-recall), and let me explain the adoption of intuitive decisions over optimal decisions at a given time, as has

been obtained by numerical economics (e.g., Sahn & von Weiszäcker, 2016), without relying on unobservable searching costs (e.g., the opportunity cost of time, the cognitive effort, and the information acquisition).

9. Conclusions

For economists who are accustomed to working with models of rational agents, it is easy to overlook how strong the assumptions of von Neumann-Morgenstern expected utility are. Experimental and empirical evidence show that agents typically act in some, but not all, circumstances according to what is predicted by this theory. The question of when the assumptions of von Neumann-Morgenstern expected utility may have greater empirical success must be addressed; in other words, the question of what leads to easy representation of the observed behaviour by those assumptions must be asked (see, for example, Klüppelberg et al., 2014).

This paper identified the contexts in which the expected-utility model can be taken as a behavioural model with the circumstances when learning is more likely to happen. In particular, learning is more probable in more realistic contexts, where a smaller number of observations are needed to reduce biases in predicted preferences than in estimated probabilities. Moreover, some individuals learn in the short-run (i.e., only unlucky individuals will learn), but all individuals learn (on average) in the long-run, whenever the probabilities of unlucky events are larger than those of lucky events. In other words, this study provides insights about the utility of misfortune. Finally, learning is more probable in more lifelike contexts, in which individuals are affected by the most frequently observed biases (i.e., anticipated or perceived regret and rejoice, over-assessment of small probabilities and under-assessment of large probabilities).

Note that the expected-utility model turns out to explain or predict individual behaviour, on average, when individuals know their degree of risk aversion. However, this feature is only weakly challenged by the economic and psychological literature.

Therefore, by applying a learning mechanism conceptually borrowed from neuroscience, although formally related to Bayesian updating, to the degree of regret or rejoice and to the degree of realism, this paper demonstrates that individuals who *repeatedly* make the same decision will behave procedurally rationally according to a non-expected-utility model at the beginning, and will behave fully rationally according to the expected-utility model at the end of the learning process. However, behaving fully optimally (on average) over time does not imply fully optimal behaviour at each time. Thus, non-standard decision-making theories seem to be most relevant only when the learning process has not been properly developed due to little experience with decisions (Birnbbaum and Schmidt, 2015; Bonnefon and Sloman, 2013).

Note that the theoretical assumptions presented in this paper may appear to be *daunting*. Indeed, the fact that individuals do not know their degree of regret or rejoice and their degree of realism implies that individuals fail to predict the satisfaction that will arise from their choices: the standard assumption that preferences are exogenous is undermined, and the alternative assumption that preferences are constructed is supported. However, the theoretical findings presented in this paper appear *encouraging*. Indeed, the fact that each individual learns their degree of regret or rejoice and becomes steadily more realistic suggests that individuals refer to a finely tuned and fully assembled kit of preferences, although they can and will apply it only to already experienced decision problems (Ortner, 2016).

Needless to say, the findings presented in this paper require further investigation, both within the same contexts (e.g., by allowing the degree of risk aversion to be updated according to experience, or the degree of regret to differ from the degree of rejoice) and in alternative contexts. For example, research should consider unknown probability distributions (Palley and Kremer, 2014), unobserved outcomes of foregone alternatives, the additional possibility of choosing actions for exploration (Gonzalez, 2013; Cox, 2015), and learning from other people's experience.

Appendix

In the case of a uniform distribution, the Bayes updating is given by:

$$p[E|Hi] = (1/n)^{\beta_{u1}} = \frac{(1/n) p[Hi]}{(1/n)^{\beta_{u0}}} = \frac{p[E|Hi] p[Hi]}{p[E]}$$

By applying logarithms to both sides, we obtain:

$$\beta_{u1} \ln[(1/n)] = (1 - \beta_{u0}) \ln[(1/n)] + \ln[p[Hi]]$$

By subtracting β_{u0} from both sides, we obtain:

$$\Delta\beta_u = (1 - 2\beta_{u0}) + \ln[p[Hi]]/\ln[(1/n)]$$

In the case of a uniform distribution, the learning rule adopted in the present study is given by:

$$\Delta\beta_u = -\varepsilon_u [1 - (1/n)^{\beta_{u0}-1}]$$

By equating the right-hand sides of the last two equations, we obtain:

$$\varepsilon_u = \frac{\ln[(1/n)](2\beta_{u0} - 1) - \ln[p[Hi]]}{\ln[(1/n)][1 - (1/n)^{\beta_{u0}-1}]}$$

In the case of an exponential distribution, the Bayes updating is given by:

$$p[E|Hi] = \exp[-y]^{\beta_{e1}} = \frac{\exp[-y] p[Hi]}{\exp[-y]^{\beta_{e0}}} = \frac{p[E|Hi] p[Hi]}{p[E]}$$

By applying logarithms to both sides, we obtain:

$$-y \beta_{e1} = -y (1 - \beta_{e0}) + \ln[p[Hi]]$$

By subtracting β_{e0} from both sides, we obtain:

$$\Delta\beta_e = (1 - 2\beta_{e0}) - \ln[p[Hi]]/y$$

In the case of a uniform distribution, the learning rule adopted in the present study is given by:

$$\Delta\beta_e = -\varepsilon_e [1 - \exp[-y]^{\beta_{e0}-1}]$$

By equating the right-hand sides of the last two equations, we obtain:

$$\varepsilon_e = \frac{y (2\beta_{e0} - 1) - \ln[p[Hi]]}{y [1 - \exp[-y]^{\beta_{e0}-1}]}$$

Note that both ε_u and ε_e are decreasing in $p[Hi]$ (i.e., individuals learn to a greater extent if they trust their *a priori* estimation to a smaller extent), with both ε_u and ε_e tending to infinity if $p[Hi]$ tends to 0 (i.e., individuals learn to the greatest extent if they trust their *a priori* estimation to the smallest possible extent), and increasing in β_u if $\beta_u > 1$ and in β_e if $\beta_e > 1$, respectively (i.e., individuals learn to a greater extent if they are affected by the most frequently observed over-assessment of small probabilities and under-assessment of large probabilities).

References

- Agastya, M., Slinko, A. (2015) Dynamic choice in a complex world, *Journal of Economic Theory* 158: 232-258
- Analytis, P.P., Kothiyal, A., Katsikopoulos, K. (2014) Multi-attribute utility models as cognitive search engines, *Judgment and Decision Making* 9: 403-419
- Balcombe, K. Fraser, I. (2015) Parametric preference functionals under risk in the gain domain: a Bayesian analysis, *Journal of Risk and Uncertainty* 50: 161-187
- Bault, N., Wydoodt, P., Coricelli, G. (2016) Different attentional patterns for regret and disappointment: an eye-tracking study, *Journal of Behavioral Decision Making* 29: 194-205
- Ben-Elia, E., Avineri, E. (2015) Response to Travel information: a behavioural review, *Transport Reviews* 35: 352-377
- Ben-Elia, E., Ishaq, R., Shifan, Y. (2013) "If only I had taken the other road...": regret, risk and reinforced learning in informed route-choice, *Transportation* 40: 269-293
- Birnbaum, M.H., Schmidt, U. (2015) The impact of learning by thought on violations of independence and coalescing, *Decision Analysis* 12: 144-152

- Bisière, C., Décamps, J.-P., Lovo, S. (2015) Risk attitude, beliefs updating, and the information content of trades: an experiment, *Management Science* 61: 1378-1397
- Blavatsky, P. (2016) Probability weighting and L-moments, *European Journal of Operational Research* 255: 103-109
- Bonnefon, J.-F., Sloman, S.A. (2013) The causal structure of utility conditionals, *Cognitive Science* 37: 193-209
- Boos, M., Seer, C., Lange, F., Kopp, B. (2016) Probabilistic inference: task dependency and individual differences of probability weighting revealed by hierarchical Bayesian modeling, *Frontiers in Psychology* 7: art. no. 755
- Bordley, R., Uberti, M. (2015) A target-oriented approach: a "one-size" model to suit humans and econs [sic] behaviors, *Applied Mathematical Sciences* 9: 4971-4978
- Bracha, A., Brown, D.J. (2012) Affective decision making: a theory of optimism bias, *Games and Economic Behavior* 75: 67-80
- Buckert, M.A., Schwieren, C.B., Kudielka, B.M.C., Fiebach, C.J. (2014) Acute stress affects risk taking but not ambiguity aversion, *Frontiers in Neuroscience* 8: art. no. 82
- Capra, C.M., Jiang, B., Engelmann, J.B., Berns, G.S. (2013) Can personality type explain heterogeneity in probability distortions? *Journal of Neuroscience, Psychology, and Economics* 6: 151-166
- Cavagnaro, D.R., Pitt, M.A., Gonzalez, R., Myung, J.I. (2013) Discriminating among probability weighting functions using adaptive design optimization, *Journal of Risk and Uncertainty* 47: 255-289
- Charles-Cadogan, G. (2016) Expected utility theory and inner and outer measures of loss aversion, *Journal of Mathematical Economics* 63: 10-20
- Charupat, N., Deaves, R., Derouin, T., Klotzle, M., Miu, P. (2013) Emotional balance and probability weighting, *Theory and Decision* 75: 17-41
- Chechile, R.A., Barch, D.H. (2013) Using logarithmic derivative functions for assessing the risky weighting function for binary gambles, *Journal of Mathematical Psychology* 57: 15-28
- Chorus, C.G. (2014) Risk aversion, regret aversion and travel choice inertia: an experimental study, *Transportation Planning and Technology* 37: 321-332
- Chorus, C.G., Bierlaire, M. (2013) An empirical comparison of travel choice models that capture preferences for compromise alternatives, *Transportation* 40: 549-562
- Cox, L.A.T., Jr. (2015) Overcoming learning aversion in evaluating and managing uncertain risks, *Risk Analysis* 35: 1892-1910
- Cubitt, R., Ruiz-Martos, M., Starmer, C. (2012) Are bygones bygones? *Theory and Decision* 73: 185-202
- Dekker, T. (2014) Indifference based value of time measures for random regret minimisation models, *Journal of Choice Modelling* 12: 10-20
- Di Cagno, D., Neugebauer, T., Rodriguez-Palmero, C., Sadrieh, A. (2014) Recall searching with and without recall, *Theory and Decision* 77: 297-311
- Di Caprio, D., Santos-Arteaga, F.J., Tavana, M. (2014) The optimal sequential information acquisition structure: a rational utility-maximizing perspective, *Applied Mathematical Modelling* 38: 3419-3435
- Duffy, J., Puzzello, D. (2014) Experimental evidence on the essentiality and neutrality of money in a search model, *Research in Experimental Economics* 17: 259-311
- Easley, D., Rustichini, A. (1999) Choice without beliefs, *Econometrica* 67: 1157-1184
- Eberhardt, F., Danks, D. (2011) Confirmation in the cognitive sciences: the problematic case of Bayesian models, *Minds and Machines* 21: 389-410

- Fennell, J., Baddeley, R. (2012) Uncertainty plus prior [sic] equals rational bias: an intuitive Bayesian probability weighting function, *Psychological Review* 119: 878-887
- Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., Dolan, R.J. (2013) The anatomy of choice: active inference and agency, *Frontiers in Human Neuroscience* September: art. no. 598 (<http://dx.doi.org/10.3389/fnhum.2013.00598>)
- Gilboa, I., Schmeidler, D. (1996) Case-based optimization, *Games and Economic Behavior* 15: 1-26
- Glaser, C., Trommershäuser, J., Mamassian, P., Maloney, L.T. (2012) Comparison of the distortion of probability information in decision under risk and an equivalent visual task, *Psychological Science* 23: 419-426
- Gonzalez, C. (2013) The boundaries of instance-based learning theory for explaining decisions from experience, *Progress in Brain Research* 202: 73-98
- Hayashi, T. (2012) Expanding state space and extension of beliefs, *Theory and Decision* 73: 591-604
- Hensher, D.A., Greene, W.H., Chorus, C.G. (2013) Random regret minimization or random utility maximization: an exploratory analysis in the context of automobile fuel choice, *Journal of Advanced Transportation* 47: 667-678
- Hey, J.D. (1987) Still searching, *Journal of Economic Behavior and Organization* 8: 137-144
- Hinvest, N.S., Brosnan, M.J., Rogers, R.D., Hodgson, T.L. (2014) fMRI evidence for procedural invariance underlying gambling preference reversals, *Journal of Neuroscience, Psychology, and Economics* 7: 48-63
- Hopfensitz, A., Van Winden, F. (2008) Dynamic choice, independence and emotions, *Theory and Decision* 64: 249-300
- Horvath, P.A., Sinha, A.K. (2015) Asymmetric reaction is rational behavior, *Journal of Economics and Finance* October, 1-20 (10.1007/s12197-015-9344-4)
- Huang, C., Burris, M., Shaw, W.D. (2015) Differences in probability weighting for individual travelers: a managed lane choice application, *Transportation* August: 1-19
- Huang, Y., Hutchinson, J.W. (2013) The roles of planning, learning, and mental models in repeated dynamic decision making, *Organizational Behavior and Human Decision Processes* 122: 163-176
- Jouini, E., Napp, C. (2012) Behavioral biases and the representative agent, *Theory and Decision* 73: 97-123
- Keskin, K. (2016) Inverse S-shaped probability weighting functions in first-price sealed-bid auctions, *Review of Economic Design* 20: 57-67
- Kloosterman, A. (2016) Directed search with heterogeneous firms: an experimental study, *Experimental Economics* 19: 51-66
- Klüppelberg, C., Straub, D., Welppe, I.M. (eds.) (2014) *Risk - A Multidisciplinary Introduction*. Springer, New York
- Kogut, C.A. (1990) Consumer search behavior and sunk costs, *Journal of Economic Behavior and Organization* 14: 381-392
- Koster, P., Verhoef, E.T. (2012) A rank-dependent scheduling model, *Journal of Transport Economics and Policy* 46: 123-138
- Krähmer, D., Stone, R. (2013) Anticipated regret as an explanation of uncertainty aversion, *Economic Theory* 52: 709-728
- Lattimore, P.K., Baker, J.R., Witte, A.D. (1992) The influence of probability on risky choice: a parametric estimation, *Journal of Economic Behavior and Organization* 17: 377-400
- Leong, W., Hensher, D.A. (2014) Relative advantage maximisation as a model of context dependence for binary choice data, *Journal of Choice Modelling* 11: 30-42

- Li, Z., Hensher, D.A., Rose, J.M. (2013) Accommodating perceptual conditioning in the valuation of expected travel time savings for cars and public transport, *Research in Transportation Economics* 39: 270-276
- Loomes, G. et al. (2002) A micro-econometric test of alternative stochastic theories of risky choice, *Journal of Risk and Uncertainty* 24(2): 103-130
- Loomes, G., Sugden, R. (1982) Regret theory: an alternative theory of rational choice under uncertainty, *The Economic Journal* 92: 805-825
- Lu, Y., Stafford, T., Fox, C. (2016) Maximum saliency bias in binocular fusion, *Connection Science* March: 1-12
- Matsumoto, B., Spence, F. (2016) Price beliefs and experience: do consumers' beliefs converge to empirical distributions with repeated purchases? *Journal of Economic Behavior and Organization* 126: 243-254
- Mengel, F., Rivas, J. (2012) An axiomatization of learning rules when counterfactuals are not observed, *B.E. Journal of Theoretical Economics* 12: art. no. 25
- Norman, A., Berman, J., Brehm, K., Drake, M., Dyer, A., Frisby, J., Govil, C., Hinchey, C., Heuer, L., Ke, J., Kejriwal, S., Kuang, K., Keyburn, S., Ler, S., Powers, K., Robertson, A., Sanghai, J., Schulze, C., Schieck, J., Sussman, J., Tan, L., Tello, A., Wang, R., Yan, K., Zeinullayev, T. (2012) Repeated price search, *Computational Economics* 39: 243-257
- Ortner, R. (2016) Optimal behavior is easier to learn than the truth, *Minds and Machines* February: 1-10
- Oyazun, C., Sarin, R. (2013) Learning and risk aversion, *Journal of Economic Theory* 148: 196-225
- Palley, A.B., Kremer, M. (2014) Sequential search and learning from rank feedback: theory and experimental evidence, *Management Science* 60: 2525-2542
- Petrova, D.G., Van der Pligt, J., Garcia-Retamero, R. (2014) Feeling the numbers: on the interplay between risk, affect, and numeracy, *Journal of Behavioral Decision Making* 27: 191-199
- Prelec, D. (1998) The probability weighting function, *Econometrica* 66: 497-527
- Qu, X. (2015) Purely subjective extended Bayesian models with Knightian unambiguity, *Theory and Decision* 79: 547-571
- Quiggin, J. (1982) A theory of anticipated utility, *Journal of Economic Behavior and Organization* 3: 323-343
- Reyna, V.F. (2012) A new intuitionism: meaning, memory, and development in fuzzy-trace theory, *Judgment and Decision Making* 7: 332-359
- Riddel, M. (2012) Comparing risk preferences over financial and environmental lotteries, *Journal of Risk and Uncertainty* 45: 135-157
- Riella, G., Teper, R. (2014) Probabilistic dominance and status quo bias, *Games and Economic Behavior* 87: 288-304
- Romm, A.T. (2014) An interpretation of focal point responses as non-additive beliefs, *Judgment and Decision Making* 9: 387-402
- Rutledge, R.B., Skandali, N., Dayan, P., Dolan, R.J. (2015) Dopaminergic modulation of decision making and subjective well-being, *Journal of Neuroscience* 35: 9811-9822
- Sahm, M., von Weizsäcker, R.K. (2016) Reason, intuition, and time, *Managerial and Decision Economics* 37: 195-207
- Schwartenbeck, P., FitzGerald, T., Dolan, R.J., Friston, K. (2013) Exploration, novelty, surprise, and free energy minimization, *Frontiers in Psychology* 4: art. no. 710
- Shen, Y., Tobia, M.J., Sommer, T., Obermayer, K. (2014) Risk-sensitive reinforcement learning, *Neural Computation* 26: 1298-1328

- Sherstyuk, K., Tarui, N., Saijo, T. (2013) Payment schemes in infinite-horizon experimental games, *Experimental Economics* 16: 125-153
- Shoji, I., Kanehiro, S. (2012) Intertemporal dynamic choice under myopia for reward and different risk tolerances, *Economic Theory* 50: 85-98
- Sonnemans, J. (1998) Strategies of search, *Journal of Economic Behavior and Organization* 35: 309-332
- Traczyk, J., Fulawka, K. (2016) Numeracy moderates the influence of task-irrelevant affect on probability weighting, *Cognition* 151: 37-41
- Turi, Z., Mittner, M., Opitz, A., Popkes, M., Paulus, W., Antal, A. (2015) Transcranial direct current stimulation over the left prefrontal cortex increases randomness of choice in instrumental learning, *Cortex* 63: 145-154
- Tversky, A., Kahneman, D. (1992) Advances in prospect theory: cumulative representation of uncertainty, *Journal of Risk and Uncertainty* 5: 297-323
- Van De Kuilen, G. (2009) Subjective probability weighting and the discovered preference hypothesis, *Theory and Decision* 67: 1-22
- Wolpert, D., Leslie, D.S. (2012) Information theory and observational limitations in decision making, *B.E. Journal of Theoretical Economics* 12: art. no. 5
- Zagonari, F. (1995) Decision making processes under uncertainty: an econometric analysis, *The Economic Journal* 105: 1403-1414



Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
<http://www.dse.unibo.it>