Limits of the foreign language effect: intertemporal choice

Michał Białek¹, Artur Domurat², Mariola Paruzel-Czachura³, & Rafał Muda⁴

1 Institute of Psychology, University of Wrocław, Poland

2 Department of Economic Psychology, Kozminski University, Poland

3 Institute of Psychology, University of Silesia in Katowice, Poland

4 Faculty of Economics, Maria-Curie Skłodowska University, Lublin, Poland

Abstract

Intertemporal choice requires one to decide between smaller sooner and larger later payoffs and is captured by discount rates. Across two preregistered experiments testing three language pairs (Experiment 1) and with incentivized participants (Experiment 2), we found no evidence that using a foreign language benefited intertemporal choices. On the contrary, there was some evidence of stronger discounting when a foreign language was used. Our results confirm that more reflective individuals tend to discount less strongly. Thinking in a foreign language did not affect consistency of decisions between payoffs with different defaults in general but can distort the cognitive reflection effects on such consistency. Finally, our findings show that the effects of using a foreign language are robust individual differences in cognitive reflection.

Keywords: foreign language effect, delay discounting, intertemporal choice, cognitive reflection test

Limits of the foreign language effect: intertemporal choice

Introduction

People make suboptimal choices. Some errors are random, like when asking people to guess how much is 124 times 47. However, some decision errors occur predictably because people engage in heuristic thinking (Gilovich et al., 2002; Tversky & Kahneman, 1974). Heuristics are mental short-cuts designed to help us with some reoccurring problems. However, they are rigid, and when adapted to different tasks than designed, they sometimes cause systematic errors (West et al., 2008). To illustrate, intuitive heuristic thinking leads to people disbelieving in science (Majima, 2015), believing in “pseudo-profound bullshit” (Pennycook, Cheyne, et al., 2015) and fake news (Pennycook & Rand, 2019), holding paranormal beliefs (Shenhav et al., 2012), being impatient and failing to accomplish long-term goals (Białek, Bergelt, et al., 2019), and being overly risk-averse (Noori, 2016). The greatest challenge to helping people improve their decisions is that heuristic thinking is automated, intuitive, and difficult to modify (Evans & Stanovich, 2013). Some people are less susceptible to a particular class of decision errors simply because they tend to reflect more by default (Frederick, 2005; Pennycook, Fugelsang, et al., 2015b; Toplak et al., 2011). When reflecting, people can decouple a problem to give greater visibility to its abstract structure or flesh out salient mental models that represent the problem¹. This allows them to override irrelevant aspects of the problem and improve their decision making. The benefits of cognitive

¹For example, consider a situation where Anne looks at Bart, and Bart looks at Catherine. Anne is married, and Catherine is not. Is a married person looking at an unmarried person? Most people say that nothing can be inferred because nothing is known about Bart and his marital status. Yet, if Bart is married he is looking at single Catherine; if Bart is single, married Anne is looking at him, so the answer must be “Yes”. Such a process of considering all the possibilities in a problem is critical in rational thinking (Johnson-Laird, 2012).

reflection are relatively independent of cognitive abilities (Stanovich & West, 2008). While research has sought ways of promoting cognitive reflection, this has had little or no effect (Janssen et al., 2020; Raelison & De Neys, 2019). However, the foreign language effect may be promising avenue of improving decision making.

The foreign language effect relates to a change in judgment and decision making when using foreign vs. native language. Recent findings suggest that people make less biased decisions when thinking in a foreign language (Costa et al., 2014; Keysar et al., 2012). A far-reaching implication of research into this foreign language effect is that ordinary people are, under favourable circumstances, able to make unbiased decisions. Understanding whether and how this is possible can facilitate programs seeking to help people make better choices. Thus, while the foreign language effect is interesting in itself, its primary value seems to be that it provides evidence that people can decide better than they do normally given the right circumstances. If this is true, the effect provides an excellent opportunity to gain insights into how people can be helped to make better decisions and improve their lives.

We explored the effects of using a foreign language to debias thinking by focusing on a new domain of inquiry in foreign language effect studies: intertemporal choice. This domain was selected because of the great importance of these decisions on real-life outcomes (Mischel et al., 1989; Watts et al., 2018).

In intertemporal choice (also called delay discounting), one has to decide whether one prefers to receive a smaller sooner or larger later payoff. People do not make these choices in a financially advantageous manner. For example, when deciding about their pension plan, a person might need to decide whether to trade a smaller immediate income for a larger but greatly delayed income and people often display a bias toward immediate payoffs (Benartzi

& Thaler, 2013). People often prefer immediate consumption of smaller payoffs over much larger, but also much more delayed payoffs. Such preferences make people poorer in the long run, preventing them from long-term-oriented actions such as saving or investing. Discount rates capture the strength of a preference for smaller sooner over larger later payoffs. Greater discount rates correlate to adverse outcomes such as lower wealth and health scores, higher divorce rates, greater body mass indices, and greater drug and alcohol dependency (Bickel et al., 1999; Bickel & Marsch, 2001; Madden et al., 1997; Moffitt et al., 2011). Our aim was to test whether using foreign language can reduce the discount rates, benefiting the decision-makers.

Intertemporal choice requires one to evaluate different, often contradictory, goals, like satisfying current needs vs. satisfying long-term goals. When making such decisions,, people do not always consider all of the relevant goals, but often focus only on more vivid goals (Weber et al., 2007). To illustrate, when first deciding to diet, people focus on future outcomes. They think that the future perspective of losing weight justifies avoiding immediate consumption of junk food. However, things change when they are exposed to junk food. Now, the temptation is more vivid, and they think that enjoying a meal is worth more than the prospect of losing weight. Hence, they stop dieting (Ludwiczak et al., 2020). As this example shows, the more vivid and thus first considered option is preferred in a subsequent choice (Appelt et al., 2011). Thus, reference points biases intertemporal choices: when a reference point lies in the future, people discount less strongly than when the reference point lies in the present. This is problematic in accomplishing real-life long-term goals.

Our second goal was to see whether in situations such as these – where intertemporal choices might naturally be framed with different reference points – using a foreign language can increase the consistency of discount rates across various reference points. If the answer to any of our questions is “yes”, this would constitute an avenue for reducing societal problems such as the accumulation of excessive debt and obesity: to do this, we only need to understand how using a foreign language affects human cognition, and adopt analogous strategies inducing a similar processing mode in people’s everyday decision making. The benefits can be twofold: firstly, we show that in a predominantly multilingual world, using a foreign language can be largely beneficial to people; secondly, it shows that improving intertemporal choice is feasible, and researchers could find methods that push the same cognitive buttons as the foreign language, but that would work in monolinguals, too.

Aiming to better understand the potential mechanisms the foreign language effect provides, we included measures of individual differences relevant for the intertemporal choice. The goal was to see whether all people benefit the same amount when using the foreign language- or whether the benefits depend on the decision maker’s traits.

Past research established that people who tend to reflect more discount less strongly (Białek, Bergelt, et al., 2019; Białek & Sawicki, 2018; Cheng & Janssen, 2019; Frederick, 2005; Littrell et al., 2020; Noori, 2016; Welsh et al., 2013), and are less susceptible to externally imposed reference points, this making their decisions more stable across decision frames (Białek & Sawicki, 2018). This provides us with an opportunity to see whether cognitively reflective individuals are more or less affected by using a foreign language. To date, almost no studies focused on who is more affected by the foreign language, with only a few examples of relevant studies including measurement of individual differences in their design.

For example, Białek, Paruzel-Czachura, et al., (2019) only report the moderating role of the cognitive reflection on the moral foreign language effect in their supplementary materials.

To sum up, our main goal was to investigate the foreign language effect as a means of “debiasing” intertemporal choice. We also explored whether individual differences in cognitive reflection moderate these potential benefits. We aimed to answer these questions in two experiments, using different methods of estimating discount rates in individuals.

Experiment 1 was designed to test whether people discount less strongly when they use a foreign language. We also explored whether this effect depends on the cognitive reflection of decision-makers. Experiment 2 was designed to additionally test whether people using a foreign language are more consistent in their intertemporal choices when payoffs have default options in the present rather than in the future. This issue was investigated for both gains and losses.

Experiment 1

This experiment, including the number of participants, materials, and statistical techniques used, has been preregistered, <http://aspredicted.org/blind.php?x=xj7d98>. The data and materials are available at:

https://osf.io/uejcr/?view_only=374724baeec740368908a53ec2523ac7

Participants

Data were analysed for 293 participants (178 females, $M_{\text{age}} = 22$ years, $SD_{\text{age}} = 3.5$ years, range = 18-65 years). The participants were recruited by an external company and reimbursed 25 PLN (~\$6.5) for their time. We initially recruited 334 participants, but 34 were excluded

because of self-reported poor understanding of the materials, two did not report their understanding, and five because of high inconsistency in their discount rates².

Materials and procedure

Participants were randomly assigned to respond to questionnaires either in their native language (Polish, $n = 119$) or in their common foreign language (English, $n = 174$)³.

Participants filled the MCQ first and then the cognitive reflection test (CRT). Next, participants reported their age of acquiring the foreign language, and estimated their proficiency in the foreign language using a 10-point Likert scale on four dimensions: general, reading, writing, speaking). The anchors were 1 – *unproficient*, 10 – *fully proficient*.

Participants also reported their results of the national standardized foreign language proficiency test and estimated the extent to which they think they understood the task. This last measure was used to eliminate participants who could not comprehend the study materials. Debrief and instructions were presented in Polish, but all measurements, were administered in one language, consistent with the condition participants were assigned to⁴.

Discount rates

Participants responded to the Monetary Choice Questionnaire (MCQ; Kirby et al., 1999) battery, on which they made 27 choices between smaller sooner and larger later payoffs. All questions referred to gains and had no imposed reference point. For example, a participant

² In all our sample of participants who declared enough understanding of the materials to qualify, there was just one person that failed the consistency requirement in the native language condition, and four in the foreign language condition. Remaining participants did not differ in their consistency of responding to the MCQ, $t(291) = -0.61$, $p = .543$.

³ A probability of 60% was used to assign participants to the foreign language condition because greater drop out was expected in this condition. Participants tend to quit the more demanding tasks (i.e., the foreign language condition) more often. This was not needed, as the drop-out rates were identical across experimental conditions, $\chi^2 = 0.85$, $p = .652$

⁴ Participants also rated the believability of several fake and true news items during the experiment. These data are not reported here since they related to a separate line of research which was combined with the present research to reduce costs.

had to decide whether they would prefer to receive \$25 “now” or \$30 in 80 days. The following formula was used to calculate discount rates:

$$V = \frac{A}{1+kD} \quad (1)$$

where V is the subjective value of a reward of amount A that is available after a delay of D time units (Mazur, 1987). Each of 27 choices had a discount rate k , and three choices had the same k , but differed in their delays and amounts. For example, participants had to decide between \$25 “now” or \$30 in 80 days. Participants also decided between \$69 “now” or \$85 in 91 days, and between \$49 “now” or \$60 in 89 days. All of these choices have the same k (.0025), and for these choices, participants would be expected to respond consistently, i.e., either always select the smaller sooner or always select the larger later payoff. A consistency score can be calculated to assess conformity with this rule using Kaplan et al., (2014) autoscorer, and, to maintain the integrity of experiments, it is necessary to remove participants with low consistency scores. We adopted the recommended threshold of 80% conformity, resulting in the removal of only five participants.

The Cognitive Reflection Test

The Cognitive Reflection Test (CRT) consists of three numerical tasks with strong intuitive but incorrect responses (Frederick, 2005). For example, one task asks a person to calculate the price of a ball, knowing that a bat and ball cost \$1.10 together, and that the bat costs \$1 more than the ball. In this case, the first answer that comes to one’s mind is 10 cents, but in this case the ball would cost \$1, and the difference between the bat and ball would be \$0.90. Hence, the correct response is 5 cents, leaving the bat costing \$1.05 and \$1 more expensive. A person’s CRT score is simply the sum of their correct responses on the three tasks, ranging from 0 to 3.

Results

In this study, we complimented the standard p-value based analyses with Bayesian analysis. Its advantages are that they compare models (rather than effects) to each other, reporting their relative likelihood given the data. So, $B_M = 10$ suggests that a particular model is 10 times more supported by the data compared to a subset of all alternative models, and $B_{10} = 5$ means that the alternative hypothesis is 5 times more likely given the data compared to the null hypothesis, and $B_{01} = 5$ means that the null is 5 times more supported than the alternative model. $B \approx 3$ is comparable evidence strength as $p < .05$ (Benjamin et al., 2018). Bayesian analyses also allow to incorporate one's prior belief about the effect size they are expected to observe. However here we used default, uninformed prior proposed by JASP: r scale for fixed effects $r = 0.5$, for random effects $r = 1$, and for covariates $r = 0.354$ (Wetzels et al., 2012)⁵.

Descriptive statistics for this study are presented in Table 1. We used a dedicated autoscorer (Kaplan et al., 2014) which allows to estimate the individual discount rates, expressed by $\log(k)$, to normalize their distribution. Greater negative values refer a discount rate close to zero, describing weak delay discounting. The average discount rate we observed in the native language condition was $\log(k) -1.70$, and therefore $k \approx 0.020$. to put these values in the context, one would expect 2% bonus for waiting a day for their payoff.

To explore our data, we first tested possible language effects on cognitive reflection to evaluate whether there is greater cognitive engagement in the processing of information when information is presented in a foreign language. Consistent with previous literature (e.g., Białek, Paruzel-Czachura, et al., 2019; Costa et al., 2014), CRT scores did not differ across conditions, $t(291) = -0.71$, $p = .48$, $d = -0.08 [-0.32, 0.15]$, $B_{01} = 6.03$. Hence, participants

⁵ Prior distribution of $r = .5$ means that we expected 50% of the effects to lie between $r = -.5$ and $r = .5$, with most of its distribution centered over zero. In other words, if there is no effect the default prior expects to observe even quite large r 's to occur from time to time. Higher prior values mean a wider distribution of the effects under null to be expected.

using a foreign language did not engage in greater cognitive reflection. However, regardless of the language used, greater cognitive reflection was associated with weaker delay discounting, $r(293) = -.161 [-.27, -.047]$, $p = .006$, $B_{10} = 3.25$.

Table 1.

Descriptive statistics of main variables collected in the Experiment 1.

	Native language	Foreign language
<i>N</i>	119	174
Female (%)	39%	39%
Age	22.6 (4.99)	21.7 (2.05)
Foreign language proficiency	7.12 (1.30)	7.10 (1.40)
Understanding	7.63 (1.08)	7.52 (1.18)
CRT	1.08 (1.10)	1.18 (1.13)
Discount rates ($\log(k)$)	-1.70 (0.67)	-1.79 (0.84)
Discount rates consistency	97.6%	97.9%

To assess the effects of language on intertemporal choice, we first estimated individual discount rates, fitting MCQ choices to a hyperbolic curve using a dedicated auto-scorer (Kaplan et al., 2014). This returns $\log k$, with lower values signifying weaker discounting.

The data were then submitted to a two-way ANOVA, with $\log k$ as a dependent variable and language and CRT scores as predictors (Figure 1). The effect of CRT scores was significant, showing weaker discount rates in more reflective individuals, $F(3, 285) = 3.23$, $p = .023$, $\eta_p^2 = .033$, $B_{10} = 1.45$, while neither the effect of language nor a CRT score by language interaction were significant, p 's $> .268$. To quantify the evidence against the effects of language on discount rates we ran an analogous two-way Bayesian ANOVA with a default prior. Results suggested that a one-way model with CRT scores as a predictor was the best among all considered models, $B_M = 5.35$. The model was around twice as strongly supported by the evidence than the null model, $B_{10} = 2.17$, and almost six times more strongly supported than a

model assuming main effects of language and cognitive reflection, $B_{10} = 5.61$. In short, greater cognitive reflection was associated with weaker delay discounting.

One can be curious whether the foreign language effect can be simply explained by people's difficulties to comprehend the materials in their second language. If so, one would expect that people low in proficiency will have it harder to engage in reflective thinking, in turn affecting subsequent decision making. There are some hints this can be true, as the performance in the CRT correlated positively with self-rated proficiency, $r(173) = .16, p = .031$. So, participants who are more proficient in their second language are either less overwhelmed when thinking in their foreign language and therefore perform better, or they are simply more reflective or intelligent. The latter is a more likely explanation, because the foreign language proficiency correlated with the CRT score with similar strength also in participants in the native language condition, $r(118) = .11, p = .229$, Fisher- $z = -0.42, p = .336$. And in this condition, there was no cognitive resources taken by the need to comprehend the task in a foreign language.

Introducing proficiency as a covariate did not produce its main effect nor interacted with any of the predictors. Finally, the effect sizes or significance of all remaining predictors remained unaffected. Hence, the proficiency issue, whether confounding the foreign language effect or not, had no significant impact on our current study. All the above analyses suggest that more reflective individuals discount less strongly (see Figure 1). However, nothing definitive can be said regarding the foreign language effects because the Bayesian evidence is classified as weakly conclusive (Lee & Wagenmakers, 2014).

Figure 1. Discount rates as a function of cognitive reflection scores and language. Lower values mean weaker delay discounting. Coloured points represent observed discount rates, and darker points represent a greater number of identical $\log(k)$'s. Error bars represent 95% confidence intervals.

Discussion

There was no evidence that using a foreign language affected discount rates. Hence, the first of the potential positive effects of using a foreign language, i.e., the notion that thinking in a foreign language reduces discount rates, was not confirmed. However, as previously mentioned, the evidence was far from decisive. Moreover, a possible explanation as to why no effects of using a foreign language were observed is that the task we used was highly numerical, with few verbal cues to trigger a foreign language effect. Therefore, we thought it possible that a foreign language effect might have occurred if we had used vignettes

describing choices instead of numerically presenting choice options. Experiment 2 addressed this possibility.

As discussed in the Introduction, discount rates may be affected by using a foreign language in another way. Specifically, discount rates are malleable and contextual factors strongly affect their strength. For example, in scenarios of the type used in Experiment 1, people choose between smaller sooner and larger later rewards. This choice has no reference point: we asked participants, e.g., whether they would “prefer \$100 now or \$120 in a month”. But another way of presenting this problem would be to ask participants to indicate the smallest reward they would accept in a month rather than immediately receive \$100. Such a choice imposes a reference point in the present: the smaller sooner payoff is the default payment, and the larger later payment is compared to this. One could also reverse the framing and ask what is the smallest amount a person would accept immediately instead of waiting a month for \$120. Such a choice imposes a reference point in the future: the larger later payoff is the default payment, and the smaller sooner payoff is compared to this.

Defaults as these strongly affect intertemporal choices. When presented with similarly attractive options, and when none of these options is default, people tend to randomly select an option. However, people usually select a default option if they are provided with one (Appelt et al., 2011; Weber et al., 2007). So, a person can be indifferent when deciding between receiving \$100 immediately and \$120 in a month. But if asked to exchange an immediate \$100 for \$120 in a month, such a person is likely to refuse the offer and stay with the default. This effect refers not to people’s preferences *per se* but to how defaults distort people’s consistent application of their preferences. Critically, this inconsistency in discount rates is moderated by cognitive reflection, with more reflective individuals tending to be more consistent in their decisions (Białek & Sawicki, 2018).

If, as suggested by the previous literature, using a foreign language debiases decision making, we would expect individuals making decisions in a foreign language to be more consistent in their intertemporal choices. Specifically, we would expect that the effect of imposing reference points would decrease when decision-makers use a foreign language.

Experiment 2 was designed to test new hypothesis along with further inquiry of differences in discounting when using a foreign language. Other changes made in Experiment 2 to improve upon the methodology used in Experiment 1 involved using more text-rich descriptions of the discounting problems posed, testing across three different foreign languages (Spanish, English, and German), and incentivizing participants by using decisions with real financial outcomes.

Experiment 2

This experiment has been preregistered on <https://aspredicted.org/yd5rm.pdf>. Materials for this experiment and the data collected are available at <https://osf.io/uejcr>.

Participants

Data were initially collected from 281 participants (234 females, two self-classified as other/unspecified, $M_{\text{age}} = 21$ years, $SD_{\text{age}} = 1.4$ years, aged 18-27 years), who were English, German, and Spanish philology students. Participants were tested in classrooms in groups of 20-30 using a pen and paper procedure. They were randomly assigned to respond to scenarios presented either in their native language ($n = 139$ in total, 114 females, two genders unspecified) or a foreign language ($n = 142$, 120 females).

Our preregistration specified data exclusion for participants self-reporting their understanding of the task to be lower than 5 on a 10-point Likert scale. Surprisingly, among the 99 participants who self-reported such low understanding, almost half ($n = 46$) were participants in the native language condition. Hence, as in our other recent set of experiments (Białek et

al., 2020), participants most likely reported how competent they felt in solving the task rather than their competence in linguistic comprehension.⁶ Given this, instead of excluding the aforementioned 99 participants, we focused on self-reported proficiency and excluded data for participants who did not answer the proficiency question ($n = 1$) or who self-reported low foreign language proficiency ($n = 18$).

Ultimately, the data for 262 participants were analysed, $n = 128$ participants in the native language condition (Polish), and $n = 134$ participants in the foreign language condition (Spanish $n = 31$, English $n = 59$, and German $n = 44$). The relatively small subsample for each language did not permit cross-language comparisons, and therefore all foreign language participants were collapsed into one cluster.

Incentives

We decided to increase the stakes of decisions made by participants but paying the actual money to some of them depending on their decisions. So, it was made obvious that careless responses could negatively affect their actual payoffs. We explained that when the group is smaller than 20, then one person will be drawn and paid real money. If the group is larger than 20, then two persons will be drawn. We told them that: (1) one of their choices from the gain domain would be selected; (2) there is a maximum bonus that we can pay, and which they do not know, and; (3) they would receive either an immediate or delayed reward, depending on their choice. To rephrase the experiment, participants knew that in tasks with default payment NOW, they will be paid now if they will expect too much bonus for waiting.

⁶This is likely to be the case because linguistic comprehension should have been almost perfect in the native language condition. Additional evidence comes from the observation that understanding and proficiency were weakly correlated in participants in the foreign language condition, $r_B(141) = .16$, $p = .008$. If low self-reported understanding were an indicator of poor comprehension, understanding and foreign language proficiency should have been strongly correlated.

They also knew that in tasks with default payment LATER, they would be forced to wait until a specified time unless they give us enough discount for receiving the payoff immediately.

It was decided in advance that participants would be paid depending on their choice on the trial involving the gain of 100 PLN⁷ delayed for one month. A minimum discount of 20 PLN was required. So, if a participant wanted 80 PLN or less “now”, they received this amount immediately after completing the study. But if they wanted more than 80 PLN “now”, they had to wait a month to receive a full 100 PLN bank transfer.

Procedure and Materials

The study was conducted using a paper-and-pencil technique. General introduction of the study, as well as debrief were provided verbally, in Polish. All experimental questions were asked in the same language: either their native or their foreign, depending on the condition they were assigned to. Participants filled-in blanks, providing subjective equivalents of future payoffs: they participated in a so-called matching procedure (Hardisty, Thompson, Krantz, & Weber, 2011). Four blocks of intertemporal choice were presented in which the signs and default payoffs were manipulated. The four blocks appeared in a fixed order: accelerating gains, delaying gains, accelerating losses, delaying losses. Each block was composed of four trials, ordered by ascending delays: 1, 3, 6, and 12 months.

To illustrate the procedure, let us consider a now-default gain condition. Participants had to declare the smallest amount they would accept in a month instead of immediately receiving 100 PLN. Next, they answered the same questions for 100 PLN delayed by 3 months, 6 months, and 12 months, respectively. In the future-default loss condition, participants read that they were about to lose 100 PLN with a given delay (e.g., in a month), and they had to declare the largest amount that they would agree to pay immediately instead of waiting. This

⁷To put this in context, 100 PLN is equivalent to the Polish 8-hour minimal wage job salary, or \$25.

amount was typically smaller than the delayed one. After responding to the discounting items, participants completed the Cognitive Reflection Test and provided the same proficiency data as in Experiment 1 (i.e., understanding of the task, proficiency, standardized test scores, age of acquisition).

Results

Discount rates were calculated using the area under the curve (AUC) method. This method has a practical advantage in that it departs from the theoretical debate as to whether it is better to model discounting with exponential or hyperbolic functions (Myerson et al., 2001). In other words, unlike $\log k$ used in Study 1, the AUC method is assumption free.

First, we calculated subjective values of payoffs for a particular delay (V_s) by dividing the present, immediate payoff by the future payoff. Next, we standardized delays $t_i, i = 0, \dots, k$, by dividing delays by the longest delay, t_k . As the delays we used were 1, 3, 6, and 12 months, the sequence of standardized delays equalled .083, .25, .5, and 1.0, respectively. We entered each participant's results into a coordinate system, with V 's on the y-axis, and t 's on the x-axis. For instance, consider payment of 500 PLN, which is expected to be received in a month. If a person answers that they would be indifferent between receiving 400 PLN immediately and waiting, then their subjective value for the future gain is $V = \frac{400}{500} = 0.8$, and the standardized delay would be $t = \frac{1}{12} = 0.083$. There are five such subjective values: $V_0 = 1$ for an immediate payoff at t_0 , and V_1 to V_4 for the four delays t_1 to t_4 , respectively. Having put all of these into the coordinate system $t \times V$, we connected them with lines to create four trapezoids (Figure 2).

Figure 2. An example area under the curve (AUC), with scaled delay on the x-axis, and corresponding scaled subjective value of a payoff on the y-axis.

The area under the curve (AUC) is a measure of discounting strength. The AUC summates areas of k trapezoids between subsequent pairs of adjacent subjective values. The adjacent pairs, V_i and V_{i+1} , $i = 0, \dots, k - 1$, are trapezoids' bases, and standardized periods between them, $(t_{i+1} - t_i)$ serve as the heights of trapezoids. The area under the curve (AUC) is the sum of the areas of all k trapezoids:

$$AUC = \sum_{i=0}^{k-1} \frac{(V_i + V_{i+1})}{2} (t_{i+1} - t_i)$$

The AUC is a standardized measure that ranges from 0 to 1. A lower AUC indicates a greater decrease in value as a function of delaying and is therefore interpreted as a sign of stronger discounting.

Discounting usually corresponds to a loss of the subjective value of a payoff over time. For example, a person is expected to ask for an increasingly greater bonus for waiting as the delay of a future gain increases. Conversely, a person would be expected to demand a greater discount when immediate payments are required of them instead of future payments.

Therefore, temporally progressing losses in value should be expressed by monotonic, non-rising sequences of V_s for increasing delays in both the gain and loss domains, as depicted in Figure 2. Non-monotonic choices, e.g., requiring greater bonus for waiting 6 months compared to waiting 12 months, are therefore considered outliers (i.e., typos or other noise), and not included in the analysis.

Data reduction

Prior to analysis, we excluded all the data for 19 participants who either did not answer the proficiency question or who self-reported their foreign language proficiency to be lower than 5 on a 10-point scale. This was a deviation from the preregistered protocol and is explained in the Participants section. Given these assumptions of monotonic discounting, we further excluded non-monotonic sequences of V_s from our analyses, these appearing, for instance, where a participant requested a greater bonus for waiting 6 months than for waiting 12 months (the latter V_s should not exceed the former V_s). Ultimately, 615 data points for 241 participants were analysed.

Here, we explored whether there were more non-monotonic responses in any of the conditions. This could be a preliminary test of the foreign language effect: if people are deliberating more in the foreign language condition, we expect fewer non-monotonic

responses. We, however, observed the opposite: there was greater number of non-monotonic responses in the foreign language condition, $M_{NL} = 1.34$, $M_{FL} = 1.96$, $t(260) = -4.10$, $d = -0.51$ $[-0.76, -0.26]$, $p < .001$,], resulting also in a lower frequencies of complete monotone AUC sequences, $M_{NL} = 31.3\%$ vs $M_{FL} = 15.7\%$, $\chi^2(1) = 8.90$, $p < .01$, $\phi = -.18$. Hence, on average, participants in the native language condition provided non-monotonic discount rates in just over one condition out of four, whereas in the foreign language condition, participants provided non-monotonic discount rates almost twice as often. This decrease in monotonic responding hints at the opposite direction of the foreign language effect on reflective decision making.

Confirmatory analysis

Descriptive statistics for the entire data are presented in Table 2. Not all cells have similar number of datapoints, e.g., because some participants provided monotonic discount rates in some but not all conditions. The exact number of each valid data points can be found in the Supplement to this paper but varies between 39 and 99.

Table 2.

Descriptive statistics for Experiment 2

	Default payoff	Native language	Foreign language
<i>N</i>		128	134
Female (%)		83.3	85.1
Age		21.32 (1.20)	21.51 (1.63)
Foreign language proficiency		7.37 (1.46)	7.62 (1.38)
Understanding		5.48 (2.28)	5.57 (2.51)
CRT		1.63 (1.16)	1.31 (1.15)
AUC gains	Now	0.63 (0.21)	0.57 (0.29)

AUC gains	Later	0.76 (0.19)	0.69 (0.27)
AUC losses	Now	0.76 (0.18)	0.73 (0.23)
AUC losses	Later	0.77 (0.22)	0.69 (0.26)
RPDI Gains		0.18 (0.20)	0.20 (0.25)
RPDI Losses		0.11 (0.13)	0.12 (0.15)

Data presented as mean (*SD*). CRT = cognitive reflection test scores, AUC = Area under the curve, RPDI = Reference point dependency index.

We preregistered a Bayesian mixed ANOVA, with a 2 (language, between-subjects) x 2 (default payoff, within-subjects) x 2 (sign of payoff, within-subjects) design. In contrast to the Experiment 1, we adopted a prior with a Cauchy distribution of a width equivalent to .5 of an r scale of .33 (we did this since we were now expecting to find smaller effects than predicted by default prior). Because most participants had at least one AUC missing, this analysis considered data for only 61 participants.

Figure 3. Discounting strength (AUC) as a function of default, sign, and language for 61 participants. Lower values suggest stronger discounting. Points correspond to real observed

data, with greater mass denoting more identical observations. Error bars represent 95% confidence intervals.

According to the data, the best model included the well-known effects of sign, default, and their interaction, $B_M = 14.68$ (see Figure 2). Specifically, people discounted gains more strongly than losses when the default payoff was the present one, but losses more strongly when the default payoff was the future one (Loewenstein, 1988; Loewenstein & Prelec, 1992; Shelley, 1994). However, there was only weak evidence supporting this model in preference to a model that also included the effect of language, $B_{10} = 3.13$. Consistently with Experiment 1, this analysis provided some weak evidence against the idea that using a foreign language influences discount rates. Details of this analysis can be found in supplementary Materials.

Exploratory analysis

The fact that we preregistered an analysis of data only for participants responding correctly to all trials resulted in a severe loss of data points. We, therefore, reperformed the analyses using linear mixed-effects (LME) models. This allowed us to retain most of our data, providing us with enough power to add CRT scores as a covariate of the effects. With this method, we were able to analyse 615 observations from the data set of all the participants self-reporting enough proficiency to comprehend the materials (at least 5/10) instead of examining data for only 61 participants with 244 observations. We contrast-coded factors as -1 and +1 and mean-centred covariates. The data were analysed in several different ways. Because of the analyses' complexity, we report each analysis's details in a Supplement and report only the effects of interest below.

Language as a mechanism reducing the strength of delay discounting

First, differences in the discount rates produced by participants using native vs. foreign languages were tested. This was done by building a model in which the dependent variable was the AUC, the sign, default, their interaction, and the main effect of language were fixed effects, and subject was a random effect. There were main effects of language, $\beta = -0.13$ [-0.23, -0.03], $t(237.1) = -2.51$, $p = .013$, of sign, $\beta = 0.19$ [0.12, 0.25], $t(432.9) = 5.68$, $p < .001$, of default, $\beta = 0.10$ [0.04, 0.16], $t(461.9) = 3.18$, $p = .002$, and a sign by default interaction, $\beta = -0.11$ [-0.17, -0.05], $t(440.4) = -3.50$, $p = .001$ ⁸. The sign by default interaction hints at greater effects of defaults on gains than on losses (Figure 4), possibly because of losses attracting greater on-task attention (Yechiam & Hochman, 2013). The most interesting effect in this analysis suggested that people using a foreign language discounted more strongly than those using their native language (Figure 4).

To gain further insight into the moderating function of the cognitive reflection, we extended this analysis by further including CRT score and its interaction with language. We observed stronger discounting for the default in the present, $\beta = 0.10$ [0.04, 0.17], $t(466.2) = 3.28$, $p = .001$, and of gains, $\beta = 0.19$ [0.13, 0.25], $t(439.9) = 5.79$, $p < .001$, and when using a foreign language, $\beta = 0.14$ [0.04, 0.25], $t(236.9) = 2.68$, $p = .008$. The main effects were qualified by a previously described sign by default interaction, $\beta = -0.11$ [-0.17, -0.05], $t(448.3) = -3.50$, $p = .001$. CRT did not produce a main effect, nor interacted with the language used, $t < 1$, $p > .551$. Details of this and previous LME analysis can be found in Supplementary Materials.

⁸ We only incentivized gains, but not losses. One could be curious, whether this fact interacted with the foreign language effect, e.g., motivating participants to reflect only in the incentivized condition. We found the effects of foreign language to be similarly pronounced, $t = 2.43$, $p = .015$ in gains, and $t = 1.95$, $p = .052$ in losses. Note that although these effects differ at significance, this does not mean they are different from each other (Gelman & Stern, 2006; Nieuwenhuis et al., 2011). Hence, regardless of whether the condition was or not incentivised, the foreign language was observed in almost identical magnitude.

Figure 4. Discounting (AUC) as a function of default, sign, and language. Shaded areas represent 95% confidence intervals. Points are jittered to avoid overplotting.

The current analysis provided novel evidence that peoples' temporal discounting is affected by using a foreign language. However, the foreign language effect was in an opposite direction to what would be expected given the assumed debiasing nature of using a foreign language: people using a foreign language discounted more strongly. Similar effect was observed in parallel by an independent team of researchers (Xu & Liang, 2021). As previously mentioned, strong discounting is associated with undesirable real-life outcomes, including outcomes in the areas of health and finance (Moffitt et al., 2011; Watts et al., 2018). Because of the null effects reported in the Experiment 1, discussion of this effect can only be tentative: using a foreign language either has no effect on discounting strength, or it increases discounting strength. Whichever is true, using a foreign language does not seem to help people make better intertemporal choices.

Language as a mechanism reducing cognitive biases

This section reports a critical test of the idea that the foreign language effect should be considered through the lens of individual differences. We have established that using a foreign language does not help in decreasing discount rates, but that, surprisingly, it can increase them. Hence, our hypothesis suggesting potential benefits of using a foreign language for delay discounting, was not supported. Studies on the foreign language effect suggested that another possible benefit of using a foreign language could be that it makes people more consistent in applying their discount rates. We expected that people would be less affected by externally imposed reference points when using their foreign language skills. To test this hypothesis, as in Białek and Sawicki (2018), we calculated a Reference Point Dependency Index (RPDI) as an absolute value of the difference between areas under the curve for decisions made with default payoffs in the present vs. in the future. This index captures the extent to which people's decisions are affected by externally imposed reference points.⁹ The RPDI only tests for consistency between reference points in the present vs. the future because being affected by reference points is a bias. The RPDI does not test for consistency across discounting of gains and losses because being more sensitive to losses than to gains can be considered adaptive rather than being a bias (McDermott et al., 2008). Hence, the only debiasing effect of using a foreign language would be for it to reduce the impact of default payoffs now vs. later.

Because calculating the RPD requires one to have a monotonic discount rates in both default conditions of corresponding types of payoffs (e.g., for the default payoff now and later in

⁹The difference between areas under the curve obtained with reference points in the present vs. the future has opposite signs for gains and losses. By considering the RPDI as an absolute value of this difference, we can account for these reversed signs and boost statistical power without losing information about reference point dependency.

gains) the available data pool was reduced to 196 observations. We then proceeded with an LME analysis attempting to predict the size of the RPD (dependent variable) by using the language in which the discounting task was presented (a fixed effect) while controlling for the sign of payoff (a random effect). Contrary to the predictions derived from previous research on the foreign language effect, language had no significant impact on the RPD, $\beta = 0.04$ [-0.11, 0.18], $t(129) = 0.53$, $p = .596$: it might be concluded that people are not less biased by a reference point when they use a foreign language. This may, however, only be true when averaging over individual differences in cognitive reflection. As previously mentioned, the possibility remains that thinking in a foreign language provides the greatest benefit to people who would normally be biased by imposed reference points. Conversely, individuals who are resilient to such reference points would not benefit because there would be no bias to be corrected in the first place. The debiasing effects of thinking in a foreign language would depend on decision maker's cognitive reflection.

The above reasoning was tested in an analysis in which we predicted RPD with CRT scores, language, and their interaction (fixed effects) while controlling for the sign of payoff (a random effect). Here, neither cognitive reflection, nor using a foreign language related to the RPD, both t 's < 1 , p 's $> .593$. However, there was some hint of a cognitive reflection by foreign language use interaction, $\beta = 0.13$ [-0.01, 0.27], $t(143.4) = 1.78$, $p = .077$. Even though this interaction was nonsignificant, we decomposed it to gain further insight. The interaction occurred because the CRT scores were associated with a decrease in reference point dependency in people using their native language, $\beta = -0.20$ [-0.36, 0.01], $t(80.5) = -1.87$, $p = .065$, but not in people using a foreign language, $\beta = 0.10$ [-0.13, 0.33], $t(59) = 0.83$, $p = .407$ (Figure 5). In short, cognitive reflection seems to be a mechanism that prevents people from falling prey to externally imposed reference points in the native language, but not when a foreign language is used.

Figure 5. Reference point dependency as a function of cognitive reflection score split by the native and foreign language conditions.

General discussion

Across two preregistered experiments, we found no evidence that using a foreign language benefitted intertemporal choices. On the contrary, there was some evidence of stronger discounting when a foreign language was used. In general, foreign language users accepted smaller sooner payoffs instead of larger delayed payoffs, and larger later losses instead of smaller sooner losses. The foreign language users were also similarly affected by the default payoffs as the native language users. High discount rates are considered problematic because they are predictive of several behavioural problems, including obesity, addictions, excessive debt, and divorce (Moffitt et al., 2011). The importance of discount rates is also supported by the fact that a child's ability to resist immediate temptation is one of the strongest predictors of their quality of life (Mischel et al., 1989; Shoda et al., 1990), especially for those of low socioeconomic status (Watts et al., 2018).

Our results confirm past research showing that more reflective individuals tend to discount less strongly compared to less reflective individuals. Their intertemporal choices are also more consistent across different reference points and perspectives. In turn, this allows for greater consistency in long-term planning, benefitting a decision-maker. Thinking in a foreign language did not affect such consistency. This raises the questions as to why and how using a foreign language helps in some decisions but not in others.

Our findings add to the understanding of the foreign language effect, which has previously been reported to debias decision-makers. The long list of potential benefits has been said to include more accurate risky choices, diminished framing effects, and greater honesty. Recent findings, however, suggest that this view can be biased by the selection of decision problems used (Muda, Pieńkosz, et al., 2020). Our findings are in-line with other recent research demonstrating that people using a foreign language do not always benefit from using their second language: their vividness of mental imagery is reduced (Hayakawa & Keysar, 2018; but see Montero-Melis et al., 2020 for a different perspective), perform no better in some judgment and decision-making tasks (Mækelæ & Pfuhl, 2019), perform worse in logical reasoning tasks (Białek et al., 2020), and do not decrease their use of irrelevant information when gambling (Muda, Walker, et al., 2020). The current research adds another example to the list of situations where using a foreign language is of no benefit, and we urge researchers and practitioners not to assume that using a foreign language is widely beneficial to bilinguals. The conditions under which foreign language usage is likely to be beneficial to bilinguals' decision making are complex and context-dependent. This change in our understanding of the foreign language effect (whereby foreign language effects are being re-evaluated from being viewed as always positive to being viewed as simply affecting decisions) can be found in two seminal papers: one by Keysar et al. (2012) suggesting that foreign language *debases*

decisions, and more recent one by Hayakawa et al. (2016) opining that it merely *changes* decisions.

The foreign language effect, cognitive reflection, and rational decision making

The foreign language effect is often explained using the dual-process theory (De Neys & Pennycook, 2019; Evans & Stanovich, 2013; Kahneman, 2011). In its most recent version, people facing a problem tend to produce one or several intuitions guiding their subsequent decisions. If the intuitions clash, a conflict is detected, and reflection is triggered. Cognitive reflection also helps reflective individuals to invest their resources strategically, so that they do not need to reflect on problems which have an intuitively correct response (De Neys & Pennycook, 2019; Pennycook, Fugelsang, et al., 2015b; Stanovich, 2018). Once multiply intuitions clash, and the possibility of bias is detected, people invest cognitive resources into thinking, increasing their chances of arriving at a normatively correct response (De Neys & Pennycook, 2019; Pennycook, Fugelsang, et al., 2015b). In short, people's decision making benefits from reflecting (Frederick, 2005; Juanchich et al., 2016; Pennycook, Fugelsang, et al., 2015a), especially when they have appropriate knowledge and skills (Erceg et al., 2019; Šrol & De Neys, 2020).

In the current research, we observed that reflective reasoners were discounting less strongly in both languages. Stronger, or weaker discounting is however just an expression of preferences and could not be easily classified as rational or irrational. The reference point independency, on the contrary, is clearly rational. A decision to accept smaller sooner or larger later payoff should be the same regardless of whether one or another payoff is default. And in this regard, we observed that the effects of cognitive reflection could have been reduced when thinking in a foreign language. Specifically, more reflective individuals showed greater reference point

independency only in their native language. No benefits of cognitive reflection on the reference point independency were observed in the foreign language. This way, we show that individuals using their foreign language could not use their trait cognitive reflection to guide their consistency in the intertemporal choice. Note that the CRT by language interaction was not significant ($p = .077$), and thus this conclusion is only tentative.

Alternative explanations of the foreign language effect

Using a foreign language cannot have a direct effect on thinking, but only an indirect one instead. For example, thinking in the second language is less fluent, and this way can either impose cognitive load (and distort thinking) or motivate people to invest more cognitive resources in thinking (and benefit thinking). Indeed, low comprehension can explain some of the effects, as illustrated on the mental imagery in the foreign language (Montero-Melis et al., 2020). However, both hypotheses are hard to reconcile with the available evidence: not only people often benefit from using their foreign language (thus, the language-as-load hypothesis is not supported), but also their cognitive reflection does not increase (thus, the language-as-facilitator-of-reflection is not supported). Moreover, the previously suggested mechanism of disfluency as a way to motivate people to engage in greater reflection (Alter, 2013; Köhl & Eitel, 2016) failed to gain support in large, preregistered replication efforts (Meyer et al., 2015; Sirota et al., 2020).

Another explanation for the foreign language effect is that languages differ in their syntax, semantics, and structure. Rather than the difference between being the first vs. the second language, these language-specific differences can be responsible for the foreign language effect. An illustrative example comes from work closely related to our project. Empirical evidence links the languages with obligatory future-time reference (vs. non-obligatory) with

less future-oriented behaviour, such as saving less, smoking more, less commitment to safe sex, and greater obesity (M. K. Chen, 2013; Roberts et al., 2015, but see J. I. Chen et al., 2019 for contradictory findings). In such a case, the foreign language effect would be confounded with simply switching between languages that differ in their grammatical structure. However, we argue this is not the case. The evidence for the foreign language effect being independent of the particularities of languages is that it is observed in counterbalanced pairs (Hayakawa et al., 2017; Keysar et al., 2012) and consistently across many language pairs (Białek, Paruzel-Czachura, et al., 2019; Costa et al., 2014; Hayakawa et al., 2019), as in this project. However, the two effects can interact, supporting or cancelling each other out, which would explain at least part of the inconsistent findings regarding the foreign language effect. Future research, especially when using a single non-counterbalanced language pair, should consider this potential confound when interpreting their findings.

Conclusions

Further research is required to validate our speculations concerning how the foreign language effect works. But, if our understanding of the effect is correct, there is little support for claims that using a foreign language is universally beneficial for decision making. Instead, when using a foreign language, people are discounting just as strongly as in their native language (Experiment 1), or even more strongly (Experiment 2).

We also provide evidence that the effects of using a foreign language does not increase cognitive reflection, and its effects on intertemporal choice are robust to individual differences in cognitive reflection. Hence, using a foreign language is unlikely to affect decision making by increasing deliberation. Instead, it could be that foreign language users have weaker access to their gut intuitions that help them make better decisions in some cases, but which can distort their decisions in other cases (Białek et al., 2020).

Acknowledgments

The current research was supported by grant 2017/26/D/HS6/01159 from the National Science Centre (NCN, Poland) to Michał Białek. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Alter, A. L. (2013). The benefits of cognitive disfluency. *Current Directions in Psychological Science*, 22(6), 437–442.
- Appelt, K. C., Hardisty, D. J., & Weber, E. U. (2011). Asymmetric discounting of gains and losses: A query theory account. *Journal of Risk and Uncertainty*, 43(2), 107–126.
<https://doi.org/10.1007/s11166-011-9125-1>
- Benartzi, S., & Thaler, R. H. (2013). Behavioral economics and the retirement savings crisis. *Science*, 339(6124), 1152–1153.
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., & Camerer, C. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2(1), 6.
- Białek, M., Bergelt, M., Majima, Y., & Koehler, D. J. (2019). Cognitive reflection but not reinforcement sensitivity is consistently associated with delay discounting of gains and losses. *Journal of Neuroscience, Psychology, and Economics*, 12(3–4), 169–183.
- Białek, M., Muda, R., Stewart, K., Niszczoła, P., & Pieńkosz, D. (2020). Thinking in a foreign language distorts allocation of cognitive effort: Evidence from reasoning. *Cognition*.
- Białek, M., Paruzel-Czachura, M., & Gawronski, B. (2019). Foreign language effects on moral dilemma judgments: An analysis using the CNI model. *Journal of Experimental Social Psychology*, 85, 103855. <https://doi.org/10.1016/j.jesp.2019.103855>
- Białek, M., & Sawicki, P. (2018). Cognitive Reflection Effects on Time Discounting. *Journal of Individual Differences*, 39(2), 99–106. <https://doi.org/10.1027/1614-0001/a000254>

- Bickel, W., & Marsch, L. (2001). Toward a behavioral economic understanding of drug dependence: Delay discounting processes. *Addiction*, 96, 73–86.
- Bickel, W., Odum, A. L., & Madden, G. J. (1999). Impulsivity and cigarette smoking: Delay discounting in current, never, and ex-smokers. *Psychopharmacology*, 146(4), 447–454.
<https://doi.org/10.1007/PL00005490>
- Chen, J. I., He, T.-S., & Riyanto, Y. E. (2019). The effect of language on economic behavior: Examining the causal link between future tense and time preference in the lab. *European Economic Review*, 120, 103307. <https://doi.org/10.1016/j.euroecorev.2019.103307>
- Chen, M. K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review*, 103(2), 690–731.
- Cheng, J., & Janssen, C. (2019). The relationship between an alternative form of cognitive reflection test and intertemporal choice. *Studia Psychologica*, 61(2), 86.
- Costa, A., Foucart, A., Arnon, I., Aparici, M., & Apesteguia, J. (2014). “Piensa” twice: On the foreign language effect in decision making. *Cognition*, 130(2), 236–254.
- De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: Advances in dual-process theorizing. *Current Directions in Psychological Science*, 28(5), 503–509.
- Erceg, N., Galic, Z., & Bubić, A. (2019). Who detects and why? Individual differences in abilities, knowledge and thinking dispositions among different types of problem solvers and their implications for the validity of reasoning tasks. *PsyArxiv*.
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Theory and Metatheory in the Study of Dual Processing: Reply to Comments. *Perspectives on Psychological Science*, 8(3), 263–271.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42.
- Gelman, A., & Stern, H. (2006). The Difference Between “Significant” and “Not Significant” is not Itself Statistically Significant. *The American Statistician*, 60(4), 328–331.
<https://doi.org/10.1198/000313006X152649>

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press.

Hayakawa, S., & Keysar, B. (2018). Using a foreign language reduces mental imagery. *Cognition*, 173, 8–15. <https://doi.org/10.1016/j.cognition.2017.12.010>

Hayakawa, S., Lau, B. K. Y., Holtzmann, S., Costa, A., & Keysar, B. (2019). On the reliability of the foreign language effect on risk-taking. *Quarterly Journal of Experimental Psychology*, 72(1), 29–40. <https://doi.org/10.1177/1747021817742242>

Hayakawa, S., Tannenbaum, D., Costa, A., Corey, J. D., & Keysar, B. (2017). Thinking More or Feeling Less? Explaining the Foreign-Language Effect on Moral Judgment. *Psychological Science*, 28(10), 1387–1397. <https://doi.org/10.1177/0956797617720944>

Janssen, E., Raelison, M., & De Neys, W. (2020). “You’re Wrong!”: The Impact of Accuracy Feedback on the Bat-and-Ball Problem. *Acta Psychologica*.

Johnson-Laird, P. N. (2012). Inference with mental models. *The Oxford Handbook of Thinking and Reasoning, Journal Article*, 134–145.

Juanchich, M., Dewberry, C., Sirota, M., & Narendran, S. (2016). Cognitive reflection predicts real-life decision outcomes, but not over and above personality and decision-making styles. *Journal of Behavioral Decision Making*, 29(1), 52–59.

Kahneman, D. (2011). *Thinking, Fast and Slow*.

Kaplan, B. A., Lemley, S. M., Reed, D. D., & Jarmolowicz, D. P. (2014). *21-and 27-item Monetary Choice Questionnaire automated scorers*.

Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science*, 23(6), 661–668.

Kirby, K., Petry, N., & Bickel, W. (1999). Heroin Addicts Have Higher Discount Rates for Delayed Rewards Than Non-Drug-Using Controls. *Journal of Experimental Psychology: General*, 128(1), 78–87.

- Kühl, T., & Eitel, A. (2016). Effects of disfluency on cognitive and metacognitive processes and outcomes. *Metacognition and Learning*, 11(1), 1–13.
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university press.
- Littrell, S., Fugelsang, J., & Risko, E. F. (2020). Not so fast: Individual differences in impulsiveness are only a modest predictor of cognitive reflection. *Personality and Individual Differences*, 154, 109678.
- Loewenstein, G. (1988). Frames of mind in intertemporal choice. *Management Science*, 34(2), 200–214.
- Loewenstein, G., & Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *Quarterly Journal of Economics*, 107(2), 574–597.
<https://doi.org/10.2307/2118482>
- Ludwiczak, A., Osman, M., & Jahanshahi, M. (2020). Redefining the Relationship Between Effort and Reward: Choice-Execution Model of Effort-Based Decisions. *Behavioural Brain Research*, 112474.
- Madden, G. J., Petry, N. M., Badger, G. J., & Bickel, W. (1997). Impulsive and self-control choices in opioid-dependent patients and non-drug-using control patients: Drug and monetary rewards. *Experimental and Clinical Psychopharmacology*, 5(3), 256.
- Mækelæ, M., & Pfuhl, G. (2019). Deliberate reasoning is not affected by language. *PLOS ONE*, 14(1), e0211428. <https://doi.org/10.1371/journal.pone.0211428>
- Majima, Y. (2015). Belief in pseudoscience, cognitive style and science literacy. *Applied Cognitive Psychology*, 29(4), 552–559.
- Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In *Quantitative analyses of behavior. Vol. 5. The effect of delay and of intervening events on reinforcement value* (pp. 55–73). Erlbaum.

- McDermott, R., Fowler, J. H., & Smirnov, O. (2008). On the Evolutionary Origin of Prospect Theory Preferences. *The Journal of Politics*, 70(2), 335–350.
<https://doi.org/10.1017/S0022381608080341>
- Meyer, A., Frederick, S., Burnham, T. C., Guevara Pinto, J. D., Boyer, T. W., Ball, L. J., Pennycook, G., Ackerman, R., Thompson, V. A., & Schuldt, J. P. (2015). Disfluent fonts don't help people solve math problems. *Journal of Experimental Psychology: General*, 144(2), e16.
- Mischel, W., Shoda, Y., & Rodriguez, M. I. (1989). Delay of gratification in children. *Science*, 244(4907), 933–938.
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., & Ross, S. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, 108(7), 2693–2698.
- Montero-Melis, G., Isaksson, P., van Paridon, J., & Ostarek, M. (2020). Does using a foreign language reduce mental imagery? *Cognition*, 196, 104134.
<https://doi.org/10.1016/j.cognition.2019.104134>
- Muda, R., Pieńkosz, D., Francis, K., & Białek, M. (2020). The moral foreign language effect is stable across presentation modalities. *Quarterly Journal of Experimental Psychology*, 174702182093507. <https://doi.org/10.1177/1747021820935072>
- Muda, R., Walker, A. C., Pieńkosz, D., Fugelsang, J. A., & Białek, M. (2020). Foreign Language does not Affect Gambling-Related Judgments. *Journal of Gambling Studies*, 36(2), 633–652.
<https://doi.org/10.1007/s10899-020-09933-6>
- Myerson, J., Green, L., & Warusawitharana, M. (2001). Area under the curve as a measure of discounting. *Journal of the Experimental Analysis of Behavior*, 76(2), 235–243.
- Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience*, 14(9), 1105–1107.
<https://doi.org/10.1038/nn.2886>

Noori, M. (2016). Cognitive reflection as a predictor of susceptibility to behavioral anomalies.

Judgment and Decision Making, 11(1), 114–120.

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2015). On the reception and detection of pseudo-profound bullshit. *Judgment and Decision Making*, 10(6), 549–563.

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015a). Everyday consequences of analytic thinking. *Current Directions in Psychological Science*, 24(6), 425–432.

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015b). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, 80, 34–72.

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50.

<https://doi.org/10.1016/j.cognition.2018.06.011>

Raoelison, M., & De Neys, W. (2019). Do we de-bias ourselves?: The impact of repeated presentation on the bat-and-ball problem. *Judgment and Decision Making*, 14(2), 170–178.

Roberts, S. G., Winters, J., & Chen, K. (2015). Future Tense and Economic Decisions: Controlling for Cultural Evolution. *PLOS ONE*, 10(7), e0132145.

<https://doi.org/10.1371/journal.pone.0132145>

Shelley, M. (1994). Gain/loss asymmetry in risky intertemporal choice. *Organizational Behavior and Human Decision Processes*, 59(1), 124–159.

Shenhav, A., Rand, D., & Greene, J. (2012). Divine Intuition: Cognitive Style Influences Belief in God. *Journal of Experimental Psychology: General*, 141(3), 423–428.

Shoda, Y., Mischel, W., & Peake, P. K. (1990). Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions.

Developmental Psychology, 26(6), 978–986. <https://doi.org/10.1037/0012-1649.26.6.978>

Sirota, M., Theodoropoulou, A., & Juanchich, M. (2020). Disfluent fonts do not help people to solve math and non-math problems regardless of their numeracy. *Thinking & Reasoning*, 1–18.

- Šrol, J., & De Neys, W. (2020). Predicting individual differences in conflict detection and bias susceptibility during reasoning. *Thinking & Reasoning*, 1–31.
- Stanovich, K. E. (2018). Miserliness in human cognition: The interaction of detection, override and mindware. *Thinking & Reasoning*, 24(4), 423–444.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, 39(7), 1275–1289.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.
- Watts, T. W., Duncan, G. J., & Quan, H. (2018). Revisiting the marshmallow test: A conceptual replication investigating links between early delay of gratification and later outcomes. *Psychological Science*, 29(7), 1159–1177.
- Weber, E. U., Johnson, E. J., Milch, K. F., Chang, H., Brodscholl, J. C., & Goldstein, D. G. (2007). Asymmetric discounting in intertemporal choice: A query-theory account. *Psychological Science*, 18(6), 516–523. <https://doi.org/10.1111/j.1467-9280.2007.01932.x>
- Welsh, M., Burns, N., & Delfabbro, P. (2013). The cognitive reflection test: How much more than numerical ability? *Proceedings of the Annual Meeting of the Cognitive Science Society*, 35(35).
- West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology*, 100(4), 930.
- Wetzels, R., Grasman, R. P. P. P., & Wagenmakers, E.-J. (2012). A Default Bayesian Hypothesis Test for ANOVA Designs. *The American Statistician*, 66(2), 104–111. <https://doi.org/10.1080/00031305.2012.695956>
- Xu, Y., & Liang, Z.-Y. (2021). *The foreign-language discount effect: Using English increases intertemporal discount rates through more distant future perception* [Preprint]. Open Science Framework. <https://doi.org/10.31219/osf.io/vrtf4>

Yechiam, E., & Hochman, G. (2013). Losses as modulators of attention: Review and analysis of the unique effects of losses over gains. *Psychological Bulletin*, 139(2), 497–518.
<https://doi.org/10.1037/a0029383>