2023

Neo-Whorfian Examination of Cross-Linguistic Temporal Discounting Behavior

Piper Connelly

NEO-WHORFIAN EXAMINATION OF CROSS-LINGUISTIC TEMPORAL DISCOUNTING BEHAVIOR

by
PIPER M. CONNELLY

SUBMITTED TO SCRIPPS COLLEGE IN PARTIAL FULFILLMENT OF THE DEGREE OF BACHELOR OF ARTS

DR. SEAN MASAKI FLYNN
DR. GALIA BAR-SEVER

MAY 3, 2023

Acknowledgements

This thesis has been an absolutely wild journey – yes, it is my independent work, but I couldn't be more grateful and proud of the help I've had along the way. To Dr. Sean Masaki Flynn (Scripps College) and Dr. Galia Bar-Sever (Pomona College) – thank you for having faith in how zany this proposal seemed at first glance and supporting my ambition of combining the fields of economics and psycholinguistics. I deeply appreciate all the last-minute meetings, furious white board scribbling, and never-ending encouragement from both of you. It's been an honor to continue soaking in your pedagogy beyond the traditional classroom experience, and I'm humbled to have your approval on my work.

Resounding and warm thanks go to Dr. Mathias Sutter (Max Planck Institute) and Dr. Robin Melnick (Pomona College) and for help in fine tuning the participant experience of my study and strategy for wrangling Amazon Mechanical Turk. Your notes of advice, both big and small, have truly gone far – even past the confines of this work.

A loving shout out to Dr. Allison Harris (Claremont McKenna College) for introducing a new light of consideration to economic behavior though your Neuroeconomics class. The direction of the procedure in my experiment was incredibly inspired my experimental economic literature we read in that class.

Also important is what sits aside from the technical, intellectual interworkings of academia. I owe so much of the actual writing process to my friend, Kat Ford. Kat is not an economics major, nor a linguistics major. In fact, she doesn't even like numbers a lot. However, she understands my work just the same through critique of my writing along the way. When I felt warpped up in dense jargon, her voice of reason helped steer my prose to be approachable.

ABSTRACT

This study examines differences in temporal discounting tendencies in German and French participants (recruited through Amazon Mechanical Turk) through the lens of Neo-Whorfianism and the Linguistic Savings Hypothesis (Chen 2013). The LSH proposes that tendencies towards future-oriented economic decisions can be cognitively explained by literal morphosyntactic conventions of one's native language. Our experiments (sooner-smaller/larger-later choices, endowment-investment task) failed to produce results aligning with the LSH, but uncovered the importance of controlling for risk appetite when specifically investigating intertemporal choice. There are several fruitful improvements to consider for the future, such as stricter sampling, taking richer detail of time preferences, and more robust risk controls.

1: Theory and Literature

The Sapir-Whorf hypothesis (SWH) makes the claim that specific features of one's native language affect one's behavior and one's general world perspective. These effects can range from the way you and your language refer to directions[1], how to name colors[2], or subconsciously categorize objects[3]. Sapir's and Whorf's original application claimed that features of a language, such as lexicon or grammar, deterministically predict cognitive limits and worldviews for native speakers of that language[4]. It garnered some academic merit in the 1970's, but the foundational

[1] Haviland, John. (1998) "Guugu Yimithirr Cardinal Directions". Guugu Yimithirr is an Australian Aboriginal language and does not contain words akin to "left" and "right" in English. Instead, it describes position/location through cardinal directions – north, south, east, west. (Ex: Pass me that book *northwest* of the cup.) This is called "geographic directionality", compared to "egocentric directionality". Native speakers have a remarkable sense of direction/navigation and spatial awareness.

[2] Brown and Lenneberg, (1954) "A Study in Language and Cognition." Ethnopsychologists considered how different languages categorize colors. Zuni, a Native-American language spoken by the Zuni people of western New Mexico, does not have separate words to describe (what English speakers call) "orange" versus "yellow". In this experiment, Zuni subjects showed some difficulty in recalling small changes in hue close to the category boundary of yellow and orange.

[3] Boroditsky, L., Schmidt, L., & Phillips, W. (2003). *Sex, Syntax, and Semantics*. Researchers found that the grammatical gender of an object affects how the native speaker subconsciously comprehends it. Spanish speakers and German speakers were asked to record the first three adjectives that came to mind when shown pictures of various inanimate objects. When asked to describe a picture of a bridge, the Spanish speakers used adjectives of a masculine connotation, such as "big, dangerous, strong, towering". German speakers used adjectives of a feminine connotation, like "elegant, peaceful, beautiful, slender". This experiment was conducted in English (which lacks grammatical gender) to avoid priming effects of using participants' native (gendered) languages. This reflects that native Spanish speakers consider the word *and* notion of "bridge" to be masculine, and native German speakers consider the word *and* notion of "bridge" to be feminine. This aligns with the Neo-Whorfian approach that subjects' perception of objects is influenced by arbitrary grammatical properties assigned by native language.

[4] Whorf et al., 2012. *Language, thought, and reality: selected writings of Benjamin Lee Whorf.*

research by Sapir and Whorf themselves presented clear downstream threats of attaching morality, marginalizing, or claiming intrinsic superiority of one language - therefore individuals speaking that language - over another. The hypothesis has been disproven and largely discarded since the 1990's, but cognitive psychologist Steven Pinker reawakened it through coining the term "Neo-Whorfianism" in his 2007 book, *The Stuff of Thought: Language as a Window into Human Nature.* His resurrection is a weaker, more nuanced application of the SWH – language *influences* the speaker's thoughts and worldview but doesn't limit or define it. Researchers including Boroditsky[3,5], Fausey[5], and Schmidt[3] have empirically demonstrated intriguing cross-linguistic behavior supporting Neo-Whorfian psycholinguistic effects quite neatly. (See footnotes for references and annotations.)

The domain of Neo-Whorfian cognition I focus on here is future time perception and delayed gratification behavior. The psychological mechanisms at play and further discussion of these decisions will come briefly, but it is vital to first understand the linguistic premise underlying the experiment reported on in this thesis. The cross-linguistic novelty of interest in this study is a language's grammatical obligation of a verb to reference the future via either a morphosyntactic[6] change or a semantic[7] reference.

The European Science Foundation ran a large-scale project that catalogued typological[8] traits of languages in Europe, known as the EUROTYP Project[9]. Osten Dahl (2000) draws upon two large prevalent grammatical constructions of verb temporality (indication of the past, present, or future) in his EUROTYP volume, *Tense and Aspect*. He refers to this paradigm as "future time reference", or "FTR", where he sees two modalities that hinge on what changes are required of a verb in any given sentence to express an action in the present versus future tense. Some languages encode verbal FTR morphologically, and others rely on context and semantics.

---

[5] Fausey and Boroditsky (2011) found that native English and native Spanish speakers remember different aspects of intentionality, accident, and fault when shown the same incident. The differences in their recollection of events directly mirror the way each language syntactically expresses agentive vs. non-agentive actions, as well as intentional vs accidental actions.

[6] Morphosyntax – key feature of grammar – the interaction of how words are made in a language (morphology), and how a language arranges words into meaningful expressions (syntax).

[7] Semantic – what words mean, how we humanly interpret and comprehend words and sentences.

[8] Typology (as used in Linguistics) - the study of classifying languages according to structural features

[9] See citations for European Science Foundation and Dahl (2008)

Certain languages, including the Greek, Polish, Swahili, and those of the Romance family, require a specific word form of a verb to convey whether an action occurs in the present or future. (1) and (2) from French compares a situation in the present tense and in the future.

(1) Les enfants jouent aujourd'hui.
the children play-**PR** today.
'The children play/are playing today.'

(2) Les enfants joueront demain.
the children play-**FUT** tomorrow.
'The children will play tomorrow.'

In this example, the FTR of when the children will play is marked explicitly by the verb changing from "*jouent"* in (1) to "*joueront"* in (2). Even if the markers specifying "today" and "tomorrow" were not part of the sentence, the speaker and listener would still fundamentally understand whether children are playing presently or not yet. Dahl calls this "strong future time reference" (sFTR), by virtue of these verbs *strongly* referencing the future – they intrinsically contain information on tense. By choosing a tense for each verb, speakers must categorize the action they want to express according to the (presumably finite) span of options available in the language. Hypothesized psychological effects will be discussed in detail shortly.

To contrast, languages such as Finnish, Mandarin, Hawaiian, and most of the Germanic family[10], do not mandate[11] verbs referencing the future to change into a unique form (or even distinguish verb tense at all). Instead, position in time is often announced using adverbial (or prepositional) references, as seen in these German examples.

(3) Die Kinder spielen heute.
the children play-**PR** today.
'The children play/are playing today.'

(4) Die Kinder spielen morgen.
the children play-**FUT** tomorrow.
'The children will play tomorrow.'

[10] English is an outlier of the Germanics. In the rest of the family, grammatical future-time reference is optional when making predictions that have no intentional component,

[11] *Mandate* is the operative word here. German contains grammatical cases which do contain auxiliary future markers, akin to English (words like "will" in "will go"). However, these constructions are not the only permitted way to designate future. See *Tense and Aspect* (Dahl 2000) for in depth typology.

Between (3) and (4), we see that the verb "*spielen"* takes on the same word form regardless of tense. It does not independently contain information on temporality; indication that the action takes place in the future is delivered separately by the presence of "*morgen*" in (4) (where, syntactically, it is an adjunct!). Time reference is semantically applied to the verb upon declaring a specific point in the future for this to exist (today, tomorrow). Dahl named this trait "weak" future time reference (wFTR); these languages' verbs *weakly* express time, thus rely on sentence context to acquire a time reference. Speakers are not required[12] to declare temporality using verbs, and ambiguity is addressed by sentence-level semantic. (As stated previously, psycholinguistic literature will be discussed in detail shortly.)

This typological pattern of FTR is the linguistic paradigm of my experiment. Next, I pivot to economic inquiry of time preferences and irrational decision making, which is my experimental mechanism to evoke behavior aligned with Neo-Whorfian hypotheses.

Aside from language, this study focuses on delay aversion and temporal discounting. Objectively, it seems irrational to opt out of a big payout in favor of a smaller one if all you have to do is wait for it. However, the cost of patience is too high for some folks. Samuelson (1937) consolidated the Discounted Utility model to describe how the subjective expected utility that one places on a future reward decreases as the waiting time for the reward increases. Thus, people may be willing to accept a smaller outcome that arrives immediately instead of waiting for a larger outcome in the future, even if the larger reward has a higher overall utility. However, attitudes towards patience and time value of money are individual preferences – some people are more tolerant of delayed rewards, others less so. For example, would you rather receive (for certain) $50 right now, or $75 in 2 weeks? What about $50 now vs $100 in 2 weeks? Or, $50 now vs $100 in a month? Given a large enough sample and, ideally, several more decision points in order to create a gradient, a spectrum for degree of delay aversion arises. These are all questions that tease out one's personal discounting preferences, which also reflects the magnitude of patience or tolerance one has for waiting for a reward. Lifestyle conditions such as income, education, current savings level, and employment status are documented influences on

[12] Careful note- these languages lack the grammatical *requirement* for tense markers. Across all the world's languages, there exists a gradient of how future time is acknowledged which is richer than the conservative binary here. Some languages considered wFTR for this experiment can sometimes express verb tense morphologically, but these constructions can usually be equivalently re-expressed with weakly grammaticalized FTR. Dahl (2000), Thieroff (2000), and Chen (2013) adhere to the condition of *obligatory* time reference by verbs and consider strong and weak FTR a binary.

discounting preferences (Watts et al., 2017). Individual traits such as age/sex (Olson et al. 2007), intelligence, and risk appetite (Dohmen et al. 2010), have been demonstrated to influence discount preferences as well.

Neo-Whorfian hypotheses can integrate the FTR duality and decision preferences of intertemporal choice. Chen (2013) applies the FTR duality to how people comprehend the abstract concept of present and future based their native language. Chen posits that, because sFTR languages require more precise, literal references of future (via the verb), then native speakers should inherently comprehend future events as distinct from the present. To complement, wFTR languages' lack in necessity for verbs to define temporality should result a less precise separation of the abstract "future" from the "present". He introduces his original "Linguistic Savings Hypothesis" (LSH), claiming that, because wFTR speakers lack a mandatory expression of future reference, they perceive the future as less distant from the present, or "more imminent", than sFTR speakers. Therefore, wFTR speakers should be more inclined to make economic decisions that are more future-oriented than sFTR speakers. Because sFTR languages strongly separate present from future (i.e., the future is viewed as something distinctly other than the present), sFTR speakers are considered to view the future as less imminent, and therefore make less future-oriented economic decisions. See Appendix for visualizations by Thoma and Tytus (2017). Chen found impressively robust empirical evidence in this favor as well from high-level survey data – within countries containing large populations of both native sFTR and native wFTR citizens, wFTR populations exhibited far more future-conscious lifestyle and financial habits (after appropriate controls), such as lower smoking and obesity rates, more wealth at age of retirement, as well as savings rates at a national level (Chen, 2013). That is not to mean the sFTR population is at a deficit; there were no specific areas of concentrated sFTR deficiency. Chen's research consisted of processing preexisting responses to large-scale panel data taken on sFTR and wFTR populations. Even though he finds the results he was hoping for, his procedure does not allow for evidence of a causal relationship between FTR modality and tendency for future-oriented choices.

Sutter et al. (2015, 2018) applied the Linguistic Saving Hypothesis to decision tasks of intertemporal choice to investigate causality. They recruited elementary school-aged children in Mehran, Italy, which is a town with a demographically even split of native Italian-speaking families (sFTR language) and native German-speaking families (wFTR language). Elementary schools either teach in German *or* Italian, and parents send their children to a school that teaches in their native language.

Researchers put the children through a series of economic decision scenarios to measure inclination for making patient decisions. They set up a currency of tokens, which could be exchanged for a little trinket of the child's choosing. There were two patience tasks with these tokens: choices between a smaller sooner reward vs larger later reward, and an endowment - investment task. For the first task, the kids chose between receiving 2 little tokens immediately, or 3, 4, or 5 tokens in 4 weeks. After the 4 weeks, the kids received their reward. In the second task, kids were given 5 tokens, and they could either take them immediately (to exchange for a trinket), or they could "invest" their tokens, which would double in value to be worth two trinkets (instead of 1) in 4 weeks. Children had to judge how many trinkets to give up in the present in exchange for more in the future. Again, the kids received the appropriate rewards at the end of the waiting period. They found that, on the whole, German kids made more patient decisions than the Italian children. This correlation became visible after controlling for age, IQ, and family background, which already are well-documented informants of self-control (Watts et al., 2018), subjective value estimation (Olson et al., 2007), and abstract perception of the future (Steinberg et al., 2009).

Researchers also ran a control for risk attitudes. This was a simple lottery task – children were again given 5 tokens, but they had to choose how many tokens to safely keep (to redeem for a trinket), or to enter into the lottery. The lottery had a random chance of doubling the amount of tokens entered, and participants received the gambling results immediately. In the behavioral control, there was no correlation found between gambling behavior and native language, indicating there were no underlying discrepancies in risk-seeking behavior. Thus, variation in delay discounting was explainable by the LSH.

2: Methods & Data

My experiment uses the LSH to capture differences in individuals' time preferences relative to their native language. According to Chen (2013), languages which do NOT require tense-related modifications to a verb expressing a future prediction from the present (weak FTR) are thought to induce more future-oriented behavior than languages that *do* require distinct verb tenses to express a future action and present actions (strong FTR). Chen's idea applies economic consequences to a Neo-Whorfian proposition that native speakers of wFTR languages may have the psychological implication of considering future events to be imminent as a result of the literal syntactic features. Both mechanisms would theoretically cast downward pressure on individuals' personal temporal discounting tendencies, hypothesizing a lower magnitude of delay aversion in native wFTR speakers than native sFTR speakers.

Sutter et al. (2015, 2018) gathered incredibly clear results in support of Chen's LSH using relatively low-level decision tasks on children. Using methods parallel to Sutter et al. (2015), I hope to find corroborating evidence for Chen's LSH in adults.

I constructed my experiment in the format of an online survey using Qualtrics that adults took independently, anonymously, and remotely using their own computer. It consisted of 3 sections: economic decision tasks, intelligence metrics, and a questionnaire for personal opinions and demographic information. I designed the economics decision tasks to mirror the tasks implemented by Sutter et al. (2015). Subjects were rewarded for successfully completing the survey and passing all human validation & cleaning checks.

One specific decision I made for this experiment was to force subjects to take the test in their native language. (The execution of this will be explained in greater detail briefly) Keysar et al. (2011) found a strong trend among Japanese native speakers in the USA in which using a foreign language "reduces decision-making biases" in choices of risk and loss. The lack of native intuition and fluency in a second language could cognitively lead to more analytical, rational decision making. Sutter et al (2015) and Fausey and Boroditsky (2011) both draw conclusions on the effects of Neo-Whorfianism through matching the native language of respective participants for administering their experiments.

In the realm of subject recruitment, the most feasible way to readily access a body of native French and native German speakers for this experiment ended up being via Amazon Mechanical Turk. Because participants took my survey remotely, the best I could do for "reward currency" throughout the experiment was to instruct participants to consider an imaginary currency of tokens instead of real money. I described the tokens to subjects such that I asked the subjects to "…not consider them [tokens] as real money. Imagine a token is guaranteed to be redeemable for equally valued items of your choosing such as food, toys, trinkets, etc." This clearly beckons the "house money", or "funny money" phenomenon, where participants in psychological or economics experiments do not feel a true attachment or scarcity to fake money used in experiments, and therefore do not behave as if their own resources were on the line. Unfortunately, this is an unavoidable issue. By nature of the MTurk platform and the cash flow mechanics of “Turkers” completing tasks, I was unable to grant monetary consequence (via compensation) specific to their unique decisions across the experiment. The subjects themselves were also aware of this - they were as a result (very reasonably) only focused on the fixed reward that comes upon completion.

During the experiment, there were three main decision tasks – two pertaining to delay aversion and intertemporal discounting and one measuring risk attitudes.

In Task 1, participants were presented with a series of decisions prompting them to choose between smaller-sooner rewards or larger-later rewards. Participants were shown two boxes containing tokens for each scenario - “NOW”, or “4 WEEKS”, and they were presented with a series of questions – would you rather have 6 tokens now, or 7 tokens in 4 weeks? 6 tokens now, or 8 tokens in 4 weeks? 9 tokens? 10 tokens? 11 tokens? 12 tokens? I assessed the point at which participants switch from preferring 6 tokens to preferring the later option, at whichever outcome that may be. I refer to this roughly as the “indifference boundary” – the minimum patient reward which is more attractive than the immediate reward. Thus, having a lower indifference boundary means making more patient choices, therefore a lower discounting rate. According to the LSH, native wFTR languages should encourage downward pressure on discount rate, encouraging more patient choices than sFTR speakers.

The second task of delay aversion was analogous to the first task by continuing to tease out individual future value-to-present value indifference points. Using tokens and boxes again, subjects were endowed 5 tokens. They were instructed to decide how many tokens they prefer to receive immediately by clicking and dragging them into the “NOW” box, and how many tokens they would rather collect for double the value in 4 weeks by putting them to the “4 WEEKS” box. In comparison to the first task, this allows participants to autonomously demonstrate their optimal endowment management, thus desired cash payout, instead of comparatively preferring one predetermined option more or less. The LSH predicts that wFTR speakers should elect to invest more tokens than sFTR speakers.

The third task assessed risk-seeking inclinations and was done with a lottery decision game. Participants began with an endowment of 5 tokens and were instructed to decide how many they prefer to securely receive immediately, using the “NOW” box, and how many they prefer to enter in the lottery – random chance of double or nothing. There should be no statistical difference in risk attitudes between language groups, which would allow any discounting patterns to be explained by the LSH.

After the decision tasks, I followed with a section of tests to generate a relative level of "intelligence" using two common non-verbal (language agnostic) tests historically included in official IQ batteries – code-digit matching and matrix reasoning. There is extensive literature correlating performance on IQ style tests as well as formal IQ to risk aversion (Dohmen et al, 2010), as well as propensity for delayed gratification (Sutter et al, 2015, 2018).

Code-digit matching is a classic test in the non-verbal portion of the Wechsler Adult Intelligence Scale (WAIS-IV), measuring processing speed and working memory. I constructed the code of symbols myself, minimizing sub patterns within the code as much as possible.[13] Participants have 90 seconds to correctly match as many numbers as possible to randomly generated[14] sequences of symbols. (See Appendix for the code of symbols used.) Participants were allowed 90 seconds to identify as many symbols as they could, with a max score of 50 symbols. To capture a similar metric of inductive and abstract reasoning applicable to adults[15] taking this exam independently on their own computer, I implemented the Short Version of the Hagan Matrices Test (HMT-S)[16]. The HMT-S is an open-source matrix reasoning test, abbreviated to 6 questions selected from the full-length HMT, which has 20 questions. Participants had 2 minutes per matrix to submit an answer before being automatically moved on to the next problem.

The nature of my execution and distribution meant I was unable to administer a formal IQ battery, nor control the immediate environment of participants while taking these time sensitive tests. An additional, but crucial consideration is that I was unable to control how motivated participants felt to perform to the best of their ability and how seriously they took it.

[13] Making the code took a little trial and error. I didn't want the participants to be able to detect patterns in the symbols that would serve as shortcuts for learning the full code. Example cases of what to avoid: triangular shape = 3, equals sign = 2, star shape = 5.

[14] Used R for generating random strings of digits 0-7. All participants got the same sequence of symbols.

[15] Sutter et al (2015) used *Raven's Coloured Progressive Matrices* to measure IQ in school children, which is designed to specifically measure level of fluid vs crystallized intelligence in children ages 6-11 and adults with cognitive impairments (Raven et al, 1998).

[16] Validation done by Heydasch et al (2020). A full-length HMT exists, but I ultimately came to sacrifice the length and thoroughness of this for constraints of a) retaining the focus and engagement of participants on MTurk, and b) monetary budget to provide market-appealing compensation for their time.

The final section of my experiment required participants to provide information about themselves – financial status, employment status, societal and personal values[17], and language background.

Overall, this experiment likely would have been better suited to the environment of classical behavioral experiments, in both psychology and economic decision making - sitting down in-person with the experimenters to understand the rules thoroughly, act with realistic judgement, be in a controlled environment, and keep the participant on task[18]. As far as trying to make use of Amazon Mechanical Turk goes, my sample size of n = 40 is clearly an initial dry run for human validation of what should eventually become hundreds of thousands of samples[19].

Subjects for this study were recruited using Amazon Mechanical Turk. The experiment was translated into a German version and a French version. The German translation was made available to "Turkers" (MTurk workers) with an IP address in Germany, Austria, Switzerland, or Belgium. The French translation was made available to Turkers with an IP address in France, Switzerland, or Belgium. These countries were chosen for their high population of German and French native speakers, respectively. Independent samples (no pairwise comparisons) of 20 German and 20 French replies passed appropriate cleaning checks.

In order to truly identify subjects whose native languages were German and French, subjects answered the following questions:

1. *List all languages you have at least an intermediate level proficiency in, beginning with your primary language.*
2. *What language(s) did the adults around you speak when you were a child?*
3. *Where are you from?*
4. *Where are you currently?*

Participants who listed anything other than French or German first in their answer to #1 were omitted[20]. #3 and #4 were included as crosschecks for MTurk's IP address filtration. Subjects

---

[17] Participants were asked to express qualitative opinions on several broad topics of morality and conduct in personal sphere. See Appendix for specific prompts. Overall, this data was very noisy and did not yield meaningful averages nor regression influence. This data overall was not used for further consideration.
[18] For example, college kids getting locked in a room for cash.
[19] If I were to run this experiment on a larger scale, I would make wider use of MTurk's international presence to test several languages across strong and weak FTR.
[20] Targeting native language is fundamentally crucial to the Sapir-Whorf hypothesis. Existing research done on business practices in bilingual workers reflects that people more frequently make rational choices when conducting business and making economic decisions while operating in a second language. (Keysar et al., 2012)

who were not from countries that each survey was distributed to were omitted. No subjects reported being currently located outside the appropriate countries.

Table 1 contains personal demographic information of the participants. French subjects were older on average, with a mean age of almost 44 years compared to a mean age of 34 among German subjects. The high standard deviations by group and in aggregate reflects the wide age range, spanning 20 to 59 years old. German subjects were between 20 and 50 years old, and French subjects spanned 28 to 59 years old. Mann-Whitney U-Test (henceforward abbreviated as "MWU test") indicates that these two groups have significantly different ages ($p = 0.0036$), which could likely indicate different behaviors or perspective on spending, saving, and risk. The MWU test indicates no significant difference in the proportion of men to women between groups ($p = 0.300$). The education category also yielded a large span of participant answers, ranging from 0 years to 22 years completed. Average education levels were not robustly different across groups as indicated by a MWU test ($p = 0.342$), with French subjects completing an average of 9 years of education and German subjects completing an average of ~11 years. Participants did not have an opportunity to specify what types of education they received. For the means of this experiment, it is fine.

Table 1: Participant Demographics - Personal

	Mean Age (yr) ± SD	# Men: # Women	Mean Years of Education ± SD
German (n = 20)	34.3 ± 7.6	16:4	10.9 ± 6.1
French (n = 20)	43.8 ± 10.1	12:8	9.0 ± 7.2
Total (n = 40)	39 ± 10	28:12	10.0 ± 6.6

Table 2 contains information reported by subjects about their personal economic footprint. Participants were asked to categorize what percentile level of income they consider themself in their country[21] by the given brackets; 1: <20%, 2: 20%-40%, 3: 40%-60%, 4: 60%-

[21] Trusting a self-assessment of "How rich are you" certainly has veritable concerns. However, this functions to be a more versatile/surface level approach than asking for raw dollar amounts of yearly income. Not only could truthfulness become more of a concern, but it would demand attention to currency exchange rates, country by country costs of living, taxation structures, etc. If this hypothesis and experiment were to be revamped, I would gather finer details about income, debt load, and other financial assets.

80%, 5: >80%. An aggregate mean of 2.7 indicates that most subjects were within the 20th-60th (categories 2 and 3) percentiles of income for their country. Similar standard deviation sizes (± 1.14 for Germans, ± 1.28 for French) suggest that there could be a comparable distributional spread of income levels across groups. There is no significant difference in average income range between language groups, with a MWU test returning a p-value of 0.555.

The next column "*Currently Employed*" summarizes binary job status by language. Slightly more German subjects currently had a job than French subjects, however with a sample size this small, that difference was only 3 people; 7 French subjects were unemployed and 4 German subjects were unemployed. To concur, a MWU test did not find a significance in the difference between employment levels ($p = 0.149$).

The next factor, Retirement Savings Status, is an ordinal variable with levels 0, 1, and 2. In both language groups, the same majority proportion of subjects reported having no savings for retirement (= 0). he "Has Retirement" and "Actively Saving for Retirement" focus on subjects' answers the following question given the possible responses:

Do you have savings for retirement?

1. *Yes, and have added to it in the last 6 months*
2. *Yes, but have not added to it in the last 6 months*
3. *No*

Equal proportions of each language have some retirement savings, and there's only 1 person of difference between Germans actively saving and French actively saving. Thus, it is very safe to say that there is no statistical difference between the retirement savings behavior between language groups. The fact less than half of all subjects have retirement savings at all could potentially be influential for regressions later on as a metric for experiential/empirical propensity for delayed gratification, but also could inform potential interactive effects with income level, age, employment status and risk appetite.

Table 2: Participant Demographics – Economic and Financial					
	Avg Income Bracket ± SD	Currently Employed	Retirement Savings Status		
			None Currently	*Has Savings*	*Added to Savings in Past 6 Mo*
German (n = 20)	2.6 ± 1.14	85%	55%	5%	40%
French (n = 20)	2.8 ± 1.28	65%	55%	10%	35%
Total (n = 40)	2.7 ± 1.20	72.5%	55%	7.5%	37.5%
Percentile Income Brackets: <20%, 2: 20%-40%, 3: 40%-60%, 4: 60%-80%, 5: >80%					

Two quantitative scales measured relative intelligence. A digit-symbol identification task assessed processing speed and working memory, and a matrix reasoning task for inductive and abstract reasoning skills. Table 3 contains the result data from each task. Overall, the French subjects performed better on both tests, with not only higher means, but less variation in performance within the group as indicated by smaller standard deviations. However, Mann-Whitney U-Tests found no statistically significant difference between groups' performance on digit identification ($p= 0.139$), as well as on matrix solving ($p = 0.149$). Matrix accuracy and digit performance had a moderately strong and statistically significant correlation ($r = 0.43$, $p = 0.0057$). It's reasonable that these do not strictly follow a perfect correlation – they test for different facets of intelligence. For regression functionality, I determined a composite score that maximizes correlation to both metrics.[22] As expected, these composites do not significantly differ between languages (MWU = 0.119).

Table 3: Intelligence Test Performance			
	Mean # Digits (± SD)	# Matrices Correct (± SD)	Intelligence Composite (± SD)
German (n = 20)	32.6 (± 9.6)	3.25 (± 2.0)	42.4 (± 13.7)
French (n = 20)	36.0 (± 7.4)	4.25 (± 1.5)	48.8 (± 9.6)
Total	34.3 (± 8.6)	3.75 (± 1.9)	45.58 (± 12.1)

[22] Intelligence Composite Score = (n digits) + 3(n matrices correct); which is highly significant to digit performance, ($r = 0.91$, $p < 0.005$) as well as raw matrix accuracy ($r = 0.76$, $p < 0.005$). I chose to summarize scores using this weighting to grant a larger reward for correctly answering the matrix questions since they do not have a 1:1 difficulty as matching one digit to one symbol. I also was sure to not over-fit the data to an extent that composite scores presented significant differences between languages.

Before considering discount preferences, it is crucial to first control for risk attitudes. Chen's Linguistic-Savings Hypothesis only applies to intertemporal choice – not choices under risk. Discount preferences have been found to relate to risk appetite (Anderhub et al., 2001; Epper et al., 2009; Dohmen et al., 2010), so we must determine within the data that there are no underlying patterns of risk appetites between French and German speakers that could confound any trends extracted from discount behavior. Figure 1 and Table 4 contain the breakdown of lottery behavior by language. There is a clear separation of how each group behaved – French individuals were far more risk-averse than Germans, with nearly opposite averages for tokens safe vs gambled. This offers a clear suggestion that there is an interaction effect. Logistic regressions will be used to determine whether there is a statistical significance in the risk behavior manifested by speakers of the two languages.

Figure 1

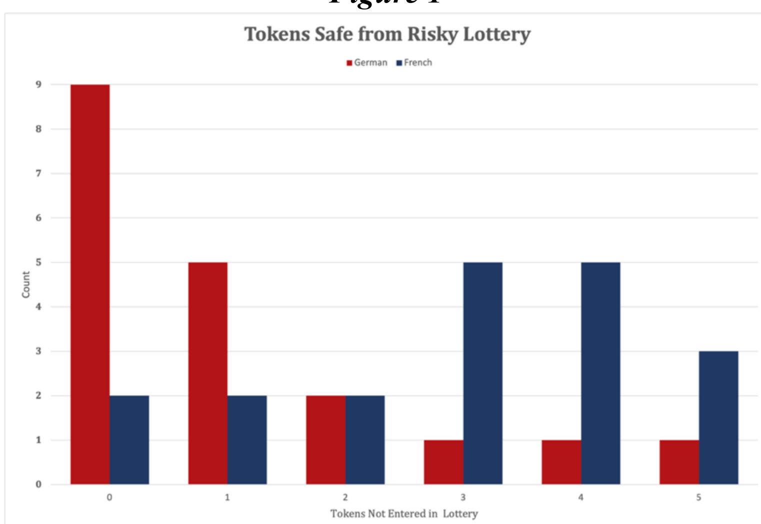


Table 4: Risk Control, Double-or-Nothing Lottery			
	Mean # Tokens to Keep	Mean # Tokens Played in Lottery	± *SD*
French (n = 20)	3.75	1.25	± *1.25*
Germans (n = 20)	1.30	3.70	± *1.30*
Total	2.5	2.5	± *1.77*

We will build regressions that use our collected data to model risk behavior. For summarizing purposes, I decided to reduce risk aversion in the lottery task from an ordinal variable (Tokens saved = 0, 1, 2, 3 4, 5), to a binary variable (# Tokens Saved < or > # Tokens Gambled) in hopes to better combat noise that arose from small sample size and unengaged/unthoughtful participants. Participants who saved more tokens than they gambled were considered risk averse, and participants who gambled more tokens than they saved are considered risk tolerant or risk-seeking.

First, we regress risk aversion solely over language to determine whether it is a significant factor. Within the scope of this diagnostic, the direction (+ or -) of β does not matter. As expressed in the null hypothesis below, the default prediction claims that language is an insignificant regression predictor of risk aversion for our sample.

For: logitP(Tokens Saved > Tokens Gambled) = β(Lang) + α + ε

H_0: $\beta = 0$

H_A: $\beta \neq 0$

Column 1 of Regression Table 1 shows the estimated beta for the logistic regression of participant language predicting whether they will behave risk-aversely by saving more tokens than they entered in the lottery. The overall model is not a strong predictor of the data, indicated by the *pseudo-*R^2 of 0.193. When regressing risk averse tendency ***solely*** over language, there exists a significant regression coefficient that is nonzero. Thus, the null hypothesis is rejected, and we must accept that there exists a significant correlation between language and lottery behavior for the participants in this study. However, though, the constant term is *also* statistically significant, which highlights that there is variation in risk aversion that *cannot* be explained by language. More analysis is required.

Regression 1: Lottery Experiment for Risk Attitudes

Logit Regression: p(Tokens Saved > Tokens Gambled)			
	(1)	(2)	(3)
	H_0	*Language-Agnostic*	*Composite*
Language	2.234^{**}		10.70^{*}
(GER = 0)	(0.742)		(4.330)
Sex		-0.847	-6.329^{*}
(M = 0)		(0.920)	(2.986)
Age		0.0145	-0.231^{*}
		(0.0363)	(0.111)
Employed (=1)		-0.811	-1.470
		(0.888)	(1.505)
Intelligence Score		-0.0349	-0.240^{*}
		(0.0337)	(0.114)
Relative Income Lvl		0.0469	-0.603
		(0.316)	(0.620)
Retirement Savings		-0.0468	-0.332
Status		(0.401)	(0.670)
Education		-0.0787	-0.104
		(0.0642)	(0.0940)
constant	-1.386^{*}	2.343	19.93^{*}
	(0.559)	(2.587)	(9.232)
cut1	0.2***	0.489***	0.365***
	(0.068)	(0.074)	(0.049)
N	40	40	40
Pseudo-R^2	0.193**	0.089	0.560***

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Significance on pseudo R^2 refers to χ^2

Next, let's examine risk aversion using the demographic information gathered. External to Whorfian effects, individuals' financial risk tolerances are highly sensitive to economic lifestyle pressures, as well as intrinsic personal traits. Grable (2000) identified several variables that undeniably affect risk tolerance, which overlap with the economic and demographic information I collected from participants – age, sex, education level, relative income bracket, and retirement savings status[24]. In Grable et al. (2000), they saw comparatively lower levels of risk aversion (i.e., higher risk tolerance) in older subjects compared to younger subjects, men compared to women, subjects with more years of education, and higher income subjects compared to lower income. I also include intelligence and employment status. Dohmen et al. (2010) administered intelligence tests standardly found on official IQ batteries and concluded that higher relative IQ scores distinctly translated to more risk-tolerant decision making. Díaz-Serrano and O'Neill (2004) successfully asserted the claim that unemployed individuals are more likely to be more risk-averse.

Predicting a well-studied (but non focal to this paper) phenomenon such as risk aversion using classic and well-documented variables serves as a check for human validation and sufficient data collection measures. There are several variables that followed preexisting beliefs on their general impact on risk attitude, but they did not achieve statistical significance. Also, the strength of the model in terms of fitting the data was very low – a *pseudo-*R^2 of 0.089 – which is actually less than half of the predictive strength of the language-only model. This is a clear indication of a poor participant sampling and/or poor data collection procedures. Please refer to concerns on participant and procedural execution quality.

Column 2 of Regression Table 1 shows results for a logistic regression of loss averse behavior demographic variables recorded from participants, excluding their native language. Overall, this data produced a very noisy and ill-fitting model. For the coefficients, there are mixed results in successfully replicating documented trends. The negative coefficient for the variable Sex in Column 2, although not statistically significant, does not align with the prior literature. This model suggests that female participants were more likely to *not* be risk-averse in the lottery task. Also, the direction of the coefficients for Age and Relative Income Level do not concur with existing research either (but neither reached statistical significance). Here, older participants and higher-earning participants were (barely) more likely to behave risk-aversely than their counterparts. Income level also seems to directly contradict the results for retirement savings; retirement savings is one of the variables here that *do* follow prior literature. Participants who have contributed to their retirement savings accounts are less likely to behave risk-aversely in the lottery task. Employment status, intelligence score, and education attainment

[24] In this context, retirement savings status seeks to function as relative dimension of wealth, alternatively to income. Retirement savings is interpreted as voluntary delayed consumption, not an inquiry to loss aversion.

also agree with existing research. Participants with greater levels of education, participants who scored higher on the relative IQ tests, and participants who were employed at the time of taking the experiment were all more likely to *not* be risk averse with their tokens in the lottery.

Column 3 of Regression Table 1 combines the language and demographic factors to predict risk attitudes, and this elicits a much stronger model than the previous two with a *pseudo-* R^2 of 0.56. Both models that include the language variable, Columns 1 and 3, calculated a positive coefficient of relatively high magnitude and statistical significance. This is double confirmation that French speaking participants were *incredibly* more likely to be risk averse in this experiment than German speaking participants. This could potentially confound downstream behavior and will be addressed as needed.

When including language, the variables Sex, Employment, Intelligence Score, Retirement Savings, and Education become far more influential (greater magnitude) and statistically significant than the non-language regression in Column 2. The variables Age and Relative Income Level flip signs when language is a covariate, and they both increase in statistical significance. (Even though Rel Income Lvl is not at the formal threshold for significance, the t-statistic increases significantly). Now, all the variables in the data set follow the hypothesized influences outlined in previously regarding Column 2. Language must capture some wide-spread patterns across these variables, but this model returns a significant constant value even larger than the previous two. Clearly, even though the model fit increased with more covariates, and more coefficients are statistically significant, there still is more variation in risk aversion that cannot be explained using the current factors.

Overall, we have identified that we cannot separate risk attitude from native language. The huge jump in prediction power when including language in the model alongside the other variables could also be a symptom of an overfitted the model. This exceedingly narrow dataset has not statistically met several hallmarks of human verification through failing to substantially demonstrate obvious and well-documented factors influencing risk appetite.

3: Task 1 Results – Indifference in Sooner-Smaller to Larger-Later Outcomes

The first batch of economic decisions consisted of 6 questions involving intemporal choice. This smaller-sooner or larger-later style forced participants to evaluate what increase in reward is worth waiting for, and we hope to detect discounting tendencies which align with the Linguistic

Savings Hypothesis. Support for the LSH would be characterized by German-speaking subjects making more patient decisions than French-speaking subjects, *ceteris paribus*.

First, we begin with the breakdown of how many people per language group made their first patient decision in Table 5. Overall, there were no statistically significant differences at any level in comparison of language groups (MWU, $p = 0.92$). As expected, there are likely to be other variables to control for in order to uncover stronger evidence of LSH effects.

Table 5: Task 1 - Future Payoffs that Motivated Subjects' *First* Patient Choice:

	Wait for 7	Wait for 8	Wait for 9	Wait for 10	Wait for 11	Wait for 12	Mean # Patient Choices (± SD)
German (n = 20)	5	6	4	4	1	0	4.4 (± 1.00)
French (n = 20)	4	8	4	3	1	0	4.45 (± 1.23)
Total	9	14	8	7	2	0	4.425

To analyze the patterns of time preference, I considered the very first patient choice taken by subjects. I refer to this as the indifference boundary/point. Lower indifference points indicate more patient decisions. Recall that according to the LSH, wFTR speakers are expected to demonstrate lower indifference points in delay aversion when compared to sFTR participants, all else equal. Higher indifference points mathematically indicate relatively larger discounting tendencies, which reflects that those individuals' subjective utility of the future outcome is cognitively less sensitive to delay aversion.

Column 1 in Regression Table 2 shows an ordinal logistic regression of language solely predicting which later reward - 7, 8, 9, 10, or 11 tokens – would first overcome the immediate attractiveness of 6 tokens. The negative direction of the coefficient disagrees with our hypothesis that German speakers should have a lower indifference boundary. Instead, this coefficient suggests that being French (Lang = 1) cast a downward pressure on what delayed reward first became preferable. That is, this coefficient says that French individuals were more likely to switch to the patient decision sooner.

Regression 2: Task 1 - Sooner- Smaller vs Larger-Later Decisions

Ordinal Logit Regression: p(Minimum Patient Reward = [7,11])			
	(1)	(2)	(3)
	H_0	*Language-Agnostic*	*Composite*
Language *(GER = 0)*	-0.0564 (0.568)		-0.287 (1.162)
Lang x Risk Avrs			-4.190^{*} (1.769)
Risk Aversion (Lotto Task)		0.784 (0.630)	4.525^{**} (1.505)
Sex *(M = 0)*		-1.483 (0.853)	-1.249 (0.916)
Age		0.0364 (0.0341)	0.0908^{*} (0.0407)
Employed (=1)		-0.00606 (0.755)	0.786 (0.834)
Intelligence Score		-0.0144 (0.0271)	0.0466 (0.0371)
Relative Income Lvl		-0.216 (0.293)	-0.182 (0.293)
Retirement Savings Status		-0.454 (0.361)	-0.675 (0.376)
Education		-0.0032 (0.0588)	-0.00186 (0.0587)
cut1	-1.266^{**} (0.482)	-1.687 (2.234)	3.637 (2.997)
cut2	0.273 (0.436)	0.104 (2.199)	5.670 (3.056)
cut3	1.208^{*} (0.475)	1.157 (2.205)	6.892^{*} (3.109)
cut4	2.917^{***} (0.777)	2.936 (2.331)	8.910^{**} (3.317)
N	40	40	40
Pseudo-R^2	0.001	0.063	0.133

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Significance on pseudo R^2 refers to χ^2

The regression coefficient itself is statistically insignificant, but the ordinal residuals below (cut1 – cut4) do contain significance. Here, language is shown to be highly significant predictors for the extremes – participants either being completely patient (immediately taking the first delayed choice of 7 tokens), or completely impatient (first patient choice being 11 tokens). Remember from Table 5 that all participants opted for 11 tokens and 12 tokens over 6, but we are interested in the point at which the delayed option becomes preferred. The negative direction of the *cut1* coefficient, which corresponds to the marginal likelihood between the first patient decision being 7 vs 8 tokens (having a very low indifference boundary; therefore, making patient choices very early on), suggests that being French *decreases* that chance within this data. In parallel, the positive direction and significance of the *cut3* and *cut4* coefficients, corresponding to the marginal likelihood of the indifference boundary being 9, 10, or 11 (therefore, making impatient choices until there is a super large reward for being patient), suggests that being French increased that likelihood. All 3 of these marginal coefficients seem to somewhat disagree with the overall regression coefficient. The ordinal margins have the statistical significance to support the original hypothesis. This *still* is rather weakly so, as indicated by the very, very low Pseudo R^2 value of 0.193 for model fit. There is no clear evidence to conclude our hypothesis. Further investigation is required.

The ordinal logistic regression in Column 2 of Regression Table 2 contains the model results of regressing each subject's indifference point on demographic/economic factors and risk tolerance (lottery behavior), without language. This model appears to capture the data even worse than the language-only model – not only is the pseudo-R^2 smaller, but none of the regression coefficients were also significant. No marginal coefficients reached significance either. Of the model results, being risk averse and higher in age increased the likelihood of making more impatient decisions, whereas all the remaining factors – being female, being employed, having retirement savings, being more educated, being more intelligent, and being a higher income bracket – were predictive of making more patient decisions. The positive coefficient that this model assigns to the factor of being risk averse does not align with preexisting knowledge – it was expected that risk-averse individuals be more patient, which would have been reflected with a negative coefficient. This coefficient did not achieve statistical significance, so it is not of serious importance overall. This is just a very noisy model, suggestive of underlying interactions, or that there exists a more informative variable that was not captured in this regression or general experiment. Further inquiry is still necessary.

Because the lottery task did not successfully pass significance controls, risk behavior ought to be considered a potential source of variation in indifference points. Figure 2 contains histograms of indifference points when subjects are sorted by lottery behavior, showing comparatively distinct trends. Indifference points of risk-seeking subjects (# tokens in lottery > # tokens safe) were most frequently in the low range of decision points, and risk-seeking subjects (# tokens in lottery > # tokens safe) display a more even spread. The average reward amount demonstrated to overcome delay aversion indifference in risk-averse subjects is 8.83 (± 1.20), and in risk-seeking subjects is 8.18 (± 1.10). The difference in these means is near, but not quite at traditional statistical significance (MWU p = 0.08)[25].

Figure 2

Column 3 of Regression Table 2 contains all an ordinal logistic regression for indifference boundary over the economic and demographics variables, risk aversion, language, and an interaction term of *Language * Risk Aversion*. Of the previous two regressions, this model is more explanative of the data than the non-language model, but not as strong as the language only model, reflected by the Pseudo-R^2 of 0.133. But now, several factors are

[25] Ostensibly, a p-value of 0.08 *could* be considered marginally significant in this context, if permitting a significance threshold of 0.1 instead of 0.05 due to this small sample size, as well as measures that this experiment is unable to control for, as explained previously.

significant. Age, risk aversion, and the interaction of language and risk are statistically significant, but job status, retirement savings status, and intelligence score became more influential than they were in the non-language model, indicated by the coefficients' larger magnitude in Column 3 vs Column 2. Although not statistically significant, sex is still a relatively influential factor as well – negative direction suggests that men were more likely to switch to patient decisions sooner than women in this study. Language became slightly more influential, but still did not achieve statistical significance when given more covariates. The negative direction of the coefficient still weakly disagrees with the original hypothesis, suggesting that French speakers, loosely, are more likely to have lower turning points of switching to patient decisions.

Risk aversion and its interaction with language are the stars of this model – they are both significant and informative. Before, risk aversion without any interaction was not significant (Column 2). Now, in Column 3, it is not only significant, but far stronger of a predictor. The positive coefficient indicates that subjects who saved more tokens than they gambled in the lottery game (behaving risk aversely) were more likely to make more impatient decisions, hence a higher indifference boundary. This does not align with preexisting literature, which holds that stating that patience and risk aversion are positively correlated (Dohmen et al, 2010). However, the interaction of language and risk aversion takes on a significant negative coefficient. Since both of these variables are binary, this term says that being French *and* risk averse (Lang = 1, Risk Averse = 1) specifically decreases the likelihood of making impatient choices. In this model, being risk averse, in general, increases the likelihood of making more patient choices. When *only* viewing language in this model, there is no statistical power behind the already weak coefficient. However, this interaction term targets the specific overlap of risk-averse French speakers, revealing them to behave very significantly distinct from risk-averse German speakers, ***and*** from risk-tolerant French speakers. The significance of this term also beckons addressing our original hypothesis – this interaction term contradicts the original hypothesis, as well as the coefficient on language within the same regression model. From here, we must consider the case that risk tolerance might have the ability to eclipse specific aspects of the psycholinguistic divide between sFTR and wFTR languages.

This is curious evidence that demands further investigation into psycholinguistic conceptions of time and certainty – economic considerations of risk are inseparable from considerations of time, even before factoring in temporal discounting preferences. This interaction term of language and risk aversion could be potentially examined deeper with a procedure that probes risk appetite across the strong/weak FTR duality by employing different syntactic constructions of verb aspect, definiteness of conditional/hypothetical future tense, or even framing manipulations. Also, for future reference, if focusing specifically on Neo-Whorfian influence on delay aversion, this reflects that it is imperative to control for risk attitudes to statistical significance, perhaps with pairwise data, or specific populations with a known/pre-controlled risk tolerance, or mathematically large enough sample sizes.

Another potential future direction is more robust inquiry of uncovering discounting preferences. This could easily be done by a longer battery of decision questions that encompassed greater variety of immediate reward values, different delay periods, and a richer gradient of delayed options relative to the immediate payoffs. This would create a more detailed profile of individual time preferences for subsequent analysis, and likely hedge against subjects making idiosyncratic inconsistent choices[26].

4: Task 2 Results: Tokens Investment

The second task complements the first task such that instead of forcing participants to select between pre-determined options for delayed consumption, the individual must now elect how much reward to purposely delay. Similar to the previous section, participants were told to consider the rewards with no uncertainty or ambiguity of whether they would imaginarily "occur". (See section 2 for procedural setbacks on structuring rewards) We hope to detect investment tendencies which align with the Linguistic Savings Hypothesis. Support for the LSH would be characterized by German subjects being more patient through investing more tokens than French subjects, *ceteris paribus*.

First, we assess token investment by language. Table 6 shows the average token behavior across language groups, and there is a very statistically significant difference in how each language group used their endowed tokens (MWU $p < 0.01$). French participants on average invested far more tokens than German subjects.

[26] One example case in the context of this experiment would be a subject choosing, 9 tokens later over 6 tokens now, but then selecting 6 tokens now over 10 later. Responses with this behavior were discarded. Increasingly large sample sizes and bodies of data collection are vital to improving the descriptive power and accuracy of the model, by nature of the Central Limit Theorem.

Table 6: Task 2 - Token Investing Task			
	Mean # Tokens to Receive Now	Mean # Tokens Invested	± *SD*
German (n = 20)	3.75	1.25	± *1.25*
French (n = 20)	1.30	3.70	± *1.30*
Total	2.5	2.5	± *1.77*

Column 1 of Regression Table 3 shows the model results of an ordinal logistic regression for the likelihood of investing 0, 1, 2, 3, 4, or 5 tokens regressed only over language. Here, participant language is shown to be a very significant predictor of investment, with the positive coefficient indicating that being French strongly increases the likelihood of investing a higher amount of tokens. The ordinal margins concur: cut3, cut4, and cut5 are significant and positive, which refer to the likelihood of investing 3, 4, or 5 tokens. This is not consistent with our original hypothesis, and there are more factors to consider.

Regression 3: Task 2 – Token Investment

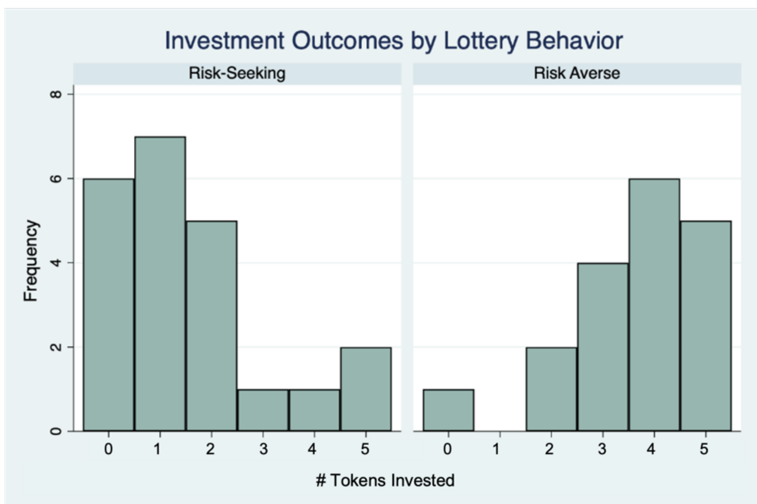
Ordinal Logistic Regression: p(Tokens Invested), [0,5] *possible*

	(1)	(2)	(3)	(4)
	H_0	*Experiment Behavior*	*Language Agnostic*	*Composite*
Language *(GER = 0)*	3.151^{***} (0.736)			4.788^{**} (1.510)
Age			0.0232 (0.0339)	-0.0115 (0.0436)
Sex *(M = 0)*			1.673^{*} (0.819)	-0.559 (1.020)
Employed (=1)			-0.473 (0.485)	-0.983 (0.570)
Retirement Savings			0.584 (0.360)	0.426 (0.383)
Intel Score			0.0186 (0.0278)	-0.0231 (0.0367)
Education			0.0541 (0.0541)	0.141^{*} (0.0646)
Relative Income Lvl			0.405 (0.288)	0.353 (0.299)
Risk Aversion (Lotto Task)		3.165^{***} (0.768)	3.157^{***} (0.796)	3.769^{*} (1.548)
Indifference Point (Task 1)		-0.629^{*} (0.267)		-0.953^{**} (0.345)
Lang * Risk Avrs				-2.501 (1.731)
/				
cut1	-0.683 (0.470)	-6.198^{**} (2.301)	2.636 (2.557)	-8.892^{*} (4.043)
cut2	0.619 (0.470)	-5.033^{*} (2.276)	4.031 (2.584)	-7.070 (3.986)
cut3	1.738^{**} (0.557)	-3.835 (2.263)	5.221^{*} (2.624)	-5.366 (3.943)
cut4	2.532^{***} (0.630)	-2.950 (2.240)	6.088^{*} (2.674)	-4.089 (3.897)
cut5	3.834^{***} (0.747)	-1.600 (2.171)	7.428^{**} (2.742)	-2.077 (3.770)
N	40	40	40	40
pseudo R^2	0.163***	0.154***	0.186***	0.320***

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Significance on pseudo R^2 refers to χ^2

Figure 3 shows the divide in investment behavior by risk attitude demonstrated through the lottery game. Among participants who acted in a risk-seeking manner, they also more frequently invested relatively fewer tokens than the participants who acted risk-aversely. The is influence of risk is important to keep in mind going forward, since this difference is statistically significant (MWU $p < 0.001$). These results also agree with existing literature that risk aversion is positively correlated with delayed gratification. We see the same pattern as Task 1 - evidence for a potential interaction of risk tolerance and delayed gratification. Regression Table 2 already demonstrated that there was a significant interaction of language and risk in Task 1 results, so we must keep this in mind going forward.

Figure 3



Columns 2 and 3 of Regression Table 3 contain ordinal logistic regressions for predicting investment using risk aversion and cofactors other than language. Column 2 predicts investment solely on the participant's previous behavior in the experiment – how they acted in the lottery game, as well as how patient they were in Task 1. In this model, both of these variables are statistically significant, and the coefficient directions agree with known literature. Here, participants who were risk averse were more likely to invest more tokens, and participants who had a higher trade-off point of switching to patient decisions in Task 1 were likely to invest less. Overall, this model serves as a check of internal consistency for participant behavior and passes.

Column 3 of Regression Table 3 predicts token investment off demographic factors and risk aversion, similarly done in Column 2 of Regression Table 2. Here, risk aversion and sex are significant – female subjects and risk averse subjects were more likely to invest a higher number of tokens. The influence of risk aversion on invested concurs with follows prior literature – being risk-averse increases propensity for patient decision making. This model is slightly better fit to the data than the language-only model, with a pseudo R^2 of 0.186.

Column 4 of Regression Table 3 shows the model of all these variables together, as well as the interaction of language and risk aversion (which was shown to be significant in the previous task). Here, language, minimum patient reward, risk aversion, and education level are significant. Sex shows a notable change in predictive power between Column 3 and 4 – it was significant when only compared to demographic variables and risk attitude, but it is not nearly as powerful in the model with more factors. Throughout these models, language has consistently been a significant factor. Also, it makes perfect sense that participants' behavior for Task 1 (token choices) is significant in predicting behavior of Task 2, since Column 2 already confirmed significant influence. No other factors seemed to overpower or confound this term, since it maintained significance and direction (negative) when other factors were introduced. The interaction term is not quite at statistical significance, but the magnitude of the coefficient still presents influence. As seen previously, being French increased the likelihood of investing more tokens, and being risk averse increased the likelihood of investing more tokens. However, the negative sign on this interaction term indicates that risk-averse *AND* French subjects invested fewer tokens. If this term were significant, it would be evidence in support of the original LSH. However, we must again consider the case that risk tolerance might have the ability to eclipse the strength of Neo-Whorfian influence on delayed gratification between sFTR and wFTR languages.

It is clear that perceptions of risk are inseparable from considerations of time, even before factoring in temporal discounting preferences. One large downfall of this dataset is that risk is not moot with respect to language. In future research, risk attitudes must be very carefully assessed. Another potential future direction is more robust inquiry of uncovering discounting preferences. This could easily be done by a longer battery of decision questions that encompassed greater variety of immediate reward values, different delay periods, and a richer gradient of delayed options relative to the immediate payoffs. This would create a more detailed profile of individual time preferences for subsequent analysis.

5: Conclusion

Chen (2013) and Sutter (2015, 2018) use Neo-Whorfianism to explain the effects of strong vs weak FTR on time preferences and delayed gratification, which they summarize as the Linguistic Savings Hypothesis. In this experiment and Sutter (2015, 2018), German speakers were hypothesized to behave more patiently, through making more patient decisions than French speakers in Task 1, and by investing more tokens in Task 2. Sutter (2015, 2018) was able to successfully identify effects of the LSH in their study of school children, but I was not able to replicate the same results with similar procedures.

Of my results, language was a significant predictive factor for decision making, but in a manner contrary to our initial hypothesis. That is, my study revealed French participants (sFTR language) to make more future-oriented, patient economic decisions than German participants (wFTR language). I was not able to isolate language influencing choices according to the LSH, but controlling for risk appetite was shown to be an exceedingly important factor to control for before investigating discounting tendencies. Risk aversion and delay aversion are so deeply intertwined from a cognitive *and* economic standpoint, so specificity is crucial for uncovering very nuanced forces, such as Neo-Whorfian cognitive biases. Interacting language and risk seemed to yield some statistical power in my regression, namely Regression 2, but far more data must be collected confirm or deny this idea with confidence.

Several places in my experiment require duplication and larger-scale implementation. My small sample size of n = 40 clearly presents worry for how noisy my data is, so consider this a dry run proof-of-concept. Another potential future direction is more robust inquiry of uncovering discounting preferences. This could easily be done by a longer battery of decision questions that encompassed greater variety of immediate reward values, different delay periods, and a richer gradient of delayed options relative to the immediate payoffs. This would create a more detailed profile of individual time preferences for subsequent analysis, and likely hedge against subjects making idiosyncratic inconsistent choices. This experiment is much better suited to the environment of classical behavioral experiments - sitting down in-person with the experimenters to understand the rules thoroughly, act with realistic judgement, in a controlled environment, and keep the participant on task. That would also allow for finer discrimination of consequences for rewards, actually following up on delayed rewards once the waiting period is completed and overcoming the “house money” effect.

All in all, this report does not prove, nor disprove, the strength of the LSH. It is an extremely sensitive pattern to isolate, and this report demonstrates that it can easily be overshadowed by idiosyncrasies, which are only exaggerated in such a small sample size.

Sources and Bibliography

Anderhub, V., & Güth, W. (2001). On the Interaction of Risk and Time Preferences: An Experimental Study. *German Economic Review*, *2*(3), 239–253. https://doi.org/10.1111/1468-0475.00036

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious Mixed Models*. https://doi.org/10.48550/ARXIV.1506.04967

Boroditsky, L., Schmidt, L., & Phillips, W. (2003). Sex, Syntax, and Semantics. In S. Goldin-Meadow & D. Getner (Eds.), *Language in Mind* (pp. 61–79). The MIT Press. https://doi.org/10.7551/mitpress/4117.003.0010

Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *The Journal of Abnormal and Social Psychology*, *49*(3), 454–462. https://doi.org/10.1037/h0057814

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22. https://doi.org/10.1037/h0046743

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. https://doi.org/10.1257/aer.103.2.690

Cobanoglu, C., Cavusoglu, M., & Turktarhan, G. (2021). A beginner's guide and best practices for using crowdsourcing platforms for survey research: The Case of Amazon Mechanical Turk (MTurk). *Journal of Global Business Insights*, *6*(1), 92–97. https://doi.org/10.5038/2640-6489.6.1.1177

Cohen, J., Ericson, K. M., Laibson, D., & White, J. M. (2020). Measuring Time Preferences. *Journal of Economic Literature*, *58*(2), 299–347. https://doi.org/10.1257/jel.20191074

Commons, M. L., Mazur, J. E., Nevin, J. A., & Rachlin, H. (2013). *The Effect of Delay and of Intervening Events on Reinforcement Value: Quantitative Analyses of Behavior, Volume V*. Psychology Press.

Copley, B. (2009). *The Semantics of the Future* (0 ed.). Routledge. https://doi.org/10.4324/9780203880258

Dahl, Ö. (Ed.). (2000). *Tense and Aspect in the Languages of Europe*. De Gruyter Mouton. https://doi.org/10.1515/9783110197099

Dawson, H., & Phelan, M. (Eds.). (2016). *Language files: Materials for an introduction to language and linguistics* (Twelfth edition). The Ohio State University Press.

Díaz-Serrano, L., & O'Neill, D. (2004). The Relationship between Unemployment and Risk-Aversion. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.571044

Difallah, D., Filatova, E., & Ipeirotis, P. (2018). Demographics and Dynamics of Mechanical Turk Workers. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 135–143. https://doi.org/10.1145/3159652.3159661

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are Risk Aversion and Impatience Related to Cognitive Ability? *American Economic Review*, *100*(3), 1238–1260. https://doi.org/10.1257/aer.100.3.1238

Epper, T. F., Fehr-Duda, H., & Bruhin, A. (2009). Uncertainty Breeds Decreasing Impatience: The Role of Risk Preferences in Time Discounting. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1416007

European Science Foundation. (n.d.). *Eurotyp Questionnaires* (Typological Tools for Field Linguistics). Max Planck Institute for Evolutionary Anthropology. Retrieved September 30, 2022, from https://www.eva.mpg.de/lingua/tools-at-lingboard/questionnaire/eurotyp_description.php

Fausey, C. M., & Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin & Review*, *18*(1), 150–157. https://doi.org/10.3758/s13423-010-0021-5

Frederick, S., Loewenstein, G., & O'donoghue, T. (2002). Time Discounting and Time Preference: A Critical Review. *Journal of Economic Literature*, *40*(2), 351–401. https://doi.org/10.1257/jel.40.2.351

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences*, *103*(2), 489–494. https://doi.org/10.1073/pnas.0509868103

Grable, J. E. (2000). Financial Risk Tolerance and Additional Factors That Affect Risk Taking in Everyday Money Matters. *Journal of Business and Psychology*, *14*(4), 625–630. JSTOR.

Haviland, J. B. (1998). Guugu Yimithirr Cardinal Directions. *Ethos*, *26*(1), 25–47.

Heydasch, T., Haubrich, J., & Renner, K.-H. (2020). Translation - The Short Version of the Hagen Matrices Test (HMT-S): A 6-Item Induction Intelligence Test. *Methods*, *data*, 23 Pages. https://doi.org/10.12758/MDA.2013.021

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, *10*(12), 1625–1633. https://doi.org/10.1038/nn2007

Kay, P., & Kempton, W. (1984). What Is the Sapir-Whorf Hypothesis? *American Anthropologist*, *86*(1), 65–79. JSTOR.

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. https://doi.org/10.1177/0956797611432178

Kidd, C., Palmeri, H., & Aslin, R. N. (2013). Rational snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability. *Cognition*, *126*(1), 109–114. https://doi.org/10.1016/j.cognition.2012.08.004

Mischel, W., & Ebbesen, E. B. (1970). Attention in delay of gratification. *Journal of Personality and Social Psychology*, *16*(2), 329–337. https://doi.org/10.1037/h0029815

Olson, E. A., Hooper, C. J., Collins, P., & Luciana, M. (2007). Adolescents' performance on delay and probability discounting tasks: Contributions of age, intelligence, executive functioning, and self-reported externalizing behavior. *Personality and Individual Differences*, *43*(7), 1886–1897. https://doi.org/10.1016/j.paid.2007.06.016

Pavlick, E., Post, M., Irvine, A., Kachaev, D., & Callison-Burch, C. (2014). The Language Demographics of Amazon Mechanical Turk. *Transactions of the Association for Computational Linguistics*, *2*, 79–92. https://doi.org/10.1162/tacl_a_00167

Raven, J. C., Court, J. H., & Raven, J. E. (1978). *Manual for Raven's progressive matrices and vocabulary scales: Vol. Section 2. The Coloured progressive matrices* (Rev). H.K. Lewis.

Samuelson, P. A. (1937). A Note on Measurement of Utility. *The Review of Economic Studies*, *4*(2), 155. https://doi.org/10.2307/2967612

Soman, D., Ainslie, G., Frederick, S., Li, X., Lynch, J., Moreau, P., Mitchell, A., Read, D., Sawyer, A., Trope, Y., Wertenbroch, K., & Zauberman, G. (2005). The Psychology of Intertemporal Discounting: Why are Distant Events Valued Differently from Proximal Ones? *Marketing Letters*, *16*(3–4), 347–360. https://doi.org/10.1007/s11002-005-5897-x

Steinberg, L., Graham, S., O'Brien, L., Woolard, J., Cauffman, E., & Banich, M. (2009). Age Differences in Future Orientation and Delay Discounting. *Child Development*, *80*(1), 28–44. https://doi.org/10.1111/j.1467-8624.2008.01244.x

Sutter, M., Angerer, S., Glätzle-Rützle, D., & Lergetporer, P. (2015). The Effect of Language on Economic Behavior: Experimental Evidence from Children's Intertemporal Choices. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2672169

Sutter, M., Angerer, S., Glätzle-Rützler, D., & Lergetporer, P. (2018). Language group differences in time preferences: Evidence from primary school children in a bilingual city. *European Economic Review*, *106*, 21–34. https://doi.org/10.1016/j.euroecorev.2018.04.003

Tang, J., Yang, J., Ye, W., & Khan, S. A. (2021). Now is the time: The effects of linguistic time reference and national time orientation on innovative new ventures. *Journal of Business Venturing*, *36*(5), 106142. https://doi.org/10.1016/j.jbusvent.2021.106142

Thaler, R. H., & Shefrin, H. M. (1981). An Economic Theory of Self-Control. *Journal of Political Economy*, *89*(2), 392–406. https://doi.org/10.1086/260971

Thoma, D., & Tytus, A. E. (2018). How Cross-Linguistic Differences in the Grammaticalization of Future Time Reference Influence Intertemporal Choices. *Cognitive Science*, *42*(3), 974–1000. https://doi.org/10.1111/cogs.12525

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. https://doi.org/10.1177/0956797618761661

Whorf, B. L., Carroll, J. B., Levinson, S. C., & Lee, P. (2012). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf* (2nd ed). The MIT Press.

Appendix

Two hypothesized explanations of how the LSH functions to distort the perception of present and future. Figure A explains that wFTR languages express less certainty over when an action might occur, which could lead to larger variation in how strongly the value of that reward is discounted. Figure B considers the idea that wFTR speakers consider the future to be "less far away from the present" than sFTR speakers. sFTR speakers are thought to consider the future as something distinctly separate from the present, and therefore it might feel less imminent, and maybe more certain.

Fig A

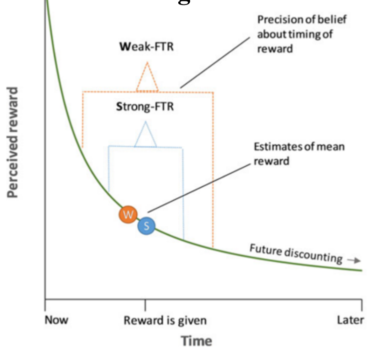
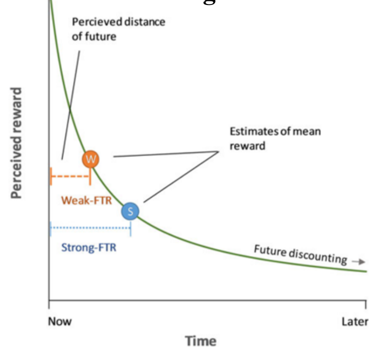
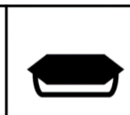
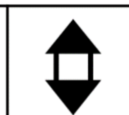


Fig B



Code symbols and digits used in the intelligence task. I made the symbols myself using random shapes in MS Powerpoint, and generated the random strings of symbols for participants to match using random sampling in R.

							
0	1	2	3	4	5	6	7

Personal Values questionnaires – this data is incredibly noisy. I had been hoping to capture an influence on person-to-person attitudes on money, but this questionnaire was either not specific enough, polarizing enough, or not distributed to a large enough participant body to yield results that would be informative in regression.

Personal and Societal Values – Mean Rankings (± SD)
1 = Less Important, 2 = Moderately Important, 3 = Very Important

	Patience		Saving Money		Taking Risks	
	Personal	Societal	Personal*	Societal	Personal**	Societal*
German (n = 20)	2.45 (± 0.69)	2.35 (± 0.59)	2.30 (± 0.65)	2.20 (± 0.70)	1.80 (± 0.77)	1.70 (± 0.66)
French (n = 20)	2.65 (± 0.49)	2.50 (± 0.61)	2.65 (± 0.49)	2.15 (± 0.81)	2.35 (± 0.67)	2.21 (± 0.64)
Total (n = 40)	2.55 (± 0.60)	2.425 (± 0.59)	2.475 (± 0.59)	2.175 (± 0.75)	2.075 (± 0.76)	1.9 (± 0.67)

MWU test by lang $p < 0.1$*; ** $p < 0.05$*