Judgment and Decision Making, Vol. 16, No. 1, January 2021, pp. 114–130

Myopia drives reckless behavior in response to over-taxation

Mikhail S. Spektor* Dirk U. Wulff†

Abstract

Governments use taxes to discourage undesired behaviors and encourage desired ones. One target of such interventions is reckless behavior, such as texting while driving, which in most cases is harmless but sometimes leads to catastrophic outcomes. Past research has demonstrated how interventions can backfire when the tax on one reckless behavior is set too high whereas other less attractive reckless actions remain untaxed. In the context of experience-based decisions, this undesirable outcome arises from people behaving as if they underweighted rare events, which according to a popular theoretical account can result from basing decisions on a small, random sample of past experiences. Here, we reevaluate the adverse effect of overtaxation using an alternative account focused on recency. We show that a reinforcement-learning model that weights recently observed outcomes more strongly than than those observed in the past can provide an equally good account of people's behavior. Furthermore, we show that there exist two groups of individuals who show qualitatively distinct patterns of behavior in response to the experience of catastrophic outcomes. We conclude that targeted interventions tailored for a small group of myopic individuals who disregard catastrophic outcomes soon after they have been experienced can be nearly as effective as an omnibus intervention based on taxation that affects everyone.

Keywords: safety enhancement, reliance on small samples, reinforcement learning, decisions from experience

*Department of Economics and Business, Universitat Pompeu Fabra, and Barcelona Graduate School of Economics. Email: mikhail@spektor.ch. ORCID: 0000-0003-0652-1993.

†Faculty of Psychology, University of Basel, and Max Planck Institute for Human Development, Berlin. Email: dirk.wulff@gmail.com. ORCID: 0000-0002-4008-8022.

We thank Laura Wiles for editing the manuscript. The analysis code is available at https://osf.io/q7pkf/.

Copyright: © 2021. The authors license this article under the terms of the Creative Commons Attribution 3.0 License.

*Department of Economics and Business, Universitat Pompeu Fabra, and Barcelona Graduate School of Economics. Email: mikhail@spektor.ch. ORCID: 0000-0003-0652-1993.

†Faculty of Psychology, University of Basel, and Max Planck Institute for Human Development, Berlin. Email: dirk.wulff@gmail.com. ORCID: 0000-0002-4008-8022.

1 Introduction

The real world comprises many situations where one is unsure about the outcomes ensuing from one's actions. These situations of risk are often structured such that a particular course of action results almost all of the time in small gains but also, on rare occasions, in catastrophic losses that can easily offset any previously accumulated gains. Choosing such courses of action is dangerous, yet in many situations people recklessly engage in them. For instance, people still text while driving or ride a bicycle without wearing a helmet. A recent paper (Yakobi et al., 2020, henceforth: YCNE) investigated the effectiveness of monetary incentives in the form of taxation as a means to regulate reckless behavior. YCNE studied situations where moderate taxation of a moderately risky option would lead to the desired effect of swaying people toward a safer option, but excessive taxation could drive people toward an even riskier, non-taxed option. Consequently, taxation was expected to produce a U-shaped pattern of reckless behavior, with increased recklessness for levels of taxation that are either too low or too high.

YCNE investigated this U-shaped pattern of taxation in two experiments using a decisions-from-experience task. In this task, participants made repeated decisions between three initially unknown options, comprising one relatively safe option, one moderately risky option that was subject to a tax, and one inferior, highly risky but non-taxed option (see Appendix for details). After each choice, participants would see the outcomes of all three options, allowing them to learn about the underlying properties of the options, but only the outcome of the chosen option affected the participant's bonus. Varying the level of taxation between three amounts (representing no, moderate, and excessive taxation), the expected U-shaped pattern emerged. YCNE put forth "reliance-on-small-samples" (Erev & Roth, 2014) as a mechanistic explanation of this result. According to this mechanism, people base their decisions on a random sample of k past outcomes from memory. Because small samples have a natural tendency to under-represent rare events, this mechanism produces (as-if) underweighting of rare events and, in turn, preference for reckless behaviors that offer the best outcome most of the time.

YCNE successfully demonstrated how, in decisions from experience (see Wulff et al., 2018, for a recent meta analysis), policies based on economic incentives can backfire. They attributed this to a specific cognitive mechanism, where people base their decisions on a small, random sample of past experiences. Building on their work, this article puts forth an alternative cognitive explanation, one that arguably rests on weaker assumptions and enables analysis of individual differences in people's response to taxation.

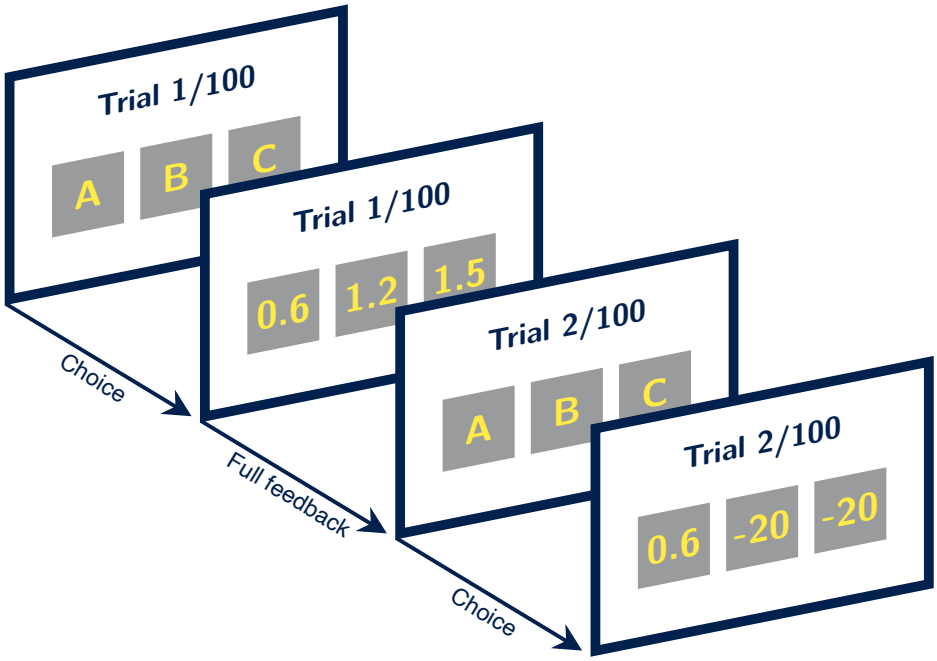

FIGURE 1: Schematic illustration of two choice trials in Experiment 2 of YCNE. Participants faced a safe option A yielding 0.60 points with certainty, a medium-risk option B yielding 2 points minus tax (here, 0.8 points) with a probability of .97 and an outcome of −20 points otherwise (the so-called accident), and a high-risk option yielding 1.5 points with a probability of .94 and −20 otherwise. The amount of tax and the properties of the safe option varied between conditions and experiments. See Appendix for details.

2 Models of reckless behavior

To identify the psychological processes that best describe people's reckless behavior, YCNE evaluated several models embodying the reliance-on-small-samples hypothesis (Erev & Roth, 2014) and a so-called full-data model.[1] The full-data model takes all previous experiences into account and deterministically predicts choice of the option that has yielded the highest average outcome. As illustrated in Figure 2, people following the full-data model should quickly develop a strong preference for the safe option as the cumulative likelihood of experiencing catastrophic events increases. However, as can also be seen, people's actual preferences developed more moderately. Furthermore, people appeared to dislike both of the two risky options less than predicted by the full-data model. Two tendencies in the data are likely responsible for these behavioral patterns: stochasticity of choices and (as-if) underweighting of rare events. Small-sample models elegantly account for these patterns using a single mechanism: A small sample of outcomes introduces stochasticity,

[1]For some comparisons, they also included the *accentuation of differences model* (Spektor et al., 2019). However, for the focus of the present investigation, this model is not of relevance.

rendering choice proportions less extreme, as well as (as-if) underweighting, accounting for higher-than-expected preference for the risky options under taxation. Consequently, the small-sample models were found to clearly outperform the full-data model across all conditions (see YCNE Tables 1 and 2).

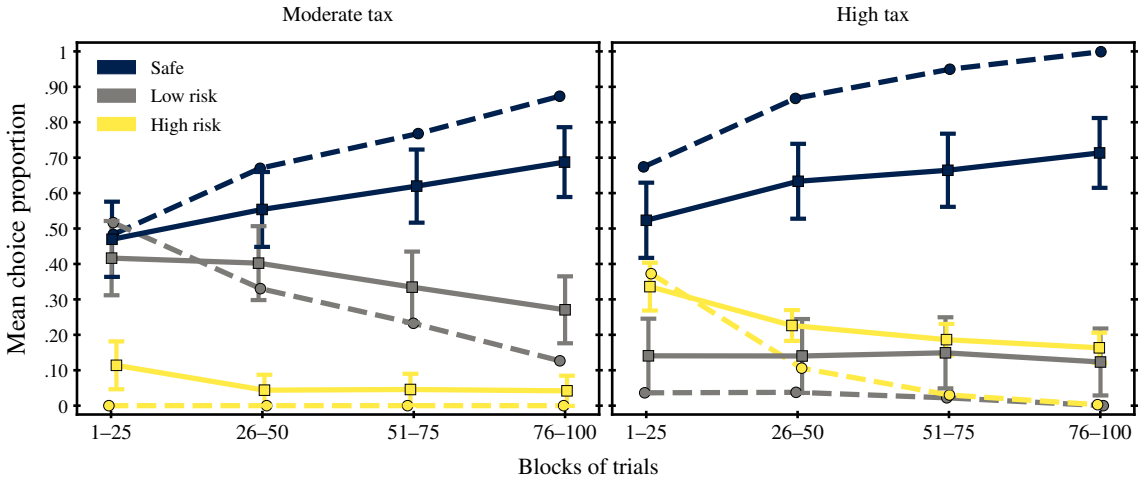

FIGURE 2: Aggregated choice proportions and predicted choice proportions of the full-data model in Experiment 1. Solid lines indicate participants' choices and dashed lines indicate the choice probabilities predicted by the model. Error bars indicate the 95% CI.

On a qualitative level, however, it is important to note that the full-data model captured the patterns of results rather well (see also YCNE Figures 3 and 5). Moreover, the sample-size parameter κ in the small-sample models was estimated to be between 24 and 47, and the overall best-performing model was an ensemble model that averages the predictions of a two-stage sampling model with those of the full-data model. These findings suggest that models that take into account many (or even all) samples might in principle be able to accurately describe people's behavior and are consistent with results from other decisions-from-experience paradigms, where the choice of the option with the higher average mean (also known as the natural-mean heuristic) is considered the benchmark model (Wulff et al., 2018).

An alternative to the full-data and small-sample models exists in recency-based models as formalized in the framework of reinforcement learning (Sutton & Barto, 1998). Recency-based models also produce probabilistic choices and (as-if) underweighting of rare events, however, via a different psychological mechanism. Such models assume that people keep track of a long-run reward expectation Q_i of option i that is updated at each time t with incoming reward (or punishment) $R_{t,i}$. If people observe a better-than-expected reward, they adjust Q upward and vice versa. A popular and simple implementation of this mechanism is given by the delta-rule model (Gershman, 2015):

$$Q_{t+1,i} = (1 - \alpha) \times Q_{t,i} + \alpha \times R_{t,i}$$

In this model, the learning rate α controls the degree to which the expectations are updated. When α is constant over time, the model inevitably produces recency, which means that recent experiences receive more weight than earlier ones. The extent of recency varies with the value of α. For instance, $\alpha = .10$ implies that an experience ten epochs ago retains about 38% of its original weight, whereas the same experience's weight essentially drops down to zero under $\alpha = .90$. Thus, α also controls the number of experiences that effectively influence choices, and with that (quite analogously to the small sample models), the degree of (as-if) underweighting of rare events. The value of α also has a limited effect on stochasticity; however, models of this class typically include extra parameters for additional sources of choice stochasticity.

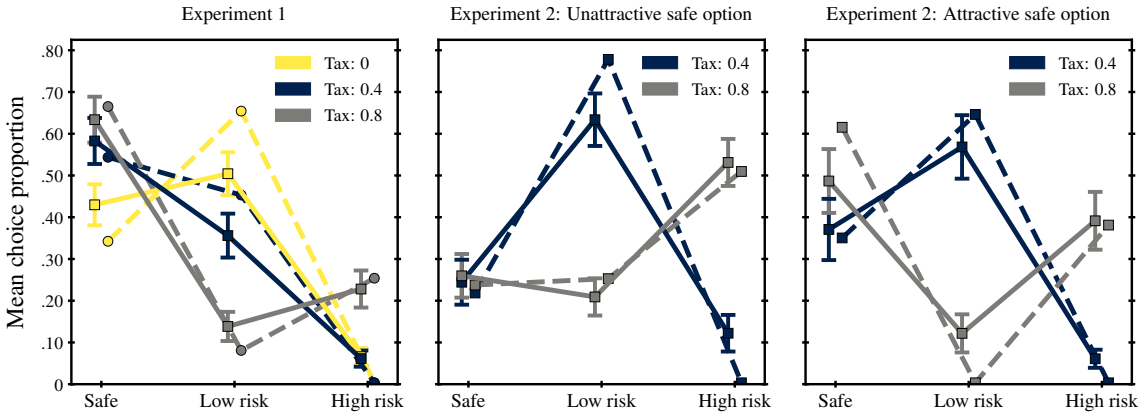

FIGURE 3: Aggregated choice proportions and predicted choice proportions of the reinforcement-learning model across the different conditions. Solid lines indicate participants' choices and dashed lines indicate the choice probabilities predicted by the model. Error bars indicate the 95% CI.

The recency-based account can be regarded as an instance of "reliance on small samples", yet it differs in important ways from the sampling-based models used by YCNE to implement this notion, which has implications for both theory and practice. First, the recency-based account can be considered more (cognitively) parsimonious. In contrast to sampling-based models, it does not require an explicit representation of all past experiences or a process of sampling from memory. Instead, people have only to memorize a single value Q and carry out only a minimal set of operations after each choice. Second, in contrast to the small-sample accounts, choices in the recency-based account always reflect all experienced information, even if their influence becomes negligible the further away they are. Third, whereas in sampling-based accounts each experience has equal sway in the long run, the recency-based account predicts that recent outcomes will influence choices more than earlier ones.

We assessed whether the recency-based account can accurately describe the data of YCNE, including people's responses to varying levels of taxation (see Appendix for technical

details and https://osf.io/q7pkf/ for the full analysis code). We fitted the delta-rule model to the aggregate choice proportions of YCNE's three between-subject conditions. The model's predictions were derived by determining, across all participants, the proportion of trials for which Q_i was highest. A single parameter (α) was used to fit, overall, 14 independent choice proportions (6 from Experiment 1, 4 from each condition from Experiment 2). A learning rate of $\alpha = .16$ yielded the best fit with a resulting mean squared error of .006. Most predictions fell within the 95% confidence interval of the observed choice proportions and the model accurately accounted for the qualitative patterns of taxation (see Figure 3). Moreover, when we used the model to predict the data of one experiment on the basis of the respective other experiment, we observed mean squared errors of .006 (Experiment 1) and .010 (Experiment 2), outperforming all sampling-based models evaluated by YCNE except for the I-SAW2 model, which achieved a slightly better performance in Experiment 2 (see Table 1 for all within- and cross-experiment predictions).

According to these aggregate-level analyses, the recency-based account given by the delta-rule model captures the aggregate data at least as well as the sampling-based accounts. However, aggregate-level analyses always bear the risk of misrepresenting the mechanisms that actually are at work at lower levels of analysis, sometimes leading to drastically wrong conclusions (e.g., Regenwetter & Robinson, 2017; Wulff & van den Bos, 2018; Birnbaum, 2011). Moreover, they can obscure crucial individual differences in both behavior and mechanism. This can be particularly problematic when a single identified mechanism serves as the basis for behavioral interventions. In the next section, we therefore use the delta-rule model to evaluate people's behavior at the individual and trial level.

TABLE 1: Aggregate-Level Comparison of Recency- and Sampling-based Models for the Data of Experiment 1 and 2 of YCNE

Experiment	Model	Fit (MSE)	Prediction (MSE)
1	reinforcement learning	.005	.006
1	full-data model	—	.014
1	naïve sampler	—	.020
1	two-stage naïve sampler	—	.009
2	reinforcement learning	.007	.010
2	full-data model	—	.040
2	extended two-stage naïve sampler	—	.007

Note. MSE = mean squared error. Fit is the MSE obtained by fitting the respective model to the choice proportions of the respective experiment. Prediction is the MSE obtained by fitting the respective model to the other experiment and predicting the respective experiment's choice proportions (in the case of non-reinforcement-learning models, the parameters were obtained by relying on other sources; see Yakobi et al., 2020, for details). Reinforcement learning = delta-rule reinforcement-learning model. Naïve sampler = small-samples model used by Yakobi et al. (2020) in Experiment 1. Two-stage naïve sampler = small-samples model that first eliminates one of the two riskier options and then compares the winner with the safe option, as used by Yakobi et al. (2020) in Experiment 1. Extended two-stage naive sampler = Two-stage Inertia, Sampling and Weighting model (I-SAW2) used by Yakobi et al. (2020) in Experiment 2.

2.1 Individual differences in recency and reckless behavior

To test the recency-based account more rigorously and address possible aggregation problems, we fitted the delta-rule model separately to each individual's trial-level choices. To achieve this, the model had to be equipped with an additional mechanism that maps subjective expectations Q to choice probabilities, accounting for the stochasticity in people's behavior (Hey & Orme, 1994). We implemented an ϵ-greedy (Sutton & Barto, 1998) choice rule which predicts the choice of the option with the highest subjective expectation with probability $1 - \epsilon$ and a randomly selected option with the error probability ϵ. In analyses reported in the Appendix, we found the ϵ-greedy choice rule to fit participants' behavior better than a popular alternative, the softmax choice rule, and, more importantly, to produce substantially lower parameter correlations, implying a cleaner separation of the psychological mechanisms.

Fitting separate learning rates α and error probabilities ϵ to each individual's choices using maximum likelihood, we observed an overall sum of Bayesian information criteria (BIC; Schwarz, 1978) of 90,101. This value was considerably lower than that of an aggregate model fitting all trial-level choices using a single learning rate α and error probability ϵ

(BIC = 109,759) and that of an aggregate baseline model assuming random guessing (BIC = 126,780). Furthermore, we found the delta-rule model to produce lower BICs for 91.1% (224 out of 246) of individuals than an individual-level baseline model.

The better performance of the individual-level models suggests meaningful individual differences, which also came through clearly in the distribution of individual-level parameter estimates: Learning rates followed a bimodal distribution (see Figure 4a), such that a vast majority of people fell into two clearly distinct groups: *myopic* and *emmetropic* learners. Myopic learners (32%) are characterized by a high learning rate of $\alpha = [.85, 1]$, implying that only the last one or two observations form the basis of their choices. Emmetropic learners (64%), on the other hand, are characterized by a low learning rate of $\alpha = (0, .15]$, implying that even the most distant experiences are still factored into their choices. The distribution of error rates, by contrast, was clearly unimodal and reflected a maximization rate of 70%, which is in line with previous research (Harless & Camerer, 1994). Furthermore, error rates barely covaried with learning rates ($r = .09$), suggesting that the estimated learning rates reflect systematic differences in people's tendency to focus on recent experiences and are not merely the result of identifiability problems known for many computational models (Spektor & Kellen, 2018).

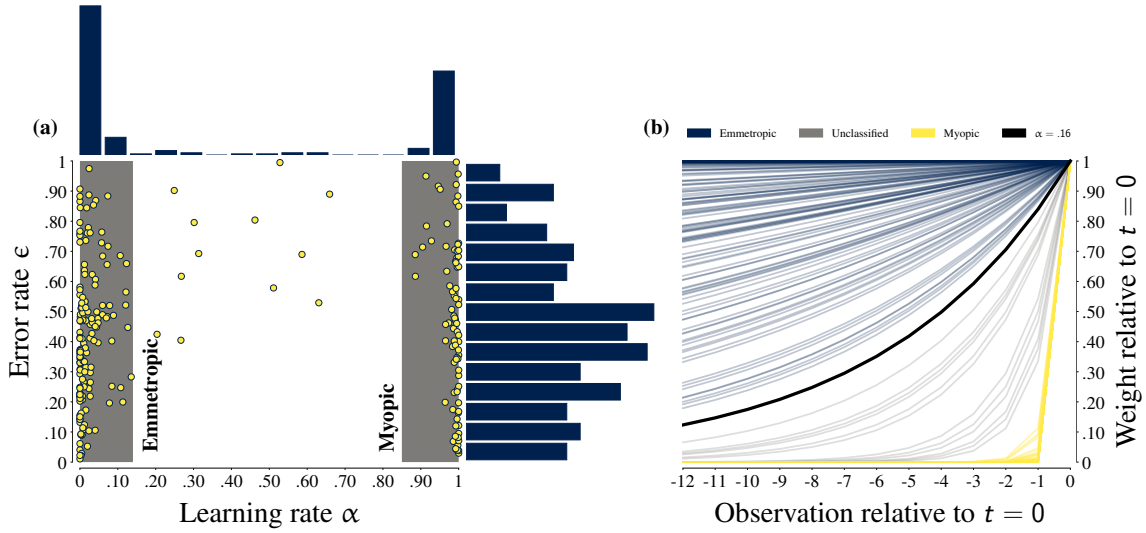

FIGURE 4: **(a)** Distribution of individual-level learning rates α and error rates ϵ across the three between-subject conditions. Parameters were estimated using maximum-likelihood estimation. Shaded areas represent classification according to $0 \leq \alpha \leq .15$ (emmetropic) or $.85 \leq \alpha \leq 1$ (myopic). **(b)** Relative weights of past experiences implied by the estimated learning rates. Each line represents one individual and the weight attached to each observation (up to 12 observations into the past).

To evaluate whether the individual differences in learning rates reflect clear and systematic differences in behavior, we plotted the modal choices of all individuals ordered by their estimated learning rate, separately for all conditions (Figure 5). This analysis revealed that

whereas most emmetropic individuals quickly learned to choose the safe option (dark blue), especially under high taxation, most myopic learners exhibited persistent preferences for whichever risky options offered the better outcome most of the time (gray = moderate risk, yellow = high risk). These patterns were most pronounced in the presence of an attractive safe option in Experiment 2.

The sustained preference for risky options observed for myopic individuals suggests that they might not have learned at all from their experiences. However, using a mixed-effects regression accounting for participant random effects nested within condition, we found preferences for the safe option to be substantially elevated immediately after the observation of an accident ($OR = 2.59, p < .001$) , but not one ($OR = 0.82, p = .45$) or two ($OR = 1.14, p = .58$) trials later, relative to all other trials. Thus, consistent with the high learning-rates estimates, myopic individuals learned about and reacted to accidents, but then discounted them very quickly as they continued. Emmetropic individuals, by contrast, showed an increased preference for the safe option not only immediately after the accident ($OR = 2.30, p < .001$), but also one ($OR = 1.38, p = .017$) and two ($OR = 1.37, p = .026$) trials later. Furthermore, consistent with the lower learning rate of emmetropic individuals, preference for the safe option right after the accident was somewhat less pronounced than for myopic individuals.

The existence of two groups of individuals has critical implications for our understanding of reckless behavior in the face of taxation. Considering only moderate- and high-taxation situations, the data showed that myopic individuals experienced, on average, 3.08 accidents, whereas emmetropic individuals experienced only 1.82 accidents (see Figure 6). More importantly, compared to moderate taxation, myopic individuals suffered 0.8 accidents more under high taxation, whereas emmetropic individuals suffered only 0.36 more accidents. These analyses suggest that myopic individuals not only suffered considerably more accidents in general, but also that they were much more susceptible to the negative effects of over-taxation.

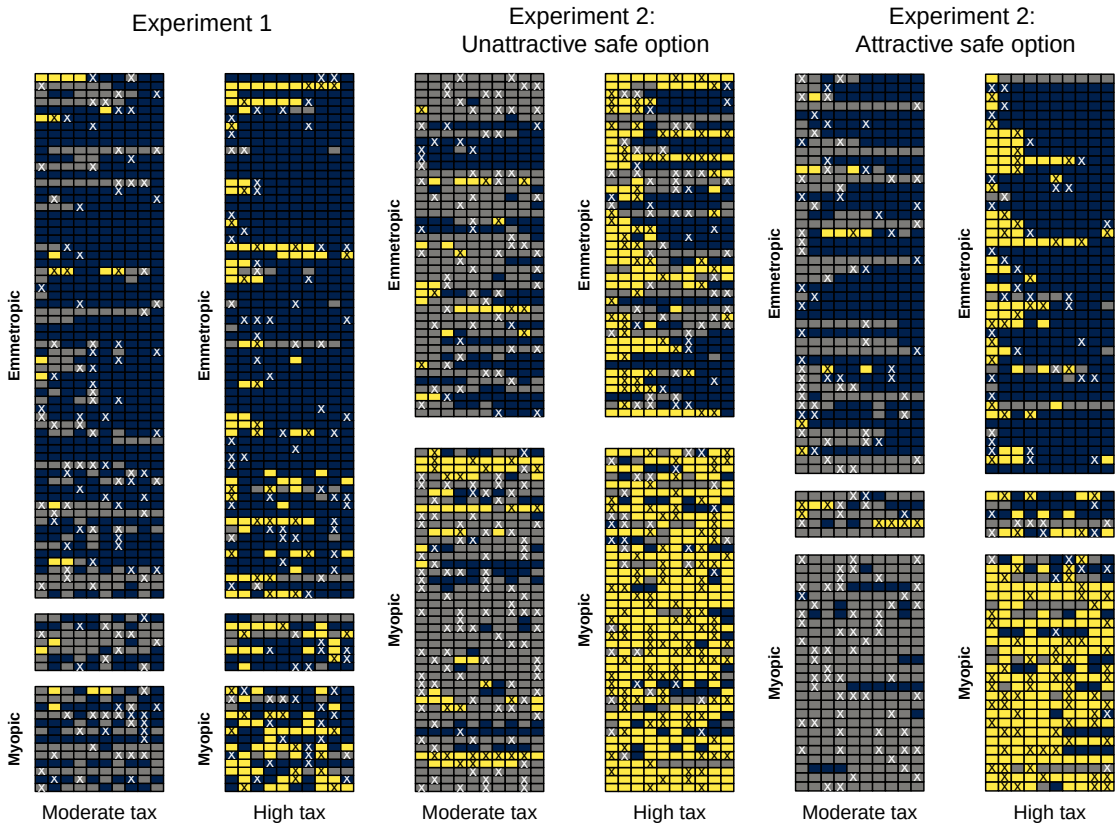

FIGURE 5: Modal choices in bins of 10 trials for each participant in each experiment, ordered by the learning rate from low (top) to high (bottom). Dark blue represents a modal choice of the safe option, gray represents of the low-risk option, and yellow of the high-risk option. Individuals are grouped according to their classification, emmetropic ($0 \leq \alpha \leq .15$), myopic ($.85 \leq \alpha \leq 1$) or unclassified. Crosses indicate that individuals suffered an accident in the corresponding bin.

3 Discussion

It is well established that people tend to choose as if they underweight small-probability events when they make decisions based on experience. This finding forms the basis of the so-called description–experience gap (Wulff et al., 2018) and it is the key to understanding people's responses to taxation in this case. As-if underweighting of rare events implies that people tend to prefer the option that yields the best outcome most of the time (Wulff et al., 2015; Erev et al., 2020). Under excessive taxation of moderately risky behaviors, an even more reckless option can suddenly become the option that is better most of the time, resulting in an increased preference for this option. The present investigation shows that different mechanisms embodying as-if underweighting can provide a good qualitative and quantitative account of how taxation affects behavior in such settings. Moreover, it

uncovered the existence of important individual differences that could be of greater import than the question of which mechanism best accounts for people's behavior. Specifically, there were two distinct groups of people, myopic and emmetropic learners, who responded to the experience of accidents in qualitatively distinct ways. Accounting for these individual differences is crucial for understanding behavior and for deriving effective policies to prevent accidents due to over-taxation.

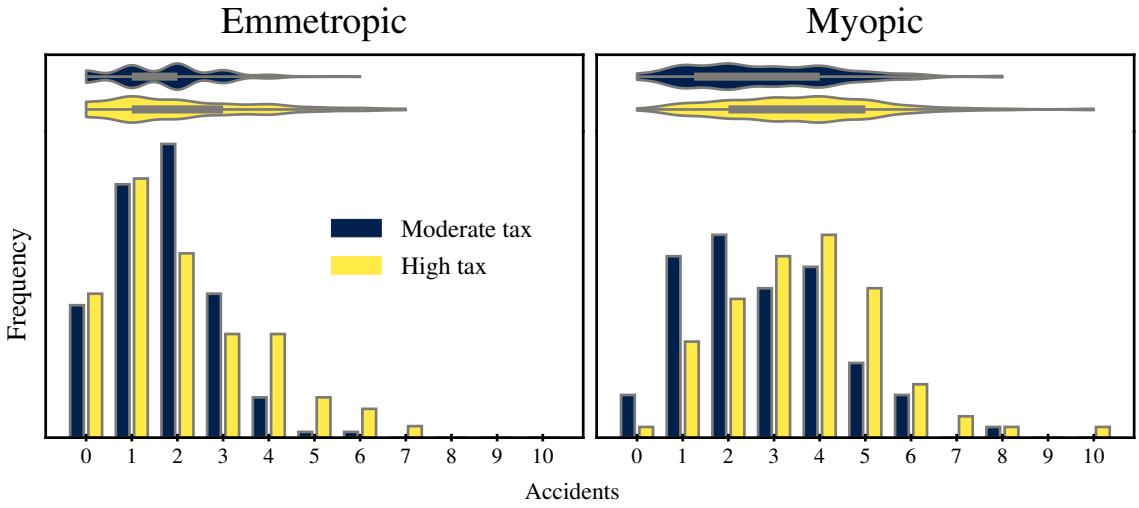

FIGURE 6: Distribution of experienced accidents, split by taxation amount. Individuals are grouped according to their classification, emmetropic ($0 \leq \alpha \leq .15$) or myopic ($.85 \leq \alpha \leq 1$). Shaded areas in the violin plots indicate the central 50% interval.

Analyses of aggregate behavior are always at risk of misrepresenting people's actual behavior (Wulff & van den Bos, 2018; Regenwetter & Robinson, 2017; Birnbaum, 2011) and, in the present investigation, this risk was real. The learning rate obtained by fitting the recency-based model to the aggregate choice proportions of both studies suggests a steady decay of the weight of past experiences, where an experience ten epochs ago receives about 20% of its original weight (see Figure 4b). However, there were almost no individuals who were accurately described by such a weighting scheme. Instead, individuals seem to assign to past outcomes a weight that is either well above that of the aggregate estimate, or one that is essentially zero. These differences imply that the groups effectively base their decision on different experiences. Furthermore, they suggest that they could have relied on different mechanisms.

Emmetropic individuals might have made their choices using a recency-based mechanism with gradually diminishing weights as formalized in the delta-rule model. However, they could have also recruited a stochastic variant of the full-data model or a sampling-based model with a large sample size. All three mechanisms actually are able to account for the behavior of emmetropic individuals equally well because, in a stable environment, a large sample of both *recent* and *random* samples will be representative of all observations.

Myopic individuals, on the other hand, cannot have relied on either the full-data model or a pure k-sampling process. For them, only the recency-based account is able to capture the high weight given to the single most recent outcome.

Despite the recency-based account's ability to fit the behavior of emmetropic and, especially, myopic individuals, we think that it does not actually provide a complete account of their psychology. For instance, recency is often attributed to either memory limitations or adaptations to assumed changes in the environment (see Wulff & Pachur, 2016; Wulff et al., 2018; Bornstein et al., 2017; Wulff & Hertwig, 2019). However, neither of these two represents a compelling account of the two extreme forms of recency observed here, namely practically no recency (emmetropic) and maximum recency (myopic). Rather, it is likely that other factors not included in the recency-based account, such as risk preferences or goals (see, e.g., Hertwig et al., 2019), also play a role. For example, the differences between emmetropic and myopic individuals can also be construed as the pursuit of short- versus long-term goals (see Wulff et al., 2015; Lopes, 1981) or as maximization versus probability matching strategies (Gaissmaier & Schooler, 2008; van den Bos et al., 2009). Moreover, there exist at least two behavioral phenomena that are difficult to reconcile with the assumptions of the delta-rule model or other reinforcement-learning accounts for that matter. First, people have been shown to possess accurate declarative memory representations of experienced samples beyond the subjective values stored by the delta-rule model (Wulff et al., 2018). Second, people have been shown to expect temporal dependencies in the outcome sequence, which can produce the wavy-recency patterns presented by YCNE. Such expectations of dependencies cannot be accounted for by any model assuming stochastic independence in the outcome distributions between choices, including the delta-rule model. To account for these phenomena, cognitive models must be equipped with mechanisms that go beyond pure small-sample or recency-based mechanics.

Notwithstanding these challenges, our findings join similar results of previous studies (Spektor & Kellen, 2018; Erev & Haruvy, 2015) in demonstrating the existence of strong individual differences in experience-based settings. These differences should be accounted for in future modeling efforts, ideally using larger and more diagnostic data sets. One promising avenue to increase diagnosticity with respect to the question of sampling- or recency-based accounts exists in reversal-learning tasks in which reward contingencies undergo sudden, drastic changes. In such situations, only recency-based accounts will allow decision makers to adaptively respond to changes in the environment (e.g., Hampton et al., 2006).

Finally, returning to the topic of reckless behavior, we believe that the presence of two groups of people relying on potentially different mechanisms has crucial implications for policy development. We have shown that myopic individuals are already at a much greater risk of suffering accidents than emmetropic individuals and that this gap widens under higher levels of taxation. A targeted policy addressing myopic individuals—for instance, by using boosts (Hertwig & Grüne-Yanoff, 2017)—might be effective over and beyond an

omnibus policy addressing everyone equally. The data suggest that had the smaller group (32%) of myopic individuals acted like emmetropic individuals, a total of 95 accidents would have been prevented. In contrast, placing everyone, myopic and emmetropic individuals, under a moderate (rather than a high) level of taxation prevented 115 accidents. Even more accidents (197) would have been prevented by the combination of both; that is, if everyone was placed under a moderate level of taxation and everyone had acted in a emmetropic fashion. This suggests that the overall best policy to prevent reckless behavior and accidents likely recruits both omnibus and targeted strategies.

References

Birnbaum, M. H. (2011). Testing mixture models of transitive preference: Comment on Regenwetter, Dana, and Davis-Stober (2011). *Psychological Review*, *118*(4), 675–683.

Bornstein, A. M., Khaw, M. W., Shohamy, D., & Daw, N. D. (2017). Reminders of past choices bias decisions for reward in humans. *Nature Communications*, *8*(1), 15958.

Erev, I. & Haruvy, E. (2015). Learning and the economics of small decisions. In J. H. Kagel & A. E. Roth (Eds.), *The Handbook of Experimental Economics*, volume 2 (pp. 638–716). Princeton University Press.

Erev, I., Plonsky, O., & Roth, Y. (2020). Complacency, panic, and the value of gentle rule enforcement in addressing pandemics. *Nature Human Behaviour*, *4*(11), 1095–1097.

Erev, I. & Roth, A. E. (2014). Maximization, learning, and economic behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 10818–10825.

Gaissmaier, W. & Schooler, L. J. (2008). The smart potential behind probability matching. *Cognition*, *109*(3), 416–422.

Gershman, S. J. (2015). Do learning rates adapt to the distribution of rewards? *Psychonomic Bulletin & Review*, *22*(5), 1320–1327.

Gershman, S. J. (2016). Empirical priors for reinforcement learning models. *Journal of Mathematical Psychology*, *71*, 1–6.

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *Journal of Neuroscience*, *26*(32), 8360–8367.

Harless, D. W. & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, *62*(6), 1251–1289.

Hertwig, R. & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, *12*(6), 973–986.

Hertwig, R., Wulff, D. U., & Mata, R. (2019). Three gaps and what they may mean for risk preference. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1766), 20180140.

Hey, J. D. & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, *62*(6), 1291–1326.

Lopes, L. L. (1981). Decision making in the short run. *Journal of Experimental Psychology: Human Learning & Memory*, *7*(5), 377–385.

Regenwetter, M. & Robinson, M. M. (2017). The construct–behavior gap in behavioral decision research: A challenge beyond replicability. *Psychological Review*, *124*(5), 533–550.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Spektor, M. S., Gluth, S., Fontanesi, L., & Rieskamp, J. (2019). How similarity between choice options affects decisions from experience: The accentuation-of-differences model. *Psychological Review*, *126*(1), 52–88.

Spektor, M. S. & Kellen, D. (2018). The relative merit of empirical priors in non-identifiable and sloppy models: Applications to models of learning and decision-making. *Psychonomic Bulletin & Review*, *25*(6), 2047–2068.

Stewart, N., Scheibehenne, B., & Pachur, T. (2018). *Psychological parameters have units: A bug fix for stochastic prospect theory and other decision models*. PsyArXiv. https://doi.org/10.31234/osf.io/qvgcd.

Sutton, R. S. & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT Press.

van den Bos, W., Güroğlu, B., van den Bulk, B. G., Rombouts, S. A., & Crone, E. A. (2009). Better than expected or as bad as you thought? The neurocognitive development of probabilistic feedback processing. *Frontiers in Human Neuroscience*, *3*, 1–11.

Wulff, D. U. & Hertwig, R. (2019). Uncovering the anatomy of search without technology. In A. Schulte-Mecklenbeck, A. Kuehberger, & J. G. Johnson (Eds.), *A Handbook of Process Tracing Methods*. Routledge, 2 edition.

Wulff, D. U., Hills, T. T., & Hertwig, R. (2015). How short- and long-run aspirations impact search and choice in decisions from experience. *Cognition*, *144*, 29–37.

Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, *144*(2), 140–176.

Wulff, D. U. & Pachur, T. (2016). Modeling valuations from experience: A comment on Ashby and Rakow (2014). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(1), 158–166.

Wulff, D. U. & van den Bos, W. (2018). Modeling choices in delay discounting. *Psychological Science*, *29*(11), 1890–1894.

Yakobi, O., Cohen, D., Naveh, E., & Erev, I. (2020). Reliance on small samples and the value of taxing reckless behaviors. *Judgment and Decision Making*, *15*(2), 266–281.

Yechiam, E. & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin & Review*, *12*(3), 387–402.

Appendix

Experimental details

Yakobi et al. (2020) conducted two experiments using a variation of the *n*-armed bandit problem (Sutton & Barto, 1998). In this task, participants repeatedly chose between three monetary lotteries for a total of 100 periods. Initially, participants had no information available about the options' outcome distributions. After each choice, participants were presented with random draws from each option, but obtained only the outcome of the chosen option (full feedback; see Figure 1 for a schematic illustration). Thereby, individuals were able to learn about the properties of the outcome distributions over time.

In every condition, participants faced a choice between a safe option, a medium-risk option, and a high-risk option. Depending on the experiment, the safe option yielded 3 points with a probability of .45 or 0 otherwise (Experiment 1), 0.60 points with certainty (Experiment 2, unattractive safe option), or 1.35 points with certainty (Experiment 2, attractive safe option). The medium-risk option yielded 2 points minus a tax with a probability of .97 and an outcome of -20 points with a probability of .03, the so-called accident. The amount of tax was implemented as a within-subject factor and varied between 0, 0.4, and .8 points in Experiment 1 and .4 and .8 points in Experiment 2. The high-risk option always yielded 1.5 points with a probability of .94 and the accident otherwise.

Eighty-five individuals (48 female, M_{age} = 35 years, SD_{age} = 11.55) took part in Experiment 1 and 161 individuals (61 female, two non-disclosures, M_{age} = 36.6 years, SD_{age} = 10.42) took part in Experiment 2, for a total of 246 participants.

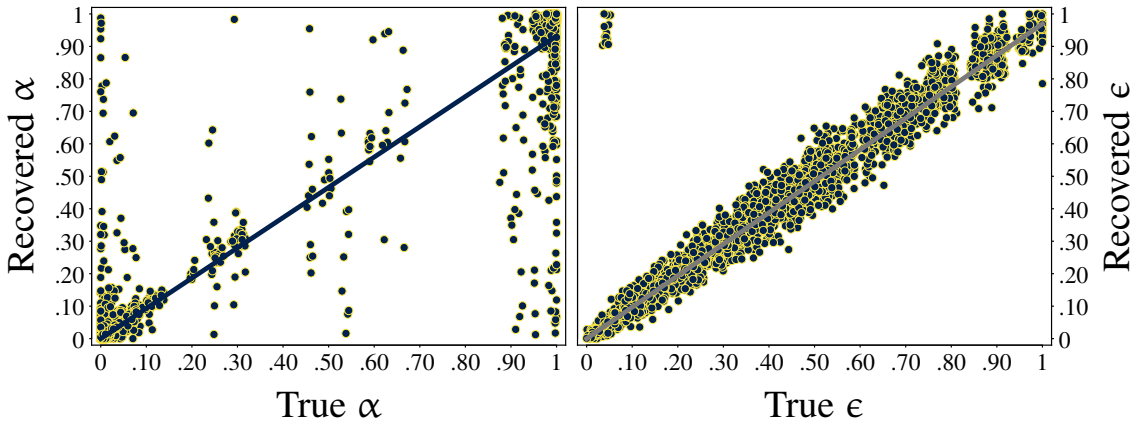

FIGURE A1: Parameter recoverability.

Modeling details

Fitting the delta-rule model to aggregate choice proportions

We fitted the delta-rule model separately to the aggregate choice proportions of each of the three between-subject conditions by minimizing mean squared error (i.e., Experiment 1, Experiment 2: Unattractive safe option, Experiment 2: Attractive safe option). Initial expectations $Q_{0,i}$ for each option i were set to 0, which is the standard procedure in the literature (e.g., Gershman, 2015; Spektor et al., 2019). On each trial, expectations are updated according to the delta-learning rule:

$$Q_{t+1,i} = (1 - \alpha) \times Q_{t,i} + \alpha \times R_{t,i},$$

where $R_{t,i}$ denotes the reward of the respective option and $\alpha \in [0, 1]$ is the only free parameter in the model and determines the degree of recency. In the case of $\alpha = 1$, individuals only consider the most recent trial, and lower values of α lead to an increasingly linear weighting scheme of all past observations. A single α was used to model chosen and non-chosen options.

Predictions for aggregate choice proportions were determined as the proportion of trials in which each of the options had the highest Q value. These predicted choice proportion were evaluated against the observed ones using mean squared error.

Fitting the delta-rule model to trial-level choices

Maximum-likelihood estimation was used to fit the delta-rule model to trial-level choices. This required that the model be endowed with a probabilistic choice rule. We used a ϵ-greedy model that chooses a randomly selected option with a probability of ϵ and the option with the highest Q value otherwise. Formally, let i on trial t be the option with the highest expectation, then the probability of choosing it will be given by $Pr(i, t) = (1 - \epsilon) + \epsilon \times \frac{1}{3}$. If there are multiple options with the highest expectation, then the model selects one of these at random. Learning followed the same implementation as for aggregate choice proportions. Parameters were estimated using the Nelder–Mead algorithm as implemented in the `scipy` Python library. We ran the algorithm 100 times with random starting values and kept the best-fitting result.

Choice of choice rule A popular alternative to ϵ-greedy, is the *softmax choice rule* (e.g., Gershman, 2015; Spektor et al., 2019):

$$Pr(i, t) = \frac{e^{\theta Q_{i,t}}}{\sum_{j=1}^{J} e^{\theta Q_{j,t}}}$$

Unlike ϵ-greedy, the softmax choice rule implements a gradual trade-off between exploration and exploitation that co-depends on the value of the sensitivity parameter θ and

differences between the options. Specifically, choices under softmax become more deterministic if the value differences between the options become large and the higher θ.

For two important reasons we relied on an ϵ-greedy choice rule rather than softmax: First, under a softmax choice rule models are known to have poor parameter identifiability (Stewart et al., 2018; Spektor & Kellen, 2018; Gershman, 2016) due to high correlations between the sensitivity and learning-rate parameters. Low parameter identifiability is highly problematic for research questions focused on interpreting parameter values, especially efforts involving the classification of individuals. Recovery analyses reported below show that identifiability was no problem under an ϵ-greedy choice rule. Second, the ϵ-greedy choice rule actually provided a better quantitative account of the individual-level data than the softmax choice rule, with 145 out of 246 individuals (59%) being better fit by the former. The ϵ-greedy choice rule also provided a better fit overall, with a sum of BICs of 90,101 compared to 97,717 for the softmax choice rule. These results are in line with reports that the ϵ-greedy choice rule is well suited to capture behavior in full-feedback paradigms (e.g., Yechiam & Busemeyer, 2005).

Parameter recovery Reinforcement-learning models are known to have poor parameter identifiability, which translates in poor recoverability (see Gershman, 2016; Spektor & Kellen, 2018, for attempts to improve identifiability). To confirm that this would not affect parameter estimation for our delta-rule model using the ϵ-greedy choice rule, we ran a parameter-recovery study. For each sequence of outcomes of an individual in the original study, we have drawn 10 random parameters from all estimated parameters across studies and simulated the choices of a "virtual" participant whose choices stemmed 100% from the reinforcement-learning model with an ϵ-greedy choice rule and the respective parameters. We re-fitted that virtual participant's choices to obtain the recovered parameters. In total, the parameter recovery was therefore based on 2,460 sets of parameters. The parameter recoverability was excellent: Both parameters yielded a near-perfect correlation between the data-generating parameters and the recovered parameters, $r_\alpha = .96$ and $r_\epsilon = .95$ (see also Figure A1).