Economic Consequences of Numerical Adaptation

Michele Garagnani*and Ferdinand M. Vieider†

February 13, 2025

Abstract

Resource constraints in neural information processing imply that numerical discriminability optimally adapts to the frequency of numerical magnitudes in a decision-maker's environment. Here, we test the economic consequences of efficient numerical range-adaptation in representative samples of the UK and Japan ($N = 2309$), as well as in a replication in Austria and Hungary ($N = 607$). We exploit natural variation in currency units and combine it with an orthogonal variation in experimental currency units to detect the effect of habitual versus non-habitual numerical ranges on the incidence of errors in decisions under risk. The results highlight the direct economic importance of numerical adaptation, thus questioning standard assumptions that choice quantities are perceived without noise.

Keywords: Numerical Adaptation · Resource constraints · Money illusion · Consistency · Errors

Introduction

Humans are adapted to living across a wide range of conditions. Given biologically-imposed limits to cognitive resources (Laughlin et al., 1998; Lieder and Griffiths, 2020; Bhui et al., 2021), optimal adaptation to specific local situations will come at the cost of increased errors for decisions that fall outside the habitual range (Louie et al., 2015). This suboptimality can be organized by a resource-constrained version of Weber's law, whereby the constraint makes adaptation to the stimuli of the environment optimal. Such optimal adaptation has long been modelled in neuroscience (Laughlin, 1981; Rustichini et al., 2017; Heng et al., 2020), psychology (Stewart et al., 2006; Stewart, 2009; Stewart et al., 2014), and economics (Robson, 2001; Netzer, 2009; Padoa-Schioppa, 2009; Khaw et al., 2017; Padoa-Schioppa and Conen, 2017; Polania et al., 2019; Khaw et al., 2021; Vieider, 2023a).

All of these approaches share the intuition that people will be more likely to misrepresent numbers which fall outside of the numerical range to which they are adapted than numbers

*Department of Finance, University of Melbourne

†RISL$\alpha\beta$, Department of Economics, Ghent University; Ferdinand Vieider gratefully acknowledges funding by the Research Foundation—Flanders under grant nr. G008021N "Causal determinants of preferences".

within that range. This, in turn, may lead to increased error rates outside the familiar range.[1] Studies investigating this issue have made great strides in accounting for hitherto unexplained paradoxes using data from laboratory experiments (Zhang et al., 2020; Khaw et al., 2021; Vieider, 2023b). The economic impact of such processes outside of laboratory settings—and hence their importance in actual economic terms—has, however, received much less attention. The question of whether range-adaptation affects real-world everyday economic decisions is, however, crucial, if we want to consider whether to modify the standard assumption that choice quantities and their value are perceived accurately.

Here, we empirically test the economic relevance of numerical range-adaptation for real-world economic decisions. We exploit natural differences in monetary denominations across countries to conduct a field test of the economic relevance of adaptation to a specific numeric range, while at the same time keeping tight experimental control, which allows us to draw causal inferences about the drivers of behaviour. This test rests on the idea that adaptation takes place with respect to *numbers*, rather than pertaining inherently to economic value (Oprea, 2022; Vieider, 2023b; Enke et al., 2023). It furthermore builds on a large literature showing how number-perception may be influenced by the specific numeric range to which a decision-maker is habituated (Piazza et al., 2004; Khaw et al., 2021; Prat-Carrabin and Woodford, 2022).

In some countries the range of prices people are exposed to on a daily basis (i.e., weekly shopping and advertisement) is between one and three digits (e.g., the Euro in many European countries, US Dollars, or GB Pounds). In other countries, however, equivalent purchases require numbers between five and seven digits (e.g., the Hungarian Forint, Czech Koruna, or Japanese Yen). We hypothesize that people in specific countries will be better at making economically relevant decisions in the numerical currency range with which they are familiar, compared to when the involved monetary amounts fall outside of their habitual range. In particular, we expect decisions falling outside of the usual currency range to be subject to more frequent and larger mistakes, which may lead to monetary losses. People used to small numerical currency ranges will thus be particularly liable to make mistakes with large numbers, whereas people habituated to large numerical values will be more liable to make mistakes with small numbers. The use of natural monetary units, in turn, ensures high external validity of our results.

In order to allow for causal inferences to be drawn, we combine the naturally-occurring variation between countries with experimentally-induced variation within countries. In particular, we compare a small-denomination currency (Great-British Pounds) to a large-denomination currency (Japanese Yen), where all monetary Pound amounts are multiplied by a factor of 180.[2] To minimize confounds deriving from economic or cultural differences, we base our tests on a symmetric setup where participants in both countries face both large and small numerical payoffs. The hypothesis entails that subjects in the UK will make more mistakes for payoffs denominated in large numerosities (i.e., when scaling up all payoffs 180-fold while keeping the underlying economic value constant). Japanese subjects, on the other hand, are predicted to make a larger number of mistakes for payoffs denominated using small numerosites (i.e., when dividing the numerical units in which a given economic value is represented by 180). To isolate the numerical effect of currency units, we conduct the experiments in both countries using identical 'experimental currency units', which can either correspond to the national currency one-to-one, or

[1]Note that while all the models we mention here make such predictions at least *in spirit*, the details of how such noise arises and over which ranges it can be detected may be different across different models. Here, we will focus on the unifying prediction of these models that the adapted range is predictive of error incidence while being agnostic about the specific model which leads to these effects.

[2]This factor is chosen to be close to the typical exchange rate between the two currencies.

have a conversion rate to map it into denominations typical for the other country (multiplying by 180 in the UK, dividing by 180 in Japan), while keeping the actual economic value of the choices constant. We chose the monetary payoffs and conversion rates so that all outcomes can be expressed using integers, in order not to introduce potential confounds deriving from the use of decimals.

We rely on lottery choices for our tests. We chose decisions under risk both for their economic relevance and because they have been used in recent publications on this topic (e.g., Webb et al., 2020; Khaw et al., 2021; Frydman and Jin, 2022; Vieider, 2023a; Garcia et al., 2023). Traditionally, choices under risk have been modelled by assuming that people's behavior can be characterized by means of stable preferences (von Neumann and Morgenstern, 1944; Kahneman and Tversky, 1979). Such *as-if* explanations, however, cannot account for recent evidence that risk taking may systematically adapt to the environment faced by the decision maker (Cohn et al., 2015; Di Falco and Vieider, 2022). They have also been criticized for making arbitrary assumptions about the origin of noise (Alós-Ferrer et al., 2021). Lottery choices are well-suited for the tests we propose because we can leverage dominance relationships between lotteries to measure error rates even in a subjective, preferential context. The use of choice under risk further allows us to quantify the "cost" of numerical range-adaptation, i.e. how much money people stand to lose. We further use representative samples of the countries to enhance the external validity of the results.

Note that any results of our manipulation could not possibly be explained away by 'calculation difficulties'. In particular, any such 'difficulties' are a manifestation of the very idea of approximate numerical judgments being aided by the adaptation we aim to study. The only principle that subjects need to follow in our setting to avoid errors is "more is better", which ought to be independent of the numerical magnitudes. The asymmetry in our design furthermore assures that smaller numbers (or for that matter, larger numbers) being generally easier to process and compare cannot possibly account for our findings, either.

Our study explicitly captures and tests predictions from several related models. One caveat with this approach is that it tests the range-adaptation hypothesis at large, rather than testing a specific model. Given that the details of the predictions can depend on fine-grained modelling assumptions, our tests are ill-suited to discriminate between different models within the adaptive class. A second caveat concerns the speed of adaptation. While a large variety of models make predictions about optimal adaptation, very few of them incorporate specific dynamic equations. We choose to see this as an opportunity. In particular, we partition the time-sequence of tasks into subsets of 40 tasks, and separately test the quantities of interest for these subsets. If adaptation is fast, we would expect the predicted patterns to occur in early batches of tasks, but to vanish over the course of the experiment. This testing sequence thus allows our experiments to be informative for future modelling exercises aiming to capture the speed of adaptation.

The numerical range-adaptation hypothesis we test may be considered optimal subject to resource constraints. Given the quick variability of numerical magnitudes in the modern world, this can nevertheless result in costly mistakes while travelling, or while trading in assets denominated in foreign currencies. Crypto-currencies, online casinos, and videogames typically entail very different numerical values than the currency denominations people are used to. People may then fall prey to "money illusion" (Shafir et al., 1997). Money illusion has been shown to not only bias individuals' decisions, but may also have profound economic consequences in the aggregate (Fehr and Tyran, 2001, 2007). Hence, this study also promises to identify the underlying mechanism leading to money illusion and explain the mixed empirical results found in the literature (e.g., Raghubir and Srivastava, 2002; Desmet, 2002; Gamble et al., 2002). The hypothesis we test further has implications for the design of experiments. Experimental currency units are often used to give the illusion of larger stakes, with the aim of increasing attention and rationality. Numerical range-adaptation, however, suggests that such procedures could have the exact opposite effect, as they may actually *increase* error rates rather than reducing them.

Experimental Design and Methods

We recruited a representative sample according to geographical location, age, gender, and income of N=2,309 participants, with 1152 respondents from Japan and 1157 respondents from the UK through Bilendi, a company providing representative subject pools for survey and experimental studies. The Japanese instructions were translated from the original English. We then had the instructions back-translated into English, and discrepancies were eliminated in discussions to arrive at the final version of the instructions.

For each country half of the participants were randomly assigned to the LOW condition and the other half to the HIGH condition. Participants took about 15 minutes to complete the study. The study was pre-registered at OSF[3]. Those failing the attention check at the beginning of the study and those who did not complete the study were excluded from the required sample size. We collected data until a sample size of at least $N = 1136$ for each country was reached (for a total of at least $N = 2272$). The sample size was determined by a power calculation in order to guarantee an a priori power of 0.95, for a non-parametric, between-subject Mann-Whitney test, for a small effect size, and a one-tailed distribution.[4] This sample size is also sufficient to guarantee that the Bayes factor is at least 10 times in favour of the experimental hypothesis over the null hypothesis, starting from an uninformative prior. Participants were blind to the aim of the study and the condition they were in.

At the beginning of the study participants were informed about the general nature of the study and we collected their informed consent to participate. They were also informed that their data would be treated completely anonymously, and that they could leave the study at any moment in case they wanted to withdraw their consent. They then received more detailed instructions informing them about the structure of the choice problem, the payment procedures, and the exchange rate between the experimental currency units in which outcomes were represented. After being presented the complete instructions and some examples of the choice tasks, participants needed to pass a simple attention check. Participants who did not pass the attention check could not continue the study and were hence excluded from the analysis.

The main part of the study required participants to make 160 binary lottery choices (see the Supplementary Information for the complete list of lotteries). Figure 1 depicts the structure of the experiment.[5] Every participant saw the same lotteries. The lotteries are constructed from 16 ordered sequences, but were presented to participants as forced binary choices in randomised order. The binary choices involve a comparison between lotteries from List A (or B) and those from List C. The latter contains only lotteries which involve sure outcomes (probabilities of 100% to get a certain amount of money). Lotteries belonging to the List A or B, on the other hand,

[3] See pre-registration document at the following link:
https://osf.io/c27jx/?view_only=e0daf412321143da899dad1bbdb9ebc8

[4] Here and throughout, we measure effect sizes using Cohen's d as a standardized measure. We refer to effect sizes of $d < 0.5$ as *small*, effect sizes between 0.5 and 0.8 as *medium*, and effect sizes of 0.8 or greater as *large*.

[5] Figure 1 reports the preregistered numbers, not the realized numbers of participants.

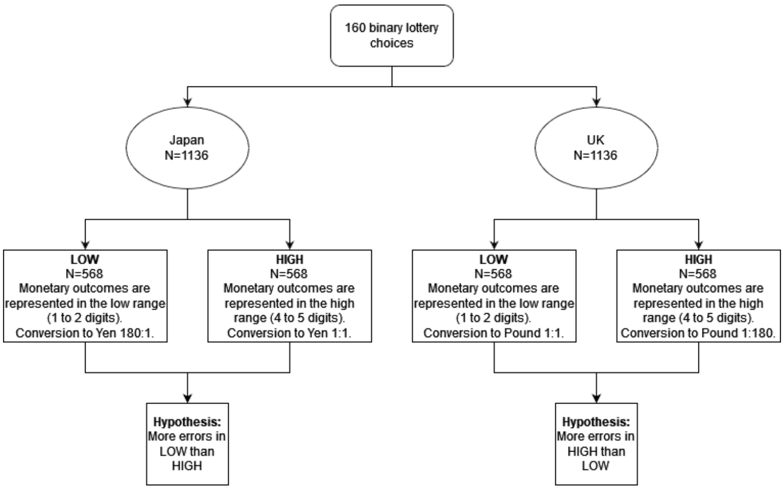

Figure 1: Treatment assignment Flow Chart

All participants see the same set of 160 lotteries. Participants randomly assigned to the LOW groups see the outcomes of the lotteries represented using 1 to 2 digits, while participants in HIGH see them represented with 4 to 5 digits. We expect participants from the UK to exhibit less errors in LOW (their familiar range) than HIGH. The opposite should obtain for participants in Japan.

have different probabilities across the different sequences, but the same probability within the same sequence. For a given position in the ordered sequence, lotteries belonging to List B involve larger outcomes than those in List A, while having the same probability. For a fixed position in the sequence lotteries in List B thus stochastically dominate lotteries in List A. For instance, take a lottery from List A, $(50\%, Out_{A_i})$, paying Out_{A_i} with probability 50%, and where i from 1 to 10 indicates the position in the list. The stochastic dominance relation implies that $Out_{A_i} < Out_{B_i}$ for each i, so that (indirect) choices between such lotteries allow us to quantify the incidence of stochastic dominance violations. As an example, the first row of the list of stimuli in Table 1 and 2 of the Supplementary Information, where $i = 1$, presents a lottery from List A, $(50\%, 26)$, paying ECU 26 with probability 50%, a lottery from List B, $(50\%, 28)$, and a lottery from List C, $(100\%, 3)$. Because B is clearly better than A, choices of A over C are considered errors if they are not accompanied by choices of B over the same C. That is, if A is preferred to C on line i, then B also needs to be preferred to C on that line. Vice versa, if C is chosen over B, then C must be also preferred to A, and all other instances are considered errors.

Another way in which we assess errors is the number of inconsistent choices within an ordered sequence of 16 lotteries (e.g., Holt and Laury, 2002; Garagnani, 2023). If a participant chooses C_n over A_n at the nth position of the ordered sequence it implies that she should choose also C_m over A_m for all $m > n$, since the lotteries improve systematically as one moves down the list. Inconsistencies in choice patterns are taken to be indicative of errors. For example, within the first list of stimuli in Table 1 and 2 of the Supplementary Information, if a participant chooses A $(50\%, Out_{A_i} = 41)$ over C $(100\%, Out_{A_i} = 18)$ at row 4 $(n = 4)$, but C $(100\%, Out_{A_i} = 23)$ over A $(50\%, Out_{A_i} = 46)$ at row 5 $(n = 5)$, then she will also choose C $(100\%, Out_{A_i} = 28)$ over A $(50\%, Out_{A_i} = 51)$ at row 6 $(n = 6)$ and for all subsequent rows. Given the construction of the lists of lotteries, the errors we can infer from participants' choices ought to be independent of their risk attitudes. In particular, based on economic principles of rationality everyone should avoid choosing dominated options (lotteries which give lower outcomes with the same probability).

In sum, avoiding errors in our setting requires no explicit calculation, but only abiding by the standard principle in economics that "more is better".

In sum, we quantify two types of decision errors, as described above:

1. The proportion of first order stochastic dominance violations at the individual level. That is, the number of times a participant indirectly stated a preference for an option in List A at line i compared to the corresponding option in List B on the same line, hence violating stochastic dominance, divided by 80, the total number of indirect comparisons.

2. The proportion of inconsistencies in choice patterns at the individual level. That is, the number of times larger than 1 a participant switched from choosing A_n over C_n (or B_n over C_n) in the ordered list divided by the total number of choices, 160.

We apply the same experimental design in two different countries, the UK and Japan (see Figure 1). We selected these countries because they have high vs. low monetary denominations, reflecting different numerical ranges people are adapted to in their daily life. The UK has low monetary denominations, while Japan has high monetary denominations (at the time we designed this study, the exchange rate was approximately 180 Yen to the Pound, which we thus adopted as the exchange rate for the experiment). In each of the two countries we randomly assigned participants to two groups. For one group (LOW) the possible outcomes of the lotteries were sampled from a range of low values (matching the one to two digits range) commonly experienced in the UK. The other group experienced only outcomes in the high range of values (HIGH, with four to five digits) commonly experienced in Japan. Note that the LOW range is constructed in such a way as to only use integers, to avoid confounds deriving from the use of amounts defined down to decimals. The low outcomes in Japan were thus obtained by dividing the Yen amounts by 180; the high outcomes in the UK were obtained by multiplying the Pound amounts by 180.

Participants were informed that there was a conversion rate, which was different depending on the country (UK vs. Japan) and the range of the outcomes (HIGH vs. LOW). That is, the payoff-relevant, randomly selected lottery determining earnings was converted to the actual currency denomination at the end of the experiment. Using four different exchange rates between experimental currency units and real currencies allowed us to keep the expected earnings of all participants the same, hence avoiding issues regarding differential expectations or effort levels between groups. The exchange rate between the HIGH and the LOW condition was based on the Pound to Yen conversion rate, which was around 1:180 when we designed the experiment. This exchange rate clearly differentiates the conditions by several orders of magnitudes. The LOW-UK group thus experienced a familiar range of outcomes between 1 and 100, while HIGH-UK saw the same lotteries with all outcomes multiplied by 180, i.e. experienced the same monetary outcomes represented by numbers in the range 180 to 18,000. The opposite happened for the two groups of Japanese participants. The design is therefore a 2×2 between subject-design (see Figure 1).

At the end of the experiment participants answered the risk elicitation task of Dohmen et al. (2011). Further demographics, such as gender, age, and income, were automatically collected as we employ a representative sample. Participants received a fixed payment of EUR 3.40 for participation to the study which took about 15 minutes. In addition, 1 out of 10 participants was randomly selected to play one of their choices for real money.

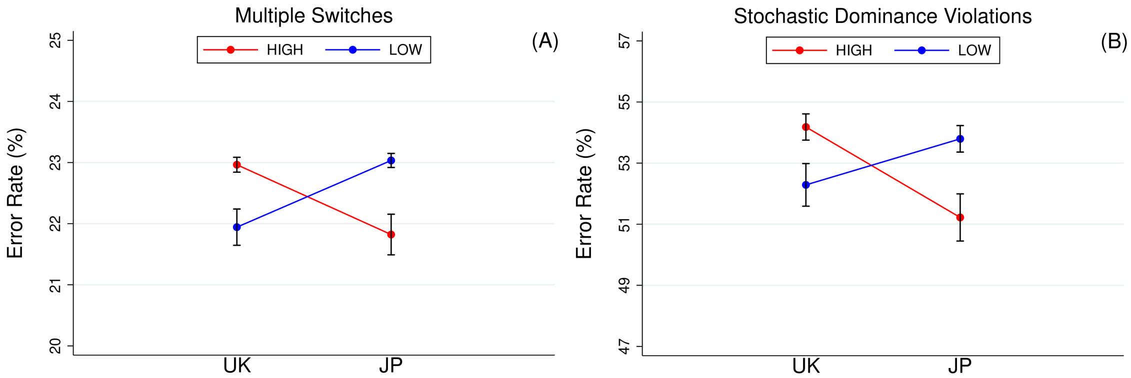

Figure 2: Error rates by country and condition

Error rates between conditions (HIGH vs. LOW) and countries (UK vs. JP) defined as the proportion of multiple switches (A) or as the proportion of stochastic dominance violations (B). According to both definition, error rates are larger for the non-adapted range of outcomes, HIGH-UK and LOW-JP, compared to the adapted range, LOW-UK and HIGH-JP. The asymmetry in error rates between countries can only be explained as range adaptation.

Results

Figure 2 shows the main results. UK participants displayed larger error rates in the HIGH condition than in the LOW condition. The exact opposite can be observed for participants from Japan, where those in the HIGH group committed fewer errors than those in the LOW group. These results support our predictions—error rates are highest outside of the habitual range, regardless whether that range entails large nominal payoffs or small nominal payoffs. Remarkably, these results are robust regardless of what definition of errors we use. If we look at error rates as the proportion of multiple-switches within each list, shown in panel A of Figure 2, we observe an error rate of 22.96% for UK-HIGH compared to 21.94% for UK-LOW, which is an increase of 4.65% in error rates from the adapted to the non-adapted range. The difference is statistically significant according to a two-sample Wilcoxon rank-sum (Mann-Whitney) test (WRS; $N = 1157$, $z = 2.71$, $p = 0.007$, CI $[0.700, 1.343]$, log Bayes factor (BF) 7.71), and corresponds to a small effect size of Cohen's $d = 0.36$ (according to an equivalent t-test $p < 0.001$). This difference in error rates translates into a statistically significant drop in subjects' earnings from 14.44 to 13.60 (5.82%, 0.84 Pounds; WRS, $N = 1157$, $z = 2.87$, $p = 0.004$, CI $[0.267, 1.411]$, BF 1.69, $d = 0.17$). We observe an error rate of 21.82% for JP-HIGH compared to 23.03% for JP-LOW, which is again an increase of about 5.55% in error rates from the adapted to the non-adapted range. The difference is again statistically significant (WRS; $N = 1152$, $z = -2.81$, $p = 0.005$, CI $[-1.559, -0.866]$, BF 11.77) and corresponds to a small effect size of $d = 0.397$. This difference in error rates translates again into a statistically significant drop in earnings from 6.74 to 5.87 (12.91%, -0.87 Pounds; WRS, $N = 1152$, $z = 2.07$, $p = 0.038$, CI $[0.048, 1.684]$, BF 3.30, $d = 0.12$).

We can directly test for the asymmetry between countries with a difference-in-difference test, which reproduces the results of the single tests. In particular, a panel regression with errors as dependent variable, dummies for country (UK vs. JP) and condition (HIGH vs. LOW), and their respective interactions tells us that the asymmetry is statistically significant (Model 1 of Table 1; mean -2.234, sd 0.241, $p < 0.001$, CI $[-2.706; -1.762]$). Model 2 of Table 1 shows that the results remain qualitatively unchanged when we control for demographic factors such as age, income, gender, and for risk attitudes.

	Multiple Switches		Dominance Violations	
Error Rate	Model 1	Model 2	Model 3	Model 4
HIGH	1.022***	0.999***	1.895***	1.975***
	(0.170)	(0.171)	(0.432)	(0.435)
JP	1.092***	1.046***	1.508***	1.352***
	(0.169)	(0.177)	(0.430)	(0.449)
HIGH × JP	-2.234***	-2.192***	-4.468***	-4.630***
	(0.241)	(0.245)	(0.611)	(0.622)
Age		-0.000		0.042***
		(0.004)		(0.010)
Female		0.005		0.027
		(0.125)		(0.318)
Income		0.001		-0.001**
		(0.000)		(0.001)
Risk Attitude		-0.027		-0.118*
		(0.027)		(0.069)
Constant	21.940***	22.090***	52.290***	50.860***
	(0.120)	(0.305)	(0.305)	(0.775)
N (subjects)	2,309	2,221	2,309	2,221
R^2	0.0362	0.0362	0.0253	0.0379

Table 1: Random-effect regression of error rates

defined as the number of multiple switches (models 1 and 2) or the number of stochastic dominance violtions (models 3 and 4). Standard errors in parentheses are clustered at the participant level, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The error rates of the two countries when comparing their respective adapted and non-adapted range are virtually indistinguishable (UK-LOW vs. JP-HIGH; WRS, $N = 1147$, $z = 0.37$, $p = 0.712$; UK-HIGH vs. JP-LOW; WRS, $N = 1162$, $z = -0.64$, $p = 0.525$) indicating that the groups are otherwise comparable. Participants are very similar with regards to other characteristics. We find no difference in declared risk tolerance between conditions within each country (UK: WSR; $N = 1157$, $z = 0.82$, $p = 0.412$; JP: WSR; $N = 1152$, $z = 1.19$, $p = 0.236$). We do observe a significant difference in declared risk attitudes (Falk et al., 2018) between countries, with JP participants reporting less risk aversion than UK subjects (WSR, $N = 2309$, $z = -14.53$, $p < 0.001$), which can however not explain our results. While it is not a requisite for our hypotheses that the countries be very similar, it strengthens the conclusion that the main difference between the asymmetry of error rates is due to the different numerical ranges to which people in the different countries are adapted.

If we look at error rates as the proportion of time participants violated stochastic dominance between the paired lists, we observe qualitatively identical results (right-hand panel of Figure 2). That is, an error rate of 54.18% for UK-HIGH compared to 52.29% for UK-LOW. The difference is statistically significant (WRS; $N = 1157$, $z = 2.46$, $p = 0.014$, CI $[1.075, 2.714]$, BF 0.12, and corresponds to a small effect size of $d = 0.27$). We once again observe the opposite pattern for Japanese participants, with an error rate of 51.22% for JP-HIGH compared to 53.79% for JP-LOW (WRS; $N = 1152$, $z = -3.44$, $p < 0.001$, CI $[-3.450, -1.700]$, BF 6.36, $d = 0.34$). We again find that the baseline error rates for adapted and non-adapted ranges is comparable between the two countries (UK-LOW vs. JP-HIGH; WRS, $N = 1147$, $z = -1.96$, $p = 0.050$;

UK-HIGH vs. JP-LOW; WRS, $N = 1162$, $z = 1.16$, $p = 0.245$). A test of the difference-in-difference confirms that the asymmetry is significant (Model 3 of Table 1; mean -4.468, sd 0.611, $p < 0.001$, CI $[-5.667; -3.270]$). Model 4 of Table 1 shows that the results are again robust to controlling for demographic factors such as age, income, gender, and risk attitudes.

The number of stochastic dominance violations may seem high, falling in the range of 51% to 54%. Typical figures reported in the literature tend to be significantly smaller at about 15% (Khaw et al., 2021; Alós-Ferrer and Garagnani, 2022). This might suggest that participants were behaving randomly. It is, however, important to keep in mind that we explicitly designed our tasks to detect such violations. In the context of our study, stochastic dominance violations are indirect, being calculated from the comparisons between A vs. C and B vs. C, where B stochastically dominates A. Because the payoff difference between A and B is small, the high error rates are not too different from inconsistencies between "repetitions" of identical choice tasks. Following this interpretation, the obtained results are similar to those reported in the literature (e.g., Alós-Ferrer and Garagnani, 2021). Another piece of evidence which tells us that people did not behave randomly is that subjects reacted strongly to the expected value difference of the choice options. In UK-LOW participants chose the option with the highest expected value in 72.02% of cases, in UK-HIGH in 72.90%, in JP-LOW is 72.34%, in JP-HIGH it is 70.93%. All four proportions are different from random choice according to a non-parametric test, (WRS, $p < 0.001$), indicating behaviour that is clearly distinct from pure noise.

Finally, we investigate potential differences in risk taking due to numerical adaptation. This comes with the caveat that our experiment was not designed to be diagnostic of differences in risk taking per se, so that we may have low power to detect potential differences in risk taking (if any exist). In particular, we investigate potential differences between the adapted and the non-adapted range in the proportion of choices for the safer option, the lottery which involves getting for sure some money compared to taking a risk and potentially earning nothing. We observe no differences in risk attitudes between treatments within each country (UK: HIGH 51.20% vs. LOW 50.98%, WSR; $N = 1157$, $z = 0.18$, $p = 0.859$; JP: HIGH 50.52% vs. LOW 50.65%, WSR; $N = 1152$, $z = -0.09$, $p = 0.927$).

Do we see adaptation over time?

One interesting question concerns whether we observe any adaptation to the numerical magnitudes used in the experiment over the 160 tasks used. Our experimental design used 4 different blocks of 40 tasks each, which give us equal power to test for the occurrence of errors during different phases of the experiment. We can do this because both definitions of errors are based on an ordered lists of 10 trials which were presented in pseudo-randomized order to participants such that within each block the order in which participants saw a trial was random but blocks were kept separate. Tasks within each of these blocks were randomized.

Figure 3 shows the results. In both the UK and Japan, inconsistencies within lists decline between blocks 1 and 2 (HIGH-UK: 22.847% vs. 18.759% Wilcoxon Signed-Rank test; $N = 576$, $z = 16.02$, $p < 0.001$; LOW-UK: 21.915% vs. 18.386% WSR; $N = 581$, $z = 14.93$, $p < 0.001$; HIGH-JP: 21.418% vs. 17.999% WSR; $N = 566$, $z = 14.34$, $p < 0.001$; LOW-JP: 22.367% vs. 18.959% WSR; $N = 586$, $z = 14.01$, $p < 0.001$) and between blocks 3 and 4 (HIGH-UK: 27.625% vs. 22.626% WSR; $N = 576$, $z = 16.99$, $p < 0.001$; LOW-UK: 26.213% vs. 21.256% WSR; $N = 581$, $z = 17.15$, $p < 0.001$; HIGH-JP: 26.303% vs. 21.568% WSR; $N = 566$, $z = 17.05$, $p < 0.001$; LOW-JP: 27.747% vs. 23.063% WSR; $N = 586$, $z = 16.994$, $p < 0.001$), but jump up markedly between blocks 2 and 3, as shown in panels A and B (HIGH-UK, WSR; $N = 576$,

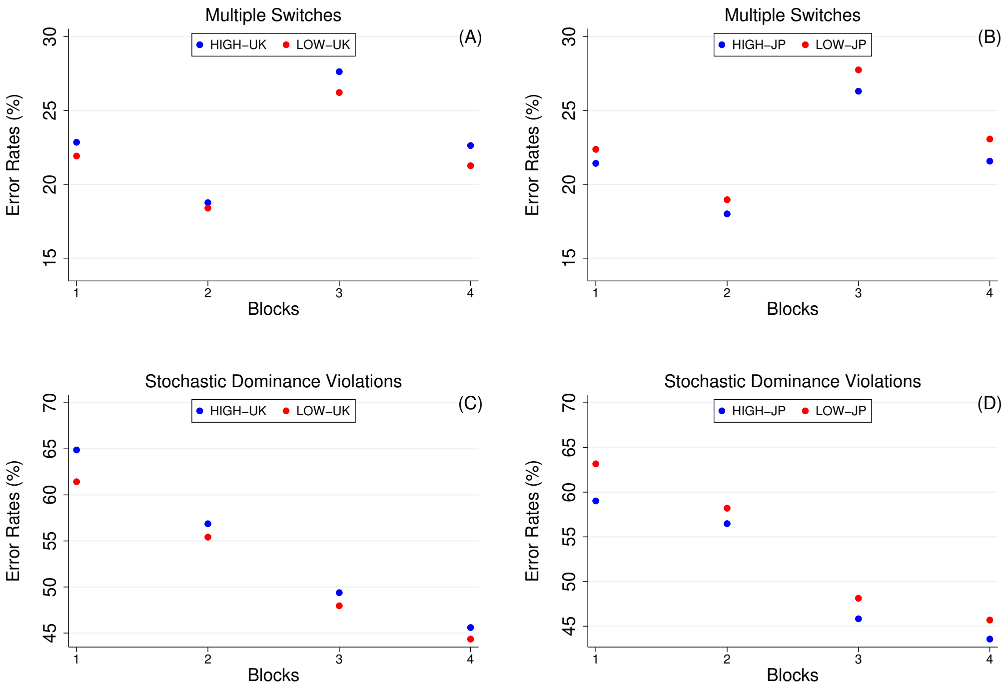

Figure 3: Error rates across blocks of 40 tasks.

Error rates between conditions (HIGH vs. LOW) across blocks of the experiment and countries (UK in the left column vs. JP in the right column) defined as the proportion of multiple switches (top row) or as the proportion of stochastic dominance violations (bottom row).

$z = 19.15$, $p < 0.001$; LOW-UK, WSR; $N = 581$, $z = 18.49$, $p < 0.001$; HIGH-JP, WSR; $N = 566$, $z = 18.35$, $p < 0.001$; LOW-JP, WSR; $N = 586$, $z = 18.82$, $p < 0.001$). We conclude from this that there is no clear trend overall. We can also not detect a clear trend in differences across treatments within either country—our main interest here. While multiple switching appears to decrease between blocks 1 and 2 in the UK (WRS; $N = 1157$, $z = 2.23$, $p = 0.026$), there is no equivalent tendency in Japan (WRS; $N = 1157$, $z = -0.83$, $p = 0.409$), and errors seem to increase again in subsequent rounds. We conclude from this that there is no systematic adaptation visible in within-list inconsistencies.

Stochastic dominance violations, shown in panels C and D, clearly decrease over time in both the UK and Japan and across treatments. That is, for all comparisons such that block $j > i$ lead to $v_j < v_i$ error rates with $p < 0.010$. The evolution of the difference between treatments, however, is again less clear. In particular, while we do see a decline in the difference between treatments from the first to the second block in both countries (UK, WSR; $N = 1157$, $z = 2.33$, $p = 0.020$; JP, WRS; $N = 1152$, $z = -2.27$, $p = 0.023$), the difference appears to remain stable thereafter (block 2 vs. 3, UK, WSR; $N = 1157$, $z = -0.25$, $p = 0.807$; JP, WRS; $N = 1152$, $z = 0.74$, $p = 0.458$; block 3 vs. 4, UK, WSR; $N = 1157$, $z = 0.07$, $p = 0.946$; JP, WRS; $N = 1152$, $z = 0.16$, $p = 0.870$). To the extent that there may be any adaptation, this adaptation thus seems to be quick initially, but to slow down over time and to be highly incomplete. Longer test runs may thus be needed to conclusively show adaptation at work.

Replication Study in Austria and Hungary

Even though we took good care to identify our main effects of interest using a diff-in-diff design, systematic differences between two countries as diverse as Japan and the UK could still result in confounds. We thus replicate the results presented in the first study in two additional countries. Specifically, we replicate our study in Austria and Hungary. These two countries share a border and are closely linked culturally and historically, but once again their currency units are vastly different. Austria has the Euro and Hungary the Fiorint, with 1 Euro corresponding to about 400 Florints (1 Fiorint is 0.0025 Euros). This gives us a similar variation in natural currency units as the one between Japan and the UK exploited in our main study.

The design of the replication study is the same as the main study, with only some minor differences to fit the study situation. While we use the exact same stimuli, we now adapt the multiplier to fit the difference in currency units. Specifically the outcomes of the lotteries in HIGH is obtained by multiplying the outcomes in LOW by 400. Since all effect sizes in the first study are above 0.3, in order to achieve a power of 0.8, for a between-subject test, we require only $N = 290$ participants (145 in each condition and country). We conservatively aimed to recruit $N = 300$ per country (150 in LOW and 150 in HIGH), and conducted the study on Prolific—an online platform for survey and experimental studies. All instructions were in English as participants are used to this language while doing experiments on Prolific.

We ended up obtaining data from $N = 607$ participants (HIGH-AU $N = 151$, LOW-AU $N = 151$, HIGH-HU $N = 154$, LOW-HU $N = 151$). Participants received a fixed payment of Pounds 3 for participation to the study which took about 18 minutes. In addition, for each participant one choice among the 160 was randomly selected to be played for real money.

Results of the Replication

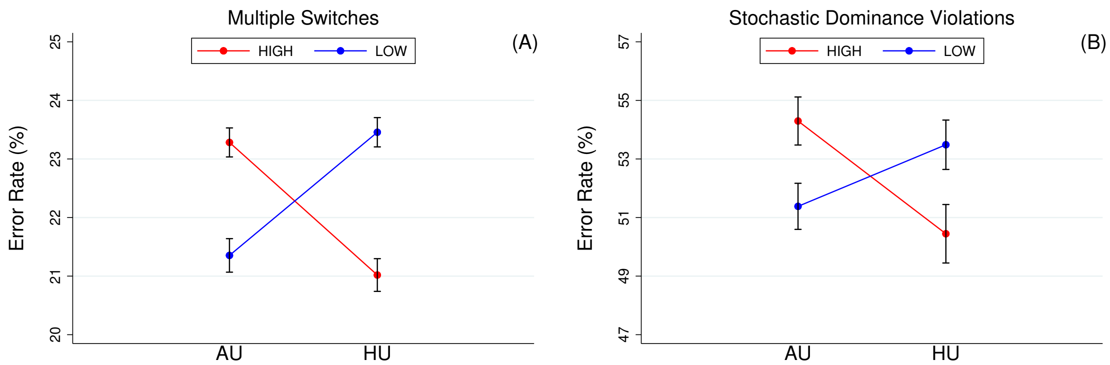

Figure 4: Error rates by country and condition

Error rates between conditions (HIGH vs. LOW) and countries (AU vs. HU) defined as the proportion of multiple switches (A) or as the proportion of stochastic dominance violations (B). According to both definition, error rates are larger for the non-adapted range of outcomes, HIGH-AU and LOW-HU, compared to the adapted range, LOW-HU and HIGH-AU. The asymmetry in error rates between countries can only be explained as range adaptation.

Figure 4 shows the main results. AU participants displayed larger error rates in the HIGH condition than in the LOW condition, which fully replicates the results obtained for UK in the previous study. The exact opposite can be observed for participants from Hungary, where those in the HIGH group committed fewer errors than those in the LOW group, fully replicating the results for Japan. Our key predictions thus replicate—error rates are highest outside of the habitual range, regardless whether that range entails large nominal payoffs or small nominal payoffs. These results are once again robust regardless of what definition of errors we use. If we look at multiple-switches within each list, shown in panel A of Figure 4, we observe an error rate of 23.28% for AU-HIGH compared to 21.35% for AU-LOW, which is an increase of 9.04% in error rates from the adapted to the non-adapted range. The difference is statistically significant (WRS; $N = 302$, $z = 8.88$, $p < 0.001$, CI $[1.552, 2.306]$, BF 25.54), and corresponds to a large effect size of $d = 1.00$. This difference in error rates translates into a statistically significant drop in subjects' earnings from 16.97 to 16.22 (4.42%, 0.75 Pounds; WRS, $N = 302$, $z = 2.23$, $p = 0.026$, CI $[0.092, 1.418]$, BF 3.17, $d = 0.26$). We observe an error rate of 21.02% for HU-HIGH compared to 23.46% for HU-LOW, which is an increase of about 11.61% in error rates from the adapted to the non-adapted range. The difference is again statistically significant (WRS; $N = 305$, $z = -10.67$, $p < 0.001$, CI $[-2.811, -2.063]$, BF 44.46) and corresponds to a large effect size of $d = 1.19$. This difference in error rates translates again into a statistically significant drop in subjects' earnings from 19.74 to 16.77 (15.05%, 2.96 Pounds; WRS, $N = 305$, $z = 12.93$, $p < 0.001$, CI $[2.573, 3.351]$, BF 78.09, $d = 1.30$).

	Multiple Switches		Dominance Violations	
Error Rate	Model 1	Model 2	Model 3	Model 4
HIGH	1.930***	2.132***	2.914***	2.271***
	(0.189)	(0.211)	(0.622)	(0.679)
HU	2.104***	2.176***	2.103***	1.741**
	(0.189)	(0.239)	(0.622)	(0.770)
HIGH × HU	-4.367***	-4.545***	-5.953***	-5.262***
	(0.267)	(0.292)	(0.878)	(0.940)
Age		-0.000		0.024
		(0.010)		(0.025)
Female		0.221		-0.024
		(0.157)		(0.505)
Income		0.001		-0.001
		(0.001)		(0.004)
Risk Attitude		-0.035		-0.445***
		(0.035)		(0.114)
Constant	21.354***	21.334***	51.382***	53.478***
	(0.135)	(0.372)	(0.440)	(1.197)
N	607	520	607	520
R^2	0.306	0.327	0.077	0.096

Table 2: Random-effect regression of error rates

defined as the number of multiple switches (models 1 and 2) or the number of stochastic dominance violations (models 3 and 4). Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

We can directly test for the asymmetry between countries with a difference-in-difference test, which reproduces the results of the single tests. In particular, a panel regression with errors as dependent variable, dummies for country (UK vs. JP) and condition (HIGH vs. LOW), and their respective interactions tells us that the asymmetry is statistically significant (Model 1 of

Table 2; mean -4.367, sd 0.267, $p < 0.001$, CI $[-4.891; -3.844]$). Model 2 of Table 2 shows that the results remain qualitatively unchanged when we control for demographic factors such as age, income, gender, and for risk attitudes.

The error rates of the two countries when comparing their respective adapted and non-adapted range are virtually indistinguishable (AU-LOW vs. HU-HIGH; WRS, $N = 305$, $z = -1.42$, $p = 0.154$; AU-HIGH vs. HU-LOW; WRS, $N = 302$, $z = -1.00$, $p = 0.316$) indicating that the groups are otherwise comparable. Participants are very similar with regards to other characteristics. We find no difference in declared risk tolerance between conditions within each country (AU: WSR; $N = 302$, $z = 0.14$, $p = 0.891$; HU: WSR; $N = 305$, $z = -0.77$, $p = 0.440$). We do not observe a significant difference in declared risk attitudes (Falk et al., 2018) between countries (WSR, $N = 607$, $z = 1.59$, $p = 0.112$).

If we look at error rates as the proportion of time participants violated stochastic dominance between the paired lists, we observe qualitatively identical results (right-hand panel of Figure 4). That is, an error rate of 54.30% for AU-HIGH compared to 51.38% for AU-LOW. The difference is statistically significant (WRS; $N = 302$, $z = 4.83$, $p < 0.001$, CI $[1.782, 4.046]$, BF 2.27, and corresponds to a medium effect size of $d = 0.56$). We once again observe the opposite pattern for Hungarian participants, with an error rate of 50.45% for HU-HIGH compared to 53.49% for HU-LOW (WRS; $N = 305$, $z = -4.30$, $p < 0.001$, CI $[-4.342, -1.735]$, BF 0.40, $d = 0.51$). We again find that the baseline error rates for adapted and non-adapted ranges is comparable between the two countries (AU-LOW vs. HU-HIGH; WRS, $N = 305$, $z = -1.41$, $p = 0.158$; AU-HIGH vs. HU-LOW; WRS, $N = 302$, $z = 1.37$, $p = 0.170$). A test of the difference-in-difference confirms that the asymmetry is significant (Model 3 of Table 2; mean -5.953, sd 0.878, $p < 0.001$, CI $[-7.674; -4.231]$). Model 4 of Table 2 shows that the results are again robust to controlling for demographic factors such as age, income, gender, and risk attitudes.

Finally, we investigate again potential differences in risk taking due to numerical adaptation. We observe statistically significant differences in risk attitudes between treatments within each country in the direction of more risk tolerance for the non-adapted range (AU: HIGH 51.88% vs. LOW 61.74%, WSR; $N = 302$, $z = -14.61$, $p < 0.001$, CI $[-10.56; -9.18]$, BF 176.12, $d = 1.70$; HU: HIGH 61.44% vs. LOW 51.35%, WSR; $N = 305$, $z = 14.93$, $p < 0.001$, CI $[9.45; 10.74]$, BF 196.14, $d = 1.74$). In the Supplementary Materials we report the results of the adaptation analysis for the second study. We reproduce qualitatively the same results obtained in the previous study.

Overall, we fully replicate the results of the first study with two different countries using a different range of adapted numbers.

Discussion

Our ability to discriminate between stimuli depends on how familiar we are with them, a phenomenon which we referred to as “efficient numerical range-adaptation”. This principle, which derives from an efficient use of our constrained neural resources, is a robust finding in psychophysics, e.g. when examining perceptual discrimination of colors or lengths. Recently, it has been applied to decision-making in economics (e.g. Polania et al., 2019; Khaw et al., 2021; Vieider, 2023a; Frydman and Jin, 2022; Prat-Carrabin and Woodford, 2022). However, most empirical tests have been limited to relatively small samples of standard subject pools in laboratory settings, such as the test of Decision-by-Sampling predictions about loss aversion by Walasek and Stewart (2014), and the test of efficient coding adaptation presented by Frydman and Jin (2022).

The question of how important range-adaptation is for real-world situations, and whether it can be found as a by-product of naturally occurring variation in economically relevant quantities remains largely open. Here we quantified the importance and economic magnitude of range-adaptation in the field. In particular, we exploited natural variation in numerical currency units across countries as a proxy for the numerical ranges people are adapted to for economically relevant choices. This allowed us to show that people make substantially more mistakes in their non-adapted range. At 5% or more, the difference in error rates when making choices in the habitual numerical range versus a non-adapted range were of medium size (they were large in the replication in Austria and Hungary), and clearly significant from an economic as well as a statistical point of view. At between 5% and 13% of potential earnings foregone because of these additional mistakes, the economic effects of these tests were also highly significant. That is, people's quality of economically-relevant choices strictly depends on the range they are adapted to, which we showed to depend on their country of residence.

It seems quite natural that currency units would play a major role for the numerical range to which people adapt. Small currency transactions are ubiquitous, and will be encountered dozens of times a day (e.g., during grocery shopping, or when checking one's bills or account balance). Stewart et al. (2006) indeed implicitly included narrow framing into their model of Decision-by-Sampling, whereby credits count for the valuation of gains, and debits towards the valuation of losses (whereas time delays count towards discounting, and probability statements toward likelihood-assessment). Other numerical units, such as those used for weights or distances, are arguably more similar between countries. While there may be some degree of difference between countries due to the use of different measurement units—e.g. due to the use of pounds (or stones) versus kilograms, yards (or feet) versus metres, miles versus kilometres, or inches versus centimetres—these differences fall far short of the two orders of magnitude we observe for currency units. They also often go in opposite directions (e.g. feet versus metres and inches versus centimetres), meaning that they are unlikely to confound our treatment (and they do not apply in the case of Austria versus Hungary). We thus argue that such small numbers being present in both countries—and being relatively similar on average—might make it more difficult to find the type of results we document, making it all the more remarkable that they do indeed emerge in the data.

We found the asymmetry in mistakes using two definitions of errors. We looked at the number of inconsistencies in an ordered list of choices where people's preferences normatively imply only one switch. We documented that people are clearly less consistent in their non-adapted range. The literature has interpreted such "multiple switches" as misunderstanding of the task (Holt and Laury, 2005) or lower cognitive abilities (Yu et al., 2021). We can exclude this as an explanation for our results since we randomly assigned participants drawn from one and the same representative subject pool to adapted versus non-adapted ranges. An alternative explanation for "multiple switches" proposed by the literature is preference imprecision e.g., Chew et al. (2022). This is in line with our findings if imprecision in the perception of the numbers directly translates into imprecision in the implied preferences (Khaw et al., 2021). Note once more that avoiding both types of errors in our setting does not require calculations of the expected (subjective) values of the options. The only element needed for avoiding errors is numerical comparison, which entails that "more is better" in all our experimental treatments.

Our experiments were directly geared at investigating whether adaptation to a certain numerical range would impact error rates. Such effects are not foreseen by standard models of decision making, such as expected utility theory or prospect theory, which have traditionally taken little interest in errors. They are, on the other hand, a direct implication of a wide class of adaptive models which postulate that the mind leverages adaptive mechanisms as a way of dealing with cognitive limitations. Our test was explicitly devised to examine the broad implication of this whole class of models in the natural context of adaptation to the numerical units of national currency units.

This is not to say, however, that these models are identical. Even when it comes to errors, adaptive models may differ with respect to where the error occurs exactly. For instance, in the Decision-by-Sampling model of Neil Stewart and colleagues, the error occurs at the level of utility attribution due to the comparison to a small number of samples drawn from memory (Stewart et al., 2006; Stewart, 2009). Similar errors occur at the level of utility assignment in Robson (2001) and Netzer (2009) because people are unable to assign arbitrarily small utility differences, and thus allocate their discrimination thresholds where they matter the most. In the noisy cognition model of Khaw et al. (2021), on the other hand, errors will occur at the level of numerical perception itself.[6]

These adaptive models furthermore differ in respect to the predictions they make in terms of the risk taking patterns one ought to observe in different situations. Our tasks were not devised to test such differences, which are often intricate and very specific—or for that matter—to test differences in risk taking at all (see Online Appendix for a discussion of the predictions of Decision-by-Sampling as an example). While showing general support for adaptive models of choice, our data are thus ill-suited to discriminate between the different models in this class and their specific predictions.

Our results highlight the direct and significant economic importance of numerical adaptation for real-world settings. Our findings are thus of direct relevance to decisions which are typically made in non-adapted ranges from investments in foreign stock markets into crypto-currencies, gambling in on-line casinos, to cross-boarder shopping. In this sense, these results also provide a micro-foundation for the “money illusion” phenomenon (Shafir et al., 1997) and its economic consequences (Fehr and Tyran, 2001, 2007). Our results suggest that a possible reason why the literature on money illusion has found mixed empirical effects (e.g., Raghubir and Srivastava, 2002; Desmet, 2002; Gamble et al., 2002) was that its determinants were unclear. People are not expected to make worst choices just because the monetary amounts are labeled differently, as usually assumed in this literature. Our results specifically predict money illusion to be stronger when people make choices using nominal currencies units falling into non-adapted ranges. Finally, our results may also provide an explanation for why error rates have been found to *increase* in stakes even when such stakes are real (Enke et al., 2023)—a finding that may otherwise appear paradoxical from an economic point of view. The wider implications of our findings are that cognitive resource-limitations clearly impact economically relevant decisions. Integrating such limitations into economic models should thus be a priority for future research.

[6] Note also that in this particular model, observed decision errors are a non-monotonic function of *coding* errors—the mistakes underlying the mental representations of choice quantities that underlie the model. This means indeed that decision noise would only increase over a certain range, in particular, as long as the variance of the baseline coding noise is small relative to the variance of the prior, which is indeed typically the case (see e.g. Vieider, 2023a).

References

Alós-Ferrer, C., Fehr, E., and Netzer, N. (2021). Time will tell: Recovering preferences when choices are noisy. *Journal of Political Economy*, 129(6):1828–1877.

Alós-Ferrer, C. and Garagnani, M. (2021). Choice Consistency and Strength of Preference. *Economics Letters*, 198:109672.

Alós-Ferrer, C. and Garagnani, M. (2022). The Gradual Nature of Economic Errors. *Journal of Economic Behavior & Organization*, 200:55–66.

Bhui, R., Lai, L., and Gershman, S. J. (2021). Resource-Rational Decision Making. *Current Opinion in Behavioral Sciences*, 41:15–21.

Chew, S. H., Miao, B., Shen, Q., and Zhong, S. (2022). Multiple-Switching Behavior in Choice-List Elicitation of Risk Preference. *Journal of Economic Theory*, 204:105510.

Cohn, A., Engelmann, J., Fehr, E., and Maréchal, M. A. (2015). Evidence for Countercyclical Risk Aversion: an Experiment with Financial Professionals. *American Economic Review*, 105(2):860–885.

Desmet, P. (2002). A Study of the Potential Effects of the Conversion to Euro. *Journal of Product & Brand Management*, 11:134–146.

Di Falco, S. and Vieider, F. M. (2022). Environmental Adaptation of Risk Preferences. *Economic Journal*, 132(648):2737–2766.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, 9(3):522–550.

Enke, B., Graeber, T., and Oprea, R. (2023). Complexity and Time. Technical report, National Bureau of Economic Research.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global Evidence on Economic Preferences. *Quarterly Journal of Economics*, 133(4):1645–1692.

Fehr, E. and Tyran, J.-R. (2001). Does Money Illusion Matter? *American Economic Review*, 91(5):1239–1262.

Fehr, E. and Tyran, J.-R. (2007). Money Illusion and Coordination Failure. *Games and Economic Behavior*, 58(2):246–268.

Frydman, C. and Jin, L. J. (2022). Efficient Coding and Risky Choice. *The Quarterly Journal of Economics*, 137(1):161–213.

Gamble, A., Gärling, T., Charlton, J., and Ranyard, R. (2002). Euro Illusion: Psychological Insights into Price Evaluations with a Unitary Currency. *European Psychologist*, 7(4):302.

Garagnani, M. (2023). The Predictive Power of Risk Elicitation Tasks. *Journal of Risk and Uncertainty*, pages 1–28.

Garcia, M. B., de Hollander, G., Grueschow, M., Polania, R., Woodford, M., and Ruff, C. C. (2023). Individual risk attitudes arise from noise in neurocognitive magnitude representations. *Nature Human Behaviour*.

Heng, J. A., Woodford, M., and Polania, R. (2020). Efficient Sampling and Noisy Decisions. *Elife*, 9:e54962.

Holt, C. A. and Laury, S. K. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, 92(5):1644–1655.

Holt, C. A. and Laury, S. K. (2005). Risk Aversion and Incentive Effects: New Data Without Order Effects. *American Economic Review*, 95(3):902–904.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, 47(2):263–291.

Khaw, M. W., Glimcher, P. W., and Louie, K. (2017). Normalized value coding explains dynamic adaptation in the human valuation process. *Proceedings of the National Academy of Sciences*, 114(48):12696–12701.

Khaw, M. W., Li, Z., and Woodford, M. (2021). Cognitive Imprecision and Small-Stakes Risk Aversion. *The Review of Economic Studies*, 88(4):1979–2013.

Laughlin, S. (1981). A Simple Coding Procedure Enhances a Neuron's Information Capacity. *Zeitschrift für Naturforschung C.*, 36(9-10):910–912.

Laughlin, S. B., de Ruyter van Steveninck, R. R., and Anderson, J. C. (1998). The Metabolic Cost of Neural Information. *Nature neuroscience*, 1(1):36–41.

Lieder, F. and Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and brain sciences*, 43:e1.

Lipkus, I. M., Samsa, G., and Rimer, B. K. (2001). General Performance on a Numeracy Scale Among Highly Educated Samples. *Medical Decision Making*, 21(1):37–44.

Louie, K., Glimcher, P. W., and Webb, R. (2015). Adaptive Neural Coding: from Biological to Behavioral Decision-Making. *Current Opinion in Behavioral Sciences*, 5:91–99.

Netzer, N. (2009). Evolution of Time Preferences and Attitudes Toward Risk. *American Economic Review*, 99(3):937–55.

Oprea, R. (2022). Simplicity Equivalents. Working Paper.

Padoa-Schioppa, C. (2009). Range-Adapting Representation of Economic Value in the Orbitofrontal Cortex. *Journal of Neuroscience*, 29(44):14004–14014.

Padoa-Schioppa, C. and Conen, K. E. (2017). Orbitofrontal Cortex: a Neural Circuit for Economic Decisions. *Neuron*, 96(4):736–754.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., and Dehaene, S. (2004). Tuning Curves for Approximate Numerosity in the Human Intraparietal Sulcus. *Neuron*, 44(3):547–555.

Polania, R., Woodford, M., and Ruff, C. C. (2019). Efficient coding of subjective value. *Nature neuroscience*, 22(1):134–142.

Prat-Carrabin, A. and Woodford, M. (2022). Efficient Coding of Numbers Explains Decision Bias and Noise. *Nature Human Behaviour*, 6:1–11.

Raghubir, P. and Srivastava, J. (2002). Effect of Face Value on Product Valuation in Foreign Currencies. *Journal of Consumer Research*, 29(3):335–347.

Robson, A. J. (2001). The Biological Basis of Economic Behavior. *Journal of Economic Literature*, 39(1):11–33.

Rustichini, A., Conen, K. E., Cai, X., and Padoa-Schioppa, C. (2017). Optimal Coding and Neuronal Adaptation in Economic Decisions. *Nature Communications*, 8(1):1–14.

Shafir, E., Diamond, P., and Tversky, A. (1997). Money Illusion. *Quarterly Journal of Economics*, 112(2):341–374.

Stewart, N. (2009). Eps prize lecture: Decision by sampling: The role of the decision environment in risky choice. *Quarterly journal of experimental psychology*, 62(6):1041–1062.

Stewart, N., Chater, N., and Brown, G. D. (2006). Decision by Sampling. *Cognitive Psychology*, 53(1):1–26.

Stewart, N., Reimers, S., and Harris, A. J. (2014). On the origin of utility, weighting, and discounting functions: How they get their shapes and how to change their shapes. *Management Science*, 61(3):687–705.

Vieider, F. M. (2023a). Cognitive Foundations of Delay-Discounting. IDEAS.

Vieider, F. M. (2023b). Decisions under Uncertainty as Bayesian Inference on Choice Options. *Management Science, forthcoming.*

von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press, New Heaven.

Walasek, L. and Stewart, N. (2014). How to make loss aversion disappear and reverse: tests of the decision by sampling origin of loss aversion. *Journal of experimental psychology: general*, 144(1):7.

Webb, R., Glimcher, P. W., and Louie, K. (2020). Divisive normalization does influence decisions with multiple alternatives. *Nature human behaviour*, 4(11):1118–1120.

Yu, C. W., Zhang, Y. J., and Zuo, S. X. (2021). Multiple Switching and Data Quality in the Multiple Price List. *Review of Economics and Statistics*, 103(1):136–150.

Zhang, H., Ren, X., and Maloney, L. T. (2020). The bounded rationality of probability distortion. *Proceedings of the National Academy of Sciences*, 117(36):22024–22034.

ONLINE APPENDIX

Adaptation in the Second Study

We investigate whether there is adaptation to the numerical magnitudes used in the experiment over the 160 tasks used in the second study between Austria and Hungary. As in the first study, also in the second there are 4 blocks of 40 tasks each, which give us equal power to test for the occurrence of errors during different phases of the experiment.

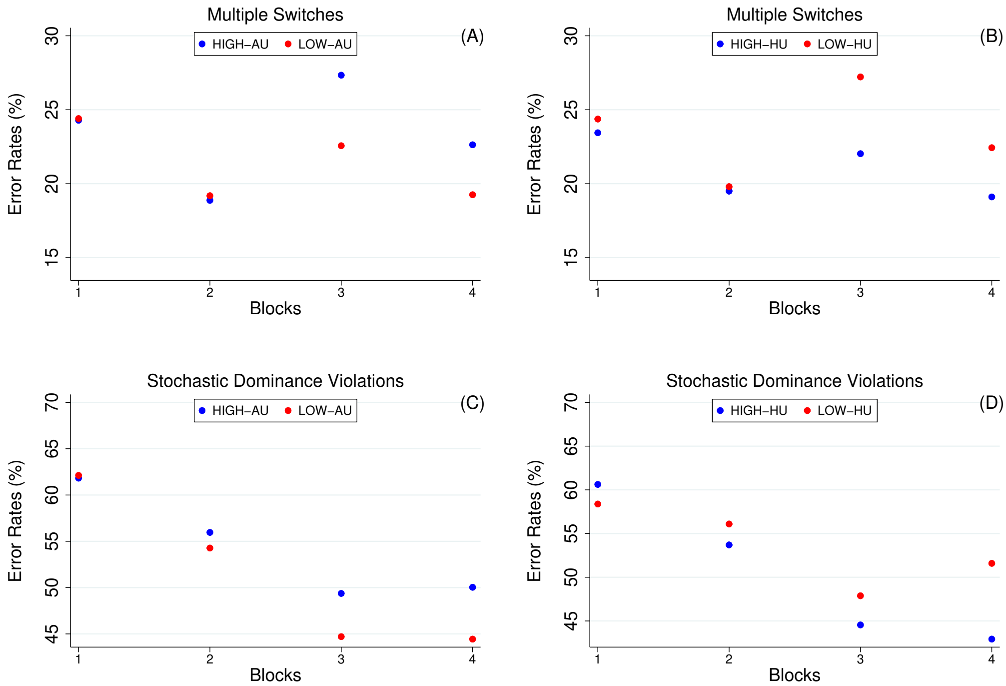

Figure 5: Error rates across blocks of 40 tasks.

Error rates between conditions (HIGH vs. LOW) across blocks of the experiment and countries (AU in the left column vs. HU in the right column) defined as the proportion of multiple switches (top row) or as the proportion of stochastic dominance violations (bottom row).

Figure 5 shows the results. In both the AU and HU, inconsistencies within lists decline between blocks 1 and 2 (HIGH-AU: 24.59% vs. 18.87%, WST; $N = 151$, $z = 9.847$, $p < 0.0001$; LOW-AU: 24.40% vs. 19.19%, WSR; $N = 151$, $z = 8.944$, $p < 0.0001$; HIGH-HU: 23.44% vs. 19.50%, WSR; $N = 154$, $z = 7.566$, $p < 0.0001$; LOW-HU: 24.37% vs. 19.80%, WSR; $N = 151$, $z = 8.908$, $p < 0.0001$) and between blocks 3 and 4 (HIGH-AU: 27.33% vs. 22.63%, WSR; $N = 151$, $z = 9.236$, $p < 0.0001$; LOW-AU: 22.57% vs. 19.25%, WSR; $N = 151$, $z = 6.256$, $p < 0.0001$; HIGH-HU: 22.03% vs. 19.11%, WSR; $N = 154$, $z = 5.766$, $p < 0.0001$; LOW-HU: 27.22% vs. 22.43%, WSR; $N = 151$, $z = 9.209$, $p < 0.0001$), but jump up markedly between blocks 2 and 3, as shown in panels A and B (HIGH-AU, WSR; $N = 151$, $z = 10.131$, $p < 0.0001$; LOW-AU, WSR; $N = 151$, $z = 6.629$, $p < 0.0001$; HIGH-HU, WSR; $N = 154$, $z = 5.195$, $p < 0.0001$; LOW-HU, WSR; $N = 151$, $z = 10.032$, $p < 0.0001$). This fully replicates the results we obtained in the first study and we again conclude from this that there is no clear trend overall. We can also not detect a clear trend in differences across treatments within either country—our main interest here. There is no statistically significant decrease between blocks 1 and 2 in both

AU (WRS; $N = 302$, $z = 0.184$, $p = 0.8539$) and HU (WRS; $N = 305$, $z = -0.900$, $p = 0.3679$), and errors seem to increase again in subsequent rounds. We conclude again from this that there is no systematic adaptation visible in within-list inconsistencies.

Stochastic dominance violations, shown in panels C and D, decrease over time in both the AU and HU and across treatments, but this is less clear than in the previous study. That is, in AU, in both treatments there is a statistically significant decrease in stochastic dominance violations between block 1 and 2 (HIGH-AU: 61.82% vs. 55.96% WSR; $N = 151$, $z = 4.733$, $p < 0.0001$; LOW-AU: 62.12% vs. 54.27%, WSR; $N = 151$, $z = 5.279$, $p < 0.0001$), and between block 2 and 3 (HIGH-AU: 49.37%, WSR; $N = 151$, $z = 6.067$, $p < 0.0001$; LOW-AU: 44.70%, WSR; $N = 151$, $z = 7.226$, $p < 0.0001$), but the difference is not statistically significant between block 3 and 4 (HIGH-AU: 50.03%, WSR test; $N = 151$, $z = -0.806$, $p = 0.4217$; LOW-AU: 44.44%, WSR; $N = 151$, $z = 0.003$, $p = 0.9981$). Instead for HU, there is a difference in stochastic dominance violations between block 1 and 2 for HIGH (60.62% vs. 53.70%, WSR; $N = 154$, $z = 4.821$, $p < 0.0001$), but not for LOW (58.38% vs. 56.09%, WSR; $N = 151$, $z = 1.670$, $p = 0.0951$). Both conditions show a decrease in dominance violations between block 2 and 3 (HIGH-HU, 44.55%, WSR; $N = 154$, $z = 6.801$, $p < 0.0001$, LOW-HU, 47.88%, WSR; $N = 154$, $z = 6.736$, $p < 0.0001$). However, there is no statistically significant difference between block 3 and 4 in HIGH (42.92%, WSR; $N = 154$, $z = 1.086$, $p = 0.2783$), but it jumps up for LOW (51.59%, WSR; $N = 151$, $z = -3.006$, $p = 0.0025$). The evolution of the difference between treatments, however, is again not clear. In particular, we do not see any statistically significant difference for AU between any blocks. For HU there is a decrease between block 1 and 2 (WSR; $N = 305$, $z = 2.558$, $p = 0.0105$), there are no differences between block 2 and 3 (WSR; $N = 305$, $z = 0.867$, $p = 0.3857$), and the difference increases for block 3 and 4 (WSR; $N = 305$, $z = 2.744$, $p = 0.0061$). As for study 1, also in study 2, to the extent that there may be any adaptation, this adaptation thus seems to be quick initially, but to slow down over time and to be highly incomplete.

Pilot Data

We performed a pilot study following what is described in the section Design and Analysis Plan. We recruited N=50 participants for each condition on Prolific, an online platform providing subject pools for survey and experimental studies. We used Hungary instead of Japan in the pilot, since there were not enough subjects available on Prolific Japan. We thus recruited 100 people from the UK and 100 people from Hungary, half of whom where randomly assigned to the LOW condition and the other half to the HIGH condition. Participants took about 15 minutes to complete the study, which consisted of 160 binary choices. Average payment was £23.88 plus a £2.25 fixed fee to cover participants' time in accordance with Prolific rules.

Figure 6 shows the main results. UK participants displayed larger error rates in the HIGH condition than in the LOW condition. The exact opposite can be observed for participants from Hungary, where those in the HIGH group committed fewer errors than those in the LOW group. These results support our predictions—error rates are highest outside of the adapted range, regardless whether that range entails large nominal payoffs or small nominal payoffs. Remarkably, these results are robust regardless of the definition of errors. If we look at error rates as the proportion of multiple-switches within each list, shown in panel A of Figure 6, we observe an error rate of 23.71% for UK-HIGH compared to 22.74% for UK-LOW, which is an increase of about 4% in error rates from the adapted to the non-adapted range. The difference is statistically significant according to a two-sample Wilcoxon rank-sum (Mann-Whitney) test

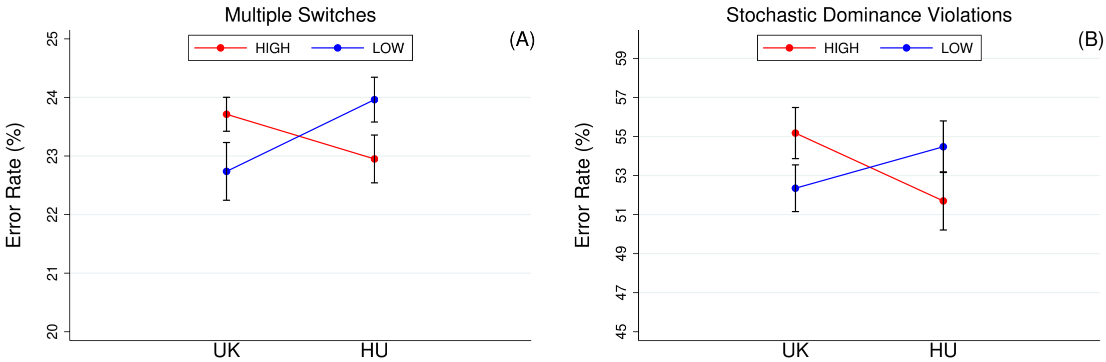

Figure 6: Pilot Results. Error rates between conditions (HIGH vs. LOW) and countries (UK vs. HU) defined as the proportion of multiple switches (A) or as the proportion of stochastic dominance violations (B). According to both definition, error rates are larger for the non-adapted range of outcomes, HIGH-UK and LOW-HU, compared to the adapted range, LOW-UK and HIGH-HU. The asymmetry in error rates between countries can only be explained as range adaptation.

(WRS; $N = 100$, $z = 2.513$, $p = 0.0116$, CI $[0.003, 0.016]$, and corresponds to a medium effect size of 0.579 (according to an equivalent t-test $p = 0.0016$). However, we observe an error rate of 22.95% for HU-HIGH compared to 23.96% for HU-LOW, which is again an increase of about 4% in error rates from the adapted to the non-adapted range. The difference is again statistically significant (WRS; $N = 100$, $z = -2.964$, $p = 0.0028$, CI $[-0.016, -0.003]$) and corresponds to a medium effect size of 0.612. We can directly test for the asymmetry between countries with a difference-in-difference test, which reproduces the results of the single tests. In particular, a panel regression with errors as dependent variable, dummies for country (UK vs. HU) and condition (HIGH vs. LOW), and their respective interactions tells us that the asymmetry is statistically significant (mean 0.020, sd .004, $p < 0.001$, CI $[0.011; 0.029]$).

Interestingly, the error rates of the two countries when comparing their respective adapted and non-adapted range are virtually indistinguishable (UK-LOW vs. HU-HIGH; WRS, $N = 100$, $z = 0.253$, $p = 0.8022$; UK-HIGH vs. HU-LOW; WRS, $N = 100$, $z = -0.042$, $p = 0.9684$) indicating that the groups are otherwise comparable. Participants are very similar with regards to other characteristics. We find no difference in declared risk tolerance between conditions within each country (UK: WSR; $N = 100$, $z = -1.500$, $p = 0.1345$; HU: WSR; $N = 100$, $z = -0.831$, $p = 0.4087$). Moreover, there are no statistically significant differences in declared risk attitudes and numerical literacy (Lipkus et al., 2001) between countries (risk: WSR, $N = 200$, $z = -0.435$, $p = 0.6646$; numerical literacy: WSR, $N = 200$, $z = -1.476$, $p = 0.1405$).

If we look at error rates as the proportion of time participants violated stochastic dominance between the paired lists, we observe qualitatively the same results (right-hand panel of Figure 6). That is, an error rate of 55.18% for UK-HIGH compared to 52.35% for UK-LOW. The difference is statistically significantly different according to a two-sample Wilcoxon rank-sum (Mann-Whitney) test (WRS; $N = 100$, $z = 2.573$, $p = 0.0097$, CI $[0.008, 0.048]$ and corresponds to a medium effect size of 0.545. We observe again the opposite pattern for Hungarian participants with an error rate of 51.70% for HU-HIGH compared to 54.48% for HU-LOW. The difference is statistically significantly different (WRS; $N = 100$, $z = -2.074$, $p = 0.0379$, CI $[-0.050, -0.005]$ and corresponds to an effect size of 0.481. Even with this definition of errors we find the baseline error rates for adapted and non-adapted ranges is comparable between the two countries (UK-LOW vs. HU-HIGH; WRS, $N = 100$, $z = 0.845$, $p = 0.4009$; UK-HIGH vs. HU-LOW; WRS, $N = 100$, $z = -0.415$, $p = 0.6805$). A direct tests of the difference-in-difference confirms that the asymmetry is statistically significant (mean 0.056, sd .015, $p < 0.001$, CI $[0.027; 0.085]$).

List of Lotteries

Table 3: The list of lotteries for the LOW condition. The list of lotteries for the HIGH condition are obtained by multiplying all outcomes by 180. Participants chose either from elements of List A and C or from elements of List B and C. "List number" is the identifier for the indirect comparison between the corresponding List A and List B.

		List A		List B		List C	
ID	List Number	Prob1	Out1	Prob2	Out2	Prob3	Out3
1	1	50	26	50	28	100	3
2	1	50	31	50	33	100	8
3	1	50	36	50	38	100	13
4	1	50	41	50	43	100	18
5	1	50	46	50	48	100	23
6	1	50	51	50	53	100	28
7	1	50	56	50	58	100	33
8	1	50	61	50	63	100	38
9	1	50	66	50	68	100	43
10	1	50	71	50	73	100	48
11	2	40	35	40	37	100	2
12	2	40	40	40	42	100	7
13	2	40	45	40	47	100	12
14	2	40	50	40	52	100	17
15	2	40	55	40	57	100	22
16	2	40	60	40	62	100	27
17	2	40	65	40	67	100	32
18	2	40	70	40	72	100	37
19	2	40	75	40	77	100	42
20	2	40	80	40	82	100	47
21	3	60	34	60	38	100	6
22	3	60	37	60	42	100	11
23	3	60	40	60	46	100	16
24	3	60	43	60	50	100	21
25	3	60	46	60	54	100	26
26	3	60	49	60	58	100	31
27	3	60	52	60	62	100	36
28	3	60	55	60	66	100	41
29	3	60	58	60	70	100	46
30	3	60	61	60	74	100	48
31	4	55	24	55	26	100	2
32	4	55	29	55	31	100	7
33	4	55	34	55	36	100	12
34	4	55	39	55	41	100	17
35	4	55	44	55	46	100	22
36	4	55	49	55	51	100	27
37	4	55	54	55	56	100	32
38	4	55	59	55	61	100	37
39	4	55	64	55	66	100	42
40	4	55	69	55	71	100	47

Table 4: The list of lotteries for the LOW condition. The list of lotteries for the HIGH condition are obtained by multiplying all outcomes by 180. Participants chose either from elements of List A and C or from elements of List B and C. "List number" is the identifier for the indirect comparison between the corresponding List A and List B.

		List A		List C		List B	
ID	List Number	Prob1	Out1	Prob2	Out2	Prob3	Out3
41	5	45	33	45	35	100	1
42	5	45	38	45	40	100	6
43	5	45	43	45	45	100	11
44	5	45	48	45	50	100	16
45	5	45	53	45	55	100	21
46	5	45	58	45	60	100	26
47	5	45	63	45	65	100	31
48	5	45	68	45	70	100	36
49	5	45	73	45	75	100	41
50	5	45	78	45	80	100	46
51	6	65	32	65	36	100	5
52	6	65	35	65	40	100	10
53	6	65	38	65	44	100	15
54	6	65	41	65	48	100	20
55	6	65	44	65	52	100	25
56	6	65	47	65	56	100	30
57	6	65	50	65	60	100	35
58	6	65	53	65	64	100	40
59	6	65	56	65	68	100	45
60	6	65	59	65	72	100	47
61	7	60	22	60	29	100	1
62	7	60	27	60	34	100	6
63	7	60	32	60	39	100	11
64	7	60	37	60	44	100	16
65	7	60	42	60	49	100	21
66	7	60	47	60	54	100	26
67	7	60	52	60	59	100	31
68	7	60	57	60	64	100	36
69	7	60	62	60	69	100	41
70	7	60	67	60	74	100	46
71	8	50	36	50	38	100	4
72	8	50	41	50	43	100	9
73	8	50	46	50	48	100	14
74	8	50	51	50	53	100	19
75	8	50	56	50	58	100	24
76	8	50	61	50	63	100	29
77	8	50	66	50	68	100	34
78	8	50	71	50	73	100	39
79	8	50	76	50	78	100	44
80	8	50	81	50	83	100	49

Noise and Risk Attitudes in Decision-by-Sampling

In this subsection, we provide a discussion of the predictions arising from the Decision-by-Sampling (*DbS*) model.

DbS makes similar predictions to other adaptive models when it comes to error rates. The reason for this is that errors are supposed to take the form of a binomial variable and to be directly attached to utility in DbS (cfr. Stewart et al., 2006, p. 6). Since outcomes far removed from the adapted range will be closer together in terms of utility (given by their rank compared to a handful of samples from memory), this immediately implies that errors will be more influential in less frequently encountered numerical ranges.

It is less clear to what extent we should expect risk taking in general to change between adapted and non-adapted numerical ranges. DbS predicts concave utility due to the discrepancy between cardinal outcome differences and ordinal rankings, which gives rise to smaller utility differences for larger outcome differences as outcome frequencies become sparser, and hence apparent concavity in the 'utility function'. This might suggest systematic changes in risk aversion as outcomes are numerically scaled up or down. Changes in risk taking in this setting ought to be determined by the empirical distributions of outcomes to which people are adapted. Stewart et al. (2006) indeed provide an extensive discussion of empirical patterns in outcome distributions—mostly credits to UK bank accounts for the gain context we have in our data—being best fit by power laws. This will then map into a power utility function, and indeed Stewart et al. (2006), p. 6, use just such a power function to describe the rank r of a given gain g, $r(g) = \frac{cg^{\gamma+1}}{\gamma+1}$ (where c is a normalization constant, and γ governs risk aversion). This function, however, implies ***constant*** *relative risk aversion (CRRA).*

If that is the case—and if all numerical outcomes are scaled up by a multiplicative constant, as in our experimental manipulation—then we should actually not expect any differences in risk taking. This may be worth an illustration, as the statement that 'utility is more concave for small amounts' in the presence of adaptation to larger amounts, and 'relative risk aversion is the same', may intuitively seem to clash. Take a lottery offering a prize x with probability p, or else 0. Let us say the task is to choose between this lottery and a sure amount s, such as in the tasks in our experiment (comparing this to another lottery does not change things). The power utility of Stewart et al. (2006) results in the lottery being chosen whenever $px^{\gamma+1} \geq s^{\gamma+1}$ (where the normalization term $\frac{c}{\gamma+1}$ drops out; note also that additional transformation of probabilities does not affect our argument). Assume now we have a multiplier $m > 0$ by which all outcomes are shifted. The preference relation now becomes $p(mx)^{\gamma+1} \geq (ms)^{\gamma+1}$. Clearly, $m^{\gamma+1}$ will drop out of this equation—relative risk aversion is thus unaffected by the scaling of the outcomes. The 'intuitive clash' of the two statements above occurs because, for small outcomes, a smaller *absolute* increase in outcomes is required to produce any *given* change in utility. Utility being 'smaller' for some outcomes, or the utility function being 'more curved' over some range, does thus not automatically imply increases in relative risk aversion.

More in general, the DbS model is silent on the precise *shape* of utility, in the sense captured by $\frac{-xu''(x)}{u'(x)}$, the Arrow-Pratt index of relative risk aversion. In the empirical data for the UK this index is $-g$, and in this case it is independent of the outcome x (and thus its numerical scale). More generally, however, the shape of utility within the model results from the empirical sampling distribution people have in their minds, which is unobservable. While the model *could* thus account for changes in risk attitudes as numerical stakes change, it does not make any clear ex ante *prediction* on this (or to the extent that it does based on the UK data examined by Stewart et al., 2006, it predicts no change). The model thus provides foundations for concave utility for gains (convex for losses), but not for any specific way in which risk taking would change with (nominal) stakes.

Nonetheless, changes in observed risk taking could result from the interaction between as-if utility curvature and noise. The 'binomially distributed noise' (bottom of p. 6 in Stewart et al., 2006) is attached to *utility*, implying that while noise increases with distance to the adapted range, risk attitudes per se will remain unaffected. Depending on the size of noise relative to the ranks of the outcomes, however, noise may well overpower the relative ranking of outcomes in infrequently encountered numerical ranges. In such cases, one could thus expect a convergence toward random (50-50) choice, which would be due to noise over-powering utility assessment.