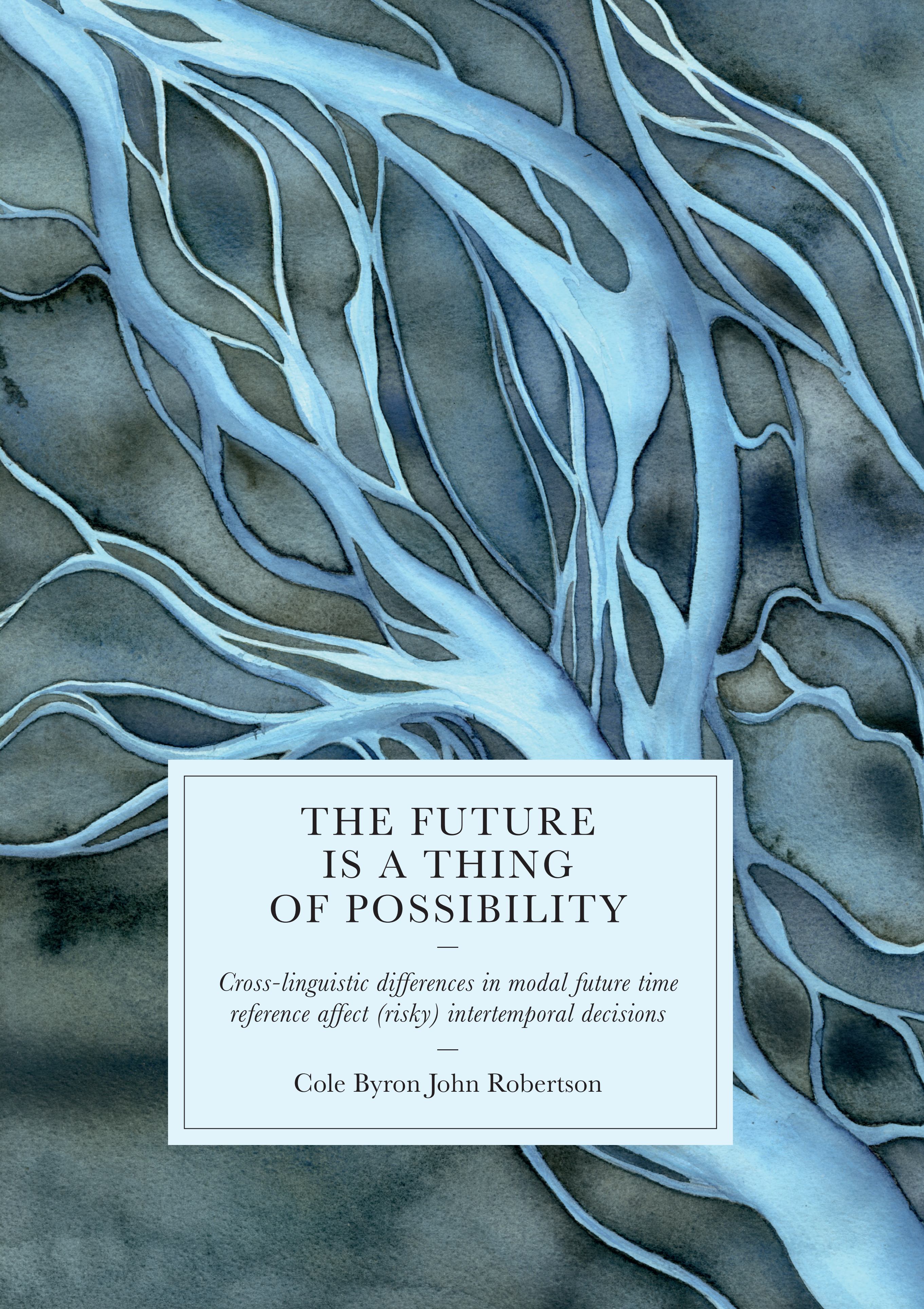


THE FUTURE IS A THING OF POSSIBILITY

—

Cross-linguistic differences in modal future time reference affect (risky) intertemporal decisions

—

Cole Byron John Robertson

The future is a thing of possibility

Cross-linguistic differences in modal future time reference affect (risky) intertemporal decisions

Cole Byron John Robertson

The research reported in this thesis was carried out at the Center for Language Studies at Radboud University, the Max Planck Institute for Psycholinguistics, the Experimental Psychology Department at the University of Oxford, and at the College of Business Arts and Social Sciences at Brunel University. This thesis was supported by a Natural Sciences and Engineering Council of Canada doctoral fellowship held by the author (PGSD2-517110-2018), by the EU's Horizon 2020 From Individual Behaviour to the Socio-Technical Man project (FET Open RIA 662725), by a Leverhulme Early Career Fellowship held by Dr. Seán G. Roberts (ECF-2016–435), and by the Wellcome Trust Necessary Fictions project (205493/Z/19/A).

Cover art
Genevieve Robertson

Jacket design
Stuart Youngs

An electronic version of this dissertation is available at `cbjrobertson.github.io/thesis.pdf`

A note to readers

All figure, table, section, chapter, appendix, citation, equation, and linguistic example references in this thesis are hyperlinks. Clicking them will take readers to the referenced material. In Apple Preview, you can go back to where you came from by pressing CMD + [. On Adobe Reader and a number of other popular .pdf viewers, it is CMD + ←. This may depend on your operating system. The above works on OSX; Windows users may try ALT in place of CMD. I hope this helps readers navigate with ease.

The future is a thing of possibility

Cross-linguistic differences in modal future time reference affect (risky) intertemporal decisions

Proefschrift
ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op
dinsdag 23 november 2021
om 16:30 uur precies

door

Cole Byron John Robertson
geboren op 30 oktober 1984
te Vancouver, Canada

Hollywell Press Limited, Oxford, UK

Promotoren:
Prof. dr. Asifa Majid
Prof. dr. Robin Dunbar (University of Oxford, Verenigd Koninkrijk)

Copromotor:
Dr. Seán G. Roberts (Cardiff University, Verenigd Koninkrijk)

Manuscriptcommissie:
Prof. dr. Jos Hornikx
Prof. dr. Falk Huettig
Dr. Laura Speed
Prof. dr. Phillip Wolff (Emory University, Verenigde Staten)
Dr. Monique Flecken (Universiteit van Amsterdam)

I thought I still had time. Why do people say that anyway? To "have time."
How can you have time when it clearly has you?

–Jantje Friese & Martin Behnke
Translated by Nathan Fritz

Contents

Chapter 1

General Introduction

The division of time into past, present, and future seems to be an irreducible fact of human experience: We remember the past, we perceive the present, and we hope, dream, imagine, plan, guess, reckon, gauge, envision, augur, intend, schedule, desire, and predict the future. Why? The answer is both trivial and mysterious: We have not yet experienced it. The future has not happened. *It never comes.* Why this should be the case, or whether it is necessarily the case in some counterfactual cosmic sense, is unclear, perhaps unknowable. It *is* the case. Consequently, we experience the future as a thing of possibility: While only one thing *will* happen, anything *could*.

Philosophers from Aristotle to Augustine have attempted to understand whether temporal perception—usually assumed to be universally experienced by all humans—represents a veridical model of reality (Le Poidevin, 2019). But is temporal perception universal? I[1] am comfortable in asserting that all humans have memories, experience an ongoing stream of perceptions, and do not know precisely what they will perceive next. The reason for this epistemic imbalance is simple in philosophical terms. Accepting that all causes precede their effects, and our perceptions are caused by real events, it follows that our perceptions (effects) cannot precede their causes (events) (Le Poidevin, 2019). We cannot see the future.

But do all cultures construe these universal experiences in the same way? Is human experience *tensed*? We can ask whether people from all cultures refer to our memories as "past", our perceptions as "present", and the perceptions which have yet to arise as "future". One way to answer this question would be to investigate whether all languages grammatically divide time in the same way. They do not. While most if not all languages afford speakers the ability to refer to distant times (Hockett, 1963), the grammatical marking of time differs. Some languages have no future tense, simply using the same unmarked forms to refer to the present and the future (Dahl, 2000a). Others collapse the present with the past and

[1] I use the first-person singular pronoun, *I*, to invoke authorial perspective in this dissertation's introduction (Chapter 1) and general discussion (Chapter 6). The plural, *we*, is used in Chapters 2–5, which present empirical work done collaboratively with my co-authors and supervisors. See Appendix A for author contributions.

only distinctly mark the future (Malotki, 1983). Yet others mark near future events with one "future tense", and far future events with another (Cable, 2013). Complexities abound. Yélî Dnye, a Papuan language spoken on Rossel Island 450 km offshore of Papua New Guinea grammatically distinguishes six tenses, which, relative to the time of speech, shift the temporal frame of reference between: earlier today; yesterday; the day before yesterday, or further in the past; later today; tomorrow; or the day after tomorrow, or later in the future (Levinson & Majid, 2013). Does this tell us anything about the way speakers perceive the world? Is the human experience of time more or less universal, or is it as varied as linguistic tense? One way to investigate this question is to use language: Do tense systems themselves influence the way speakers conceive of temporal experience? The linguistic relativity hypothesis predicts that languages influence the way speakers perceive the world, and therefore expects that cross-linguistic differences in linguistic tense impact speakers' representation of time, and their pursuant beliefs and decisions (Whorf, 1956).

There are multiple ways in which language is predicted to impact thought. John Lucy (1997) provides a useful ternary framework for navigating such questions (see also Gumperz & Levinson, 1996). In *functional* relativity, learning domain-specific vocabularies aids speakers in navigating privileged knowledge domains within a given language (Lucy, 1997). Anyone who has listened to two lawyers speak may understand what I mean. This vein of work is characterised by difficulties unpacking language-thought correlations, since people self select to join social (and therefore speech) communities. As a result, work on this question involves the isolation of exposure effects which can be quantitatively evaluated, such as eduction (Vygotsky, 1987). For instance, wine experts were found to have more vivid imagery for the smell and taste of wine and this was due specifically to training (Croijmans, Speed, Arshamian, & Majid, 2020). In *semiotic* relativity, the human linguistic endowment—the ability to speak at all—affects human cognition (Lucy, 1997). Investigations into this question involve comparative work with prelinguistic infants, deaf individuals with (unfortunately) very little language exposure, or non-human animals (for review, see Majid, 2018). Finally, in *structural* relativity, cross-linguistic differences affect non-linguistic cognition between speakers of different languages. The relevant question is: What is it like to speak one language compared with another? Although Chapter 5, touches on *functional* relativity, this thesis principally focusses on *structural* relativity. I will follow widely-used nomenclature conventions by using the term "linguistic relativity" to refer to *structural relativity* (Bloom & Keil, 2001; Lucy, 1997). While evidence has been steadily accruing in support of the linguistic relativity hypothesis (see Everett, 2013; Evans & Levinson, 2009; Boroditsky, 2006; Casasanto, 2016; Bohnemeyer, 2020; Majid, 2018; Wolff & Holmes, 2011), questions remain as to the scope of the effects (Montemayor, 2019; Firestone & Scholl, 2015), and critics have tended to dismiss results as trivial or negligible (Bloom & Keil, 2001; McWhorter, 2014), or have argued that languages merely reflect deeply-held cultural beliefs shared by communities of speakers (McWhorter, 2014). I detail these issues in Section 1.2, such that my approach may be understood in context.

In this thesis, I address some of these concerns by zooming in on one particular question about linguistic relativity and time: Do differences in the way languages oblige speakers to talk about the future influence how they make decisions *now* which affect them *later*? In particular, I focus on untangling *time* from *probability* in three related Western Germanic languages: English, Dutch, and German. The future cannot be known at the time of speech, and therefore must be guessed at, gauged, and predicted: Probabilities must be attached, risks assessed, rewards weighted by the chance of success, and decisions made in the face of uncertainty. Language about the future therefore tends to involve the expression of probability, desire, intention, obligation, hope, and the *if... then...* logic of conditionality. The somewhat capacious term for the linguistic expression of such notions is *modality*. *Tense*, on the other hand, encodes temporal location (past, present, future) relative to the time of speech (Lyons, 1968), while "aspect" encodes the internal temporal structure of an event (i.e. durative vs. punctual) (Broekhuis & Verkuyl, 2014). "Future Time Reference" (FTR) is a broader term encompassing any linguistic statement which refers to future events, whether modal, tense, aspectual, or unmarked. "Intertemporal decision making" refers to decisions with delayed outcomes. Think of the trade-offs between present efforts and future results involved in purchasing a house, starting a company, pursuing educational goals, planting crops, migrating to a new country, or fighting climate change. In this thesis, I explore whether cross-linguistic differences in FTR grammar affect the way speakers make such decisions—their "time preferences".

In this introduction, I provide a brief history of work on linguistic relativity, time, and modality. This initially involves a discussion of how the Hopi language represents future time: Are modal or temporal notions involved? I then outline several common critical perspectives on linguistic relativity research, such that my work can be evaluated in context. Following this, I discuss contemporary research on the effects of tense on intertemporal decision making, critical background to this thesis. I then outline my empirical approach. Linguistic relativity research projects comprise, at a minimum, three related steps: (a) Relevant cross-linguistic differences need to be established; (b) correlated differences in non-linguistic cognition need to be identified; and (c) causal evidence linking (a) to (b) needs to be developed. I use this framework to structure my approach. Finally, I motivate the key questions answered in this dissertation.

1.1 A history of linguistic relativity and future time: Modality, FTR, and the Hopi time controversy

For most of the mid 20th century, the question of whether cross-linguistic differences impact speakers' beliefs about time involved the evaluation of controversial claims made by the linguistic anthropologist Benjamin Lee Whorf. At issue was the relationship between the language and beliefs of the Hopi, an indigenous North American people from the high desert *mesas* of Arizona, Colorado, and New Mexico. At least for the purposes of this review, this is where we will begin (for a detailed work on the historical development of linguistic relativity, see Leavitt, 2011). Whorf was a chemical engineer by training, and a linguistic autodidact who moonlighted—and sometimes lectured—in the Yale graduate seminars of Edward Sapir—himself a student of Franz Boas. In its modern conception, the linguistic relativity hypothesis is so associated with Whorf's writings that it is sometimes called the "Whorfian" or "Sapir-Whorf" hypothesis.

Whorf cut his teeth as a linguist in the 1930s, working with a Hopi man who lived in Manhattan, and it is from interviews with this informant that he developed his thinking on the relationship between the Hopi language and Hopi beliefs about time. Whorf claimed the Hopi did not share modern Western conceptions of time (Whorf, 1956, pp. 57–8). However, his claim about Western concepts, tended to be misinterpreted as the more general claim that the Hopi had no concept of time at all (Leavitt, 2011, pp. 179–87). Critics pointed to examples of Hopi temporal language as clear evidence that Whorf was confused (e.g. Malotki, 1983; Pinker, 1994).

While the debate was often framed in terms of whether the Hopi could conceive of time, much disagreement centres around the complex semantics of Hopi FTR; in particular, on whether the Hopi suffix, *-ni*, is a temporal marker, encoding futurity (Malotki, 1983, p. 624), or a modal marker, encoding notions of desire, expectation, and hope (Whorf, 1956, pp. 60–1, 113). That the semantics of the suffix should be ambiguous enough to permit extended disagreement is actually stereotypical of future "tenses", which often encode notions not just of futurity, but also of aspect, modality, volition, desire, prediction, ability, obligation, motion, and change. Dahl (2000a) notes that a central issue to future tense typology is, "the distribution of labour between temporal, modal, and aspectual elements in their meanings and whether to subsume them under the traditional categories of tense, mood/modality, or aspect" (p. 313). In this section, I argue that much debate about Hopi "time" is best understood as arising from this characteristic ambiguity of future tenses (see also Lyons, 1968; Fries, 1956).

Whorf's ideas on Hopi time are most clearly articulated in his essay "An American Indian model of the universe" (Whorf, 1956, pp. 57–68). Many of his claims are difficult to evaluate, as they are not accompanied by much linguistic evidence. They rather form an ethnographic treatise on what Whorf termed Hopi "metaphysics" (Whorf, 1956, p. 58). Only in light of his ideas about Hopi beliefs do Whorf's claims about Hopi grammar make sense. According to Whorf:

> The Hopi language is seen to contain no words, grammatical forms, constructions, or expressions that refer directly to what we call "time," or to past, present, or future, or to enduring or lasting, or to motion as kinematic rather than dynamic (i.e. as a continuous translation in space and time rather than as an exhibition of dynamic effort in a certain process).... Hence, the Hopi language contains no reference to "time" either explicit or implicit. (Whorf, 1956, pp. 57–8)

This statement has been misconstrued almost as much as it has been ridiculed, in both cases most thoroughly by Ekkehart Malotki in *Hopi Time* (see Levinson, 2012). Malotki's approach is well summarised by two quotes which appear on the same page in the opening material of *Hopi Time*. The first is an excerpt from Whorf's passage above. The second is a Hopi sentence glossed as, "then indeed, the following day, quite early in the morning at the hour when people pray to the sun, around that time then he woke up the girl again" (Malotki, 1983, p. xii). The contrast is clearly intended to indicate that Whorf should have stuck to his day job. The rest of *Hopi Time* extends this basic approach. Conceived as a protracted refutation of Whorf's claims, the 677-page monograph is evidenced throughout with linguistic data from several years of fieldwork. For example, it details Hopi spatio-temporal metaphors, the Hopi horizon-based time system, the Hopi ceremonial calendar, Hopi time-keeping devices, Hopi temporal adverbials, and the Hopi "tense" system (Malotki, 1983). It is enough evidence for Steven Pinker to (gleefully) report Malotki's winking juxtaposition and conclude that Whorf's "limited, badly analysed, sample of Hopi speech and his long-time leanings towards mysticism" must have contributed to the formation of his "outlandish claims" (Pinker, 1994, p. 63). Evaluating Whorf's ideas against this body of work is beyond the scope of this review. Suffice it to say *Hopi Time* is a meticulous testament to the fact that Hopi has a nuanced system for talking about time.

But does this refute Whorf's claims? Possibly not. In the sentence immediately following the excerpt above, Whorf writes: "At the same time, the Hopi language is capable of accounting for and describing correctly, in a pragmatic or operational sense, all observable phenomena in the universe" (Whorf, 1956, p. 58). This seems impossible to reconcile with the assertion that the Hopi can not talk about capital-T universal Time.[2] In fact, close reading indicates Whorf is not talking about Time at all. Leavitt (2011, p. 180), and also Hinton (1988), point out that Whorf doubly qualifies his statements with both scare quotes and the cryptic "we call". What, then, does Whorf mean by *what we call "time"*? In fact, he defines

[2] I will use the present tense to refer to Whorf's claims about Hopi, but the reader should be aware that the effect of language drift and the influence of colonial language and culture (English, American) may mean that the debate only applies in a historical sense to the Hopi people and language.

it quite explicitly. He posits that a standard Newtonian model undergirds modern European conceptions of space and time:

> The metaphysics underlying our own language, thinking, and modern culture... imposes upon the universe two grand COSMIC FORMS, space and time; static three-dimensional space, and kinetic one-dimensional uniformly and perpetually flowing time.... The flowing realm of time is, in turn, the subject of a three-fold division: past, present, and future. (Whorf, 1956, p. 59, emphasis original)

This is more specific than any sweeping reference to Time in a general sense.

How does the Hopi conceptualisation of time differ from this? Firstly, Whorf claims the Hopi language does not have a ternary division of time into past, present, and future, but only grammatically disambiguates future from non-future (i.e. past + present). This seems fairly uncontroversial: Malotki argues for a future vs. non-future system in Hopi (1983, p. 626), as does Gipper (1977). However, Malotki is satisfied with concluding that high-frequency use of the Hopi suffix *-ni* in FTR contexts indicates its "referential force is futurity" (Malotki, 1983, p. 624). Whorf sees it as indexing a broader division of experience in which the contrast between future vs. non-future is subsumed:

> The Hopi metaphysics.... imposes upon the universe two grand cosmic forms, which as a first approximation in terminology we may call MANIFESTED and MANIFESTING (or, UNMANIFEST) or, again, OBJECTIVE and SUBJECTIVE. The objective or manifested comprises all that is or has been accessible to the senses, the historical physical universe, in fact, with no attempt to distinguish between present and past, but excluding everything that we call future. The subjective or manifesting comprises all that we call future, BUT NOT MERELY THIS; it includes equally and indistinguishably all.... mentality, intellection, and emotion. (Whorf, 1956, pp. 59–60, emphasis original)

According to Whorf, it is these grand categories which Hopi grammatical "tense" indexes. The zero form indicating the manifested, a suffix—*-ŋʷï*—marks gnomic (timeless, general) statements, and *-ni* marking the unmanifest, or subjective. He refers to these grammatical distinctions not as tenses, but assertions:

> Hopi verbs have three assertions: reportive (zero form), expective (suffix *-ni*), nomic (suffix *-ŋʷï*). These translate, more or less, the English tenses. But they do not refer to time or duration. They distinguish three different kinds of information. Assertion, in other words, is a classification that refers the statement to one of three distinct realms of validity. (Whorf, 1956, p. 113)

Thus, according to Whorf, *-ni* does not indicate simply future time, but rather the grand cosmic forms he saw as essential to understanding the Hopi language and culture.

His evidence for this is rather light. He analyses the semantics of *-ni* through its use across a variety of contexts. He concludes it does not index purely temporal notions, but is used in specific contexts which imply aspectual and modal meanings (typical for future tenses Lyons, 1968; Giannakidou & Mari, 2018). For instance, *-ni* marks the initiation of events the duration of which are unknown: *The man begins to chop wood* would be translated with *-ni* only if the speaker does not know whether the man intends to continue for more than one chop. On the other hand, when the speaker knows the man intends to keep at the chopping for some time, *-va* and a reduplication are used (Whorf, 1956, pp. 109–10). This suggests that *-ni* has aspectual as well as temporal semantics. More importantly, Whorf argues that its primary meaning is modal in character. While he does not use the term, it is clear he thinks *-ni* encodes *bouletic* modality, which has to do with the expression of hopes, wishes, and desires (see Section 1.3.1). Whorf writes, "the 'future' tense asserts that expectancy of the event is present, [and] that the subject's will to the event, if it is a voluntary act, is present" (Whorf, 1956, p. 109). That Whorf sees the semantics of the Hopi "assertion" system as fundamentally modal is further evidenced in his discussion of modality in Hopi:

> I would say that the modalities of Hopi are moduli of moduli, that is, they are methods of further modifying and amplifying the three-assertion system that distinguishes three basic realms of validity, so that in effect many more than three realms and subrealms of validity are distinguished. (Whorf, 1956, p. 118)

In fact, the modal semantics of the Hopi future "tense" marker are not really controversial (see Levinson, 2012). Malotki himself admits that in addition to temporal marking, *-ni*, "also takes on a number of secondary, atemporal functions which essentially belong to the modal category" (Malotki, 1983, p. 624). However, whereas Malotki views these as "secondary", Whorf sees them as indicating that *-ni* is something other than a future tense *sensu stricto*. Malotki's dismissal of these modal complexities draws criticism from the likes of Bernard Comrie, who notes that he finds Malotki's claim that Hopi has a "tense" system based on future vs. non-future opposition to be questionable: "Given the wide range of modal uses of the so-called future, it is at least plausible that this is a modal rather than a temporal distinction, with the result that Hopi would have no tense distinctions" (Comrie, 1984, p. 133).

The difference may be a fundamental matter of approach: Typically, for an ethnolinguist in the orbit of Frans Boas, Whorf attempts to understand Hopi temporal language in the context of their system of cultural beliefs (see Leavitt, 2011). Malotki may be guilty of projecting his own Western conceptions of time onto Hopi semantics. According to Hinton, "Malotki never wrestles with the problem of defining the basis of the concept of 'time', either universally or language-specifically. Instead, he makes the error of attributing temporality to any Hopi sentence that translates in English with a temporal term" (1988, p. 363). As a consequence, he fails to "deeply grapple" with Whorf's central concern: understanding "what the Hopi actually *meant* by the word 'time'" (Hinton, 1988, p. 364).

What, then, does Whorf understand the Hopi to mean by "time"? He argues that they construe desire as the causal source of the ongoing manifestation of existence. This is conceived and realised in the internal desires of all living things—not least the Hopi (Whorf, 1956, pp. 59–62). Thus, Whorf's "subjective realm" encompasses not just the future but all the hopes and desires out of which the future emerges. He writes: "If we were to approximate our metaphysical terminology more closely to Hopian terms, we should probably speak of the subjective realm as the realm of HOPE or HOPING" (Whorf, 1956, p. 61, emphasis original). This Hopi notion is best expressed, he writes, by the Hopi verb *tunátya*, which contains elements of "thought" and "desire", but also of "cause" (Whorf, 1956, p. 61). This, according to Whorf, is the Hopi term for the subjective:

> It refers to the state of the subjective, unmanifest, vital and causal aspects of the Cosmos, and the fermenting activity toward fruition and manifestation with which it seethes—an action of HOPING; i.e. mental activity, which is forever pressing upon and into the manifested realm.... The Hopi see this burgeoning activity in the growing of plants, the forming of clouds and their condensation into rain, the careful planning out of the communal activities of agriculture and architecture, and in all human hoping, wishing striving, and taking thought; and it is most especially concentrated in prayer, the constant hopeful praying of the Hopi community, assisted by the esoteric rituals in the underground kivas—prayer which conducts the pressure of the collective Hopi thought and will out of the subjective and into the objective. (Whorf, 1956, pp. 61–2)

This is a fascinating picture of the progress of time: Hope, prayer, and desire are construed as the causal source of the ongoing emergence of existence. Surely, this is a different picture from a Newtonian model of time and Western conceptions of causality? It makes a sharp contrast against a universe in which the future emerges from random but intelligible physical processes, and must be predicted, quantified, and assessed, not conditionally upon hope, but upon knowledge of the present and best approximations of the deterministic laws and probabilistic tendencies which govern causal processes embedded in one-dimensional time.

It certainly seems so, but Whorf's claims are difficult to assess, as they are often made without extensive linguistic evidence. They may be impossible to verify now, as the influence of colonial language and culture on the present day Hopi must be taken seriously (Levinson, 2012). Moreover, Whorf's argument may be circular. Language-specific meanings are argued for in the context of ethnography, while ethnographic conclusions are justified with reference to analysis of the Hopi language—a problem with much early work on linguistic relativity (Bloom & Keil, 2001; Everett, 2013, p. 23). Additionally, the cultural contrast may be overstated, made to seem more intriguing by the exotic nature of the Hopi. Is the Hopi notion of bouletic causality very different from the Christian practice of prayer, or the self-help truism that our futures emerge from the power of our intentions? What might a Hopi person from the 1930s have to say about the phrase *the power of positive thinking*? Who knows. In the end, my principal intention is to draw attention to the entanglement of time and modality which arises in language about the future, such that my exploration of cross-linguistic cognitive effects of FTR grammar are seen in this historical context. I also want to point out that, according to Whorf, the principal modal "base"—which is to say the domain which is modally quantified (see Section 1.3.1)—in Hopi FTR is bouletic, having to do with hopes and desires, a point I will come back to in the general discussion (Section 6.5).

1.2 Critical perspectives on contemporary linguistic relativity research

In the 1990s, there was a surge of interest in linguistic relativity stemming from seminal work by Lucy (1992) and a volume edited by Steven Levinson and John Gumperz (1996). Whereas Whorf made sweeping attempts to correlate linguistic features with ethnographic investigations of culture, this work focussed on tightly-controlled experimentation which sought to understand whether—and to what extent—specific cross-linguistic differences impact non-linguistic cognition (Leavitt, 2011; Lucy, 1992, pp. 189–211).[3] Rather than gestalt attempts to describe "worldview", cognition was operationalised in terms of testable and specific cognitive processes: memory, psychophysical perception, judgement, reasoning, and categorisation (Lucy, 1992; Gumperz & Levinson, 1996).

In terms of research into cross-linguistic effects on temporal cognition, work on modality and FTR gave way to research on the relationship between space-time metaphors and temporal representations. There is good evidence it is universal for human languages to use spatial metaphors to talk about time (Clark, 1973; Gentner & Boroditsky, 2001). This is driven by general tendencies to use language drawn from more physically concrete domains (space) to metaphorically refer to more abstract domains (time) (Clark, 1973; Lakoff & Johnson, 1980). For instance, English speakers talk about time using the spatial terms *forward, back, long* and *short* (Clark, 1973; Traugott, 1978). A predominant English space-time metaphor—as pointed out by Whorf—is that *time is a line*, or alternatively *time is an arrow*. Additionally,

[3]Not to say there were no important mid-century cognitive-scientific contributions to linguistic relativity research. For example, on linguistic relativity and colour see Brown and Lenneberg (1954); Heider (1972), and Heider and Olivier (1972); and on Hopi verb semantics and classification of events see Carroll and Casagrande (1958).

English speakers—as well as speakers of a vast number of other languages—tend to think of the past as being *behind ego* on the line of time, and future as being *in front*. However, these metaphors are by no means universal, and much work on linguistic relativity and time has sought to understand how cross-linguistic differences in commonly used space-time metaphors impact how speakers represent time. For instance, work has explored whether "up vs. down" (Fuhrman et al., 2011) or "East vs. West" (Boroditsky & Gaby, 2010) spatial metaphors relate to speakers' representations of the "arrow of time".

Whereas this work focusses on space-time metaphors and temporal representations, the present thesis investigates the impacts of FTR grammar on intertemporal decision making. As such, there is little which *topically* connects relativity work on space-time metaphors to this dissertation, apart from time being involved in both. Therefore, the aim of this section is to provide *critical*, rather than *topical*, context. Below, I detail key thematic criticisms which are general to linguistic relativity research. These are illustrated, where possible, with examples from relativity work on space-time metaphors. My aim is not to provide a comprehensive history of contemporary work on linguistic relativity and time (the reader should consult Everett, 2013, pp. 109–39). Rather, I hope to summarise common critical perspectives illustrated with examples from this body of work, such that my approach to addressing these recurrent criticisms can be seen in context. I focus on three common criticisms, which I refer to as: (a) the *correlation* problem; (b) issues of *magnitude*; and (c) concerns around *scope*.

1.2.1 The correlation problem: Does language cause observed effects?

What I refer to as the correlation problem is not specific to research on linguistic relativity; correlation does not imply causation, regardless of the topic of inquiry. Applied to this context, most criticism focusses on the possibility that cultural or environmental factors are responsible for both cross-linguistic and non-linguistic cognitive differences between populations (see Bloom & Keil, 2001; McWhorter, 2014). Imagine some population of individuals, for whom research has identified a cross-linguistic difference and established a correlated non-linguistic cognitive one. It would be possible that the cognitive difference had caused the linguistic one, or that both had been caused by other factors. First, shared environments might influence language, culture, and cognition. Compared with populations from small-scale societies, populations from large-scale ones have spent their lives in very different physical environments. This has sweeping impacts (on the extensive cognitive differences between denizens of large- and small-scale societies, see Henrich, Heine, & Norenzayan, 2010). Alternatively, culture could impact cognition and language. While language is, "perhaps the most unique component of culture.... [it] does not represent the entirety of culture" (Everett, 2013, p. 24). There are surely aspects of culture—value systems, resource extraction patterns, physical technologies, built environment, etc.—which are unrelated, or very loosely related, to language. Such differences may affect both language and non-linguistic cognition (Everett, 2013, pp. 24–5). Here, a crucial point is that language may be the medium through which culture influences thought: "Every student of language or society should be familiar with the essential idea of linguistic relativity, the idea that culture, through language, affects the way we think" (Gumperz & Levinson, 1996, p. 1). Untangling these alternative accounts is the goal of much relativity work.

A useful theoretical framework to parse how correlational evidence may be used to support causal inference was developed in the 1960s by researchers grappling with whether smoking cigarettes has a causal impact on the development of lung cancer (Pearl & Mackenzie, 2018, p. 179). While this causal relationship might now be widely accepted as obvious, critics at the time insisted that the observed association between smoking and lung cancer could have been driven by a shared genetic factor. They insisted that shared genetics might be responsible for addictive tendencies and susceptibility to cancer. While it is unclear that as firm a causal connection will ever be developed between cross-linguistic differences and non-linguistic cognition, the problem is similar. Shared cultural or environmental factors might be responsible for both cognitive and linguistic differences between populations. The observed association could be the result of a confound. In 1962, the Royal College of Physicians issued a report in which they identified 5 criteria by which to judge and evaluate the causal significance of an association. Among them were: *consistency* (similar effects demonstrated in multiple populations); *strength* (a "dose response": more exposure should have a bigger effect); *specificity* (a specific exposure should have a specific effect, rather than a long list); and *coherence* (plausibility of an effect given relevant research and theory) (Pearl & Mackenzie, 2018, pp. 179–81). Even if they are not specifically mentioned, much research on linguistic relativity tacitly invokes these factors when arguing that observed associations between language and non-linguistic cognition constitute causal evidence. I outline examples below.

1.2.1.1 Specificity: A single cause should have a single effect

A good example of *specificity* can be found in work investigating whether vertical temporal metaphors in Mandarin affect speakers' temporal representations. In Mandarin, temporal events are often represented using vertically-oriented metaphors. For instance, the spatial metaphor *shàng* (上 'up') is used to refer to earlier (past) events, while *xià* (下, 'down') is used to refer to later (future) events. Fuhrman et al. (2011) investigated whether this linguistic difference was associated with a difference in temporal representations.

They compared English speakers living in the US ($n = 118$) with Mandarin speakers living in Taiwan ($n = 63$). For each trial, participants were presented with an image on a computer screen. Then, after two seconds, they were shown a second image which was either relatively earlier or later than the first. For example, they might initially be shown a picture of Woody Allen, and then a second picture of him as either an older or younger man. Their task was to indicate with a keypad whether the second image was earlier or later, using a non-linguistically-designated key (i.e. “earlier” was marked with a black sticker and “later” was marked with a white one). Reaction time was recorded. The keypad was mounted on a rotating ball which allowed the “earlier vs. later” vector to be aligned with three different axes: transverse (left-right), vertical (up-down), or sagittal (front-back). Each participant completed two blocks of 56 trials for each axis. The direction along each axis was reversed between blocks, e.g. for the transverse axis the left key was “earlier” in one block and the right key was “earlier” in the other. This meant that the horizontal block with the left key as “earlier” reflected the canonical space-time metaphor in English, while the vertical block with the top key as “earlier” reflected the canonical space-time metaphor in Mandarin.

They found that Mandarin speakers responded significantly faster in the canonical conditions for both the transverse and vertical axes, which the authors note is predicted given the relatively recent (2004) switch between top-down to left-right orthography in Taiwan (orthographic direction has previously been found to align with temporal representations (Fuhrman & Boroditsky, 2010)). On the other hand, this “canonicality” effect only emerged for English speakers in the transverse axis. Critically, no reliable patterns emerged in the sagittal axis. Since, neither language employs sagittally-oriented space-time metaphors, the canonicality effect corresponded *specifically* to the predominant space-time metaphors in each language.[4] In this pattern of effects, we see both positive and negative predictions being confirmed, which significantly strengthens the grounds for the inference that linguistic space-time metaphors have a causal influence on temporal representations (on this general point, see Firestone & Scholl, 2015). However, it does not necessarily indicate a causal relationship. Critics might reasonably claim that cultural differences caused both cognitive effects and linguistic differences, although it is difficult to conceive of what non-linguistic cultural fairy dust would need to be invoked (apart from orthography).

1.2.1.2 Strength: Effects should scale with doses

A good example of *strength*, comes from a second study in Fuhrman et al. (2011). In this context, I am treating cross-linguistic exposure as the “dose” and non-linguistic cognition as the response. Fuhrman et al. (2011) had English speakers with no Mandarin exposure ($n = 134$), as well as Mandarin-English bilinguals who resided in the USA and Taiwan ($n = 196$), complete the following task. An experimenter would stand beside a participant and indicate a point approximately one-foot in front of the participant’s chest using a palm up fingers-together gesture. The experimenter would then ask, “if this is today, where would you put yesterday?” (or “tomorrow”). Ninety-four percent of English participants used a horizontal orientation

[4] Though a critical point is that this is also the orthographic direction in each culture, which means that a non-linguistic cultural difference (orthographic orientation) might equally explain findings.

(and 88% placed the earlier events to their left). However, in the case of the Mandarin-English bilinguals, the proportion of participants who used a vertical orientation went up as their proficiency with Mandarin increased. Low-Mandarin-proficiency bilinguals tested in English indicated mostly horizontal orientations, the canonical English orientation. Conversely, high-Mandarin-proficiency bilinguals residing in Taiwan who were tested in Mandarin responded with vertical orientations approximately 50% of the time (Fuhrman et al., 2011). As speakers' exposure to Mandarin went up, so too did their vertical representation time's arrow. This indicates a dose response in the predicted direction and strengthens causal evidence, though again is open to the criticism that cultural differences might be producing linguistic and cognitive ones, particularly orthography.

1.2.1.3 Other factors: Coherence and consistency

I will not spend undue time illustrating the other factors in the Royal College framework, as *strength* and *specificity* are most germane to the interpretation of the results presented in this thesis. However, it should be noted that the accumulation of similar effects I have reviewed above contributes to the *consistency* and *coherence* of the general hypothesis that cross-linguistic differences in space-time metaphors affect temporal representations.

1.2.1.4 Strengths and weaknesses of the Royal College framework

The value of this framework is that it provides several clear criteria by which researchers can judge causal evidence. However, a significant drawback is that the factors can only be used to qualitatively access evidence. The factors help *bolster* causal inference, but are not conclusive. Even when a particular criterion is met, objections are usually possible. Additionally, the criteria can be difficult to apply in practice. What about if one theory produces *specific* effects and another exhibits *strength*? How are we to judge the relative importance of these factors (see Pearl & Mackenzie, 2018)? Nonetheless, in interpreting the work I present in this thesis, I invoke some of these factors; principally *specificity* and *strength* (Section 6.4.1). However, to further address the shortcomings of the framework, I also used path-analytic techniques to understand how language use *mediates* relationships between speaking populations and cognitive effects (see Section 1.4.3, and Chapters 3 and 4). Such approaches can be useful as a way of strengthening causal evidence using correlational data (Pearl & Mackenzie, 2018). I also manipulated language exposure within a population of bilinguals, and then tested for cognitive effects. This within-subjects design helps rule out cultural differences as a causal factor, given that the control and test population comprise individuals from one culture tested at different times (Chapter 4). Finally, I used a controlled language-production task to test whether directing participants to speak about the future in specific ways affects intertemporal decision making (Chapter 3). By experimentally controlling language production, these methods provide evidence that language use impacts cognition.

1.2.2 The issue of magnitude: Are relativity effects really very profound?

A second common criticism of linguistic relativity is that the effects of language are real but negligible. *The Language Hoax* by John McWhorter is a good contemporary example, summarised effectively in a jacket quote by Steven Pinker, who admits linguistic relativity may have a “smidgeon of truth on minor matters” (McWhorter, 2014), but no more. In general, this critical perspective evaluates contemporary relativity work against the backdrop of Whorf’s ambitious ideas, and concludes there is only mild cross-linguistic influence (Bloom & Keil, 2001). McWhorter takes such a stance in critiquing a recent study. In this study, English and Mandarin-English bilinguals completed a reaction-time task using a horizontally- or vertically-oriented keypad in response to temporally ordered pictures (Boroditsky, Fuhrman, & McCormick, 2011). Mandarin speakers responded 170 milliseconds faster when the vertical axis was arranged in the Mandarin-canonical “up is earlier” condition. McWhorter points out that they responded with even greater relative speed (232 milliseconds) in the English-canonical “left is earlier” condition, and asks, “is that layered, subordinated twinge a ‘world-view’?” (p. 27). This is a fair point. How do response-time differences of less than a quarter second amount to the profundities limned by Whorf?

Such criticisms tacitly invoke an important notion which informs my approach: representative design. This refers to the extent to which experimental settings represent real-world environments of interest (Araújo, Davids, & Passos, 2007), and therefore the extent to which experimental findings predict real behaviour (Araújo et al., 2007). The assertion that most contemporary work on linguistic relativity involves poor representative design is at the core of McWhorter’s argument:

> Even if you can, as it were, trick someone in to revealing some queer little bias in a very clever and studiously artificial experiment, that weensy bias has nothing to do with anything any psychologist, anthropologist, or political scientist could show us about how the people in question manage existence. (2014, p. 28)

What Mandarin speaker, he asks, will talk about the future while pointing animately to their feet (p. 27). In fact, they do (Gu, Mol, Hoetjes, & Swerts, 2017), but McWhorter’s point stands: Even if Mandarin speakers do point down to tomorrow, it is hard to see this impacting much about their lives.

In this thesis, I address such criticism by investigating how cross-linguistic differences may influence the way people make judgements and decisions involving intertemporal trade-offs.

Unlike the impact of spatial metaphors on construals of time's arrow, experimentally-measured intertemporal decision making has been shown to reliably predict a range of consequential real-world behaviours, including health outcomes (Vuchinich & Simpson, 1998), drug use (McKerchar & Renda, 2012), pathological gambling (Hodgins & Engel, 2002), and investment in savings (Liu & Aaker, 2007) (see also Section 1.3.3.1). By investigating cross-linguistic impacts on judgement and decision-making processes which underpin such behaviour, I emphasise representative experiments which closely reflect (and statistically predict) important behaviour outside the laboratory.

1.2.3 Concerns around scope: How deep do relativity effects go?

A third issue for linguistic relativity concerns the extent to which perceptual and cognitive processes are (im)penetrable by language. Modern approaches to investigating linguistic relativity have reframed the question, "from one of whether language shapes our thinking or not, to one that [tries] to understand the factors that contribute to the extent and nature of any observable influence of language on *perception*" (Athanasopoulos & Casaponsa, 2020, p. 393, emphasis added). The relevant question here is—not whether—but which aspects of mental activity are more or less "penetrable" by language. Below, I outline two important issues which fall under the auspices of concerns around scope: (a) the extent of linguistic penetration into perception; and (b) questions around the durability of relativity effects.

1.2.3.1 The extent of linguistic penetration into perception

I emphasised perception above because discussion around scope tends to involve the extent to which language can exert "top down" influences on "deeper" processes of—usually visual—perception (see Bloom & Keil, 2001; Firestone & Scholl, 2015; Lupyan, Rahman, Boroditsky, & Clark, 2020; Montemayor, 2019). A relevant distinction is between *perception* and *cognition*, where the former involves lower-level processes which encode relevant features of the external environment into accurate *percepts*, and the latter involves higher-level processes of remembering, inferring, reasoning, or desiring (Firestone & Scholl, 2015). This is summarised by a recent review as the difference between, on the one hand, "*seeing* a red apple and, on the other hand, *thinking about, remembering* or *desiring* a red apple" (Firestone & Scholl, 2015, p. 1). The parameters of this debate mostly involve the evaluation of evidence for the relative extent to which low-level processes of perception are "encapsulated" and therefore (im)penetrable to higher-level cognition (Firestone & Scholl, 2015). In the context of linguistic relativity research, the relevant question becomes to what extent perception is penetrable by language, and therefore (potentially) influenced by cross-linguistic differences (Lupyan et al., 2020). On one hand, some authors have argued for pervasive influences of language on fundamental perceptual processes (Lupyan et al., 2020), while others have suggested that apparent evidence for "top-down" effects can be explained by other experimental or theoretical factors (Firestone & Scholl, 2015).

With regards to temporal perception specifically, Montemayor (2019) argues that many aspects of temporal perception are encapsulated and not influencable by language. He makes a case that a relevant distinction is between "early and late stages of processing" (Montemayor, 2019, p. 137). He defines early-stage processing as being characterised by very short (sub 1 second) duration estimation and simultaneity judgements. These are constrained by the need to reliably map onto features of the environment, and—since they are shared by non-linguistic animals—language-independent (Montemayor, 2019). On the other hand, late-stage processing is characterised as occurring over longer time spans, which "integrate temporal structures in terms of linguistically based representations and symbolic formats, paradigmatically, explicit judgements" (Montemayor, 2019, p. 138). The relevant empirical findings here involve the extent to which estimations or duration are penetrable by experimental interference based on language-canonical temporal metaphors for quantities of time. For instance, in Spanish it is common to use amount-based metaphors to talk about time quantities, e.g. *mucho tiempo* 'much time', and *poco tiempo* 'little time' (Bylund & Athanasopoulos, 2017). On the other hand in Swedish—as in English—it is much more common to use length-based metaphors, e.g. *lang tid* 'long time', and *kort tid* 'short time'.

Using a psychophysical paradigm developed by Casasanto and Boroditsky (2008), Bylund and Athanasopoulos (2017) conducted an experiment where they showed Spanish and Swedish participants either lines which grew from left to right for a given duration, or containers which filled up for a given duration. The container volumes and line lengths were varied and crossed with a set of temporal durations, such as to systematically vary visual and temporal disagreement. Participants then re-produced perceived duration by clicking their mouses after each trial. Swedish speakers' estimations of duration were more biased by visual interference in the "line" condition, and Spanish speakers were more biased in the "container" condition—the visual representations which matched their language-canonical temporal quantity metaphors (for similar findings, also see Casasanto & Boroditsky, 2003; Casasanto et al., 2004; Casasanto, 2005; Casasanto & Boroditsky, 2008; Casasanto, 2008). Montemayor (2019) argues that this kind of evidence is not enough to conclude early-stage temporal processing is impacted. This is because outcomes involve explicit judgements and post-test reproductions, which are argued to be canonical "late-stage" temporal cognition, rather than "early-stage" temporal perception.

The point of this discussion is not to conclude on either side of this debate. Colour category effects *have* been found in very early stage (100ms post stimulus onset) visual processing (Forder, He, & Franklin, 2017). While this suggests low-level perceptual processes are penetrable by language, debate will surely continue. In any case, it may be less controversial to suggest that higher-level—in many cases linguistically-mediated—processes of judgement, reasoning, and memory are permeated by linguistic representations and therefore influenced by language. Montemayor (2019) writes that the purpose of late-stage temporal processing may be to, "integrate personal narrative that... build[s] long-term temporal representations based on linguistic ones" (p. 149). We might therefore expect cross-linguistic effects on such higher-level cognitive processes.

As such, my approach is not to wade into the fraught debate about linguistic penetration into visual (Firestone & Scholl, 2015; Lupyan et al., 2020) and temporal (Montemayor, 2019) perception. Rather, I take my cue from these theoretical (Montemayor, 2019) and empirical (Firestone & Scholl, 2015) arguments which suggest that higher-level cognitive processes ought to be impacted by language. Specifically, I investigate whether language affects how people assess value when outcomes are delayed and/or risky (Chapters 3 and 4). I make no claims whatsoever on the basis of my findings about linguistic influence on what Montemayor (2019) calls early-stage temporal processing. However, I gently push back against literature which views perception as "deeper" than cognition, and therefore tacitly trivialises effects on anything but carefully delimited early-stage aspects of perceptual processes. Of course, any serious impacts on the perceptual faculties which allow an organism to, for instance, temporally bind the movements of its own body with its sense of agentive action, would be catastrophic. It may be reasonable not to expect cross-linguistic differences to impact processes which are relevant to survival (Montemayor, 2019). However, for most people, deciding whether to keep smoking, exercise regularly, invest in personal education, start a college fund for their first child, or save money each month may be more salient to their daily experience of time. While such intertemporal decision making represents a restricted area of influence, it hardly seems "narrow" (cf. Montemayor, 2019).

1.2.3.2 The durability of linguistic effects on cognition

A second issue which falls under the auspices of concerns around scope has to do with the persistency of effects of language, i.e. whether and to what extent cognitive effects persist beyond the immediate context of online language production and/or comprehension.

One possibility is that linguistic effects are limited to the preparation for speech. The idea is that the grammatical or lexical system of a particular language mandates speakers notice certain things about their environment in order to communicate effectively. This drives attention during the conceptual planning required to produce a linguistic utterance. For instance, certain languages (e.g. English, German, Russian, Chinese) tend to encode manner of motion in the main verb (e.g. *hop, skip, stroll, mosey*) whereas others (e.g. Greek, Spanish, French, Japanese) encode path of motion in the main verb (e.g. *go, pass, enter*) and encode manner in an optional paraphrastic satellite[5] (Talmy, 1985; and see Ibarretxe-Antuñano, 2008; Slobin, Ibarretxe-Antuñano, Kopecka, & Majid, 2014; Slobin, 2006; Slobin, 1987). Concordantly,

[5]It was difficult to identify illustrative examples involving temporal cognition, so this section reviews relativity effects from other domains.

compared with Greek speakers, English speakers attended more closely to manner of motion in a visually-presented tableau, but only when they were instructed to watch in preparation to speak. When they were not told they would have to describe the scene, there was no difference (Papafragou, Hulbert, & Trueswell, 2008). Such "thinking for speaking" effects (Slobin, 1987) suggest that the cognitive influence of cross-linguistic differences persist—in some cases—only insofar as people are engaged in attending to experience in preparation to use language.

Another possibility is that some linguistic effects may be limited to the context of online language production and/or comprehension. In this case, linguistic representations interfere with non-linguistic ones, such that cognition is affected only while linguistic representations are being engaged in the performance of some task. When this is the case, cross-linguistic differences should disappear when participants are given an additional linguistic interference task (Wolff & Holmes, 2011). A good example comes from work on colour terms. Speakers of languages with more precise colour terminology may be faster at discriminating between colours which span their language-specific colour category boundary (Winawer et al., 2007). For instance, Russian speakers were faster than English speakers at discriminating between between light and dark blues (Winawer et al., 2007); this is attributed to the fact that Russian has different words for light blue (*goluboy*) as compared with dark blue (*siniy*) (Winawer et al., 2007). However, this effect was non-significant when a linguistic interference task was added. This suggests that linguistic representations interfere with cognitive processes of colour discrimination, and also that such interference disappears when the language production system is otherwise engaged.

In contrast to such ephemeral cases, relativity effects may also persist beyond language use, during offline cognition. Over time, the attentional demands of language production may "spotlight" certain aspects of experience, and thereby contribute to the formation of persistent language-specific offline attentional and cognitive differences which arise over a lifetime of language use. In such cases, we would expect relativity effects to persist under linguistic interference, which is sometimes the case (Boroditsky & Schmidt, 2000).

The common thematic concern has to do with whether relativity effects emerge before, during, or after language production, and whether cognitive effects become independent from linguistic representations (Wolff & Holmes, 2011). These are important questions to any specific relativity effect, i.e. to what extent some particular influence of language is (im)permanent. This thesis does not explicitly investigate such questions, but it is worth mentioning a theoretical point now. There may be important differences between studies of perception and studies of judgement and decision making (like this one). To wit, if online language production influences decision making by, for instance, engaging linguistic representations to weight the costs and benefits of a decision, then such influence would persist once a decision is made. In other words, since the downstream effects of decisions—once made—are more durable than perceptions, online effects would also endure longer. However, the extent to which linguistic reasoning is involved in intertemporal decision making is—as far as I can tell—understudied.

1.3 Cross-linguistic effects on economic behaviour: The effect of tense on saving

Over the last decade, economists have increasingly looked to cross-linguistic differences as causal factors of economic outcomes (Samuel, Cole, & Eacott, 2019; Mavisakalyan & Weber, 2018). For instance, there have been recent investigations of the effects of grammatical gender marking on labour supply (Gay, Hicks, Santacreu-Vasut, & Shoham, 2018), wage inequality (Shoham & Lee, 2018), and women's labour force participation (Jakiela & Ozier, 2018); and of grammatical mood on attitudes towards immigration (Kovacic & Orso, 2016).

Economic work focussed on intertemporal outcomes has proliferated. This work is characterised by a theoretical framework which can be traced exclusively to a single publication: "The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets" by Keith Chen (2013). This publication, and follow up in the literature, is a critical touchpoint for this thesis, so understanding Chen's theory and evidence is crucial. Chen's approach was to review the relevant literature on future tense typology, quantify what he theorised were relevant cross-linguistic distinctions, and use this quantified typological variable to predict various behaviours in which intertemporal decision making appeared to be involved. Particularly, he presents evidence that FTR grammar impacts temporal discounting processes. Temporal discounting is the subjective devaluation of delayed outcomes as the time until they occur grows longer (L. Green, Myerson, & Vanderveldt, 2014). People tend to radically discount delayed rewards. For instance, most people would prefer $20 now to $40 in twenty years time, because they devalue the delayed $40. Differences in temporal discounting rates affect a wide range of intertemporal behaviour, including health decisions (Tate, Tsai, Landes, Rettiganti, & Lefler, 2015) and spending (Bickel et al., 2010). See Section 1.3.3.1 for a general introduction.

The relevant typological work was primarily research produced in collaboration with Östen Dahl who, in the 1990s, lead the EUROTYP Theme Group on Tense and Aspect (Dahl, 2000b). Among other things, this working group sought to characterise about 30 European languages in terms of differences in <u>tense</u> marking of future time reference. My approach to measuring cross linguistic differences in FTR is deeply indebted to this work, and I outline key similarities and differences in Section 1.4.1. A critical point worth flagging is that Dahl's project focussed on tense and aspect, but not modality. This is not intended as a criticism—all researchers need to choose areas of focus. It is simply a fact, which may have resulted in Chen's theorising exclusively about <u>tense</u> and not also <u>modality</u>. Over the course of this thesis, I hope to demonstrate that this is a critical omission in the context of cross-linguistic impacts on intertemporal decision making.

I identified three steps as necessary for relativity projects: (a) the establishment of cross-linguistic differences; (b) the identification of correlated cognitive differences; and (c) the causal linking of these. In keeping with this framework, Chen's (2013) project involves: (a) a typological enterprise involving FTR grammar; and a number of statistical analysis designed to (b) establish population differences in various economic and health outcomes, and (c) provide causal evidence for the impact of (a) on (b). The typological aspects of Chen's project (i.e. [a], above) can be further divided in two distinct parts. First, Chen (2013) classifies languages in terms of how obligatory[6] it is to use the future tense. Second, he makes a general argument about the semantics of the FUTURE TENSE[7] from which he derives his hypotheses as to how obligatory FUTURE TENSE use is predicted to impact intertemporal decision making. In this section, I detail these aspects of Chen's (2013) project in turn: first the establishment of cross-linguistic differences, and then the evidence that these affect intertemporal behaviour. Before getting into the details of Chen's (2013) hypotheses and evidence, it will be useful to provide some working definitions of the linguistic elements involved: tense and modality. I do this in the next section before returning to a close analysis of Chen's (2013) work.

1.3.1 Tense and modality: A brief primer

Chen's (2013) theoretical approach involves argumentation about how the FUTURE TENSE encodes notions having to do with the timing of future events. As such, his work, and much follow up in the literature, treats time as one-dimensional, and future time as analogous with past time. I argue these assumptions are not justified due to the overlap between the grammatical categories of tense and modality when it comes to FTR. The difference between tense and modality is usually defined according to semantics: Tenses are those linguistic elements which encode temporal notions, whereas modal operators are those linguistic elements which encode "modal" notions of probability, possibility, necessity, desire, and obligation (Dahl, 2000a; Palmer, 2001; Nuyts, 2000). I expand on this distinction below.

[6]Obligatory. This term is perhaps misleading. It implies the grammar of a language permits no other options. The extensive individual differences in FTR language use which this dissertation identifies may undermine this assertion. In fact, this is expected by usage-based accounts of grammar which understand there to be differences in grammatical knowledge between speakers (Bybee, 2006). It may therefore be more appropriate to speak in terms of clines of "obligatoriness" (Dahl, 2000c, p. 9), but this is seems rather clunky. A better terminological decision might be to talk about *preferred* expression types in a given context, but this is not always strong enough. I therefore use obligatory, despite the fact that there are nearly always low-frequency alternatives which are possible. English seems to lack good term for "nearly obligatory."

[7]I will use SMALL CAPS to refer to general cross-linguistic categories; these will be omitted when I refer to language-specific realisations of such general categories, e.g. the English future tense is an example of the cross-linguistic category FUTURE TENSE. I raise this now because Chen (2013) makes an argument about the semantics of the generic FUTURE TENSE (for a discussion on the importance of distinguishing between language specific descriptive categories and comparative concepts, see Haspelmath, 2010).

1.3.1.1 Tense: Deictic encoding of relative temporal order

Tenses are usually thought of as deictic expressions which relate the time of a referenced event to the time of speech (Lyons, 1968; Mezhevich, 2008). In the Western Germanic languages in my sample, future tenses are formed using paraphrastic auxiliary modal verbs or andative—go-based—constructions (e.g. English *will, shall,* or *be going to*; Dutch *zullen* 'will', or *gaan* 'be going to'; and German *werden* 'will'). In a typical ternary account of tense, Klein (1995) proposes that tense clarifies the temporal order between the utterance time and the reference time, so e.g. the present tense indicates reference time and utterance time are the same, past tense indicates reference time precedes utterance time, and the future tense indicates reference time follows utterance time. For instance, in English:

(1.1) a. **Past:** *It rained.*

b. **Present:** *It is raining.*[8]

c. **Future:** *It will/shall/is going to rain.*

What is being expressed in example 1.1a–c is when, relative to the time of utterance, the event in question takes place. Other theoretical treatments of tense eschew ternary models inspired by properties ascribed to time by contemporary physics (Broekhuis & Verkuyl, 2014). For instance Te Winkel (1866), and later Verkuyl (2008), combine elements of tense and aspect in positing that there are 8 Dutch tense forms based on three binary oppositions: 1) present vs. past, 2) synchronous vs. posterior, and 3) imperfect vs. perfect. For instance, in the present-synchronous category, an imperfective statement would be *Elsa loopt* 'Elsa walks' (i.e. the simple present), while a perfective statement would be *Elsa heeft gelopen* 'Elsa has walked' (the present perfect); whereas in the present-posterior category, an imperfective statement would be *Elsa zal lopen* 'Elsa will walk' (simple future), and a perfective would be *Elsa zal hebben gelopen* 'Elsa will have walked' (future perfect) (examples from Broekhuis & Verkuyl, 2014). Thus, past/present distinguishes between what most would consider past and present tense, synchronous/posterior distinguishes between past + present on the one hand and future on the other, and perfective/imperfective distinguishes between the English simple and prefect aspect (which express deictic time relations relative to the time of reference rather than the time of utterance). These are just two accounts of tense. Many more likely exist, and of course languages differ in terms of how they grammatically divide up time (Broekhuis & Verkuyl, 2014; Cable, 2013; Chen, 2013; Dahl, 2000a; Levinson & Majid, 2013; Lyons, 1968; Malotki, 1983). The salient point is that tenses encode notions about when in time some event occurs, relative to the time of speech (Lyons, 1968).

[8] English uses the present progressive to refer to present time events. The zero-form simple present tense, *It rains*, is used for gnomic statements which have truth value independent of any deictic time reference, e.g. *It rains in Oxford* (Broekhuis & Verkuyl, 2014).

1.3.1.2 Modality: Possibility, necessity, obligation, ability, and desire

Modality is usually suggested to express degrees of possibility vs. necessity and/or obligation relative to a modal "base" (Palmer, 2001). Again, various schemas have been proposed (Bybee, Perkins, & Pagliuca, 1994; Nuyts, 2000; Palmer, 2001). For the sake of brevity, I only describe the schema proposed in the definitive *Mood and Modality* (Palmer, 2001). According to Palmer (2001), *deontic* modality has to do with expressing what is necessary or expected relative to social norms or obligations. For instance, a parent might say to a child, *You must brush your teeth every night before bed*, or (to an older child), *You should always make sure to get enough sleep before an examination.* This is contrasted with *dynamic* modality, which expresses what is possible or likely relative to internal properties of the entity being referred to, e.g. *Billy can run a 4-minute mile*, or *It sure can get cold in Copenhagen in the winter.* I should also mention *bouletic* modality which quantifies some state of affairs relative to the desires and hopes of the speaker, e.g. *I want to come tonight*, or *I hope Mike will agree to be my prom date.* Finally, *epistemic* modality involves speakers expressing what is possible, probable, or certain relative to what they know or believe (Nuyts, 2000). A speaker expressing their confidence in a prediction is a prototypical example, e.g. *I think the Bears have a good chance of winning the trophy tonight*, or *The Bears just might win!* Epistemic modal notions map most transparently onto notions of probability and risk known to impact intertemporal decision making (see below) so epistemic modality is the main focus of this dissertation.

A relevant question is what, exactly, do modal operators express relative to modal bases. What does it mean to say *It could be the case that X*, or *It must be the case that Y*? In answering this question, much work on modal semantics has used ideas derived from formal modal logic. In this view, modal operators "quantify" over a set of possible worlds (Kratzer, 2012). For instance, a weak modal operator like *may* serves as a marker of possibility, quantifying over at least one possible world. In this framework, to say *It may be the case that X* is to say, in effect, *There is at least one possible world in which X is the case.* A strong modal operator like *must* quantifies over all possible worlds. To say *It must be the case that Y* is to say *Y is the case in all possible worlds.* However, I am sympathetic to recent accounts which treat modal expressions as scalar operators. In the case of *epistemic* modality, scalar accounts map transparently onto notions of probability (p). Rather than Boolean quantification, epistemic modal semantics involve speakers expressing beliefs about the likelihood of events on a one-dimensional scale between high ($p = 1$) and low certainty ($p = .5$)[9] (Lassiter, 2015; Moss, 2015; Santorio & Romoli, 2017). I take the view that scalar accounts capture how real people use language to express themselves better than abstruse formal theories. This is partly motivated by my research question. Scalar accounts afford easy translation between modal semantics and intertemporal decision making involving risk (i.e. involving probability).

[9] Negated high certainty ($p = 0$) might be added, for cases where a speaker is highly certain it is NOT the case that X, but such cases are avoided in this dissertation, see Chapter 4.

They therefore provide a parsimonious theoretical framework by which the obligation to use weak modals affects (risky) intertemporal decision making. Beyond their felicitousness to the preoccupations of this dissertation, evidence suggests scalar accounts capture modal semantics better than notions of Boolean quantification, since the latter yield incorrect predictions in some linguistic contexts (Lassiter, 2015).

In English, Dutch, and German, modality is largely expressed through a system of modal verbs, which polysemously relate to different modal bases dependent on context (Nuyts, 2000). For instance, *may* can express epistemic or deontic modality:

(1.2) a. **Epistemic:** *It may get cold later.*
b. **Deontic:** *You may come in.*

On the other hand, *might* would sound fine in example (1.2a) expressing epistemic notions, but would come across as odd expressing deontic notions in (1.2b): *You might come in* manages to sound like a polite suggestion rather than expression of permission. In Dutch, *kunnen* 'may' is the only modal for which epistemic use is possible, and encodes possibility (Nuyts, 2000). The German case is complicated by the *Konjunktiv II*, a bound morphological subjunctive mood. Mood and modality relate in the following way. Modality is a functional category defined in terms of the linguistic expression of notions to do with possibility, probability, necessity, obligation, desire, and so on. Mood is, more specifically, a grammatical category which, in European languages, tends to index realis versus irrealis notions, which is to say real vs. non-real statements[10] (Thieroff, 2010b). In English, this can be seen in the loosely grammaticised subjunctive mood, usually expressed using *if... would*, or *were*, constructions, e.g. *If it rained I would come inside*, *If I were a rich man...*, or *... but he would never make it home* (Bergs & Heine, 2010). While mostly used to express irrealis notions (Thieroff, 2010a), the *Konjunctive II* can also attenuate German modal verbs, altering their modal strength, or the base to which they can refer (Mortelmans, 2000; Nuyts, 2000). For instance, *mögen* 'may' encodes possibility, but only in the present tense *mag*, while in the *Konjunktiv II*, *möchte*, it can only be used in deontic contexts (Nuyts, 2000). Conversely, some modals are only associated with a particular base; for instance *must* (Dutch *moeten*; German *mussen*), is mostly limited to deontic rather than epistemic use, e.g. *You must be home by dark*. Similarly, *can* expresses mostly dynamic ability (*He can eat 50 hot-dogs in one hour*), though its past tense form *could* can be used epistemically (*It could rain later*) (Nuyts, 2000). This is not the case in Dutch: *Kunnen*, which is a cognate of 'can' but translated as 'may', is the main modal verb used to express epistemic notions.

Additionally to the modal verbs, modality can be expressed in English, Dutch, and German through a variety of periphrastic constructions, principally mental state predicates and

[10]Though other moods such as the interrogative mood distinguish questions from assertions.

what I refer to as modal modifiers. Modal modifiers are adverbial, adjectival, or (rarely) nominal lexemes which are derived from shared roots and encode highly-specific modal notions, e.g. in English *possibly* (adverb), *possible* (adjective), and *possibility* (noun) (Nuyts, 2000). The principal modal modifiers in the languages in my sample are: English *possibly, probably, certainly*; Dutch *mogelijk(erwijze), waarschijnlijk, zeker*; German *möglich/möglicherweise, wahrscheinlich, bestimmt*. These are highly comparable between the languages, e.g. in English *possibly* indicates low certainty, *probably* indicates middling certainty, and *certainly/definitely* indicate high certainty; while in Dutch *mogelijk* 'possibly' indicates low certainty, *waarschijnlijk* 'probably' indicates middling certainty and *zeker* 'certainly' indicates high certainty; and in German *möglicherweise* 'possibly' indicates low certainty, *wahrscheinlich* 'probably' indicates middling certainty, and *sicher/sicher(lich)* 'certainly' indicates high certainty (Nuyts, 2000, p. 56). Mental state predicates are an additional way epistemic modality is expressed in English, Dutch, and German. These are psychological verbs which encode modal notions by allowing speakers to talk about their thoughts and beliefs (Nuyts, 2000). In English and Dutch, the mental state predicate prototypically used to express epistemic modality is *think* (Dutch *denken* 'think'). In German it is *glauben* 'believe' (Nuyts, 2000, p. 109). The notion being indicated is highly similar.

Additionally, Dutch and German utilise systems of modal particles while English does not. However, these constructions are peripheral enough to the main concerns of this thesis so as not to seriously impinge comparability. This thesis principally focuses on the extent to which cross-linguistic differences in FTR grammaticisation are associated with differences in the obligation to encode epistemic modality. While some modal particles have epistemic meaning (e.g. Dutch *wel eens*; German *wohl*), these mostly serve to attenuate other modal structures (Nuyts, 2000). The flavour of this can be seen in *well* in English. For instance, *That <u>could well be</u> the train arriving* communicates strengthened modality compared with *That could be the train arriving*. However, most Dutch and German modal particles are principally used to mark discourse roles and states (see Bross, 2012; Gutzmann, 2009; Karagjosova, 2004), and the epistemic modal attenuation they encode is beneath the resolution of the linguistic coding system I develop (see Section 1.4.1.4). For these reasons, they are not the main focus of this dissertation.

Modal future tense controversies: Much debate about the theoretical status of the FUTURE TENSE has to do with whether it should be subsumed under the category of modality (Dahl, 2000a). In this debate, it is important to remember that the distinction between modality and tense is a semantic one. The relevant question with regard to the FUTURE TENSE, therefore becomes whether it encodes temporal or modal notions. Scholarly consensus around *will* has been that it is more a marker of modality than time (Enç, 1996; Fries, 1956; Huddleston, 1995). Several lines of evidence have been proposed, which include the fact that *will* can be used to express *dynamic* modality without any temporal notion being encoded e.g. *Oil will float on water*, or *He will sit like that all day.* Similarly *will* can be used to mark a present time prediction, e.g. on hearing a knock, *That will be the postman.* In fact, such uses may be common to the FUTURE TENSE (Giannakidou & Mari, 2018; Lyons, 1968), which suggests the FUTURE TENSE is better described by modal semantics than tense. Additionally, some scholars have pointed out there are no formal differences between *will* and other "future-shifting" modal verbs like *could, may*, and *might* (Enç, 1996). If encoding a mixture of modal and temporal notions is a general feature of English modals, on what basis are we to single out *will* as the future tense? Other researchers have argued that *will* exhibits elements of both a tense and a modal (Sarkar, 1998), or argue it has *only* temporal semantics (Kissine, 2008) (though this is a minority view). Similar controversy surrounds *zullen* 'will', the "future tense" in Dutch. Some linguists suggest *zullen* should be considered *only* as a marker of modality (Broekhuis & Verkuyl, 2014), while others take the more mainstream opinion that it exhibits characteristics of both a modal and tense marker (Kirsner, 1969).

As with *will* and *zullen*, the German future "tense" *werden* 'will' is theoretically controversial. Like *will*, it can be used to mark present-time predictions, leading some to suggest it is a marker of prediction rather than future time (Bohnemeyer, 2000). There are key differences between *werden* and *will*. It is cross-linguistically common for future tenses to grammaticise from modal lexemes and verbs of motion and/or change (Dahl, 2000a). The semantics of these roots differ. For instance, *will*, grammaticised from the bouletic proto-Germanic *willan* 'to want' (Dahl, 1985). Glimmers of these archaic semantics can still be seen in low-frequency uses, e.g. *I will that she agrees to my proposal.* On the other hand, *werden* grammaticised from a root meaning *to become* (Petré, 2013), or perhaps *to turn* (Mortelmans, Boye, & Auwera, 2009). It still retains aspects of these semantics: *Er wird alt.* 'He is becoming old' is an acceptable use (Heine, 1995). This suggests *werden* retains some of the semantic flavours of its lexical antecedent, and can be used to indicate processes of change and/or becoming. It is hard to give an account of such semantics working within the strict framework of "tense".

In fact, given a semantic definition of the difference between tense and modality, such complications as those elucidated with reference to *will*, *zullen*, and *werden* are cross-linguistically common enough to call into question the theoretical status of the FUTURE TENSE (Dahl, 2000a). In other words, if the FUTURE TENSE tends to encode modal semantics, is it really a "tense" at all? My main point through this discussion is to flag the entanglement of temporal and modal notions in linguistic future time reference (Dahl, 1985; Bybee & Dahl, 1989; Dahl, 2000a), such that Chen's (2013) approach to characterising cross-linguistic differences can be understood with this in mind.

1.3.2 Chen's approach to establishing cross-linguistic differences

Here, I return to Chen's (2013) establishment of cross-linguistic differences. As mentioned above, this aspect of Chen's (2013) work comprises two distinct parts: first, a cross-linguistic typology based on whether languages oblige the future tense for FTR; second, an argument about the semantics of the FUTURE TENSE. In this section, I outline these parts in turn. Excepting Chen's (2013) inattention to modality—a critical oversight—I argue that his cross-linguistic typology is generally supported by review of the relevant literature (Dahl, 1985, 2000a). While Chen (2013) oversimplifies a hugely complex area—the linguistic expression of FTR—his typological criteria are a principled application of extant FTR typology, and many criticisms misunderstand Chen's (2013) criteria, or simplify cross-linguistic research on the FUTURE TENSE. On the other hand, I argue that Chen's (2013) treatment of FUTURE TENSE semantics are not well-grounded in the typological or linguistic-relativity literature. I demonstrate this through a close reading of Chen's (2013) arguments and evidence—in their own terms—a critical approach which I have not found elsewhere in the literature.

1.3.2.1 Chen's FTR typology

The essential cross-linguistic distinction of Chen's (2013) typology is that some languages oblige the FUTURE TENSE, while others do not. For example, Finnish speakers are able to use the present tense to talk about events today... or tomorrow:

(1.3) Finnish

a. *Tänään on kylmää.*
Today be.PRS cold
'It is cold today.'

b. *Houmenna on kylmää.*
Tomorrow be.PRS cold
'It will be cold tomorrow.'[11]

In contrast, English obliges speakers to use *will*, *be going to*, or *shall* when they talk about the future. Chen's (2013) typological project is to classify languages in terms of whether—like English—they oblige the grammatical marking of FTR, or whether—like Finnish—they do not. The term Chen coins for languages like English is "strong-FTR" as compared with "weak-FTR" languages like Finnish. These terms are unfortunate: FTR stands for "Future Time Reference" and all languages permit FTR whether it is grammatically marked or not (Dahl, 2013). A better convention might be weak-/strong-FTR$_g$ to indicate the distinction

[11] I use the Leipzig glossing rules to provide interlinear glosses (Comrie, Haspelmath, & Bickel, 2015).

involves FTR grammer. However, since the terms have been widely used in the literature I will not deviate from them here. I use the term "FTR status" to refer to the weak-FTR vs. strong-FTR distinction.

Detailed description of FTR status: Some further detail will be useful to forestall potential criticism. The technically specific definition Chen (2013) gives of weak-FTR is a language with no obligatory tense marking of FTR in prediction contexts in main clauses, while strong-FTR is any language which is not weak-FTR (Chen, 2013). The invocation of "main clauses" is because conditional subordinate clauses often do not oblige marking, e.g. *If the postman comes, don't answer!* is FTR but obliges no tense marking.

What is meant by prediction contexts? To answer this question, I need to introduce a set of notional categories which are critical to understanding FTR in general and Chen's FTR typology in particular. The notions are: schedules, intentions, and predictions (Dahl, 2000a). I use the term FTR-mode to refer to these distinctions. These categories arise from the epistemological deficit people face when they talk about the future. We cannot perceive or remember future events, but we "do talk about the future and there may be different grounds for our doing so" (Dahl, 2000a, p. 310). Among the most important are when we talk about *schedules* of well-known future events, when we talk about our own *intentions* for how we plan to act, and when we make *predictions* about some state of affairs. Dahl defines the schedule-based FTR as reference to well-known events for which a specific time is known:

(1.4) a. *The train gets in at 7pm.*
b. *Dinner is at 5.*
c. *I leave in the morning.*

In contrast with this, intentions are generally restricted to things under the control of the speaker, and prototypically show up in sentences where people talk about their own intentions with regard to future outcomes. As an example, Dahl (2000a) borrows a sentence from Lewis Carroll:

(1.5) ['I know SOMETHING interesting is sure to happen,' she said to herself, 'whenever I eat or drink anything;] *so I'll just see what this bottle does.*'[12]

In this example we see Alice use the English future *will* in firm possession of her own intentions about the future outcome. Predictions, on the other hand, involve reference to things beyond the control of the speaker which must be guessed at from knowledge about some current state of affairs. Dahl (2000a) provides the following example also cribbed from Lewis Carroll:

[12]Following Dahl's conventions (1985, 2000a), context is given in [brackets].

(1.6) [There was nothing else to do, so Alice soon began talking again.] *'Dinah'll miss me very much tonight, I should think!'*[13]

It would be nonsensical for Alice to talk about her *intentions* with regard to Dinah's feelings. Alice's intentions have no bearing on that event, because she is not the cat. To talk about Dinah's future feelings, Alice has to make a prediction, which she further qualifies with the modal phrase *I should think.* On this basis, Dahl (2000) distinguishes between intention-based and prediction-based FTR.

While it may be relatively rare for this notional distinction to correspond to a simple grammatical category (Dahl, 2000a), FTR usage differences often loosely track FTR-mode. For instance, when a speaker is talking about their immediate intentions, it is common for European languages to allow the present tense:

(1.7) a. Russian

Idu	*v*	*gorod.*
go:IPFV:PRS:1SG	to	town:ACC

'I go to town.'

b. French

Je	*vais*	*en*	*ville.*
I	go:PRS:1SG	to	town:ACC

'I go to town.'[14]

The Russian verb is imperfective which is typical for present time reference (Dahl, 2000a, p. 312). In English, this is a typical context for the present progressive, e.g. *I'm going to town*, or *I'm leaving for town* [sometime soon]. Dahl (2000a) notes that these uses convey a sense that preparation for the act has already begun.

Further, present-tense FTR is commonly allowed in scheduling contexts, see example (1.4). Such uses of the present tense for FTR—sometimes called "futurates" (Chen, 2013)—are highly common cross-linguistically, even in strong-FTR languages with otherwise-obligatory future tenses (Dahl, 2000a). In contrast to this, pure prediction contexts tend to oblige marking:

(1.8) a. ? *It's raining* [sometime soon].

b. *It will rain* [sometime soon].

Leaving out the temporal adverbial *sometime soon* in (1.8a) simply fails to convey futurity, whereas retaining it sounds odd. Rather, English predictions oblige a future tense marker,

[13] Dinah was the cat.

[14] This and (1.7a) are from Dahl (2000, pp. 311–2)

either *will*, *be going to*, or *shall*. Since it is common for (even strong-FTR) languages to permit present tense FTR for schedules and intentions, Chen (2013) uses prediction-based FTR as the criteria which defines FTR status.

This decision is motivated by evidence that the obligatoriness of future tense marking in prediction-based FTR is a bellwether for FTR grammaticisation more generally. Dahl (2000a) writes that it, "can be used as one of the major criteria for whether [FTR is grammaticised] in a language or not" (p. 310). As evidence, Dahl (2000a) contrasts the following two weather forecasts one from English (strong-FTR) and the other from Finnish (weak-FTR):

(1.9) a. English
Outbreaks of rain will clear on Monday to leave a mix of sunshine and showers across the country. Longer periods of rain are likely midweek, especially in the west. It will be mostly cool and windy. Cool and unsettled conditions over much of Scandinavia will extend into central and western Europe during Tuesday and Wednesday. Mediterranean coasts will remain sunny and very warm.

b. Finnish
Sää kylmenee, mutta keskiviikkona tuulee idästä ja pyryttää lunta. Lämpötila kohoaa tilapäisesti nollaan tai jopa vähän suojan puolelle. Torstain tienoilla voi olla jopa kymmenisen pakkasastetta. Viikonlopulla taas lauhtuu, pilvistyy ja alkaa sataa lunta.
(Rather literal translation:) 'The weather becomes cooler, but on Wednesday it blows from the east and there is drifting snow. The temperature rises temporarily to zero or even a little higher. By Thursday it can already be around ten degrees below zero. During the weekend it again becomes milder, overcast and begins to snow.' (2000, pp. 310–11, emphasis added)

In English *will* is systemically used throughout the text. In contrast, the Finnish example is entirely in the present tense.[15] Dahl concludes that—arguably—FTR is grammaticised "in English in a way that it is not in Finnish" (2000a, p. 311). Dahl is careful to say nothing on *what* this implies about English FTR grammaticisation. Fleshing out this question is a major goal of this thesis.

In summary, Dahl identifies a group of "futureless" European languages in Northern Europe on the basis of whether the future tense is obliged in prediction-based main-clause FTR statements. The futureless zone includes, "at least all the Finno-Ugrian and Germanic languages except for English" (Dahl, 2000a, p. 326). Dahl (2000a) argues that this is a reasonable proxy for FTR grammaticisation in general. Chen essentially follows Dahl (2000a) by defining weak-FTR languages as those languages which, like Finnish, do not require the

[15] Though temporal and modal notions are expressed using the modal *voi* 'can' and the verb *tulla* 'to become'.

future tense in main clause sentences in prediction-based contexts, and strong-FTR languages as any language which is not weak-FTR (Chen, 2013). Chen's typological project is then to use the literature (notably Bybee et al., 1994; Cyffer, Ebermann, & Ziegelmeyer, 2009; Dahl, 1985; Dahl & Kós Dienes, 1984; Nurse, 2008, but see Chen (2013) for details on sources for African languages) to extend Dahl's distinction beyond the languages in EUROTYP. This resulted in a list of $N = 129$ languages marked in terms of whether they are weak or strong-FTR (i.e. FTR status) (Chen, 2013).

Criticisms of FTR status in the literature: I have spent time explaining FTR-mode (schedules, intentions, predictions) in detail here for two reasons. Firstly, FTR-mode motivates my empirical approach to measuring cross-linguistic differences in FTR (Chapter 2). Secondly, Chen's typological project has been criticised because of misunderstandings of his criteria. For instance Pullum (2012), Pereltsvaig (2011), and McWhorter (2014, p. 98) raise examples where English permits the present tense for FTR in intention and scheduling contexts. However, such examples do not undermine FTR status since the criteria specifically delineates marking in prediction-based FTR as the class criterion. In fact, present tense use for schedule-based FTR may be a nearly meaningless cross-linguistic distinction, given how cross-linguistically common it is for languages to permit such constructions (see above, and Dahl, 2000a).

Another common criticism goes along the lines of: *Language x should not be considered strong-FTR because its future tense exhibits complex (modal, aspectual) semantics.* For instance, McWhorter (2014) argues that Russian—which Chen classifies as strong-FTR—does not have a future tense, because Russian FTR is often marked with an aspectual marker (p. 98). Similarly, Pullum (2012) argues that English should not be classed as strong-FTR because *will* is a modal auxiliary, and not a "future tense". I defer to McWhorter on the details of Russian grammar, but it is arguable that Russian does have future time marking of some sort. Dahl (1985, p. 110) classifies the Russian future copula *budet* as being an edge case of the FUTURE TENSE, since it is restricted to being used with imperfective verbs. This implies *budet* is characterised by a division of labour between aspectual and temporal semantics, whereas *will* is characterised by a division of labour between modal and temporal semantics (Enç, 1996; Huddleston, 1995; Fries, 1956). The idea that such complexities mean *budet* and *will* do not therefore qualify as exemplars of the FUTURE TENSE falls prey to essentialist reasoning where language-specific future tenses are compared to an idealised FUTURE TENSE and inevitably found lacking. These arguments exemplify the one-to-one-mapping fallacy, i.e. the idea that each unique meaning should be mapped to a unique form. The cross-linguistic data consistently fail to uphold this assumption (e.g. Dahl, 2000b). In fact, rather than an idealised prototype, cross-linguistic data indicate the FUTURE TENSE is a fuzzy, imprecise category which is (a) never fully realised in any language-specific future tense, and (b) characterised by division of labour between modal, aspectual, and temporal semantics (Dahl, 2000a). To point out such complexities with regard to a particular future tense only illustrates its prototypicality in this regard.

A final criticism is worth forestalling. Several commentators have pointed out that a more meaningful distinction than Chen's is whether a language has a bound, morphological future tense (e.g. Pereltsvaig, 2011). For instance, French verbs (and verbs in Romance languages generally) are marked with a bound morpheme when they refer to the future (Chen, 2013):

(1.10) French

a. *Il fait froid aujord'hui.*
It do/make:PRS cold today.
'It is cold today.'

b. *Il fera froid demain.*
It do/make:FUT cold tomorrow.
'It will be cold tomorrow.'

Grammars like this, since they mark the future with a more strongly grammaticised morphological marker and not a free periphrastic one like the English *will*, might more naturally lead to the kinds of effects Chen hypotheses, the argument goes (Pereltsvaig, 2011).

However, such lines of argumentation overlook the fact that, cross-linguistically, future tenses are approximately equally spit between periphrastic and morphological grammatical elements (Dahl, 2000a). There is no criteria by which to define the more basic category. While it is plausible to privilege morphosyntax—morphosyntactic elements are more grammaticised than lexical paraphrastic ones—it seems unclear, from a practical perspective, why it should matter whether FTR is marked with a bound morpheme or a periphrastic auxiliary. Surely whether *some* marking is obligatory is more likely to engender cross-linguistic cognitive effects.

In conclusion, I do not suggest there will not be language-specific complications in the application of FTR status (see Dahl, 2013), a point which Chen is careful to acknowledge (Chen, 2012). However, Chen's (2013) creation of FTR status is based on close reading of the relevant typological literature, and many criticisms either do not understand FTR-status or misrepresent modern research on the FUTURE TENSE. The simple classification of a hugely complex area is bound to sacrifice granularity to Chen's larger project which requires a binary description of FTR grammar. However, there is one critical issue with Chen's (2013) creation of FTR status, which I address in the next section.

The issue with FTR status which is addressed by this thesis: A different critical shortcoming needs to be mentioned: Chen (2013) simplifies Dahl's criteria (at least insofar as English is concerned) in one key way. Taking the contrast between English and Finnish given in example (1.9), Dahl points out that *will* is required, "unless there is another modal expression in the sentence" (Dahl, 2000a, p. 311). In a critical simplification, Chen (2013) asserts that English predictions oblige the use of the future tense. However, English does not oblige the future tense for prediction-based FTR. Rather, it obliges the future tense or another modal verb:

(1.11) a. *It will rain.*
b. *It is going to rain.*
c. *It shall rain.*
d. *It could rain.*
e. *It may rain.*
f. *It might rain.*
g. *It should rain.*
h. *It would rain.*
i. *It must rain.*

This means that—insofar as English is concerned—prediction-based main clauses must be marked not with the future tense but with a modal verb. When these encode epistemic modality, they are weak modals. Specifically, *could, may* and *might* encode weak epistemic modality (Nuyts, 2000); *should* encodes slightly stronger (but still weak) epistemic modality (Nuyts, 2000); *would* may have peripheral weak epistemic uses but mostly encodes counterfactual notions (Nuyts, 2000); and *must* generally encodes deontic modality. When used in prediction-based contexts it manages to convey bouletic notions of desire, e.g. gazing sorrowfully at his parched crops, a farming might say, *It must rain tomorrow.* Therefore, the core epistemic modals *could, may, might*, and *should* encode weak modality which may map onto notions of low probability (Lassiter, 2015). The basic contribution of this thesis is to explore the consequences of this simple insight to the relationship between FTR grammar and intertemporal decision making.

1.3.2.2 Chen's arguments about future tense semantics

The second part of Chen's cross-linguistic project is an argument about the semantics of the FUTURE TENSE. I hope to demonstrate his arguments are neither structurally sound or grounded in good evidence.

Chen (2013) argues that the notional domain encoded by the FUTURE TENSE has everything to do with *timing*, noting that FTR status indexes "*when* [languages] require speakers to specify the timing of events or when timing can be left unsaid" (p. 3), and that languages like English force speakers to "habitually divide time between the present and the future" (2013, p. 3). This account treats the FUTURE TENSE as a tense *sensu stricto*, despite the noted debate which hinges on whether the FUTURE TENSE is a tense at all, since it tends to encode modal, temporal, and aspectual notions (Dahl, 2000a). Chen (2013) hypothesises two ways that obligatory tense marking of FTR is predicted to impact intertemporal decision making. The first is that the FUTURE TENSE encodes notions of temporal distance, the second is that it encodes notions of temporal precision. Together, I refer to these temporal mechanism as the "temporal hypothesis". I outline them in turn before providing critical perspectives.

The temporal-distance mechanism: Chen (2013) first argues the FUTURE TENSE encodes notions of temporal distance. As evidence, he notes that the English present tense is often used to convey immediacy, for instance through the use of the "historical present"—where writers refer to past events using the present tense. He notes this is used to make stories more "vivid and immediate", pointing out that we tend to tell jokes in the present tense rather than the past: It is always "a man *walks (not walked)* into a bar" (Chen, 2013, p. 5). Use of the present tense for FTR is therefore hypothesised to cause future events to be construed as more temporally proximal. This seems plausible, but an important point is that Chen (2013) makes a claim about the semantics of the PRESENT TENSE in weak-FTR languages on the basis of solely English (strong-FTR) examples. It is hard to see how this is justifiable, given the fact that FTR status indexes differences in FTR grammar between weak and strong-FTR languages. It is also unclear why use of the historical present to convey (past time) immediacy would map onto the semantics of future time.

As a further line of argumentation, Chen (2013) notes that past tenses are sometimes used to convey the *modally* distant irrealis (counterfactual) mood—e.g. *I wish I had a car.* He cites Mezhevich (2008) as evidence for the (less common) use of the FUTURE TENSE to also convey modally distant irrealis notions. Here, it should be noted that Mezhevich (2008) is a treatment of tense and mood in Russian and Hebrew, so does not provide much evidence about the FUTURE TENSE. Regardless, Chen (2013) argues that since the FUTURE TENSE is sometimes used to express *modally* distant notions of irrealis, it may also encode *temporal* distance. Here, Chen (2013) makes a conceptual jump from modal to temporal semantics. This is not necessarily ungrounded. It seems reasonable that modal distance might be cross-modally mapped to temporal distance (Trope & Liberman, 2010). However, it is more parsimonious to interpret this as indicating the FUTURE TENSE encodes notions of modal distance. Regardless, on the basis of these arguments, Chen (2013) hypothesises that speakers of weak-FTR languages who commonly refer to the future with the (temporally immediate) PRESENT TENSE should construe future events as more temporally proximal, or conversely that speakers of strong-FTR languages who commonly refer to the future using the (temporally distant) FUTURE TENSE should come to construe the future as more distal.

The temporal precision mechanism: Similarly to the temporal-distance mechanism, Chen (2013) hypothesises the FUTURE TENSE encodes temporal precision. First, Chen (2013) suggests that people may have distributional representations of the timing of future events. For instance, on hearing "in two weeks" we represent "two weeks" as a fuzzy distribution around the objective date, such that the spread of this distribution might be different between individuals. He then hypothesised that weak-FTR speakers should have less precise temporal beliefs. In support, he makes an argument based on relativity work which found that Russian speakers were faster at discriminating between light and dark blue (mentioned above, Section 1.2.3.2). Chen (2013) characterises such findings as indicating Russian speakers divide the colour spectrum more precisely. He argues that strong-FTR languages analogously divide temporal space more finely (past, present, future) than weak-FTR languages (past, present). If this affected temporal perception in analogous ways to colour category terms in Russian, weak-FTR speakers might have less precise temporal beliefs. A weakness of this argument is that strong-FTR languages do not necessarily have a past tense, e.g. they could have a future vs. non-future system. Therefore, strong-FTR languages do not necessarily divide temporal space into finer gradations than weak-FTR languages. Additionally, faster discrimination across category boundaries (i.e. present vs. future) does not necessarily entail differences in the precision of beliefs within category, which is what Chen's (2013) argument amounts to. Regardless, Chen (2013) argues that weak-FTR speakers should construe future events with less temporal precision than strong-FTR speakers.

The temporal hypothesis: If either mechanism (distance, precision) were true, this would lead to increased temporal discounting. The argument goes that if the FUTURE TENSE encodes temporal distance, strong-FTR speakers will come to construe future events as more distal, and therefore to temporally discount more. A similar effect is predicted by the hypothesised precision effects, see Chapter 2, *Fig. 2.1.* If strong-FTR speakers temporally discount more, they should be less likely to save money, eat healthy foods, exercise, etc.

Critical perspectives on Chen's account of FUTURE TENSE semantics: Chen's ideas about temporal distance and precision are plausible, but they do not seriously engage with linguistic scholarship on the semantics of the FUTURE TENSE. Here, I raise two criticisms: First, FTR status may itself be a determinant of cross-linguistic differences in future tense semantics; and second, future tenses tend to encode notions not of temporal distance or proximity but of modal high certainty, which may have consequences for discounting.

I first address the idea FTR status may itself be a determinant of cross-linguistic differences in the semantics of the future tenses. One way to understand the semantic role of a linguistic unit is to analyse it paradigmatically, in terms of oppositional contrasts with the other linguistic units by which it can be replaced in a given utterance. For instance, in example (1.11), the future tense markers encode the highest certainty relative to the other modals. However, this is not the case in Dutch, where the present tense is acceptable:

(1.12) Dutch

a. *Het zal morgen regenen*
It will tomorrow rain:INF
'It will rain tomorrow.'

b. *Morgen regent het*
Tomorrow rain:INF it
'It rains tomorrow.'

c. *Het kan morgen regenen*
It may tomorrow rain:INF
'It may rain tomorrow.'

d. *Het moet morgen regenen*
It must tomorrow rain:INF
'It must rain tomorrow.'

It is difficult to say whether the future tense (example [1.12a]) or the present tense (example [1.12b]) encode higher or lower certainty. The paradigmatic role of the Dutch future tense is therefore affected by FTR status. In Chapter 3, I explore whether the English and Dutch future tense use differentially relates to intertemporal decision making as a result of these paradigmatic differences. This complexity is not addressed in Chen (2013).

The second critical objection is that Chen's (2013) arguments about FUTURE TENSE semantics overlook widely-noted tendencies of the FUTURE TENSE to encode modal high certainty when viewed in the context of paradigmatic oppositions with other modal FTR constructions. For instance, in examples (1.11a-c), the future tenses convey that a speaker is highly certain about the outcome. In examples (1.11d-g), the other epistemic modals convey low certainty. We might therefore predict that use of high-certainty future tenses would engender high-certainty construals of future events in strong-FTR speakers. Given that people devalue risky outcomes in processes analogous to temporal discounting (L. Green & Myerson, 2004), future tense use would be associated with decreased rather than increased discounting. This is the opposite to Chen's prediction, and is tested in Chapter 3 and Appendix G.

Finally, I note that an analysis of English FTR reveals that *will, be going to*, and *shall* do not encode temporal distance or precision. These notions are linguistically expressed through the use of lexical temporal adverbials, *in two weeks, tomorrow, next year,* etc. The use of *will* in *It will rain tomorrow* does not change anything about the temporal distance or precision of the referenced event.

1.3.2.3 Conclusions

Chen's (2013) establishment of cross-linguistic differences involved two aspects, a cross-linguistic typology and two arguments about the semantics of the FUTURE TENSE. The typology—FTR status—divides languages into strong- and weak-FTR based on whether they oblige a future tense for prediction-based main-clause FTR statements. Despite simplifying a complex area, FTR status appears well-grounded in FTR typology with the critical exception that—as far as English is concerned—it disregards modal FTR structures. Chen (2013) argues that the FUTURE TENSE either encodes temporal distance or precision. Regarding temporal distance, I demonstrate that conclusions are based on a limited language sample, shaky analogies with past time reference semantics, and the dubious notion that cross-linguistically limited use of the future tense to mark irrealis cross-modally maps onto construals of temporal distance. Regarding temporal precision, I demonstrate that conclusions are based on analogies with linguistic relativity work on categorical colour perception which do not withstand close scrutiny. Critically, these argument disregard extensive contemporary work on the FUTURE TENSE which revolves around its *modal* semantics. Nonetheless, I have reported Chen's (2013) arguments in their own terms as they form an important counterpoint to the empirical work of this dissertation. In the next section, I address the second stage of Chen's (2013) project, which involves testing whether obligatory future tense use impacts intertemporal behaviour.

1.3.3 Chen's establishment of cross-linguistic effects on behaviour: Psychological discounting

The next step of Chen's (2013) project is to use FTR status to predict a number of behaviours related to temporal discounting. This accords with steps (b) and (c) in the relativity framework I use: (b) the establishment of behavioural correlates to cross-linguistic differences; and (c) the development of causal evidence linking cross-linguistic predictors to behavioural outcomes. Before describing these aspects of Chen's (2013) project, it will be useful to explain a critical concept: psychological discounting. The non-linguistic behavioural outcomes identified in Chen (2013)—and in this thesis—involve psychological discounting, so understanding it is crucial.

1.3.3.1 A brief primer on psychological discounting

Psychological discounting refers to a general phenomenon whereby people tend to devalue outcomes according to various factors: The two factors focussed on in this thesis are probability and delay (on other factors, see Białaszek, Ostaszewski, Green, & Myerson, 2019). "Discounting" refers to the extent to which some outcome is devalued as a function of either delay or probability. As I have explained, "temporal discounting" refers to the subjective devaluation of an outcome as the time until it will occur grows longer. Similarly, "probability discounting" refers to the subjective devaluation of an outcome as the probability of it occurring grows lower (L. Green et al., 2014; Rachlin, Raineri, & Cross, 1991).

In the measurement of psychological discounting, outcomes are usually operationalised using hypothetical monetary transactions. Participants are given repeated choices between various monetary rewards offered under differing delays and probabilities. To measure temporal discounting, participants are asked to make repeated hypothetical intertemporal decisions, usually between immediate and delayed rewards, e.g. \$20 now vs. \$25 in two weeks. To measure probability discounting, participants are asked to make repeated hypothetical risk-preference decisions, usually between certain and risky rewards, e.g. a 100% chance of \$10 vs. a 80% chance of \$20, see *Fig. 1.1.* As with delayed rewards, people discount risky rewards: Given the option between a guaranteed \$100, and a 50% chance of \$100, most go for the sure thing.

A useful concept to understanding the measurement of discounting is the *indifference point*, i.e. the point at which someone is *indifferent* in a choice between two rewards (L. Green & Myerson, 2004). For instance, if a participant is indifferent between \$10 now and \$20 in a month, the subjective value—the indifference point—of the delayed \$20 is approximately \$10. If the discounted utility of the \$20 were higher than \$10, the participant would chose to wait. A similar framework can be used to understand *probability* discounting. In this context, the concept of an indifference point is entirely the same, but rather than capturing temporally discounted subjective value, it captures probabilistically discounted subjective value.

There are individual differences in psychological discounting. Given the choice between an immediate \$10 and delayed \$20, some wait and others do not. Such differences arise as a function of whether an individual's subjective valuation of the delayed \$20—their indifference point—is above or below \$10 (L. Green & Myerson, 2004). There are also individual differences in probability discounting: Given a choice between a guaranteed \$100, and a 50% chance of \$200, some will choose the former and others the later, dependent on whether their subjective valuation of the risky \$200—their indifference point—is more or less than \$100 (L. Green & Myerson, 2004).

Discounting is non-linear over both time and probability: The steepest discounting rates occur in nearer times and higher probabilities, and grow more shallow over more distal times and lower probabilities (L. Green & Myerson, 2004). Various theoretical functions have been suggested to best approximate real discounting rates, including hyperbolic (Mazur, 1987), exponential (Samuelson, 1937), and hyperboloid functions (L. Green & Myerson, 1996). It appears that the last—which is a hyperbolic function raised to a power—best describes empirical results (for general reviews see L. Green & Myerson, 2004; L. Green et al., 2014; McKerchar & Renda, 2012). All these functions are convex.

While psychological discounting is usually measured using hypothetical monetary rewards, the phenomenon is perfectly general: It has been observed in non-human animals

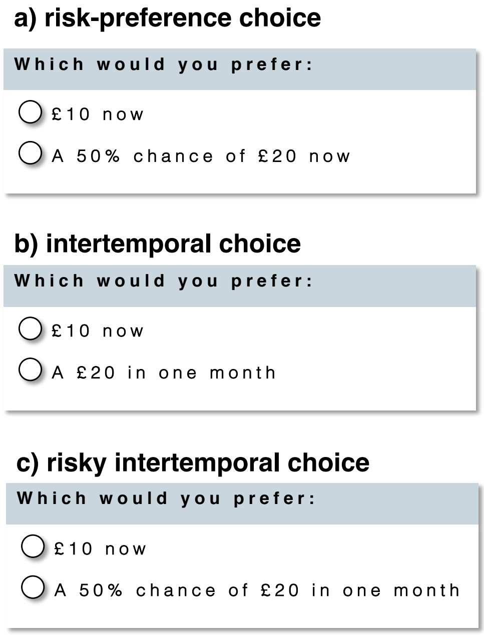


Figure 1.1: Some examples of my approach to measuring (risky) intertemporal decision making: a) a simple risk-preference choice between two immediate options awarded under different certainties, b) a simple intertemporal choice between two rewards awarded at different delays; c) a risky-intertemporal choice between two options, one certain and immediate, the other risky and delayed. I take the view in this thesis that such risky-intertemporal choice paradigms are more representative of real-world intertemporal decision making than either *a* or *b*.

(e.g. Addessi, Paglieri, & Focaroli, 2011; Ainslie & Herrnstein, 1981; Freeman, Green, Myerson, & Woolverton, 2009; Freeman, Nonnemacher, Green, Myerson, & Woolverton, 2012; J. R. Stevens, Halliman, & Hauser, 2005; J. R. Stevens & Mühlhoff, 2012), in cross-cultural studies of humans (Du, Green, & Myerson, 2002), in monetary transactions involving both gains and losses (Cox & Dallery, 2016; Murphy, Vuchinich, & Simpson, 2001), for both hypothetical (Rachlin et al., 1991; L. Green, Myerson, & Ostaszewski, 1999), real (Hinvest & Anderson, 2010), and non-monetary rewards (e.g. vacations (Raineri & Rachlin, 1993), alcohol and food (Odum & Rainaud, 2003), and social policy outcomes such as the legalisation of gay marriage (Wetherly, Plumm, & Derenne, 2011)). Both probability and temporal discounting rates have trait-like qualities: Rates were stable over moderate (three month) periods (Ohmura, Takahashi, Kitamura, & Wehr, 2006).

Additionally, hypothetical and real reward discounting were significantly correlated (Hinvest & Anderson, 2010), or were indistinguishable (Locey, Jones, & Rachlin, 2011; Madden, Begotka, Raiff, & Kastern, 2003), which suggests decisions involving hypothetical rewards are a valid measurement paradigm. Moreover, hypothetical rewards significantly correlated with real behaviour in which intertemporal decision making was involved, including real spending (Bickel et al., 2010), and financial outcomes such as income levels (Hamilton & Potenza, 2012; Xiao & Porto, 2019) and financial mismanagement (Hamilton & Potenza, 2012; Xiao & Porto, 2019). Experimentally-measured temporal discounting also predicted various health-critical behaviours such as alcohol abuse (Vuchinich & Simpson, 1998), opioid dependency (Garami & Moustafa, 2019), substance abuse in general (Mejía-Cruz, Green, Myerson, Morales-Chainé, & Nieto, 2016; Kirby, Petry, & Bickel, 1999), odds of smoking cigarettes (Bickel, Odum, & Madden, 1999), and the likelihood of exercising in older individuals (Tate et al., 2015). It is also a marker of numerous mental health disorders including social anxiety disorder (Rounds, Beck, & Grant, 2007), bipolar disorder (Hart, Brown, Roffman, & Perlis, 2019), schizo-affective disorder (Hart et al., 2019), and schizophrenia (Hart et al., 2019). In this thesis, I focus on investigating the impacts of cross-linguistic differences on psychological (temporal, probability) discounting. This is contrasted with observational work involving real-world outcomes where intertemporal decisions are involved.

There is an an extensive literature on the differences between *probability*, and *temporal* discounting. For instance, reward amount has different effects on delayed and risky outcomes. Participants discount large delayed rewards less, but large risky rewards more, i.e. they wait for large delayed rewards but do not gamble on large risky rewards (L. Green et al., 1999). Additionally, risk preferences may involve a simple preference for certainty when one option is certain and another awarded under risk, which leads to inconsistent outcomes compared to when a choice is between two risky options (Andreoni & Sprenger, 2012). This suggests that certainty may play a role in generating preference for present rewards if delayed rewards are interpreted as risky (see Andreoni & Sprenger, 2012). Such concerns are not taken up in this thesis. Because of the need to focus on the measurement of cross-linguistic differences and psychological discounting, I necessarily take a simplified approach to both. Despite noted differences, similar (enough) processes appear to govern temporal and probability discounting; for instance, both are well-described by similar discounting functions (L. Green & Myerson, 2004).

Most real-world intertemporal decisions are usually made under some degree of uncertainty. The future is unknown: Someone deciding whether to smoke a cigarette must balance the pleasurable reward against potential future health costs discounted both by delay and probability, i.e. *when* and *if* she may develop smoking-related health problems. Even returns on investment in cash savings are contingent on imponderable externalities involving inflation rates. In fact, previous experience of high inflation rates affected putative (strictly) intertemporal decision-making outcomes in Polish (Ostaszewski, Green, & Myerson, 1998) and Argentine (Macchia, Plagnol, & Reimers, 2018) participants. This suggests that people factor in previous experience with future risk when making “intertemporal” decisions. In light of these considerations, a growing number of researchers focus on discounting when rewards are both delayed and risky (Vanderveldt, Green, & Myerson, 2015; Vanderveldt, Green, & Rachlin, 2017; Yi, de la Piedad, & Bickel, 2006). Such "risky intertemporal" decisions are measured by having participants choose between immediate, certain rewards, and delayed, risky ones, e.g. $10 now or a 50% chance of $40 in three months, *Fig. 1.1.* Growing evidence attests to the utility of such approaches. For instance, probability and delay multiplicatively interacted in risky intertemporal decision making outcomes (Vanderveldt et al., 2015). This indicates delay discounting is contingent on the probability associated with an outcome and *vice versa.* Additionally, models trained on both risky and intertemporal choices fitted data better than models trained on either (just) risky or (just) intertemporal choices (Luckman, Donkin, & Newell, 2018). This suggests that both may be special cases of a unified "risky intertemporal" decision-making process (Luckman et al., 2018). In this thesis, I make the case that real-world intertemporal behaviour tends to involve decision making under risk. Risky-intertemporal-choice paradigms therefore exhibit good representative design and are suitable for understanding how cross-linguistic differences in FTR grammar affect real-world (risky) intertemporal behaviour (Chapter 4). I used the term "combined discounting" or "risky intertemporal outcomes" to refer to discounting of outcomes which are both delayed and risky.

A critical point is that "subjective value" and "discounting" are inversely related: *More* discounting entails *lower* subjective value. In intertemporal-choice tasks, increased discounting means a participant will make fewer "future-oriented" choices to wait for a delayed reward (which they perceive as less valuable) and more "present-oriented" choices to opt for an immediate reward; decreased discounting means a participant will make more future-oriented choices and fewer present-oriented ones. Accordingly, in risk-preference tasks, more discounting means a participant will make fewer risky choices (which they perceive as less valuable) and more safe ones; decreased discounting means a participant will make more risky choices and fewer safe ones. I refer to such individual differences in terms of time and risk preferences. At different points in this thesis, I analyse measures of both subjective value and discounting. It is relatively straight-forward to analytically derive a single measure of discounting in strictly risk- or time-preference decisions (see Mazur, 1987; Kirby et al., 1999). However, it is more difficult in risky-intertemporal decisions (Vanderveldt et al., 2015). Because of this, I analyse the indifference points directly for risky-intertemporal outcomes. This means that the direction of the effects of linguistic predictors will reverse between Chapters 3 and 4. I apologise in advance for the increased cognitive load this may cause.

1.3.3.2 The effects of FTR status on intertemporal behaviour

Recall, FTR status indexes whether a language obliges the future tense for prediction-based FTR (strong-FTR), or allows present tense constructions to be used (weak-FTR). Chen (2013) hypothesises that strong-FTR speakers should temporally discount more, because they either construe future events as more distal, or more precise. This entails they should engage in more present-oriented behaviour, i.e. not saving for the future, exercising less, etc. To test this, Chen (2013) conducted a number of statistical analyses using FTR status as a predictor of such outcomes. Outcome variables were taken predominantly from the World Values Survey, which is an ongoing, large-scale data gathering effort which has been running since 1981 and comprises self-reported survey responses to a range of questions. Respondents comprise nationally-representative samples of people from more than 100 countries around the world (World Values Survey Association, 2021). Among other things, respondents are asked whether they have saved money in the past year. This is Chen's (2013) principal outcome. Health outcomes and national savings rates were taken from other data sources (see Chen, 2013).

Evidence for the effect of FTR status on intertemporal behaviour: Chen's (2013) results are impressive. First, he used cross-country regression to estimate the impact of FTR status on probability of having saved in the last year. As he had predicted, people residing in weak-FTR-speaking countries were 46% more likely to have saved. However, cross-country regressions estimate the effect of FTR status across different countries. As such, they leave open the possibility that cultural differences are confounding results. Chen (2013) therefore supplemented these analyses with within-country regressions. These involved estimating the effect of FTR status within countries which have multiple national languages, e.g. Switzerland, where both French (strong-FTR) and German (weak-FTR) are spoken. Such methods address cultural confounding, since people residing in the same country arguably share a cultural environment, despite speaking different languages. Chen (2013) compared matched samples of individuals from: Belgium, Burkina Faso, Ethiopia, Estonia, The Democratic Republic of the Congo, Nigeria, Malaysia, Singapore, and Switzerland. Both weak- and strong-FTR languages are spoken in these countries. He found that weak-FTR speakers were 30% more likely to have saved money in any given year, had accumulated 39% more wealth by retirement, were 24% less likely to smoke, were 29% more likely to be physically active, and were 30% less likely to be physically obese (Chen, 2013). In a further analysis of national savings outcomes, he found that weak-FTR-speaking countries saved on average 6% more of their GDP each year. All these findings included statistical controls for numerous potential confounds, and many used a variety of versions of the FTR status variable (see Chen, 2013).

Further evidence: Most evidence for the effect of FTR status has emulated Chen's (2013) approach, using FTR status as a predictor of country-, organisation-, and individual-level behaviour in which time preferences are plausibly involved. For instance, German (weak-FTR) speaking households were found to save more, smoke less, and resort to credit less than French (strong-FTR) speaking households in Switzerland (Guin, 2017). In a sample of children of immigrants in Florida, students from weak-FTR-speaking families had better educational outcomes, as indicated by math and reading scores, graduation, absenteeism, retention, and disciplinary incidents (Figlio, Giuliano, Özek, & Sapienza, 2016). In a globally-balanced sample, speakers of weak-FTR languages self-reported they were more likely to engage in positive and negative reciprocity, as well as be more patient, more altruistic, and more trusting (Falk et al., 2018). Countries in which the predominant language is weak-FTR also had higher suicide rates, while weak-FTR-speaking individuals self-reported higher acceptance of suicide and euthanasia (Lien & Zhang, 2020). These outcomes are predicted because weak-FTR speakers are hypothesised to discount future suffering less, e.g. if one believes future suffering will be excruciating, this would result in higher support for legalised suicide and/or euthanasia.

FTR status has also been used to predict organisation-level phenomena. For instance, corporations headquartered in weak-FTR-speaking countries held larger cash reserves (S. Chen, Cronqvist, Ni, & Zhang, 2017), potentially an example of saving for the future. Firms headquartered in Hong Kong accumulated more cash holdings than a control group of South Asian firms after the predominant language switched from English (strong-FTR) to Chinese (weak-FTR) in 1997 (S. Chen et al., 2017). Similarly, firms whose official language was weak-FTR invested more in research and development in both an international sample and within-country in multilingual Belgium and Switzerland (Chi, Su, Tang, & Xu, 2018; Liang, Marquis, Renneboog, & Sun, 2018). Likewise, Hong Kong-based firms were found to invest more in research and development after 1997 (Chi et al., 2018). Within the US, firms whose Chief Executive Officer (CEO) was from weak-FTR-speaking countries invested more in research and development (Chi et al., 2018). Investment in research and development is suggested to be a future-oriented behaviour because present expenditure yields future profitability. Finally, weak-FTR negatively predicted present-orientated earnings management practices at the firm and country level, and as a function of CEO nationality in the US (J. Kim, Kim, & Zhou, 2017), though this result did not replicate when controls for linguistic history were added (Gotti, Roberts, Fasan, & Robertson, 2020).

FTR status also predicts country-level outcomes. Hübner and Vannoorenberghe (2015a) conducted a two-stage instrumental variables analysis, first regressing a measure of patience on FTR status, and then regressing a variety of macro-economic outcomes on the predictions of the first regression. Such analyses help rule out confounding by statistically simulating an exposure effect (Lousdal, 2018; Sovey, 2011). In this literature, future orientation is framed in terms of "patience", where higher patience is synonymous with higher future orientation, i.e. because patient individuals are more likely to wait for future rewards. Via measures of patience, weak-FTR-speaking countries had better economic growth, as indicated by output per worker, patents, innovation performance, and capital stock (Hübner & Vannoorenberghe, 2015b). Using the same methods to estimate inflation rates, Hübner and Vannoorenberghe (2015a) found weak-FTR countries had lower inflation rates. This suggests weak-FTR speakers are more patient and therefore perceive the long-term costs of inflation as higher (less discounted). As a function of such tendencies, administrators in these countries may choose future-oriented monetary policies which minimise long-term costs rather then present-oriented policies which use inflation to temporarily increase employment and output (Hübner & Vannoorenberghe, 2015a).

Climate change policy decisions can also be construed as intertemporal choices, because climate change mitigation efforts incur short-term costs but confer long-term benefits. A number of studies have found that weak-FTR speakers tend to support future-oriented climate policies. For instance, weak-FTR-speaking individuals had higher support for climate change mitigation policies (Mavisakalyan, Tarverdi, & Weber, 2018), higher likelihood of personally having adopted environmentally responsible behaviour (Mavisakalyan et al., 2018), and more pro-environmental attitudes (S. Kim & Filimonau, 2017; Pérez & Tavits, 2017). Weak-FTR-speaking countries were also more likely to have taken country-level climate change mitigation action (Mavisakalyan et al., 2018).

If the correlation identified in these studies is not spurious, it would constitute profound new ground for the linguistic relativity hypothesis. These findings provide a clear counterpoint to criticisms which view relativity results as trivial (e.g. Section 1.2.2). If the correlation is real, cross-linguistic differences affect behaviour far beyond tightly-controlled laboratory experiments.

1.3.3.3 Problems and issues with this body of evidence: Risk as a confounding factor

A important caveat is that most evidence for the effect of FTR status on intertemporal behaviour is correlational. It does not conclusively demonstrate causality. Research involving linguistic predictors is notorious for producing spurious relationships due to non-independence between languages resulting from historical processes of cultural inheritance which see causally unrelated bundles of traits being passed on from antecedent to descendant cultures (Roberts & Winters, 2013). In fact, when linguistic history controls have been added, results were mixed (Roberts, Winters, & Chen, 2015), or failed to replicate (Gotti et al., 2020). This suggests reported correlations could be spurious. Despite this, many analyses do not include controls for linguistic history, (the following do not Chen, 2013; Chi et al., 2018; Falk et al., 2018; Fasan, Gotti, Kang, & Liu, 2016; Figlio et al., 2016; Hübner & Vannoorenberghe, 2015b, 2015a; Liang et al., 2018; Mavisakalyan et al., 2018; Pérez & Tavits, 2017; Zhu, Hu, Wang, & Zheng, 2020). These findings should therefore be interpreted with caution.

The crucial issue addressed by this thesis is that risk may be an overarching confound. Most real-world intertemporal decisions involve some degree of risk. Much support for Chen's hypothesis therefore involves real-world intertemporal decision making involving risk. Even seemingly risk-free outcomes usually involve some degree of uncertainty. For instance, the pursuit of educational goals is fraught with uncertainty about relative rate of return (Figlio et al., 2016), the discounting of future suffering in the context of support for euthanasia is permeated with uncertainty about the relative extent of future distress (Lien & Zhang, 2020), and deciding to undertake earnings management practices involves estimating the probability of being caught (Fasan et al., 2016; J. Kim et al., 2017). Chen (2013) predicted whether survey respondents had saved in the past year, but the survey question did not specify whether savings involved investment in risky assets such as stocks and shares (World Values Survey Association, 2014). Other studies reviewed above are susceptible to such criticisms as well. Even seemingly strong experimental results may be confounded by risk. For instance, Pérez and Tavits (2017) randomly assigned 1200 Russian-Estonian bilinguals to complete a survey in either Russian (strong-FTR) or Estonian (weak-FTR). Participants who completed the survey in Estonian were more likely to express support for a hypothetical environmental tax; estimating the future benefits of such a tax involves risk and/or uncertainty. Much of the empirical support for the temporal hypothesis therefore involves not intertemporal decision making, but risky intertemporal decision making. Critically, such outcomes are affected by beliefs about time and risk (Vanderveldt et al., 2015).

In fact, as far as I know, only two studies have tested Chen's (2013) temporal hypothesis using experimentally-controlled intertemporal-choice paradigms which strictly measure temporal discounting. Results have been mixed. In one study, German-speaking (weak-FTR) and Italian-speaking (strong-FTR) school children completed an intertemporal-choice task in bilingual South Tyrol (Sutter, Angerer, Glätzle-rützler, & Lergetporer, 2015). As predicted, the German cohort made relatively more future-oriented choices. In another study, speakers of 5 different languages completed an intertemporal-choice task (Thoma & Tytus, 2018). This time, weak-FTR-speaking participants were less future-oriented (Thoma & Tytus, 2018). While the authors attribute this to differences between hypothetical and real rewards, this is not supported by studies which report little to no difference between real and hypothetical rewards, and strong correlations between hypothetical discounting and real behaviour (Bickel et al., 2010; Hinvest & Anderson, 2010; Locey et al., 2011; Madden et al., 2003).

Critically, in addition to the dependent variables, the independent variable used in these studies—FTR status—may also be confounded by risk. As I have pointed out, English does not oblige the future tense, but rather obliges a modal verb. Most of these encode low certainty (see example [1.11]), which is suggested to map onto notions of low-probability (Lassiter, 2015). If strong-FTR languages oblige speakers to use low-certainty FTR constructions, this could impact speakers' beliefs about future risk, causing them to construe future events as riskier. Given that people probability discount risky outcomes, this would cause them to make fewer future-oriented choices. This means that there is an alternative account for how FTR grammaticisation impacts (risky) intertemporal decisions: by causing speakers to be less certain about the future. This thesis is predominantly concerned with the development of this line of thinking. I refer to it as the modal hypothesis.

1.3.4 Critical summary with mitigation strategies adopted in this thesis

These, then, are the goals of this thesis: (a) to test the hypothesis that cross-linguistic differences in FTR grammar impact speakers' (risky) time preferences; and (b) to disambiguate between Chen's (2013) temporal hypothesis and the modal hypothesis I develop.

I have gone about this in such a way as to address the shortcomings in Chen's (2013) seminal work. I outline how below.

1.3.4.1 Cross-linguistic typology fails to capture modal differences

First, FTR status may be problematic. In particular, it disregards modality and the extent to which modal marking may be obliged in strong-FTR languages. To address this weakness, I used language-elicitation paradigms to quantitatively evaluate how FTR status impacts both future tense, present tense, and modal language use (Chapter 2). Such methods allowed me to understand how FTR status impacts FTR, in terms of both modal and tense structures, and thereby establish original empirically-based cross-linguistic differences in FTR structures which I theoretically expected to impact (risky) intertemporal decision making.

Second, Chen's (2013) arguments about the semantics of the FUTURE TENSE is problematic. This is principally because they disregard the possibility that future tenses encode high certainty, and additionally because they disregard the possibility that FTR status may be a determinant of cross-linguistic differences in future tense semantics (see Section 1.3.2.2). I addressed this by measuring speaker-level future tense use in a weak- and strong-FTR language. This enabled me to empirically determine whether future tense use was associated with increased discounting—as predicted by Chen (2013)—or decreased discounting—as predicted by paradigmatic analysis (Chapter 3 and Appendix G). It also allowed me to estimate how language-specific future tenses impacted (risky) intertemporal decision making differently as a function of FTR status. In other words, whereas Chen's (2013) methods estimate a singular effect of the FUTURE TENSE, the approach taken by this thesis allows for the estimation of different effects for language-specific future tenses (Chapter 3). This is important since modern typological approaches understand language-specific grammatical categories to differ to some extent (i.e. Section 1.4.1, and see Dahl, 2000a, 1985; Haspelmath, 2010).

1.3.4.2 Behavioural outcomes are confounded by probability

Third, whereas Chen (2013), and much follow up work, has found FTR status to predict real-world (risky) intertemporal behaviour, no work (that I can tell) has tested whether FTR status predicts experimentally-controlled risky-intertemporal choices. By using experimental measures of both probability and temporal discounting, I was able to identify whether FTR status had differential impacts on temporal discounting as compared with combined discounting (Chapter 4).

1.3.4.3 Poor mechanistic evidence

Chen (2013) hypothesises that temporal representations mediate the relationship between FTR status and behaviour, but no work has estimated the impact of FTR status on subjective temporal representations. Chen (2013)—and much, if not all, follow up studies—found effects of FTR status on real-world (risky) intertemporal behaviour. Such findings are interpreted to support a plausible—but untested—theoretical account. I addressed this by measuring the proposed mechanisms, using elicited FTR language, as well as measurements of subjective temporal representations. This allowed for contrastive hypothesis tests between Chen's (2013) temporal hypothesis, and the modal account I develop. These methods allowed me to answer questions such as: "Does higher English future tense use cause people to construe future events as more distal, and therefore discount more?", and "Does higher English use of low-certainty modals cause people to discount more?" (Chapters 3 and 4).

1.3.4.4 Weak causal evidence

Forth, strong causal evidence is lacking in the body of work which putatively tests Chen's (2013) hypotheses. Most work is correlational, so leaves open the possibility that observed effects are confounded in some way. To address this weakness, I used a combination of path-analytic techniques and experimental manipulation to provide causal evidence for the impact of FTR grammar on (risky) intertemporal decision making.

In the next section I outline the methodological approach taken to implementing these improvements on the existing state of the art.

1.4 The approach taken by this thesis

I have asserted above that investigations into linguistic relativity effects involve three steps: (a) Cross-linguistic differences need to be identified; (b) correlated population-level differences in non-linguistic cognition need to be established; and (c) the causal impact of (a) on (b) needs to be demonstrated. The work presented in this dissertation also followed this framework. Cross-linguistic differences were established first, and then steps (b) and (c) were largely addressed in tandem, through the use of statistical mediation models which allowed me to estimate the impact of language-level grammatical differences on speaker-level language use and therefore population-level cognitive differences. In this section I outline the empirical approach taken by this thesis, with reference to this framework.

1.4.1 Characterising cross-linguistic differences

My approach to establishing cross-linguistic differences can be broken into two aspects. First, I developed a new FTR language elicitation questionnaire (the FTR-elicitation questionnaire) to establish cross-linguistic differences in future-referring language. Second, I wrote a computer program (the FTR-type classifier) to score questionnaire responses in terms of the linguistic structures they exhibited. My development of the FTR-elicitation questionnaire is heavily indebted to Dahl's quantitative approach to language typology (Dahl, 1985, 2000a, 2000c; Bybee & Dahl, 1989). As such, I describe his project below with a view to characterising key similarities and differences. I then describe the FTR-elicitation questionnaire and the FTR-type classifier.

1.4.1.1 Background on Dahl's typological approach

For simplicity, I refer to Dahl's (e.g. Dahl, 1985, 2000a) work in terms of the FUTURE TENSE, but the work is broader than this. Specifically, Dahl (Dahl, 1985, 2000b) is concerned with understanding the cross-linguistically shared elements of TENSE, MOOD, and ASPECT, and classifying languages according to differences in the language-specific realisations of these general grammatical categories.

Data from a large set of languages are used to infer something general about the FUTURE TENSE (Dahl, 1985). In this view, the FUTURE TENSE is thought of as an "imprecise category" (Dahl, 1985, p. 3), with no clean-cut class criterion. The notion of an imprecise category is not specific to language typology. For instance, despite a clear case being obvious, it is impossible to define *baldness*: How many *n* hairs count as *bald*; what about *n+1* (Dahl, 1985, p. 3)? Like *baldness*, the FUTURE TENSE is thought of as a fuzzy set, but instead of being one-dimensionally defined (e.g. how many hairs?), it is defined in terms of multiple dimensions (Dahl, 1985); for example, its tendency to mark intentions, predictions, or future time (Bybee & Dahl, 1989). Language-specific future tenses variously exhibit the general characteristics of the FUTURE TENSE. Some are more like prototypes, while others are liminal edge cases, which exhibit very few shared characteristics, and may be difficult to conclusively classify. Since it would be biased to use any language-specific future tense as a prototype for the FUTURE TENSE, Dahl (1985, 2000b) collected data from a variety of languages in order to establish the cross-linguistically common elements of the FUTURE TENSE.

Data were collected using a translation questionnaire. The questionnaire was administered by a linguist collaborator to a native-speaking informant[16] who was tasked with translating a supplied sentence into their native language. Each of the approximately 200 items in the questionnaire had a context designed to provide some understanding of the circumstances of communication, and a target sentence, which the native speaker was asked to translate

[16]Sometimes self-administered, when the collaborator was a native speaker of the target language.

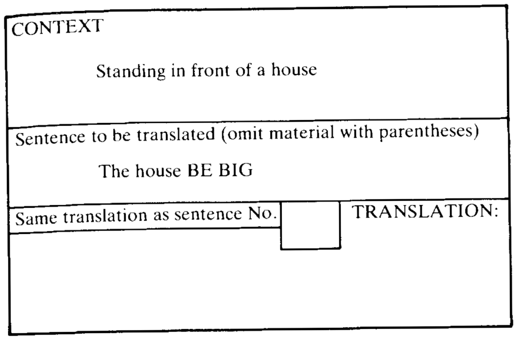

CONTEXT
Standing in front of a house
Sentence to be translated (omit material with parentheses)
The house BE BIG
Same translation as sentence No.
TRANSLATION:

Figure 1.2: An example of Dahl's translation-questionnaire paradigm, from Dahl (1985, p. 45).

into their own language. The main verb in each sentence was given unconjugated and in CAPITALS, in order not to bias the target-language translation by exposure to source language grammatical norms, *Fig. 1.2.* A handful of speakers for each of a sample 64 languages completed the questionnaire. Languages were balanced to represent a world-wide geographical and language-familial distribution (Dahl, 1985, pp. 37–43). This allows for a relatively straight forward fuzzy-set definition of future tense in terms of the set of characteristics shared by questionnaire items in which future tenses were most commonly used (Dahl, 1985, pp. 51–64). Then, by cross-tabulating contingency tables of each language-specific future tense against the others, Dahl was able to numerically score future tenses according to the extent to which they were typical of the future tense.

1.4.1.2 Critical differences between Dahl and the approach of this thesis

While Dahl's approach may have been invaluable in informing this thesis, several points of difference need to be outlined: (a) I focus more on individual differences in language use within languages; (b) this means a (much) smaller language sample is necessary; (c) I emphasis the importance of modality to the investigation of FTR; and (d) whereas Dahl is concerned with the investigation of grammatical categories, I emphasise the investigation of semantic categories. Below I detail the reasons for these differences.

Individual differences: In this thesis, I take the view that the linguistic relativity hypothesis is a *mediational* hypothesis, and that the mediating variable is language use. This is simply to say that grammar affects language use and thereby (in some cases) non-linguistic cognition. This may hardly need saying, but it is fundamental enough to belabour: Through what mechanism other than language use could cross-linguistic differences affect cognition? Nativist ideas about universal linguistic and cognitive endowments (e.g. Chomsky, 1956) go some way to explaining semiotic relativity, i.e. cognitive differences humans may possess as a function of having a language faculty at all (Lucy, 1997; Gumperz & Levinson, 1996). However, such theories cannot explain how cross-linguistic differences could impact speaking populations. Relativity researchers should therefore pay close attention to speaker-level differences (Günther, Müller, Schmid, & Geyer, 2016). This view leans heavily on usage-based approaches, which see grammar not as an abstract language-level set of rules which are identically assimilated by all speakers, but rather as a set of emergent tendencies which become entrenched as speakers use and re-use linguistic constructions (Bybee, 2006). Under this model, grammatical knowledge is naturally understood differently by different speakers of the same language (Barlow, 2013). Thus it is more appropriate to speak of English "grammars" than "grammar." Fully confronting the extent of individual differences and quantitatively integrating these into an account of linguistic relativity is a major goal of this thesis.

This is not an arbitrary decision: The research questions addressed herein resolved largely around a contrastive test between two competing accounts of how FTR status affects speaker-level (risky) time preferences. On one hand, Chen (2013) hypothesises obligatory FUTURE TENSE use drives effects; on the other, I develop the modal hypothesis that obligatory low-certainty modal verb use is responsible. To test these accounts, it was necessary to quantitatively measure which construction types drove effects. The need to do this was compounded by the fact that language-level differences were often predicted to be the same by both accounts. As a consequence of these factors, my empirical approach emphasised the importance of characterising individual differences in language use between different speakers of the same language, and on collecting data from a large number of speakers of each language. To achieve this, Dahl's elicitation paradigm was altered in several critical ways. Dahl gave speakers a context and a target sentence, and left the main verb in each target sentence unconjugated. I duplicated this basic approach, but rather than having participants translate from a shared source language into their native target language, I pre-translated my questionnaires. Participants conducted them entirely in their own languages. This allowed for online data collection, which meant the necessary sample sizes could be achieved.

Choice of languages: It was not feasible to undertake a large typological investigation *and* investigate correlated cognitive effects across a large sample of languages and populations, particularly given the data volume required by a focus on individual differences. Chen's (2013) project is admirable because it involves a large sample of languages (N = 129 languages). However, it is problematic because it does not give adequate treatment to language-specific complexities. The project of retaining Chen's (2013) sample size but giving adequate treatment to the cross-linguistic complexities of FTR would have been untenable. It was therefore appropriate to focus on a smaller set of related languages.

English, Dutch, and German were chosen. (German was dropped after initial investigations revealed that the sharper contrast in terms of FTR grammaticisation was between

English and Dutch, see Chapter 2.) Choice of this sample was motivated by the fact that English is strong-FTR and Dutch and German are weak-FTR. This was an essential requirement to investigating impacts of FTR status. An additional motivation involved comparability. In cross-linguistic comparisons, it can be difficult to identify which language-specific grammatical categories are cross-linguistically equivalent (Dahl, 1985, p. 54). However, the Western Germanic languages sampled in this dissertation are similar enough to permit the identification of shared functional categories which encode similar notional domains, and therefore permit comparison (see Section 1.3.1). This facilitates the quantitative approach I took to coding participant responses with the FTR-type classifier (see Section 1.4.1.4).

An emphasis on FTR and therefore modality: A major focus of Dahl and his collaborators' investigations (1985, 2000b) is on understanding the cross-linguistic similarities and differences in tense, mood, and aspect. In order to keep the project manageable, Dahl (1985) chose to focus on "affirmative, declarative, non-embedded, active constructions" (p. 54). This ruled out moods, as they are usually formed using conditional statements (see Section 1.3.1.2). Aa a consequence, Dahl (1985), and the EUROTYP volume (Dahl, 2000b) focus on tense and aspect. This means that Dahl's questionnaires included contexts designed to elicit past tense and aspectual categories, which were eschewed in mine. Modality was also generally disregarded, despite the important role it plays in FTR. In contrast to Dahl's approach, this dissertation exclusively investigates FTR. This leads to a critical difference: Whereas modality may not be essential to talking about the past, or expressing aspectual notions, it is critical for FTR. In fact, a major aim of this thesis is to demonstrate the extent of entanglement between modal and temporal notions in the linguistic representation of future time (which Dahl is careful to acknowledge (Dahl, 1985, 2000a)). Since being uncertain about upcoming events is a stable feature of how humans relate to the future, an investigation of FTR should elicit language under varying degrees of certainty. This is the critical difference between Dahl and his colleagues' approach and the approach taken in this thesis. It leads directly to the fourth key difference, addressed in the next section.

A focus on semantics rather than grammatical categories: Whereas Dahl and colleagues focussed on investigating grammatical categories (1985, 2000b), this thesis focusses on semantic categories. This is because modality is not a grammatical category, but a functional one, defined in terms of its semantics (Nuyts, 2000; Palmer, 2001). Moreover, in Western Germanic, it tends to be expressed through a variety of word classes: modal verbs, modal modifiers, and mental state predicates (Nuyts, 2000). In consideration of this, investigations into modality should emphasise semantic rather than grammatical categories, a priority which was reflected in the decisions underpinning the coding of elicited text responses (see Section 1.4.1.4).

Figure 1.3: An example of one of my FTR-elicitation questions. The same base item is seen in difference modality conditions and translated into a) English, b) Dutch, and c) German. Participants are instructed to write in the "target" sentence with the main verb—in this case {CRASH}—conjugated. We added {CURLY BRACES} around the main verb to make it clearer for online participants. Differing "certainty information" is displayed for each language, though certainty information was always balanced between languages.

1.4.1.3 Description of the FTR-elicitation questionnaire

Here I describe the more technical aspects of the implementation and creation of the FTR-elicitation questionnaire. As in Dahl's work, participants were given a context and asked to conjugate a target sentence, see *Fig. 1.3.* Many questions were initially borrowed from the FTR section of Dahl's (1985, 2000) questionnaire. However, my questionnaire was iteratively refined to address shortcomings in earlier versions (see Appendix I). These included: use of common FTR terms, e.g. *think*, in the context, which participants were simply copying into the target sentences, causing potential modal false positives for the FTR-type classifier; some questions being mistaken as referring to present rather than future time; and translation-related complications, e.g. "dine out" being translated to Dutch as *uitgaan* 'go out', which caused false positives for the FTR-type classifier because it uses the Dutch future time marker *gaan* 'go' in its andative rather than temporal sense. The final version is—as far as I have established—free of such issues.

To elicit FTR language use under differing degrees of speaker certainty, I included some "certainty information" in the contexts, see *Fig. 1.3.* Participants were directed to imagine they were "_____% certain" about the referenced event. In addition to certainty information, other factors in the FTR-elicitation questionnaire were temporal distance from the present and FTR-mode (intentions, predictions, schedules). The final version comprises 174 items translated into English and Dutch. Following Dahl (1985, p. 53), all items are declarative, affirmative, and active, and non-embedded. There are 12 temporal distances between "later same day" and five years in the future, three certainty conditions, and the three FTR-modes. These factors are fully balanced. Where possible, factors are created by minimally altering base sentences such that factor-level effects can be estimated as freely (as possible) from idiosyncratic item effects.

Such features represent significant alterations from Dahl's questionnaire which was not designed to facilitate the kind of complex statistical analyses undertaken in the present volume. Particularly, my FTR questionnaire allows language use to be statistically analysed in terms of interactions between factors. Additionally, the expansion of levels of temporal distance (I believe Dahl [1985] had four) and the addition of certainty information allows FTR language use to be correlated with other quantitative outcomes (i.e. [risky] intertemporal decision making) at matched levels of delay and probability. Such possibilities represent significant departures from Dahl's original questionnaire, which, for instance, did not include any quantitative or balanced factors which might have facilitated quantitative analyses of language use across FTR-modes and temporal distances. This is not a criticism. Dahl's approach was to analyse language use across such factors qualitatively and this yielded an impressive contribution to knowledge about the cross-linguistic differences and similarities in the grammatical categories of TENSE and ASPECT. Nonetheless, these innovations represent important points of departure, and the FTR questionnaire represents a large part of the methodological contribution of this thesis. Full versions of all iterations are in Appendix I.

1.4.1.4 Description of the FTR-type classifier

In order to treat free-text responses as quantitative variables in statistical analyses, it was necessary to numerically code them. As far as I can tell, Dahl (1985) hand-coded his questionnaire data, but this was not tenable given the large samples I collected. Dahl (1985) mentions mostly using only one informant per language, but being lucky enough to obtain two or more "sometimes" (p. 39). This implies the need to code approximately $1.5 \times 200_{items} \times 64_{languages} = 19,200_{responses}$. This thesis involves the analysis of approximately $244,000$ items of text data, an increase of an order of magnitude.

As such, it was necessary to automate the process. My solution was to use the programming language Python (Python Software Foundation, 2017) to write a deterministic, closed-vocabulary, keyword-based classification system called the FTR-type classifier. It comprises: (a) a number of reference lists of FTR lexemes in English, Dutch, and German; and (b) a set of rules which determine how the keyword lists are used to classify text documents. These keywords are used as criteria to classify free text data in terms of four exclusive *semantic* categories: the future tense, the present tense, low-certainty modality, and high-certainty modality, see *Table 1.1.* The last two are broken down into two further non-exclusive categories depending on whether they use a modal verb (verbal-high/low-certainty) or some other modal construction (other-high/low-certainty), i.e. a mental state predicate, a modal modifier, or a modal particle (in Dutch or German).

Table 1.1

FTR-type classification schema with example class-criterion keywords in English, Dutch, and German

language	future tense	present tense		low-certainty modality	high-certainty modality
English	It... *will/is going to/shall* rain tomorrow.	It... *rains* tomorrow.	verbal	It... *could/might/may/should* rain tomorrow.	It... *must* rain tomorrow.
			other	I... *think/believe* it will... *possibly/probably/potentially* rain tomorrow.	It will... *certainly/definitely/surely* rain tomorrow.
Dutch	Het... *zal/gaat* morgen regenen.	Het... *regent* morgen.	verbal	Het... *kan* morgen regenen.	Het... *moet* morgen regenen.
			other	Ik... *denk/geloof* dat het morgen... *misschien/waarschijnlijk* zal regenen.	Het zal morgen... *zeker/definitief* dat het regenen.
German	Es... *wird* morgen regnen.	Morgen... *regnet* es.	verbal	Es... *könnte/dürfte/mag/müßte* morgen regnen.	Es... *muss* morgen regnen.
			other	Ich... *denke/glaube* dass es morgen... *möglicherweise/wahrscheinlich* regnen wird.	Morgen wird es... *definitiv/bestimmt/sicherlich* regnen.

Evidence that particular words encode class-general semantics, and therefore justification for their use as class-criterion keywords, is grounded in review of the relevant literature. In particular, see: Nuyts (2000); Mortelmans (2000, 2009); Lee (2016); Mindt (1995); Dahl (2000a, 1985); Bouma (1973); Byloo, Kastein, and Nuyts (2007); Donaldson (2017); Huitink (2012); Karagjosova (2004); Shepherd, Rossner, and Taylor (1984); Simon-Vandenbergen and Aijmer (2007); C. M. Stevens (1995); van der Auwera and Plungian (1998); Zimmermann (2019). To see the relevant reference for each particular word, see *Tables C.4–C.6* in Appendix C. In fact, as part of this thesis's committment to transparency and open science, the FTR-type classifier code includes a bibliographic database which links class-criterion keywords to the relevant reference, such that word-level references are automatically produced for whichever class-criterion keywords are encountered in some corpus of text data.

As an example, the class-criterion keyword list for the verbal-low-certainty category in English is: *may, can, could, might, should, ought,* and *would.* Any English response which uses one of these modal verbs would be classed as verbal-low-certainty. Some of these are peripheral to the encoding of *epistemic* modality, e.g. *ought* and *would.* In practise, use of these peripheral keywords was highly infrequent. For instance, of the approximately 27,000 responses analysed in Study 1 in Chapter 2, *would* was used 187 times, and *ought* was used less than 3 times. Some high-frequency class-criterion keywords are given in *Table 1.1.*

A few things should be said about the FTR-type classifier. Firstly, it classifies according to *semantic* rather than *grammatical* categories. This is a principled and potentiality tendentious decision, with serious consequences for the classification system as a whole. However, a semantically oriented approach made the most sense for my particular research question. In addition to modality being defined by semantics, the relativity effects explored in this theses are well-described by theoretical characterisations of language as a "spotlight" (Wolff & Holmes, 2011). Such views assert that language-specific grammars bias speakers towards paying particular attention to some notional (semantic) domain during language use, i.e. an obligatory grammatical distinction requires speakers to notice a semantic distinction in order to use their language. Eventually, such "spotlighting" may cause speakers to pay particular attention to the spotlighted semantic domain, even (sometimes) in non-linguistic cognition. This is true of theoretical approaches to my particular question: Chen (2013) hypothesises that FTR status serves to direct speakers' attention toward temporal semantics through the use of future tenses (the temporal hypothesis). I develop the idea that FTR status directs speakers' attention to modal semantics through the use of low-certainty modal verbs (the modal hypothesis). In either case, the relevant question is not which grammatical categories are being used, but which semantic notions are being spotlighted. It was therefore appropriate from the perspective of my research question to orient the FTR-type classifier towards semantic rather than grammatical categories.

This was not without consequences. Principally, it means that modal semantics "dominate" tense, such that responses which use modal class-criterion words are *not also* classed in a tense category, see *Fig. 1.4*. This decision was made from a need to deal with responses which exhibited so-called modal concord, i.e. the use of both a modal verb (which could be a future tense) and modal modifier to quantify a single statement (Huitink, 2012). Modal concord is possible in English, Dutch, and German, and can lead to conflicts between the modal and tense categories. For instance, *It will probably rain* and *It will definitely rain* are both formally future tensed, but different epistemic commitments are expressed. Similar dynamics characterise present tense FTR; for instance, *They could win tonight* and *The game is definitely at 7* are both present tense but express different modal notions. Critically, such dynamics alter the semantics of the future tense. *Will* "means" two different things in the following examples:

(1.13) a. *It will rain tomorrow.*

b. *It will possibly rain tomorrow.*

In example (1.13a) *will* encodes high certainty when paradigmatically contrasted with other possible modal verbs—*may, might,* or *could.* In (1.13b), it does no such thing. The modal notion being expressed by *possibly*, and as a consequence the entire statement, is low certainty. In other words when used together, *possibly* dominates the normal high-certainty semantics of *will.* The same dynamics can be observed in present tense FTR in Dutch and German:

(1.14) German

a. *Morgen regnet es.*
Tomorrow rain:PRS it.
'It rains tomorrow.'

b. *Es regnet möglicherweise morgen*
It rain:PRS possibly tomorrow.
'Rain tomorrow is possible.'

(1.15) Dutch

a. *Morgen regent het.*
Tomorrow rain:PRS it.
'It rains tomorrow.'

b. *Mogelijk regent het morgen.*
Possibly rain:PRS it tomorrow.
'Rain tomorrow is possible.'

In examples (1.14a) and (1.15a), the present tense conveys no modal valence, which is arguably the same as high certainty (Giannakidou & Mari, 2018). In examples (1.14b) and (1.15b) the modal adverbials *möglicherweise* and *mogelijk* dominate high-certainty semantics, such that the statements as a whole encode low certainty.

The FTR-type classifier simply reflects the semantic domination of tense by modal modifiers. I therefore use the terms "future tense" and "present tense" to indicate something quite different from what linguists might understand the terms to mean: Specifically, throughout this thesis, I use these terms to indicate only future and present tense when no other modal modification is used. This limited semantic definition allows me to analyse these structures' *semantic* contribution within Western Germanic FTR systems.

Several other points should be addressed. I can foresee sensible critics questioning the FTR-type classifier's assumption that *could, may, might* necessarily indicate *epistemic* low certainty in light of the highly polysemous nature of the modal verb systems in English, Dutch, and German (Nuyts, 2000). *Could* can be used as a marker of *dynamic* modality, and *may* as a marker of deontic modality. The modals are complex and licenced to quantify over various modal bases which need to be inferred pragmatically. My solution to this complexity was to constrain the FTR questionnaire items in various ways which make inference more reliable. For instance, the polysemous nature of the modals is reduced by some significant degree when they future-shift, e.g. *could* only indicates dynamic modality when used for past time reference (see Section 1.3.1.2). When used for prediction-based FTR, it generally encodes epistemic modality, or blurs epistemic with dynamic notions, e.g. *That question could be on the test tomorrow.* Additionally, the restriction of items to declarative, affirmative, active and

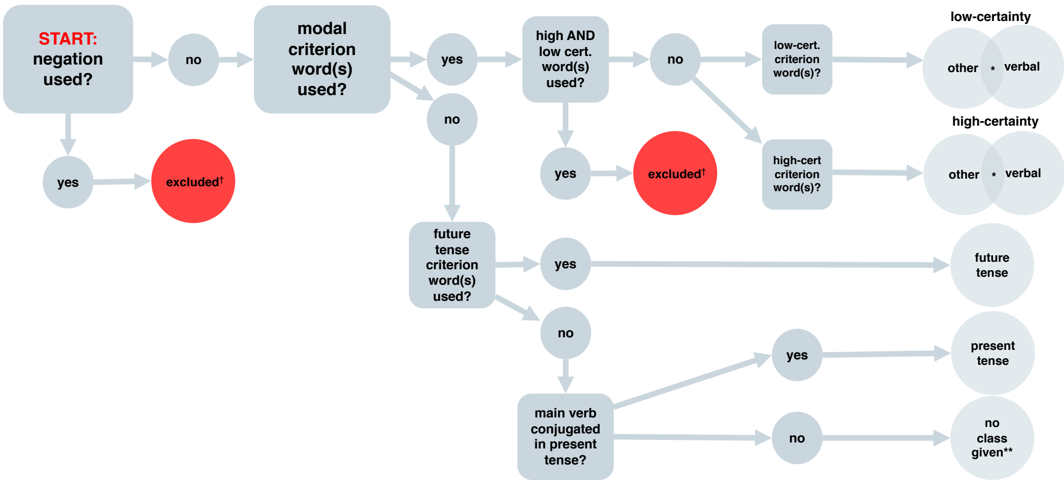


Figure 1.4: A schematic flow diagram of the FTR-type classifier. Modal categories "dominate" tense categories, so that examples such as *It will possibly rain tomorrow* or *Rain tomorrow is probable*, are classed as other-low-certainty and not also future/present tense.
*Overlap between verbal and other modal category Venn diagrams: In cases of so-called "modal concord", it is possible to use both a modal periphrastic modifier and a modal verb, e.g. *It could probably rain tomorrow* (Huitink, 2012). Such responses are classified as *both* other- and verbal-low/high-certainty.
**No class given: Inspection of the text data indicated these tended to be misspellings, illegible responses, past tense responses, etc.
† Exclusions: Both negations and what I term "mixed modal" responses make it impossible for keyword techniques to identify the intended modal polarity. For instance, in the presence of negations, modal adverbs reverse modal polarity, e.g. *I am not certain it will rain tomorrow*, would be classed as other-high-certainty because of the use of *certain*, but it expresses low certainty. What I term "mixed modal" responses are similarly complex. These are expressions which use both high- and low-certainty keywords, e.g. *There is definitely a possibility of rain tomorrow.* The FTR-type classifier cannot know which word should be used as the class criterion in such cases. As a solution, I simply used a statistical model implemented in the popular natural-language-processing package *spaCy* (Montani & Honnibal, 2020) to detect and exclude negations, as well as any mixed-modal responses which the FTR-type classifier indicated used both high- and low-certainty modal words. Such responses were low frequency; for instance of the 38,166 text items analysed in Study 2 of Chapter 2, only 472 had to be excluded (1.23%).

non-embedded sentences further reduces the polysemous nature of the modals. For instance, *may* is commonly used deontically in the interrogative mode, e.g. *May I come in?*, but less-so in the declarative mode when the context does not involve permission, e.g. *Tickets may sell out tonight.* By excluding such contexts, I can be relatively sure the FTR-type classifier is not mislabelling as verbal-low-certainty some deontic use. More broadly, in drafting the FTR questionnaire, I avoided contexts in which non-epistemic uses would be expected. There are no items such as, "[Parent to child:] ...You MAKE SURE your bag is packed by tonight!" which might be expected to produce deontic modal uses, e.g. *You should make sure...*

Another point involves the breadth of the semantic categories implemented in the FTR-type classifier. For instance, there are differences between the uses of the Dutch *gaan* 'be going to' and *zullen* 'will' (Fehringer, 2018). Likewise, there are differences between modal modifier derivations, e.g. the adjective *possible* encodes subtly different notions than the adverb *possibly* (Nuyts, 2000). Yet the FTR-type classifier treats *zullen* 'will' and *gaan* 'be going to' both as future tense, and *possible* and *possibly* both as other-low-certainty. This was motivated by the kind of cost benefit trade-offs which characterise most research choices. Specifically, having four simple FTR-type categories allowed for relatively straight-forward predictions to be made about how cross-linguistic differences should impact (risky) intertemporal decision making. To achieve this, it was necessary to overlook numerous subtleties.

1.4.1.5 Testing the accuracy of the FTR-type classifier

To ensure that the predictions of the FTR-type classifier were accurate, I—or a linguistically-trained assistant—hand-coded a tranche of data from each experiment. Responses from each category, as well as high frequency responses, were selected, so the test data had high coverage against the experimental corpora (e.g. $\approx 20\%$ in Study 2 of Chapter 4). These hand-annotated data were used as ground truth against which to test the predictions of the FTR-type classifier. In early stages, this process was iterated to fine tune the FTR-type classifier. By the end of the project, predictions closely approximated ground truth with no adjustments to the code, i.e. accuracy metrics tended to be over 95%.

1.4.2 Establishing differences in non-linguistic cognition: Psychological discounting

To establish relevant population-level differences in non-linguistic cognition, my approach was fairly simple, and mostly borrowed with little or no modification from the relevant literature. To measure psychological discounting, I followed well-established paradigms by giving participants repeated binary choices between options which were either just delayed (to measure temporal discounting), or risky and delayed (to measure combined discounting), see *Fig. 1.1* (L. Green & Myerson, 2004; Vanderveldt et al., 2015). I then used either analytic or empirical methods to approximate participants' indifference points and discounting rates. Since Chen's (2013) hypothesis involves predictions about linguistic effects on *temporal* beliefs, I also measured individual differences in the subjective representation of temporal distance. To do this, I used methods developed by econometricians to measure subjective representations of time (Zauberman, Kim, Malkoc, & Bettman, 2009; B. K. Kim & Zauberman, 2009; Bradford, Dolan, & Galizzi, 2019). Very simply, participants used a "time slider" to indicate whether they felt a given temporal distance to be "close" or "far", see *Fig. 1.5*.

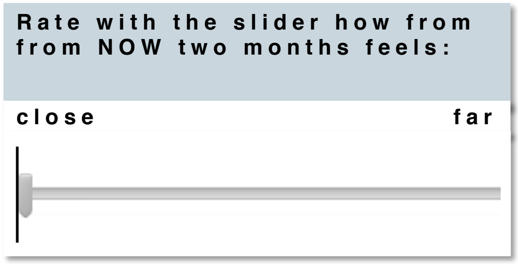


Figure 1.5: An example of the temporal distance slider-based measurement paradigm I implemented.

1.4.3 Developing causal evidence: Mediation, mechanisms, and statistical approaches to cultural confounds

Here, I outline the approach taken in this thesis to providing causal evidence that cross-linguistic differences were impacting (risky) intertemporal decision-making preferences. My approach was deeply informed by the view that the linguistic relativity hypothesis is a meditational hypothesis, and that language use is the mediator. In light of this view, I used mediation models which estimate the *indirect* effects of language-level grammatical constraints on (risky) intertemporal decisions *via* elicited language use. As a special case of path analytic approaches in general, mediation models can help researchers generate causal evidence from correlational data (Pearl & Mackenzie, 2018). In particular, mediation models allow for the investigation of the mechanisms intermediately involved in some causal relationship. A classic example is that vitamin C levels *mediate* the relationship between eating citrus fruits and developing scurvy (Pearl & Mackenzie, 2018, p. 300). In path analysis in general, arrows are used to represent conditional relationships, so this example is represented as, $citrus\ fruits \rightarrow vitamin\ C \rightarrow scurvy$. I use such mark-up conventions throughout this thesis.

There can be multiple mediators involved in most causal relationships. For instance, in this thesis, I contrast various linguistic variables against each other as potential mediators. Specifically, I test how the future tense, low-certainty modal language, and subjective temporal distance contribute to mediating effects of FTR status on (risky) intertemporal outcomes. This allows me to conduct contrastive hypotheses tests between Chen's (2013) temporal hypothesis that $FTR\,status \rightarrow future\,tense \rightarrow temporal\,beliefs \rightarrow intertemporal\,decisions$, and the contrasting modal hypothesis that $FTR\ status \rightarrow$ *low-certainty* $\rightarrow$ $(risky)$ *intertemporal decisions* (Chapter 4). Additionally, by quantitatively capturing the role that language use plays in mediating between FTR status and outcomes, I was able to show that language is involved in producing population-level differences (Chapters 3 and 4). This does not completely rule out the possibility that non-linguistic cultural differences (and not grammatical constraints of the languages involved) are driving differences in language use. However, it does suggest that language-level grammatical features are involved in the maintenance and transmission of such putative cultural differences, if not that language-level grammatical features are causally responsible for cognitive outcomes. To address these concerns, I manipulated language within a population of bilinguals, which arguably isolates the influence of language from culture (Chapter 4). Additionally, I used a constrained language-production task to test whether directing participants to use the future tense impacted time preferences (Chapter 3). By experimentally manipulating language, these methods rule out cultural factors which could influence language use in correlational data.

1.4.4 Open science principles

Throughout this thesis, I have maintained a commitment to open science principles, publishing code and data where possible, and pre-registering the later experiments (see Appendix B). In keeping with these values, I have made the FTR-type classifier open source and publicly available for download and use as a Python package (Robertson, 2019b). Additionally, I will publish the statistical future and past time reference classifiers (which estimate whether a text datum refers to future/past time at all, rather than how an item of text datum makes FTR) which I built (Chapter 5). As well, the full FTR-elicitation questionnaire is available to download and use, including some functions which translate it into text files which can be uploaded to the popular survey hosting platform Qualtrics (Robertson, 2019a). I am committed to reproducibility, which can include simple principles such as ensuring analysis scripts are conserved so relevant results can be reproduced at a later date, or subsequent data can be analysed in precisely the same ways. In this vein, the entirety of this thesis—including the production of figures and table elements—can be re-created. The outputs from analysis are piped directly into the relevant in-text reportage in results sections. This means that the relationship between analytic and reported results is transparent and reproducible.

1.5 Key questions

Below, I outline the key questions this dissertation seeks to answer, with reference to the three-step linguistic relativity framework I have been using. As a reminder, this is: (a) establish relevant prospective cross-linguistic differences; (b) identify correlated non-linguistic cognitive differences; and (c) provide causal evidence for the impact of (a) on (b).

1.5.1 Establishing cross-linguistic differences

The first step it to establish cross-linguistic differences. In this thesis, this involved having English, Dutch, and German participants complete the FTR questionnaire and analysing language use.

1.5.1.1 Question 1: Are there differences in modal FTR between weak- and strong-FTR languages?

The first step was to adequately characterise the relevant cross-linguistic difference between strong-FTR (English) and weak-FTR (Dutch, German) languages. Both Dutch and German have complicated, flexible systems of modal future time reference which allow for the expression of modal FTR. In fact, the German modal system may be even more complicated than the English one, in light of the complexities of the *Konjunktiv II* which—as I have explained—can subtly alter the German modals. Moreover, Dutch and German allow for the expression of modal notions via lexical systems of modal modifiers, mental state predicates, and modal particles. It does not necessarily follow from the fact that English modal verbs are obligatory that English speakers encode a higher proportion of low-certainty FTR. Dutch and German speakers might make up the difference through use of other expression types. A critical question was therefore whether the obligatory English modal system resulted in increased encoding of low-certainty FTR in English. I answered this question by having English, Dutch, and German participants complete the FTR-elicitation questionnaire and statistically analysing their responses (Chapter 2).

1.5.1.2 Question 2: What notional domains do future tenses encode?

An important point within a broad investigation into cross-linguistic difference is to establish not only whether some grammatical or lexical feature differs between languages, but also to understand the notional domains involved. In some cases, this is straight forward, e.g. the semantics of separate Russian words for light- and dark-blue are perfectly transparent. This is not so easy in the context of my research question, given the noted complexities of the FUTURE TENSE. In light of the division of labour between modal, temporal, and aspectual semantics, a critical question is: What do the future tenses in English, Dutch, and German "mean"? Chen (2013) argues that the FUTURE TENSE exhibits temporal semantics (precision, distance), whereas I argue it encodes modal high certainty, given its paradigmatic opposition with other available low-certainty modals. Which of these hypotheses is closer to the truth? Two lines of evidence were used to explore this question. Firstly, participants rated FTR constructions in terms of whether they encoded temporal distance (Chapter 2). For example, participants used a time slider to rate temporal distance for "It *will* rain tomorrow" compared with "It *rains* tomorrow." If Chen's (2013) account of FUTURE TENSE semantics is correct, participants should indicate the future tense statement as being more distant. Secondly, participants rated FTR constructions in terms of whether they encoded high or low certainty (Chapter 2). If future tenses encode high certainty, they should be rated accordingly.

A further relevant question is whether the semantics of future tenses differ systematically between weak- and strong-FTR languages. On one hand, Chen's (2013) account treats FUTURE TENSE semantics as universal. This means that English and Dutch future tense use should relate to time preferences identically. On the other hand, I note that Dutch permits the present tense in prediction-based FTR and English does not. I argue this alters the paradigmatic semantics of such weak-FTR future tenses. If this is true, future tense use would relate to time preferences differently in English compared with Dutch. To test this, I measured future tense use and then time preferences in English and Dutch speakers and tested whether these variables related the same way in each language—as predicted by Chen's (2013) account—or whether FTR status led to differential impacts—as predicted by a paradigmatic analysis (Chapter 3).

1.5.2 Identifying correlated differences in non-linguistic cognition

The second identified step is to establish relevant population-level non-linguistic cognitive differences.

1.5.2.1 Question 3: What non-linguistic cognitive differences are associated with FTR status?

The question was whether FTR status affected outcomes which were *just* delayed (temporal discounting), or whether it only affected outcomes which were risky *and* delayed (combined discounting). Chen's (2013) temporal hypothesis predicts an effect of FTR status on time preferences, whereas the modal account I develop predicts an effect on risky time preferences (see Chapter 4). By systematically altering intertemporal choices between delay, and risk + delay, I investigated whether English and Dutch speakers differed in terms of time preferences, or risky time preferences (Chapters 3 and 4).

A second relevant question has to do with whether strong-FTR speakers construe future events more distally. This is predicted by Chen's (2013) temporal account, whereas no such difference is predicted by the modal account. To establish whether English speakers rated future outcomes more distally than Dutch speakers, I had speakers of both languages complete time slider tasks (Chapters 2 and 4).

1.5.3 Developing causal evidence

The third step is to develop causal evidence that cross-linguistic differences causally impact the outcomes. This predominantly involved estimating how different FTR construction types (future tense, low-certainty modality, etc.) mediated between FTR status and (risky) intertemporal outcomes.

1.5.3.1 Question 4: How does obligatory future tense use in English affect intertemporal decisions?

An obvious starting point was how future tense use related to time preferences, since Chen's (2013) temporal account revolves around this question. Chen (2013) hypothesises that obligatory use of the FUTURE TENSE in strong-FTR languages should cause speakers to have more precise beliefs about the temporal location of future events, or to believe future events are temporally distal. Either theoretical account is hypothesised to cause strong-FTR speakers to temporally discount to a greater extent. This entails that if individual-level use of the future tense is correlated with individual-level temporal discounting, the correlation should be negative: More future tense use equals more temporal discounting. On the other hand, I have pointed out that a paradigmatic analysis of the English future tense suggests it encodes high certainty. Given people probability discount low-certainty outcomes, this entails that the correlation should be positive: More future tense equals less temporal discounting. Since English speakers use more future tense FTR constructions, this entails that the mediated effect of *English* → *future tense* → *temporal discounting* should be positive if Chen (2013) is correct and negative if the paradigmatic account is correct. To test which account was supported, I had participants complete the FTR questionnaire, as well as a temporal discounting task, and used mediation modelling in contrastive hypotheses tests between the two accounts (Chapter 3).

1.5.3.2 Question 5: How does obligatory low-certainty modal verb use affect combined discounting?

The modal account I develop predicts that obligatory low-certainty modal FTR in strong-FTR languages should be associated with increased combined discounting. The onus strong-FTR languages place on speakers to encode future uncertainty is predicted to cause them to construe future outcomes as riskier and therefore as having lower utility when risk and delay are both involved. This means that the mediated effect of *English* → *low-certainty* → *combined discounting* should be positive (higher uncertainty = discounting). To test this prediction, I had participants complete a version of the FTR-elicitation questionnaire, and then a risky intertemporal decision-making task (Chapter 4).

1.5.3.3 Question 6: Is language causally involved in producing observed differences between populations?

In general, the use of mediation models strengthen grounds for the causal inference that cross-linguistic differences are involved in producing observed impacts on (risky) intertemporal outcomes. However, such approaches leave open the possibility that cultural differences (and not grammatical ones) are driving differences in observed language use. This seems unlikely when language use patterns conform to well-known features of grammar. It is nonetheless possible. To address such concerns, I replicated population-level experiments with English and Dutch speakers, but with a critical difference. I implemented a within-subjects linguistic manipulation in a sample of Dutch-English bilinguals (Chapter 4). Bilingual participants completed a version of the FTR-elicitation task and then a risky intertemporal decision-making task. First, they completed the survey in one language, then, a few weeks later, in the other (order was randomised). Such methods ensure that the sample populations are comparable (they are the same people), and thereby go some way to ruling out criticisms of cultural confounding. Another strategy I employed to address cultural confounds was to direct Dutch speakers to complete an FTR-elicitation task using either the present or future tense. I then tested whether their temporal discounting was affected relative to a previously-measured baseline (Chapter 3). Such methods address cultural confounding by exogenously manipulating language production to mirror cross-linguistic differences indexed by FTR status. Intervening directly in participants language production rules out cultural factors which might have driven language use differences between English and Dutch participants in other studies.

1.5.4 Further questions: Big-data and functional relativity

With the following questions, I widen the field of enquiry beyond questions strictly involving linguistic relativity. Here I am interested in a description of naturally-occurring English FTR, and in broad questions about potential relationships between language and thought.

1.5.4.1 Question 7: What kinds of FTR structures are most common in real-world FTR?

Dahl (2013) points out that Chen's (2013) choice to use prediction-based FTR as the class criterion for FTR status may be flawed. This is because it is hard to know how much natural FTR actually involves predictions. People might simply spend more time talking about intentions, obligations, desires, or schedules. As such, Chen's (2013) prediction-based definition may fail to capture cross-linguistic differences that are more pertinent to the majority of natural FTR. In a similar vein, it is unclear to what extent future tense constructions characterise FTR. Elicitation methods like Dahl's and my FTR questionnaire inevitably generate a biased sample of language production. Moreover, keyword-based corpus methods which seek to make inferences about some linguistic unit on the basis of co-occurrence patterns lend themselves to investigating individual linguistic units, rather than broad explorations of a notional domain. To find co-occurrences, one must start with a search term... How would one find, for instance, unmarked present tense FTR constructions? As such, I take the view that conclusively inferring much about the general preoccupations of English FTR on the basis of an analysis of *will* is not possible.

I therefore explored what kinds of FTR construction types characterise naturally-occurring English FTR (Chapter 5). To do this, I used big-data methods, building a corpus of naturally-occurring language scraped from the social media website Reddit. I then built a statistical machine learning model which estimates not *how* but *whether* a piece of text refers to the future. I then passed the future-referring text items to the FTR-type classifier. This helped establish the proportions of modal, future tense, and present tense construction types in English FTR.

1.5.4.2 Question 8: Do FTR usage and (future) time perspective differ between healthy populations and people who suffer from mental health issues?

This question may be seen as a significant departure from the main concerns of the rest this dissertation. However, Lucy's (1997) framework for linguistic relativity research provides a useful tool for moving between different, related aspects of how language may impact thought. Mostly this dissertation involves investigating *structural* relativity (effects of cross-linguistic differences on non-linguistic cognition). With Question 8, I investigate *functional* relativity (the relationship between language and thought in different speech communities within the same language) (Lucy, 1997; Gumperz & Levinson, 1996).

Time perspective, in general, and future time perspective, in particular, encompasses a variety of ways in which people relate to time. Psychological discounting is one way to characterise future time perspective, but it is not the only way. For instance, "time horizons" measure how far into the future people tend to think (Wallace, 1956). Time horizons and psychological discounting have been identified as markers of anxiety (Rounds et al., 2007; Steinglass et al., 2017) and depression (Dilling & Rabin, 1967; Felton et al., 2020; Pulcu et al., 2014). Since I demonstrate that linguistic FTR also impacts discounting (Chapters 3 and 4), a relevant question becomes whether FTR usage differences also mark these mood disorders. As far as I know, no work has investigated this question.

I therefore supplemented the Reddit corpus with a matched sample of text items drawn from Reddit users who had previously posted comments to forums dedicated to anxiety and depression (Study 1, Chapter 5). I then passed these data to the statistical time-reference classifier, and the FTR-type classifier, as well as SUTime, a system for extracting time-horizons from free-text documents (Chang & Manning, 2012). This allowed me to establish whether people who self-identify as suffering from anxiety or depression differed from a control sample across these variables. I followed this up with an experimental study designed to understand whether the observed results were characteristic of sufferers of anxiety or depression, since having once posted in an online forum is a poor diagnostic criteria (Study 2, Chapter 5).

1.5.5 Chapter guide

In the next Chapters, I present the empirical results of my investigation. Chapter 2 focusses on an analysis of FTR-elicitation data with a view to characterising relevant cross-linguistic differences. Chapter 3 focusses on testing Chen's (2013) hypothesis that increased future tense use in English should result in increased temporal discounting. In Chapter 4, I develop a simple formal theoretical perspective for the modal hypothesis. I then test whether this framework, or Chen's (2013) temporal hypothesis, best predicts results using (just) intertemporal outcomes (temporal discounting) and risky intertemporal outcomes (combined discounting). Finally, in Chapter 5, I undertake an exploration of functional relativity. In this chapter, I characterise differences between healthy individuals and people who suffer with anxiety and depression in terms of FTR, time horizons, subjective representations of time, and temporal discounting. Over the course of these chapters, I hope to demonstrate the critical importance of modal notions to the way people think about, talk about, and make decisions about the future.

Chapter 2

Modality and future time reference in English, Dutch, and German

Abstract

Economists have hypothesised that cross-linguistic differences in Future Time Reference (FTR) grammar affect how people devalue (‘temporally discount”) future outcomes. Habitual use of the present tense for FTR is hypothesised to cause speakers to perceive future outcomes as temporally proximal, and thereby as more valuable. This account assumes that FTR primarily encodes when in time events occur. This overlooks the tendency for FTR structures to encode “modal” notions about the likelihood of future outcomes. People also devalue (“probability discount”) low-probability outcomes. It is therefore important to understand what FTR systems actually encode: notions of time or notions of modality. We elicited FTR language from speakers of English, which exhibits strongly grammaticised FTR, and Dutch and German, which do not (Study 1). English speakers were more likely to use modal terms indicating low certainty, e.g. *could, may, might*, *possibly*. Relative to Dutch speakers, English speakers used more low-certainty terms: when they talked about temporally distant events, when they made predictions, and when they were directed to imagine they were uncertain (Study 2). Additionally, framing an FTR statement in the present (“Ellie arrives later on”) versus the future tense (“John *will* arrive...”) did not affect estimates of temporal distance and English speakers rated statements as more temporally proximal—opposite to the predicted direction. If English obliges low-certainty FTR constructions, this could impact temporal and/or probability discounting. Researchers working on linguistic relativity and intertemporal decisions should reconsider widely-cited theoretical explanations.

2.1 Introduction

Do differences between languages change the way people think, feel, and act? The linguistic relativity hypothesis suggests that they do (Whorf, 1956; also see Everett, 2013; Gumperz & Levinson, 1996; Leavitt, 2011; Lucy, 1992). The idea is that languages force speakers to notice different things in order to communicate, and that the resultant differences in online attentional demands can grow through lifelong language use into entrenched offline cognitive differences (Wolff & Holmes, 2011). For instance, when choosing between the English demonstratives *this* and *that*, speakers need only pay attention to whether the referred-to object is located near or far from themselves. Spanish breaks this space into three degrees of distance: *este* ‘this’, *ese* ‘that’, and *aquél* ‘that’ (distant, i.e. ‘yon’ [archaic]). Malagasy breaks it into seven (Evans, Bergqvist, & San Roque, 2018). Might speakers of Spanish or Malagasy be faster or more precise at estimating distance from ego? A growing body of research attest to affects like this (see Casasanto, 2016; Everett, 2013; Lupyan et al., 2020; Majid, 2018; Wolff & Holmes, 2011).

A typical way linguistic relativity research progresses is by identifying cross-linguistic differences and then investigating whether they give rise to corollary cognitive effects (Lucy, 1997, 2016). In this vein, economists have recently begun to explore whether cross-linguistic differences in the grammaticisation of FTR[1], affect speakers’ subjective estimation of the value of delayed outcomes (for review, see Mavisakalyan & Weber, 2018). This is referred to as “temporal discounting.”

2.1.1 The linguistic savings hypothesis explained

The hypothesis that FTR grammaticisation could affect temporal discounting is referred to as the “linguistic savings hypothesis” (Chen, 2013). It is based on two observations. First, some languages oblige speakers to use future tenses when referring to the future. Others do not. For example:

(2.1) a. English
It will rain tomorrow.

b. Dutch

Morgen	*regent*	*het.*
tomorrow	rain:PRS	it

‘It will rain tomorrow.’

[1]Typically, linguists use separate but related terms for notional categories and the linguistic structures which grammatically encode them (Bybee et al., 1994). In this vein, we use FTR to refer to any statement about the future and “future tense” to refer to the linguistic structures which sometimes grammatically mark FTR, e.g. in English *will*, *shall*, or *be going to*.

c. German

Es regnet morgen.
It rain:PRS tomorrow

'It will rain tomorrow'

When English speakers reference the future, it is necessary to use *will*, *be going to*, or *shall* (Dahl, 2000a), e.g (2.1a). No such restrictions exist in Dutch or German, where constructions like examples (2.1a) and (2.1b) are acceptable. Based on typological research by Dahl (1985, 2000a) and his colleagues, Chen (2013) created a dichotomous typological variable which classes $N = 129$ languages into two categories. Languages like Dutch with non-obligatory future tenses are classed as "weak-FTR". Languages like English with obligatory future tenses are classed as "strong-FTR"[2]. Specifically, weak-FTR languages are those languages which do not oblige the future tense in <u>prediction-based</u> contexts. Strong-FTR languages are those languages which are not weak-FTR (Chen, 2013). In this context, prediction-based FTR contexts are contrasted with schedule-based and intention-based ones (Dahl, 2000a). This is relevant, because it is quite common for languages to permit the present tense when referring to schedules (Dahl, 2000a). English is a good example: *The Bears play at 7pm* is perfectly acceptable. As such, when we refer to languages "obliging" the future tense, we follow Chen (2013) in only meaning for prediction-based FTR. We refer to Chen's (2013) dichotomous weak/strong distinction as "FTR status."

The second observation is that people tend to temporally discount delayed future outcomes as the time until they will occur grows longer (L. Green et al., 2014; Rachlin et al., 1991). For example, most people would prefer \$100 immediately, rather than \$100 after a year-long delay. Temporal discounting rates are variable. Offer \$200 in a year and some people will prefer the immediate \$100 while others will wait for the \$200. Such preferences depend on individual differences in temporal discounting rates (L. Green & Myerson, 2004). Less discounting means higher estimations of delayed value. More discounting means lower estimations. Therefore, people with high temporal discounting rates tend to make "present-oriented" decisions not to wait. People with low rates tend to make "future-oriented" decisions to wait. It is common to refer to such differences in terms of "time preferences."

Time preferences are reliable predictors of real-world "intertemporal decisions." This refers to decisions in which individuals balance present versus future costs and rewards. For instance, time preferences have been found to predict real spending (Bickel et al., 2010), and financial outcomes such as income levels and financial mismanagement (Hamilton & Potenza, 2012; Xiao & Porto, 2019). Time preferences also predict substance abuse tendencies, which

[2] Dahl (2013) points out that these terms are problematic. FTR stands for "Future Time Reference", and of course it is possible in all languages to refer to future time (whether or not a tense is obliged). Better might be "strong-FTR_G" to indicate that the difference is a matter of <u>G</u>rammatical marking. However the terms are in widespread use. We will not deviate from them, despite Dahl's (2013) pertinent critique.

often incur long-term costs (professional, social), but confer short term benefits (hedonistic pleasure). This includes alcohol abuse (Vuchinich & Simpson, 1998), opioid dependency (Garami & Moustafa, 2019), and substance abuse in general (Mejía-Cruz et al., 2016; Kirby et al., 1999). Health behaviours are often impacted as well, because many heath-critical decisions involve trade-offs between immediate (dis)comfort and future (ill)health. For instance, time preferences predicted the odds of smoking cigarettes (Bickel et al., 1999), and the likelihood of exercising in older individuals (Tate et al., 2015).

The linguistic savings hypothesis predicts that speakers of weak-FTR languages like Dutch or German will temporally discount less than speakers of strong-FTR languages like English. This implies that weak-FTR speakers will tend to make more future-oriented decisions, because they construe future rewards as relatively more valuable. Chen (2013) hypothesised that there were two distinct mechanisms which might explain this. Both involve differences in underlying beliefs about the temporal "location" of future events (Chen, 2013). The first is that the present tense conveys a sense of temporal immediacy. Since speakers of weak-FTR languages often use the present tense for FTR, Chen (2013) hypothesised this would cause them to perceive future events as relatively temporally proximal, *Fig. 2.1a.* Secondly, he hypothesised weak-FTR languages might not mandate that speakers think as precisely about the temporal location of future events, *Fig. 2.1b.* The idea is that strong-FTR languages divide the "arrow of time" into three segments (past, present, future). Weak-FTR languages divide it into two (past, present + future). This is suggested to cause more precise temporal representations of future events (Chen, 2013). If beliefs were affected in either of these ways, it would lead relatively less discounting in weak-FTR speakers, see *Fig. 2.1.* This would cause weak-FTR speakers to be more future oriented. In other words, relative to English speakers, Dutch and German speakers are hypothesised to make healthier decisions, save more money for the future, and be less likely to abuse addictive substances.

Chen (2013), tested these predictions by using FTR status in regression analyses to predict a range of behaviours. He found that speakers of weak-FTR languages were more likely to have saved each year, retired with more assets, were less likely to have smoked, and were more likely to practice safe sex. He also found they were healthier, as indexed by obesity, peak blood flow, grip strength, and physical exercise levels. Since then, a number of studies have extended this basic approach. For example, speakers of weak-FTR languages engaged less in present-oriented accounting practices (Fasan et al., 2016; J. Kim et al., 2017), had better educational outcomes (Figlio et al., 2016), made healthier lifestyle choices (Guin, 2017), had greater support for future-orientated environmental policies (Pérez & Tavits, 2017; Mavisakalyan et al., 2018), and had better macro-economic performance (Hübner & Vannoorenberghe, 2015b; Hübner & Vannoorenberghe, 2015a). A number of other studies attest to the conclusion that FTR status is a reliable predictor of intertemporal behaviour (S. Chen et al., 2017; Chi et al., 2018; Galor,

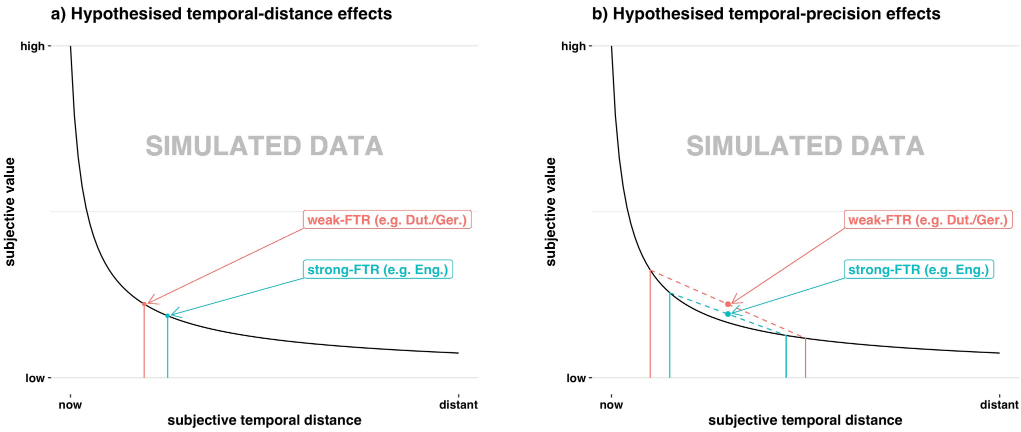


Figure 2.1: Mechanisms by which FTR grammaticisation is hypothesised to affect temporal beliefs and therefore discounting. Chen (2013) hypothesised that speakers of weak-FTR languages would construe future events as more temporally proximal (a), or less temporally precise (b). Consequently their subjective valuation of future outcomes would be higher (less discounted). We have presented the mechanisms in simplified terms. The distance mechanism is presented as a point estimate (a), and the precision mechanism is presented as the mean of a two-item uniform distribution (b). In Chen's (2013) account, temporal beliefs are represented using distributions, and subjective values are integrals. The precision mechanism (b) is contingent on two assumptions. First, weak-FTR beliefs are assumed to be a mean-preserving spread of strong-FTR beliefs. Complex empirical distributions might, in reality, mean effects of FTR status went in the opposite direction. Second, discounting functions are assumed to be strictly convex, which they generally are L. Green and Myerson (2004). If discounting functions were concave, effects of FTR status would also go in the opposite direction. The discounting function plotted is a hyperboloid function, $V = A/(1 + kD)^s$, from L. Green and Myerson (2004) where V is subjective value, A is the objective amount, D is the delay, b is a parameter that governs discounting rate, and s is a non-linear scaling factor typically less than 1. This function has been found to accurately describe empirical discounting rates in humans (L. Green & Myerson, 2004; L. Green et al., 2014; Du et al., 2002; Vanderveldt et al., 2015). Plotted values for s and k are approximately average human discounting rates for the given delay D (0–100 months) and value V ($200 in this case), i.e. $s = .7$ and $k = .4$ (from L. Green & Myerson, 2004).

Özak, & Sarid, 2016; Liang et al., 2018; Lien & Zhang, 2020; Roberts et al., 2015; Sutter et al., 2015; Thoma & Tytus, 2018).

2.1.2 Critical perspectives on the linguistic savings hypothesis

In this section, we outline three issues with the theory and evidence for the linguistic savings hypothesis. These are: (a) probability may be a confounding factor in observed effects of FTR status; (b) modal FTR expressions are disregarded despite being an import way of talking about the future; and (c) temporal accounts of the future tense disregard modal semantics of future tenses themselves.

2.1.2.1 Probability may confound observed findings

A serious issue is that (as far as we know), no work has directly tested the temporal mechanisms proposed by Chen (2013). Regression analyses which use FTR status to predict real-world intertemporal behaviour cannot identify whether temporal or probability discounting is driving outcomes. Probability discounting is analogous to temporal discounting. It refers to the subjective devaluation of outcomes as their odds of occurring grow lower (L. Green et al., 2014; Rachlin et al., 1991). For example, most people would prefer \$100, over a 50% chance of \$100. However, offer a 50% chance of \$200, and some will choose to gamble while others will chose the guaranteed \$100. Differences like this are referred to in terms of “risk preferences”. Recently, there has been an increasing interest in investigating outcomes which are both delayed and risky, e.g. \$100 or a 50% chance of \$200 in a year (Luckman et al., 2018; Vanderveldt et al., 2015, 2017). These are referred to as “risky intertemporal decisions.”

Many (if not all) of the behaviours found to be predicted by FTR status seem to involve risky intertemporal decision making. Even seemingly risk-free outcomes usually involve some degree of uncertainty. For instance, the pursuit of educational goals is fraught with uncertainty about their relative rate of return (Figlio et al., 2016), the discounting of future suffering in the context of support for euthanasia is permeated with uncertainty about the relative extent of future suffering (Lien & Zhang, 2020), and earnings management practices must be undertaken having given consideration to the probability of being caught (Fasan et al., 2016; J. Kim et al., 2017). Even the main finding in Chen (2013) involves predicting whether survey respondents had saved in the past year, which clearly could have involved investment in risky assets such as stocks and shares (World Values Survey Association, 2014).

In fact, since the future cannot be known at the time of speech, most real-world intertemporal decisions involve some degree of risk. For instance, someone deciding to smoke a cigarette must balance the immediate pleasure of smoking against her beliefs about *when* and *if* she may develop smoking-related health problems. Critically, probability and delay have been found to interactively predict subjective estimations of future value (Vanderveldt et al., 2015, 2017). Models which combine these factors fitted empirical results better than models which isolate them (Luckman et al., 2018). The probability of a reward had a greater impact on temporal discounting rates than delay has on probability discounting (Vanderveldt et al., 2015). These results support the conclusion that probability and delay interact complexly in how people make intertemporal decisions. This is a critical issue for the linguistic savings hypothesis. FTR status has been found to predict a range of behaviours. However, the nature of the outcomes makes it unclear why this is the case. Is probability or temporal discounting driving results?

2.1.2.2 FTR-status and modal future time reference

Chen (2013) uses obligatory tense marking of prediction-based FTR as a proxy for FTR grammaticisation (Chen, 2013). This may be reasonable (Dahl, 2000a), but *what* is it a proxy for? The expression of future time is very complex and often involves the expression of modal notions of ability, desire, (un)certainty, probability, volition, intention, and obligation (see Palmer, 2001; Bybee & Dahl, 1989; Bybee et al., 1994; Fries, 1956). Modality involves quantifying what is likely–unlikely, or possible–necessary, relative to various modal "bases". For instance, *deontic* modality involves expressing what is desirable or necessary relative to social norms, taboos, and institutions (Palmer, 2001), e.g. *One should always get up early.* In *epistemic* modality, speakers express what is likely relative to what they know or believe (Palmer, 2001), e.g. *I really think he's got a chance!*

The grammaticisation of FTR can involve multi-dimensional obligatorisation processes, which see many of these domains simultaneously become more grammaticised (Hopper, 1996). Epistemic modality may be of critical relevance to questions of psychological discounting. Risky intertemporal preferences are impacted by the perceived likelihood of a future outcome. If epistemic modality is more grammaticised, it could impact such perceptions. This is problematic, because FTR status affects the extent to which languages oblige the encoding of low-certainty epistemic modality. For instance, *will* is not actually obligatory for prediction-based FTR. Rather, English obliges speakers to use a modal verb:

(2.2) a. *The Bears will win (tonight).*
b. [If they get their defence together...][3] *...the Bears would win (tonight).*
c. *The Bears can win (tonight).*
d. *The Bears could win (tonight).*
e. *The Bears may win (tonight).*
f. *The Bears might win (tonight).*
g. *The Bears shall win (tonight).*
h. *The Bears should win (tonight).*

Any of example (2.2a–h) are perfectly acceptable. These modal verbs all encode futurity, but express differing speaker commitment to the probability of the event occurring. If the English case generalises, the salient difference between strong- and weak-FTR languages might be that strong-FTR languages oblige speakers to use a modal verb to encode *whether* they think an event will occur. If this results in net higher encoding of uncertainty, such linguistic spotlighting might cause increased probability discounting in strong-FTR speakers.

Critically, it is unclear whether there actually is higher encoding of low certainty in English FTR. It is plausible that the English obligatory modal verb system obliges higher encoding of low-certainty modality. However, weak-FTR speakers might be making up for language-level grammatical constraints by expressing low certainty in other ways. For instance, in English, Dutch, and German, epistemic modality can be expressed using modal modifiers, e.g. English *possibly, probably, certainly*; Dutch *mogelijk(erwijze)* 'possibly', *waarschijnlijk* 'probably', *zeker* 'certainly'; German *möglich/möglicherweise* 'possibly', *wahrscheinlich* 'probably',

[3] Context is given in [brackets].

bestimmt ‘certainly’. Mental state predicates might also help weak-FTR speakers make up the difference. These are psychological verbs which allow speakers to express modal notions by talking about their thoughts and beliefs (Nuyts, 2000). In English and Dutch, the mental state predicate prototypically used to express epistemic modality is *think* (Dutch *denken* ‘think’). In German it is *glauben* ‘believe’ (Nuyts, 2000).

There may also be modal FTR differences which cut across the FTR status dichotomy. For instance, Dutch and German utilise systems of modal particles which can express epistemic notions about the future. English does not. Additionally, the German modal system may be more complex than English or Dutch. In English, *may*, *might*, and *could* encode possibility, while *should* encodes slightly stronger notions of probability (Nuyts, 2000). In Dutch, *kunnen* ‘may’ is the only modal for which epistemic use is possible, and encodes possibility (Nuyts, 2000). The German case is complicated by the *Konjunktiv II*, a bound morphological subjunctive mood. While mostly used to express irrealis notions (Thieroff, 2010a), the *Konjunctive II* can also attenuate German modal verbs. It can alter their modal strength, or the modal base to which they are licensed to refer (Mortelmans, 2000; Nuyts, 2000). For instance, *mögen* ‘may’ encodes possibility, but only in the present tense *mag*. The *Konjunktiv II*, *möchte*, can only be used in deontic contexts (Nuyts, 2000). Conversely, *dürfen* ‘may’ encodes probability in the *Konjunktiv II*, *dürfte*, but not in the present, *darf*, which is deontic. *Müssen* ‘must’, expresses probability in the *Konjunctive II*, *müßte* ‘should’. In the present tense, *muss*, it is deontic (Mortelmans, 2000). Finally, *Können* ‘may’, expresses epistemic possibility in both the present, *kan*, and *Konjunktiv II*, *könte*, but the latter indicates a comparative weakening of commitment (Mortelmans, 2000). Do these complexities mean that German FTR is characterised by higher obligation to encode low certainty? It is unclear.

2.1.2.3 Do future tenses encode time or modality?

A second issue is that future tense markers tend to be characterised by a division of labour between temporal and modal semantics (Dahl, 2000a). For instance, some linguists have suggested *will* encodes modal weakening because it is used to mark predictions (Dahl, 2000a; Enç, 1996; Fries, 1956; Huddleston, 1995; Klecha, 2014). Others have suggested it encodes high certainty and is closer to a prototypical future tense (Salkie, 2010). These contrasting accounts entail obliging its use would impact risky intertemporal preferences in different ways.

To understand this discussion, it is necessary to understand what we refer to as “FTR-mode”. FTR-mode is a set of notions which are essential to understanding future time reference. They delineate the contexts in which it is possible to refer to future events. As we have mentioned, these are: (a) intentions, (b) predictions, and (c) schedules. We follow Dahl’s (2000a) useful schema by defining these categories as follows. Intentions are statements about our own or other people’s intentions for the future. Speakers can usually be fairly certain about their own intentions, e.g. *I shall see what's behind that door* expresses an intention. The speaker can be sure about the outcome because he has access to the internal contents of his own mind. Schedules are statements about well-known scheduled events. This FTR-mode is the highest certainty. When a speaker says *the game is at 6pm*, she can be highly certain that kick-off will be at the scheduled time (notwithstanding exceptional circumstances). Predictions are statements about less well-known events about which the speaker cannot be sure. For instance, *that coin will land on heads* is a prediction. This is the least certain FTR-mode. Outcomes are usually indeterminable at the time of speech, though they can be more or less uncertain. For instance, *Sea ice will melt by at least a meter in July* and, on looking out the window, *it will rain soon*, are both predictions, but they differ in implied certainty.

The modal semantics of the English *will*: The case that *will* encodes modal weakening usually involves pointing out that it becomes increasingly obligatory as the implied certainty decreases from schedules, to intentions, to predictions:

(2.3) a. *Sun rise is/?will be at 6am.*
b. *I set out/am setting out/will set out for the coast soon.*
c. *The bomb ?explodes/?is exploding/will explode at 6pm.*[4]

These sentences are syntactically similar, and all refer to the future. However, *will* becomes obligatory as the FTR-mode grows increasingly uncertain. In example (2.3a), *will* sounds out of place. It manages to convey an overly formal register, as though the speaker were announcing a hotel breakfast. While it might be acceptable, it does not seem standard. In example (2.3b), *will* does not serve strictly as a marker of future time. Rather, the meaning changes as a matter of stress. In *<u>I</u> will set out for the coast...*, the speaker will go (as opposed to someone else). In *I <u>will</u> set out for the coast...*, the speaker in fact going (as opposed to not going at all). Absent the expression of such notions, the present tense is likely more common. On the other hand, in example (2.3c), neither the present or the present progressive is grammatical. On the basis of acceptability judgments like these, it is usually suggested that *will* marks *prediction* rather than FTR (Enç, 1996; Dahl, 2000a; Huddleston, 1995; Fries, 1956; Klecha, 2014).

Such a conclusion is supported by the fact that it is perfectly acceptable to use *will* to mark a prediction in present time contexts. For instance, on hearing a knock at the door, it is grammatical to say either of:

(2.4) a. *That will be the postman.*
b. *That is the postman.*

[4] Bouma (1975).

In example (2.4a) *will* marks a present time prediction. This suggests that the semantics of *will* are not strictly temporal. Rather, *will* tends to mark predictions regardless of the time frame (*inter alia:* Enç, 1996; Giannakidou & Mari, 2018; Huddleston, 1995; Huddleston & Pullum, 2002; Klecha, 2014). Some commentators have pointed out that *will* may also operate as a marker of modal necessity, similarly to *must* (Giannakidou, 2017; Giannakidou & Mari, 2018). For instance, in example (2.4a), *will* expresses something similar to *that must be the postman.* A relevant point here is that such statements actually also express modal weakening relative to statements of fact (Giannakidou & Mari, 2018). In other words, *that must be the postman* implies that the speaker is inferring this, perhaps on the basis of relevant knowledge. If they knew it was the postman, the would just use example (2.4b).

For these and other reasons, most scholars agree that a purely temporal interpretation of *will* is inadequate, though the precise modal semantics of *will* are debated (Broekhuis & Verkuyl, 2014; Cariani & Santorio, 2018; Dahl, 2000a; Enç, 1996; Fries, 1956; Huddleston, 1995; Huddleston & Pullum, 2002; Klecha, 2014; Sarkar, 1998; Salkie, 2010). For instance, obliging the use of *will* for predictions may spotlight the uncertainty associated with this FTR-mode. The meaning of *will* may be associated with its use, i.e. it may "mean" epistemic weakening. On the other hand, there do not appear to be any convincing demonstrations that the modal weakening of *will* in example (2.4a) "carries over" when *will* is used to mark future predictions. It seems unclear that it would, given it is not possible to use the present tense for prediction-based FTR in English. In fact, as we have pointed out, it is not actually obligatory to use *will* in example (2.4a). English speakers are rather obliged to use one of *The bomb will/is going to/could/may/might/should/shall explode at 6pm.* A paradigmatic analysis of the options available indicates that *will* therefore encodes high certainty: It is among the highest certainty options available. This echoes suggestions that it is a marker of epistemic necessity (Giannakidou & Mari, 2018; Klecha, 2014).

The modal semantics of the Dutch *zullen*: Similar debates are had about the theoretical status of the Dutch future, *zullen* 'will'. Is it a modal or a tense? Broekhuis and Verkuyl (2014) make the case that its semantics are only modal. The authors point out that the Dutch present tense can be used to refer to a time span encompassing both before (using the present perfect) and after the time of speech. On this basis it is concluded that the contribution of *zullen* must be *purely* modal (Fehringer, 2018; Giannakidou, 2014, 2017). They give the following examples. The uncontroversial modal auxiliaries of possibility, *kunnen* 'may', and necessity, *moeten* 'must', are contrasted with *zullen* 'will':

(2.5) Dutch

a. *Dat huis op de hoek <u>moet</u> instorten.*
that house on the corner <u>must</u> collapse:PRS
'That house on the corner <u>must</u> be collapsing.'

b. *Dat huis op de hoek <u>kan</u> instorten.*
that house on the corner <u>may</u> collapse:PRS
'That house on the corner <u>may</u> be collapsing.'

c. *Dat huis op de hoek <u>zal</u> instorten.*
that house on the corner <u>will</u> collapse:PRS
'That house on the corner <u>will</u> be collapsing.'

According to Broekhuis and Verkuyl (2014) examples (2.5a–c) are all compatible with a future reading. However, given concurrent evidence of a collapse actually occurring (i.e. rumbling, visible instability), they can also refer to a present time event (Broekhuis & Verkuyl, 2014). If both present and future time interpretations are possible for *zullen*, they suggest that its primary contribution cannot be temporal and must be modal. Other linguists have suggested that *zullen* encodes both temporal and modal notions (Kirsner, 1969; also see: Janssen, 1989; Fehringer, 2018; Olmen, Mortelmans, & Auwera, 2009; Sluijs, 2011).

The modal semantics of the German *werden*: Similar controversies surround *werden* 'will' (see Bochnak, 2019; Janssen, 1989; Mortelmans et al., 2009; Olmen et al., 2009). As in Dutch, it is acceptable in German to use the present tense to talk about future intentions, schedules, and predictions (Bohnemeyer, 2000):

(2.6) German

a. *Wenn du diesen Pilz (jetzt) ißt, bist du morgen mittag*
if you this mushroom now eat:PRS be:PRS you tomorrow noon
tot.
dead
'If you eat this mushroom (now), you will be dead by noon tomorrow.'

b. *Nächstes J*ahr *schreibe ich ein Buch über Semantik.*
next year write:PRS I a book on semantics
'Next year I'm going to write a book on semantics.'

c. *Laut Fahrplan fährt der Zug um zehn.*
According schedule leaves:PRS the train at ten
'According to the schedule, the train leaves at ten.'

d. *Ich nehme dich mit nach Glasgow.*
I take:PRS you along to Glasgow
'I'll give you a lift to Glasgow.'

How can *werden*, the "future tense", be marking futurity, if such statements are acceptable? It must be marking something else, likely a modal notion.

Additionally, similarly to *will* in example (2.4a), *werden* constructions can also be used to mark present time predictions. For instance (Bohnemeyer, 2000):

(2.7) German

Er wird gerade an seinem Buch arbeiten.
He will:FUT right now on his book work:INF

'He will be working on his book right now.'

On this basis, it is possible to make the case that *werden* monosemously expresses prediction, or inferential modality, rather than future time reference (Bohnemeyer, 2000).

However, *werden* can also be used to mark future time (Heine, 1995):

(2.8) German

Er wird kommen.
He will:FUT come:INF

'He will come.'

Since *werden* can be used as in example (2.8), a natural argument is to treat it as a future tense (Heine, 1995). However, critics have pointed out that it is unreasonable to expect monosemous semantics (see Bohnemeyer, 2000). Polysemous analyses may better capture general qualities of language (Bohnemeyer, 2000). It may be reasonable to conclude that, like *will*, the semantics of *werden* are characterised by a division of labour between temporal and modal notions. In fact, this is a general characteristic of the future tense (Dahl, 2000a).

Comparing future tense semantics in English, Dutch, and German: As with *will*, the exact nature of the semantic contribution of *zullen* and *werden* are difficult to pin down. Broekhuis and Verkuyl (2014) suggest *zullen* constitutes marking of an expected or "projected" future. A paradigmatic analysis is useful. In Dutch it is possible to use the present tense for prediction-based FTR (Dahl, 2000a). It therefore seems unclear whether Dutch present or future tense FTR encodes greater certainty. If even epistemic necessity is modally weaker than statements of fact, this suggests that *zullen* encodes modal weakening relative to present tense FTR. On the other hand, relative to *kunnen* 'may', *zullen* appears to encode higher certainty. *Werden* is similarly characterised by multiple possible meanings. For instance, it can be used to express epistemic modality, intentions, and imperatives, e.g. similarly to *will*: *Du wirst jetzt schlafen gehen!* 'You will go to bed now!' (Heine, 1995).

There are key differences. For instance *will*, grammaticised from the proto-Germanic *willan* 'to want' (Dahl, 1985). Glimmers of these archaic semantics can still be seen in low-frequency uses, e.g. *I will that she agrees to my proposal.* On the other hand, *werden* grammaticised from a root meaning *to become* (Petré, 2013), or perhaps *to turn* (Mortelmans et al., 2009). It still retains aspects of these semantics (Heine, 1995):

(2.9) German

Er *wird* *alt.*
He will old.

'He is growing old / he is becoming old.'

This suggests *werden* retains some of the semantic flavours of its lexical antecedent, and can be used to indicate processes of change and/or becoming. What such differences might mean for the linguistic savings hypothesis is unclear.

It is beyond the scope of this discussion to evaluate the various positions linguists have taken in these debates, or present new theoretical arguments about the semantics of *will*, *zullen*, or *werden*. The relevant point is simply that relativity accounts of how FTR grammaticisation may impact (risky) intertemporal decisions need to confront evident complexities. Critically, Chen's (2013) arguments ignore the division of labour between temporal and modal semantics—whatever they are—which often characterises future "tenses". We have outlined plausible arguments that *will*, *zullen*, and *werden* could encode both modal strengthening and weakening. Which of these accounts is closer to reality has important implications. If modal weakening is encoded, obligatory future tenses should cause speakers to perceive the future as less certain. They would therefore discount more. If modal strengthening is encoded, future outcomes might be construed as more certain. Speakers would therefore discount less. Additionally, cross-linguistic differences in future "tense" semantics undermine Chen's (2013) argument that obligatory use of the future tense should impact speakers of different languages in the same way. In some languages the future tense might encode modal strengthening. In others, it might encode modal weakening. In fact, FTR status may be a relevant determinant of such differences. The English future tense is the highest certainty construction possible for future predictions. On the other hand, *zullen* and *werden* may be paradigmatically contrasted present tense FTR. Relative to such unmarked statements of fact, any modalisation is weaker. The paradigmatic oppositions of future tenses may therefore differ as a function of the cross-linguistic differences indexed by FTR status.

2.1.3 Implications for linguistic relativity

In summary, probability and delay impact subjective estimations of value (Białaszek et al., 2019; L. Green et al., 1999; L. Green & Myerson, 2004; Rachlin et al., 1991). There is a robust typological tendency for future tenses to encode modal notions (Broekhuis & Verkuyl, 2014; Bybee et al., 1994; Enç, 1996; Fries, 1956; Giannakidou, 2014; Giannakidou & Mari, 2018; Huddleston, 1995; Nuyts & Vonk, 1999; Sarkar, 1998). Modals are cross-linguistically common FTR structures which allow speakers to encode degrees of epistemic commitment to future events. Strong FTR languages may oblige the use of such structures.

In other words, FTR tends to entangle the notional domains of time and probability, and both domains impact subjective estimations of value. Research which isolates only one of these factors (time) may be producing biased results due to unmeasured confounding variables (probability). Alternatively, the grammaticisation of modality may actually be driving reported results. At the same time, the extent to which the encoding of future probability is obligatory in strong-FTR languages is not known (as far as we know). As we have pointed out, modal systems are flexible enough to permit lexical workarounds. Additionally, arguments among linguists have not resolved questions as to the modal semantics of future tenses, despite this having implications for the linguistic savings hypothesis. Therefore, these factors should be studied in a sample of both weak- and strong-FTR languages. This is what we undertook to do in Study 1.

2.2 Study 1

To establish FTR language use we created an FTR-elicitation task based on Dahl's (1985, 2000a) FTR questionnaires. Chen's (2013) FTR status dichotomy is largely based on work by Dahl and colleagues in the EUROTYP Working Group on Tense and Aspect (Dahl, 2000b), so this was an appropriate starting point. In this task, participants were given a context and a target sentence. The main verb in the target sentence was unconjugated, and participants were asked to render the target sentence given the context. All contexts referred to future events. We made several modifications to the original questionnaire. In addition to creating many new items, we modified the contexts to include information of the likelihood of the referenced event occurring. This change was made in order to elicit modal future-referring language. We refer to this as the "modality condition". In order to elicit language from a wide variety of contexts, we included a range of temporal distances from time of speech, as well as examples from each FTR-mode.

The principal hypothesis we wanted to test was whether English speakers used more low-certainty constructions. English modal verbs may be generally more obligatory than Dutch or German modals (Bybee et al., 1994; Nuyts, 2000). Additionally, English FTR is generally more grammaticised (Chen, 2013; Dahl, 1985, 2000a). Notwithstanding noted complexities of the German modal system, it seemed reasonable to hypothesise that English speakers would use more low-certainty language. We refer to this as the English uncertainty hypothesis.

2.2.1 Methods

2.2.1.1 Participants

The final Study 1 sample was $N = 1069$ participants ($n = 459$ in English [$n = 257$ female, $n = 195$ male, $n = 7$ other], $n = 446$ in Dutch [$n = 298$ female, $n = 147$ male, $n = 1$ other], and $n = 164$ in German [$n = 93$ female, $n = 71$ male]). Two versions of the same experiment were conducted with different groups of participants ("Version A" and "Version B"). See below. Data for Study 1 was generated by combining data from Versions A and B (Version A sample size: $n = 45$ English, $n = 19$ Dutch, and $n = 164$ German; Version B sample size: $n = 414$ English, and $n = 427$ Dutch). Some participants needed to be excluded because their answers were obviously given in bad faith, see Appendix C, and 6 participants needed to be excluded because of missing demographic data. Sample sizes above are given after these exclusions. All participants were native speakers of the survey language. Data collection was between July 2016 and May 2020. Native English speaking participants were recruited from the University of Oxford and from Prolific Academic. Native Dutch speaking participants were recruited from the participant research panel of the Max Planck Institute (MPI) for Psycholinguistics, Nijmegen, from Radboud University's research panel, and from participant panels at Prolific Academic, and Qualtrics. Native German speaking participants were recruited from the MPI participant research panel, and using a Qualtrics panel. Ethical approval for the study was granted by the University of Oxford internal review board, ref. no. R39324/RE001, and by the MPI ethical review committee. All participants were remunerated.

2.2.1.2 Materials

Study 1 consisted of one task: an FTR-elicitation task designed to establish free-text future-referring language.

The FTR-elicitation task: In this task, participants are given a "context" and a "target" sentence. The main verb in the target sentence was not conjugated. The participants' task was to conjugate these given the context. For example,

Context: Chris's brother {SEND} him some money next month. When he get it...

Target: ...he {SPEND} it at the bar.

A typical response might be *He is going to spend it at the bar.* Prior to starting, participants were given three learning items with example responses. These were given in the past tense in order to avoid biasing participants. Because linguistic relativity effects are most likely to arise as a function of how people actually use language, not as a function of how they believe it is supposed to be used, participants were advised that there "were no correct answers," and that they should complete the questionnaire sentences, "as though they were speaking to a close friend." After this, participants completed the FTR questionnaire. In Version A, participants completed the full 48-item task. In Version B, participants completed a subset of 29 of these items (see below). Item order was randomised and one item was displayed per page.

The FTR-elicitation task had three within-subjects factors. The first factor was distance from present, i.e. temporal distance from the time of speech (today, tomorrow, one week, one month, six months, one year, two years, ten years, twenty five plus years, ongoing prediction, and indeterminate). Temporal distance was operationalised using temporal adverbials in the contexts, e.g. "one month", "in a year", etc. An example of an ongoing/indeterminate question is "If you eat that mushroom you DIE". The second factor was modality condition (high-certainty, low-certainty, and neutral). In the high-certainty condition, extra information was added to the context indicating that the speaker was supposed to be highly certain about the event, e.g. "I leave tomorrow (100% decided)", or "I leave tomorrow (definitely)". This information was included in (parentheses). In order to avoid participants simply copying this material, they were directed beforehand that they should "not include material within parentheses". In the low-certainty condition, this material indicated the speaker was supposed to be uncertain, e.g. "It {RAIN} tomorrow (probably)", or "It {RAIN} tomorrow (50% chance)." In the neutral condition, no certainty information was included, e.g. "It {BE} cold next week." The third factor was FTR-mode (i.e. whether the context involved a prediction, or an intention[5]). We followed Dahl's (2000a), schema by defining predictions as situations in which the future outcome is not under the control of the speaker, intentions as statements about how the speaker intends to behave in the future, and schedules as about well-known firmly-scheduled events. Some ambiguity exists when a speaker is talking about a third person's intentions. This is because a speaker can make a prediction about, or simply report, a third person's intentions. We therefore defined third-person intentions as those situations where the speaker knows the third person's intentions, e.g. "(John told me) he {WALK} next weekend." We defined third-person predictions as those situations where the speaker does not know the third person's intentions, e.g. "Q: 'What will Mary do about it, do you think?', A: 'She {RESIGN}!"' See *Table 2.1* for examples. See Appendix I for items. Because cross-linguistic differences are expected to be most marked for prediction-based FTR, most contexts were predictive. This meant that conditions were not balanced (see Appendix C).

Text classification: There were $N = 28,110$ text responses in Study 1. It was therefore necessary to automate the scoring of responses in terms of whether they used the present tense,

[5] After the item exclusions described above, no scheduling items remained.

Table 2.1

Examples of FTR-elicitation task conditions (Study 1)

variable	level	distance	item
FTR-mode	intention	1 year	[Q: You {STAY} here, next year?] A: No, I {LIVE} in San Francisco (next year).
	prediction	1 month	[Father to son: If you {PUT} your allowance in a savings account⋯] ...next month it {BE} worth more.
	scheduling	1 day	The sun {RISE} at six O'clock tomorrow.
modality condition	neutral	2 years	[Don't buy those shoes...] ...they {WEAR OUT} in a couple of years.
	low-certainty	10 years	[Don't bother investing in real estate...] ...the housing market {COLLAPSE} (in the next ten years, probably).
	high-certainty	1 week	[Q: When does your mother arrive? A: Ug. In a week. She {INSIST} on {BUY} the kids too much sugar, like always⋯] ...I {SAY} no this time (definitely).

Following Dahl's (1985, 2000a) conventions, contexts are given in [brackets], target setentences thereafter.

future tense, or some kind of modal expression. To accomplish this, we wrote a keyword-based, deterministic, closed-vocabulary classification program written in Python (Python Software Foundation, 2017). We refer to it as the FTR-type classifier. It comprises a number of word lists which are used in combination with a set of rules to classify text items according to which tense and/or modality words they contain. The FTR-type classifier categorises text data into four exclusive semantic categories: future tense, present tense, low-certainty, and high-certainty. The latter two are further divided into two non-exclusive categories based on whether a modal verb or some other construction type is used (see below). Each category is coded with (1) to indicate a response is a positive example, otherwise (0). These comprise the dependent variables for this task.

In English, Dutch, and German, modal words can be used in combination with the future tense and present tense. For example, *It will probably rain* and *It will definitely rain* are both future tense, but different epistemic commitments are expressed. Similarly, *They could win tonight* and *The game definitely is at 7* are both present tense but different modal notions are expressed. Since we were not interested in formal tense structure, and were rather attempting to explore differences in marking of the notional domains involved, it was appropriate to have epistemic modal morphemes "dominate" tense morphemes. Specifically, responses which used both tense and modal words, were classed as low-certainty (or high-certainty), and <u>not also</u> as future or present tense. We outline the FTR-type classifier categories classification system below (see Appendix C for details).

Present tense: Responses were classed as present tense if they conjugated the main verb in the target sentence using the present tense, and also failed to be classed as any of the other categories.

Future tense: Responses were classed as future tense if they used commonly accepted "future" auxiliaries or explicit temporal adverbials (English *will, shall, be going to, about to*; Dutch *zullen* 'will', *gaan* 'be going to', *staat op* 'about to'; German *werden* 'will'). Any response exhibiting these words, without additional modal epistemic words, was counted as future tense.

Verbal-low-certainty: Responses which used low-certainty modal verbs were classed as verbal-low-certainty (English *can, could, may, might, should*; Dutch *kunnen* 'may'; German *mag* 'may', *dürfen* 'may', *müßte* 'should', *können* 'may'). A prototypical example is *This team might/may/could/should win tonight.*

Verbal-high-certainty: Responses which used modal verbs which encode high certainty were classed as verbal-high-certainty (English *must*; Dutch *moeten* 'must'; German *müssen* 'must'). A prototypical example is, *I must remember to take in the laundry*, although this suggests a deontic or bouletic base (i.e. having to do with obligations or desires, respectively). In fact, clearly epistemic contexts in which *must* sounds natural in English are difficult to find, e.g. *The test tonight must be difficult* seems to again suggest a bouletic rather than epistemic base. We nonetheless include it as the only criteria for the verbal-high-certainty category.

Other-low-certainty: Responses which used modal expressions indicating low-certainty (apart from modal verbs) were classed as other-low-certainty. This includes low-certainty modal modifiers (English *possibly, probably, potentially*, etc.; Dutch *misschien* 'perhaps', *mogelijk* 'possibly', *waarschijnlijk* 'probably', *wellicht* 'maybe', etc.; German *wahrscheinlich* 'probably', *möglicherweise* 'possibly', *möglich* 'possibly', etc.). A prototypical example of a modal modifier encoding low-certainty FTR is *It will possibly rain tonight.* It also includes low-certainty mental state predicates (English *think, believe, reckon*, etc.; Dutch *denken* 'think', *annehm* 'assume', *veronderstellen* 'suppose', etc.; German *meinen* 'mean', *glauben* 'believe', *denken* 'think', etc.). A prototypical example might be, *I think it's going to be hard win.* Finally, it also includes low-certainty epistemic modal particles (Dutch *wel eens, wel*, approximately 'well be', 'well'; German *wohl* approximately 'well').

Other-high-certainty: Responses which used modal expressions which encode high certainty (apart from modal verbs) were classed as other-high-certainty. This includes modal modifiers (English *certainly, definitely, absolutely*, etc.; Dutch *zecker* 'certainly', *definitief* 'definitely', etc.; German *definitiv* 'definitely', *bestimmen* 'certainly', *sicherlich* 'certainly', etc.). A prototypical example is *The storm will definitely hit the east coast this week.* It also includes high-certainty modal particles (Dutch *toch*, approximately 'fixed', 'firm').

To check whether the FTR-type classifier was accurate, expert human raters classified a subset of the responses, $N = 1443$ ($n = 675$ in English, $n = 675$ in Dutch, and $n = 93$ in German). When systematic differences were found, the FTR-type classifier was adjusted to better replicate human judgement. After this process was complete, accuracy metrics against the test data were all $> .96$ (see Appendix C).

The FTR-type classifier cannot accurately classify responses which use negations, or responses which use words from two conflicting class criteria keyword lists. We refer to these as "mixed modal" responses. In the first instance, modal keywords switch polarity in the presence of negations. For instance, *I'm <u>not certain</u> it will rain tomorrow*, expresses low certainty. However, because of the presence of the high-certainty class-criterion keyword *certain*, it would be classed as high-certainty. Similar in-determinability characterises mixed modal responses. For instance, *Rain tomorrow is <u>certainly possible</u>* expresses moderate certainty, but would be classed as both other-high-certainty and other-low-certainty because of the present of the class-criterion keywords *certainly* and *possible*. Since such responses were in practice low-frequency, our strategy was simply to exclude them from data analysis. We therefore detected the presence of negations using an averaged perceptron tagger following Collins (2002) but with Brown cluster features as described by Koo, Carreras, and Collins (2008) and using greedy decoding (implemented in *spaCy* (Explosion AI, 2020)). We excluded any sentences which used negations ($n = 649$), and any responses which the FTR-type classifier classed as both high- and low-certainty ($n = 137$). This left a final sample of $N = 27,394$ text responses ($n = 70$ items used both mixed modality and a negation).

2.2.1.3 Procedure

The questionnaire was hosted on the Qualtrics survey platform, and all participants completed the survey online. The within-subjects factors for each task are described above. There was one between-subjects factor: survey language. There were three (otherwise identical) versions of the survey: one in English, one in Dutch, and one in German. Following research practice to ensure translational equivalency, Dutch and German versions surveys were created by first translating from (British) English into Dutch and German, and then back-translating into English. Translation was done by trained experts. Where this process produced any substantial differences between the original and back-translated English versions, Dutch and German translations were adjusted (Brislin, 1970). At the beginning of the surveys, participants confirmed their native language matched the language of the survey. English speakers confirmed they were native English speakers, Dutch speakers confirmed they were native Dutch speakers, and German speakers confirmed they were native German speakers. If they did not, they were immediately ejected from the survey. Following this, they answered some demographic questions (age, sex, income, education, marital status, and employment status), which were recorded as control variables. After this, they completed the FTR-elicitation task.

2.2.1.4 Version A details

The full FTR-elicitation task comprises 48 questions, but for various reasons some items needed to be excluded. In Version A, human errors uploading the questionnaire to Qualtrics resulted in several questions being accidentally not included. Since these did not always match between language conditions, we created a matched sample of $n = 42$ items by taking the intersection of items across language conditions. Additionally, some items needed to be excluded from analyses for several reasons. Firstly, inclusion of tense or modal words in the contexts (*go, think*, etc.) made it impossible to identify whether participants had used the words themselves or had copied the context. Secondly, for one item, it was impossible to tell whether participants were referring to present or future time.[6] After these exclusions, the Version A data comprised a matched sample of $n = 35$ items. Additionally, for German data collection which occurred in 2019, 3 items needed to be excluded because a specific temporal adverb ("by 2018") was in the past at the time of data collection (see Appendix C).

2.2.1.5 Version B details

A critical difference between versions is that Version A sampled English, Dutch, and German participants. Version B sampled only English and Dutch participants. This is because preliminary analysis of Version A indicated that English and Dutch most differed in terms of FTR grammaticisation. Version B introduced no new items. However, some were removed, meaning the items in Version B were a subset of (29) items from Version A. Given that no new items were introduced, we combined data from Version A and B for a unified analysis in Study 1. To check whether this was justified, we included a fixed effect for study version. We found this did not improve model fit, $p > .05$, which indicates that study version did not impact outcomes. The subset of items in Version B were selected because preliminary analysis of Version A indicated responses for these items were least driven by "item effects" and most by "participant effects". In other words, these were the language-matched sample of questions for which answers were least constrained by context and most driven by differences between participants (see Appendix C). Five items needed to be excluded from Version B for the same reasons as explained in Version A. The final number of items included in analysis was therefore $n = 24$ (see Appendix C for items).

2.2.2 Results

We present an overview of results in *Fig. 2.2.* English speakers used more future tense and fewer present tense constructions. This reflects well-known differences between English

[6]The item was: "[Q: Do you think your dad will go to sleep?] A: Yes, he {BE} tired." (context is given in [brackets]). Participants overwhelmingly responded with e.g. *Yes he is tired.* It is impossible to tell whether this means that he is tired *now* or that he will be tired *later*.

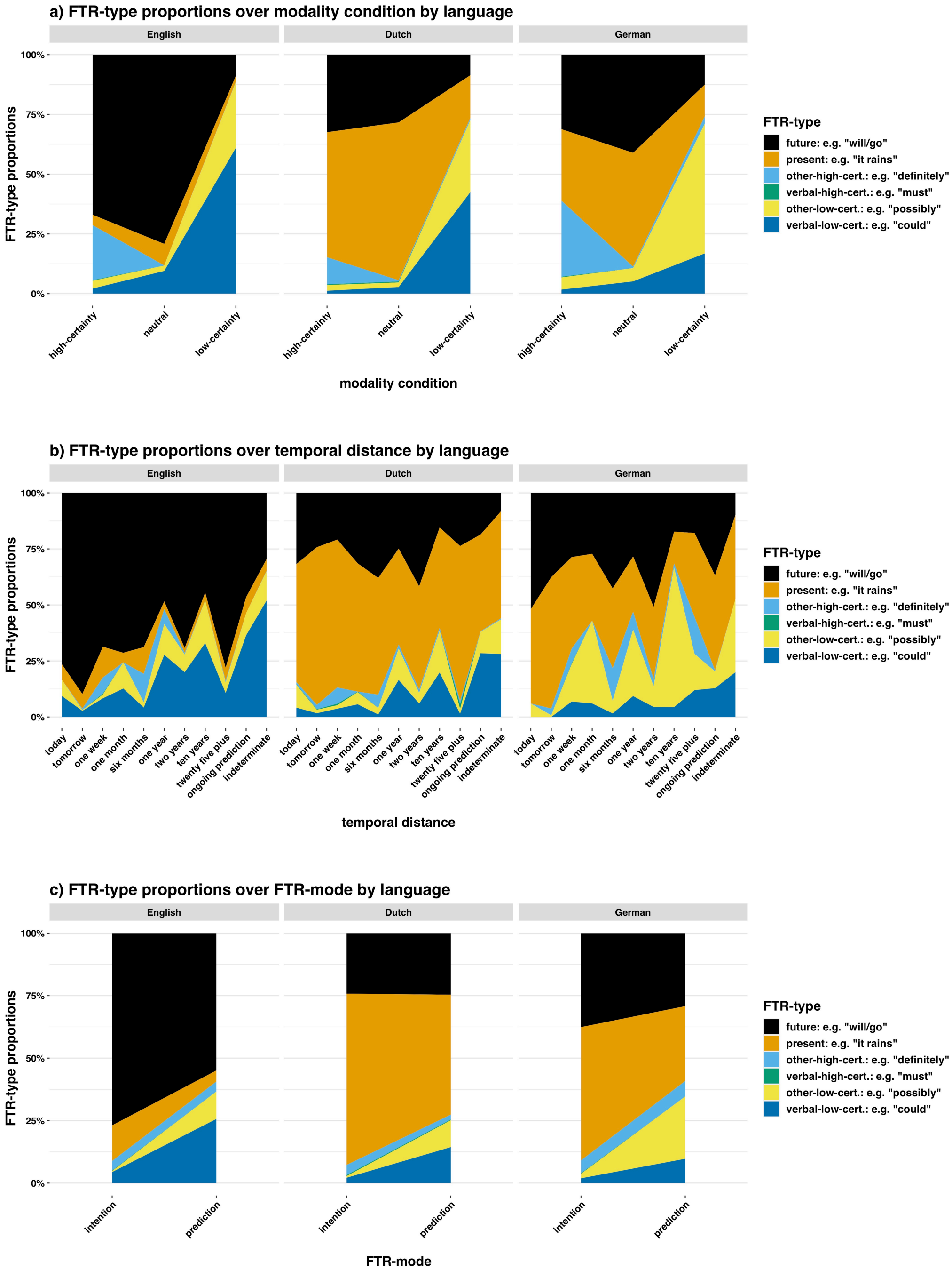


Figure 2.2: FTR-type proportions over modality condition, temporal distance, and FTR-mode (Study 1). Depicted values above and in *Fig. 2.3* are the proportions per language and FTR-task factor, i.e. means per factor level normalised to sum to 1 ($x = x_i / \sum_n x$).

on the one hand and Dutch and German on the other, i.e. FTR status. Additionally, English speakers appeared to use more low-certainty language than Dutch or German speakers. This was mostly driven by modal verb use, e.g. *It could/may/might rain.* German speakers made up for this to some degree through use of other-low-certainty constructions, *Fig. 2.2.* All three languages appear to use more low-certainty language for predictions, though this seemed most pronounced for English, *Fig. 2.2c.* It appeared to be the case that low-certainty language use increased with temporal distance, particularly in English and German, *Fig. 2.2b.* However, a clear pattern is difficult to discern: In the 25+ years category, very few low-certainty terms were used.

2.2.2.1 The English uncertainty hypothesis

With the English uncertainty hypothesis we had predicted that English speakers would use more low-certainty language than Dutch or German speakers. To test this, we combined verbal-low-certainty and other-low-certainty into a single dichotomous variable ("low-certainty") which was (1) for any response which encoded low certainty, otherwise (0). For example, responses like *I think/believe/guess it will rain*, *It will possibly/probably/potentially*, and *It could/might/may/should/can rain* would all be classed as low-certainty (1).

Multilevel modelling was appropriate, as responses from a single participant were likely to be similar across different items, and responses to a single item were likely to be similar across different participants. We followed research practice by building models sequentially and using log-likelihood ratio tests to ascertain whether adding variables improved model fit (Raudenbush & Bryk, 2002; Twisk, 2006; Aguinis, Gottfredson, & Culpepper, 2013; Legler & Roback, 2019). Using generalised linear regression with a logit link function (logistic regression), we regressed binary (0,1) low-certainty language over a fixed intercept, and then allowed intercepts to randomly vary by item and participant. We added a fixed effect for language, and allowed slopes for language to vary by item. All of these steps were significant, $p < .001$. By modelling such variance, we were able to estimate parameters of fixed effects independently of item-by-language-level and participant-level idiosyncrasies.[7]

As we had predicted, we found that both Dutch, $e^{\beta} = 0.39$, $SE = 0.2$, $z = -4.65$, $p < .001$, and German, $e^{\beta} = 0.38$, $SE = 0.33$, $z = -2.97$, $p = .003$, speakers were significantly less likely to use low-certainty language. Coefficients are exponentiated, so represent changes in the odds ratio of using a low-certainty term. This means that Dutch speakers used low-certainty terms approximately 1/3 as much as English speakers. German speakers did so a little less than 2/3 as much. These results support the English uncertainty hypothesis.

[7]We found there were significant between-language differences for a number of the demographic covariates (see Appendix C). In unreported analyses, we nonetheless included them as covariates but found that their inclusion did not improve model fit, $p > .05$, or substantively change parameters of interest. We therefore disregarded them.

2.2.3 Discussion

English speakers used more low-certainty language, and this was mostly due to higher low-certainty modal verb use. This suggests that English grammatical constraints which oblige modal verb use for prediction-based FTR are responsible for higher net encoding of low-certainty modality. This has obvious implications for the linguistic savings hypothesis. Such differences could cause English speakers to probability discount, rather than temporally discount. This could be responsible for observed results.

However, there were subtleties which we were not able to test due to the imbalanced factors of the FTR-elicitation task. In particular, we had to exclude scheduling items, and there were only a small number of intention items. Moreover, these factors were not balanced with temporal distance, or modality condition. This meant that it was impossible to identify which factor was driving results. Intriguing patterns in the descriptive analyses could not be statistically explored. In particular, it appeared to be the case that English and German—but not necessarily Dutch—speakers used more low-certainty terms as temporal distance increased, *Fig. 2.2b*. However, factor imbalance made it difficult to conclude much.

Finally, there were differences between Dutch and German results. Dutch speakers used more present tense constructions than did Germans, *Fig. 2.2*, despite the fact that both languages are weak-FTR. Relative to Dutch speakers, Germans also used more other-low-certainty constructions, and fewer verbal-low-certainty ones, *Fig. 2.2*. Evidently, the noted complexities of the German modal verb system had not resulted in higher use of modals to encode low certainty. Despite this, German encoding of low certainty did not appear massively different from English, even if it was significantly so. Other-low-certainty structures were driving this effect, *Fig. 2.2*. These are less grammaticised than modal verbs: They are not obligatory. This suggests that cultural rather than grammatical differences are involved. Whatever the case, the sharper contrast in terms of FTR grammaticisation appeared to be between English and Dutch speakers.

We therefore decided to conduct a follow up, focusing on English and Dutch speakers. We used a re-written version of the FTR-elicitation task designed to address the various noted weaknesses.

2.3 Study 2

We made several important changes in Study 2. Firstly, we balanced the factors of the FTR-elicitation task and addressed weaknesses which had necessitated excluding some items. These changes allowed us to estimate the unbiased effect of the factors of the FTR-elicitation task, as well as estimate interaction effects between the factors.

We also added two novel measures which allowed us to establish whether future tenses encode temporal or modal notions. In the first instance, participants rated FTR structures in terms of whether they perceived them to be temporally distal or temporally proximal. In the second, they rated FTR structures in terms of whether they perceived them to encode high- or low-certainty. We made several predictions.

Predictions about FTR-mode: It had appeared in Study 1 that, relative to Dutch speakers, English speakers were more likely to mark predictions with a low-certainty modal marker, *Fig. 2.2c.* Factor imbalance had made this impossible to test. We therefore predicted that English—but not Dutch—speakers would be more likely to use low-certainty language for prediction-based FTR. We refer to this as the uncertain predictions hypothesis.

Predictions about modality condition: It had appeared in Study 1 that English speakers were more sensitive to modality condition: Relative to Dutch speakers, they used more low-certainty constructions in the low-certainty condition, *Fig. 2.2a.* Again, experimental design had made this difficult to test. We therefore predicted that use of low-certainty language would be higher for English participants in the low-certainty condition. We refer to this as the low-certainty-sensitivity hypothesis, i.e. because English speakers are predicted to be more sensitive to the low-certainty condition.

Predictions about effects of temporal distance in the FTR-elicitation task: The possible effect of temporal distance on low-certainty language appeared more pronounced in English than in Dutch, *Fig. 2.2b.* Did English speakers also use more low-certainty constructions when they talked about temporally distant events? This was an intriguing possibility. We reasoned that if English speakers use more low-certainty FTR language, this could, over time, lead to stronger cross-modal mapping between temporal distance and low certainty. We therefore predicted that English speakers would use more low-certainty language as a function of temporal distance, but this would not be true of Dutch speakers. The idea here is not that English speakers use modal markers to encode notions of temporal distance. Rather, they might perceive temporally distant events as uncertain, and use low-certainty language in light of this. We refer to this as the English cross-modal-mapping hypothesis.

Predictions about temporal-distance ratings: We made two predictions about temporal-distance ratings. On the basis of the linguistic savings hypothesis, we predicted that: (a) Future tenses would be rated as more distant than present tenses; and (b) that Dutch participants would construe future events as more proximal. We refer to these predictions as the linguistic-savings-distance hypotheses.

Exploratory analyses: With regard to ratings of high- vs. low-certainty, we did not make any hypotheses. Rather, we chose to conduct exploratory analyses.

2.3.1 Methods

2.3.1.1 Participants

The final Study 2 sample was $N = 651$ participants ($n = 330$ in English [$n = 165$ female, $n = 162$ male, $n = 3$ other], $n = 321$ in Dutch [$n = 161$ female, $n = 159ccc$ male, $n = 1$ other]). This is after one participant was excluded because their age datum was missing. Data were collected between September and November 2019. English participants were recruited from Prolific Academic and Dutch participants were recruited from Qualtrics. Participants were native English and Dutch speakers currently residing in the UK and the Netherlands. The sample was matched to UK population norms for age and sex. Ethical approval for the study was granted by the Oxford internal review board, ref. no. R39324/RE001. All participants were remunerated.

2.3.1.2 Materials

Study 2 comprised three tasks: (1) the updated FTR-elicitation task, (2) a subjective-temporal-distance task designed to establish whether the tense of an FTR statement (future vs. present) impacted participants' construals of future temporal distance, and (3) a subjective-certainty task designed to establish whether participants construed common FTR structures as encoding high or low certainty.

The FTR-elicitation task: This task was similar to the FTR-elicitation task in Study 1. Participants were again given a context and a target sentence and were tasked with typing in the conjugated target sentence. Before starting, participants were advised that there "were no correct answers," and that they should complete the questionnaire sentences, "as though they were speaking to a close friend." They were given two training items with example responses, and one trial item where they typed in a response. These were in the past tense in order to avoid biasing participants. See Appendix I for items. There was one attention check: At a random point, participants were instructed to enter the word "dance" (Dutch "dans"). If they failed to do this, there were ejected from the survey immediately.

There were three within-subjects factors in the task: FTR-mode (predictions, intentions, scheduling); modality condition (high-certainty, low-certainty, neutral); and temporal distance (one month, two months, three months, six months, one year, five years). As in Study 1, temporal distance was operationalised using temporal adverbials in the contexts. The operationalisation of modality condition was different. Rather than including verbal cues in parentheses, participants were given numerical "certainty information" above each target sentence, for example:

Table 2.2

Example of the revised FTR-elicitation task conditions (Study 2)

modality condition	FTR-mode	certainty information	item
high-certainty	intention	100% decided	[Jen's uncle {SEND} her some money next month. She just loves skiing. When she gets it⋯] ⋯she {BUY} a new pair of skis.
	prediction	100% certain	[A to B: Don't invest in commodities. The market is very shaky...] ⋯it {CRASH} within a month.
	scheduling	100% certain	[In June: Q: When do you and Jen fly to France? A: She {SAID} it {BE} the 17th, but I just {CHECK} the schedule...] ⋯we {LEAVE} on the 15th.
neutral	intention	-	[Chris's father {SEND} him some money next month. When he {GET} it⋯] ...he {BUY} a present for Amelie.
	prediction	-	[A to B: Don't invest in derivatives. The market is fraudulent...] ...It {CRASH} within a month.
	scheduling	-	[In November: Q: When do you fly to Mexico?] A: My flight {LEAVE} on 15 December!
low-certainty	intention	50% certain	[Chris's brother {SEND} him some money next month. You never know with him. When he gets it⋯] ...he {SPEND} it at the bar.
	prediction	40% certain	[A to B: Don't invest in Latin America. Conditions are unstable⋯] ...Brazil {CRASH} within a month.
	scheduling	60% decided	[In February: Q: When do you fly to Spain? A: I have to check the ticket...] ⋯my flight {LEAVE} 15 March.

Context: Chris's brother {SEND} him some money next month. You never know with him... When he get it...

Certainty: 50% certain.

Target: ...he {SPEND} it at the bar.

Prior to starting, participants were told "there will be some 'certainty information' included in the context". They were informed that, "this indicates how certain you are about what you are saying". They were then directed to, "please imagine you are this certain and write down what you would say". For schedules and predictions, participants were told they were supposed to be "____% certain", and for intentions they were told they were supposed to be "____% decided" (this was because it was difficult to make "certain" agree with all intention contexts). In the low-certainty condition, certainty information varied between 40%, 50%, and 60%. This was implemented to try to maintain participant engagement. In the high-certainty condition, certainty information was invariably 100%. In the neutral condition no certainty information was given. We changed our definition of FTR-mode slightly, and counted as intention any intention statement whether it was first or third person. This was to try to isolate elicited language in more prototypical prediction contexts. *Table 2.2* presents example items from a single frame distance—one month—with one example from each modality condition and FTR-mode. See Appendix I for questionnaire items.

A critical difference is that the updated FTR-elicitation task had balanced factors. In each temporal-distance by modality condition there were 5 items: 3 prediction items, 1 intention item, and 1 scheduling item. This means there were 15 items per temporal distance, $3_{prediction} \times 3_{prob.cond.} + 1_{intention} \times 3_{prob.cond.} + 1_{scheduling} \times 3_{prob.cond.} = 15$, and 90 items in total, $5_{temp.dist} \times 5_{FTR-mode} \times 3_{prob.}$. In order to address the possibility that idiosyncratic aspects of items were driving elicited language, we systematically varied certainty information across highly similar items (unavoidable differences between FTR-modes prevented base items from being altered across FTR-mode). Semantic details (nouns, names, pronouns) were altered, but other details were only changed to ensure the certainty information did not clash with the certainty implied by the context of the item. Because of time constraints, each participant answered 60 randomly selected items. Item order was randomised and one item was displayed per page.

The classification categories of the FTR-type classifier were the same as in Study 1, and mixed-modality responses and negations were again excluded. Of the total responses, $n = 472$ were excluded ($n = 192$ mixed-modality responses, $n = 229$ negations, and $n = 51$ because they were in both of these categories). This left a final sample of $n = 38,166$ responses. To test the reliability of the FTR-type classifier on the updated FTR-elicitation task, linguistically trained coders annotated $N = 1006$ responses ($n = 504$ in English, and $n = 501$ in Dutch). Where systematic errors were found, the FTR-type classifier was adjusted. After this process, all accuracy metrics were $> .99$ (see Appendix C).

The subjective-temporal-distance task: In this task, participants were given two phrases. One used the future tense (English "Ellie *will* arrive later on"; Dutch "Ellie *zal* later aankomen."), and the other used the present tense (English "John *is arriving* later on"; Dutch "John *arriveert* later"). We refer to this manipulation as "tense condition". Both used the temporal adverbial "later on" to ensure that participants construed the present tense frame as referring to the future. Participants rated subjective temporal distance using a slider between "close to now" (0) and "far from now" (10). Numbered slider intervals were not displayed. Prior to starting, participants were told "you will also be asked to indicate how far away from you a length of time feels." For each item, they were told to "Indicate with the slider how far away from NOW the given time feels to you." Before beginning, participants were given one example involving past time reference ("nine months ago"). As a distraction task, participants also rated 10 objective future distances (later today, one week, one month, two months, three months, six months, nine months, one year, two years, and five years). Item order was randomised and one item was displayed per page.

The subjective-certainty task: In this task, participants used a slider to rate between "uncertain" (0) and "certain" (100) how much certainty they construed a given FTR statement as expressing. FTR statements were created imputing different common FTR constructions types into the same "base" sentence: "It {RAIN} next week." We chose representative examples from each of the coding categories of the FTR-type classifier: future tense ("It will rain..."), present tense ("It is raining...")[8], verbal-low-certainty ("It could rain"), other-low-certainty ("It will

[8]We used the present progressive, because the simple present tense *it rains* is not grammatical for English predictions.

possibly rain...”), and other-high-certainty (“It will definitely rain...”). Verbal-high-certainty was excluded because Study 1 indicated *must/moeten* is not used to express epistemic high certainty about the future, *Fig. 2.2* (also see Nuyts, 2000). For a complete set of the items, see *Fig. 2.6.* Prior to beginning, participants were told “you will be asked to indicate how much certainty each statement expresses in YOUR eyes.” For each item, they were told, “Indicate how much certainty YOU would be expressing in the following statement.” Before starting, they were given one training example involving past time reference: “I think Pete picked up bread yesterday.” Item order was randomised and one item was displayed per page.

2.3.1.3 Procedure

Study 2 was hosted on the Qualtrics survey platform and was conducted online. It had a mixed-design. The within-subjects factors for each task are described above. There was one between-subjects factor: survey language. At the beginning of the surveys, participants confirmed their first language and current residence. English speakers confirmed they were native English speakers residing in the United Kingdom and Dutch speakers confirmed they were native Dutch speakers residing in the Netherlands. If they did not, they were immediately ejected from the survey. Participants then answered the same demographic questions as in Study 1. To understand whether multilingualism was affecting language elicited language, we added a second-language proficiency measure. After answering demographic questions, participants self-rated their proficiency for up to three second languages. Ratings were between, “can ask directions and answer simple questions” (1), and “very fluent, can use the language as well as a native language” (5) (see Appendix C). Following this, participants completed the FTR-elicitation task, the subjective-temporal-distance task, and then the subjective-certainty task.

2.3.2 Results

The patterns we found in Study 1 were again broadly apparent. Dutch speakers used more present tense and fewer future tense constructions. English speakers used more low-certainty modal verbs, *Fig. 2.3.* This supports the English uncertainty hypothesis we had previously tested.

To test our hypotheses we again created a low-certainty language variable by combining the verbal-low-certainty and other-low-certainty categories. We regressed this over fixed effects for language, FTR-mode, modality condition, and temporal distance. For temporal distance we used the natural log of the number of days from time of speech. Since effects might be expected to vary interactively, we also included all two-way interactions between these variables. As in Study 1, we allowed intercepts to vary randomly by participant and item, and allowed slopes for language to vary by item. Each of these steps was again significant, $p < .001$.

Table 2.3

Low-certainty language use in Study 2

	e^{β}	SE	z-score	p	
(Intercept)	0.18	0.12	-13.91	<.001	***
language: Dutch	0.24	0.14	-10.2	<.001	***
FTR-mode: prediction	2.56	0.07	13.21	<.001	***
FTR-mode: intention	0.9	0.09	-1.18	.237	
modality condition: high-certainty	0.09	0.1	-24.06	<.001	***
modality condition: low-certainty	28.79	0.08	41.7	<.001	***
temporal distance	1.08	0.05	1.79	.074	·
item order	1.06	0.02	2.66	.008	**
language: Dutch * FTR-mode: prediction	0.7	0.1	-3.61	<.001	***
language: Dutch * FTR-mode: intention	1.16	0.12	1.21	.228	
language: Dutch * modality condition: high-certainty	2.08	0.11	6.8	<.001	***
language: Dutch * modality condition: low-certainty	0.72	0.09	-3.59	<.001	***
FTR-mode: prediction * modality condition: high-certainty	1.19	0.11	1.64	.102	
FTR-mode: intention * modality condition: high-certainty	0.9	0.14	-0.7	.487	
FTR-mode: prediction * modality condition: low-certainty	0.5	0.09	-7.77	<.001	***
FTR-mode: intention * modality condition: low-certainty	1.19	0.11	1.49	.137	
language: Dutch * temporal distance	0.98	0.06	-0.24	.807	
modality condition: high-certainty * temporal distance	0.86	0.07	-2.13	.033	*
modality condition: low-certainty * temporal distance	1.06	0.06	0.95	.345	
language: Dutch * item order	0.86	0.04	-4.29	<.001	***
age	1.15	0.07	2.2	.028	*
employment: employed	0.83	0.1	-1.93	.054	·
employment: student	1.22	0.18	1.12	.264	

Coefficients are exponentiated, so represent changes in the odds ratio of using a low-certainty term. Age was mean centered at 0, and scaled such that $SD = 1$. Employment was sum-coded. Only those demographics the addition of which improved model fit were included, $p < .05$. We also tested whether multilingualism affected elicited language. We operationalised this as $S_i = \sum_k p$, where S is the sum for participant i, of self reported proficiency p (1–5) for up to k (0–3) second languages. In no case did adding this improve model fit, $p > .1$. Generally, English speakers used more future and low-certainty constructions as the task progressed, and Dutch speakers used more present constructions, suggesting speakers trended towards to language-level norms as they progressed. For random components see Appendix C.
$^{***}p < .001$; $^{**}p < .01$; $^{*}p < .05$; $\cdot p < .1$

This time, some demographic variables significantly predicted low-certainty language use. We included these, see *Table 2.3*. We also included effects of item order, which we had not recorded in Study 1.

2.3.2.1 The uncertain predictions hypothesis

We had predicted that relative to intentions and schedules, English speakers would use more low-certainty terms when making predictions. We predicted that this would not be the case for Dutch speakers. To test this we used the *emmeans* package to conduct planned comparisons for the effect of FTR-mode by language averaged across modality condition. We found that compared with intentions, English speakers used significantly more low-certainty constructions when making predictions, $e^{\beta} = 2.86,\ SE = 0.13,\ z = 8.1,\ p < .001$. Contrary to our prediction, we found that Dutch speakers did as well, $e^{\beta} = 1.75,\ SE = 0.19,\ z = 3,\ p = .032$. However, they did this to a much lesser extent, and inspection of *Fig. 2.4* suggests significant effects were driven by very small confidence intervals for low frequency low-certainty

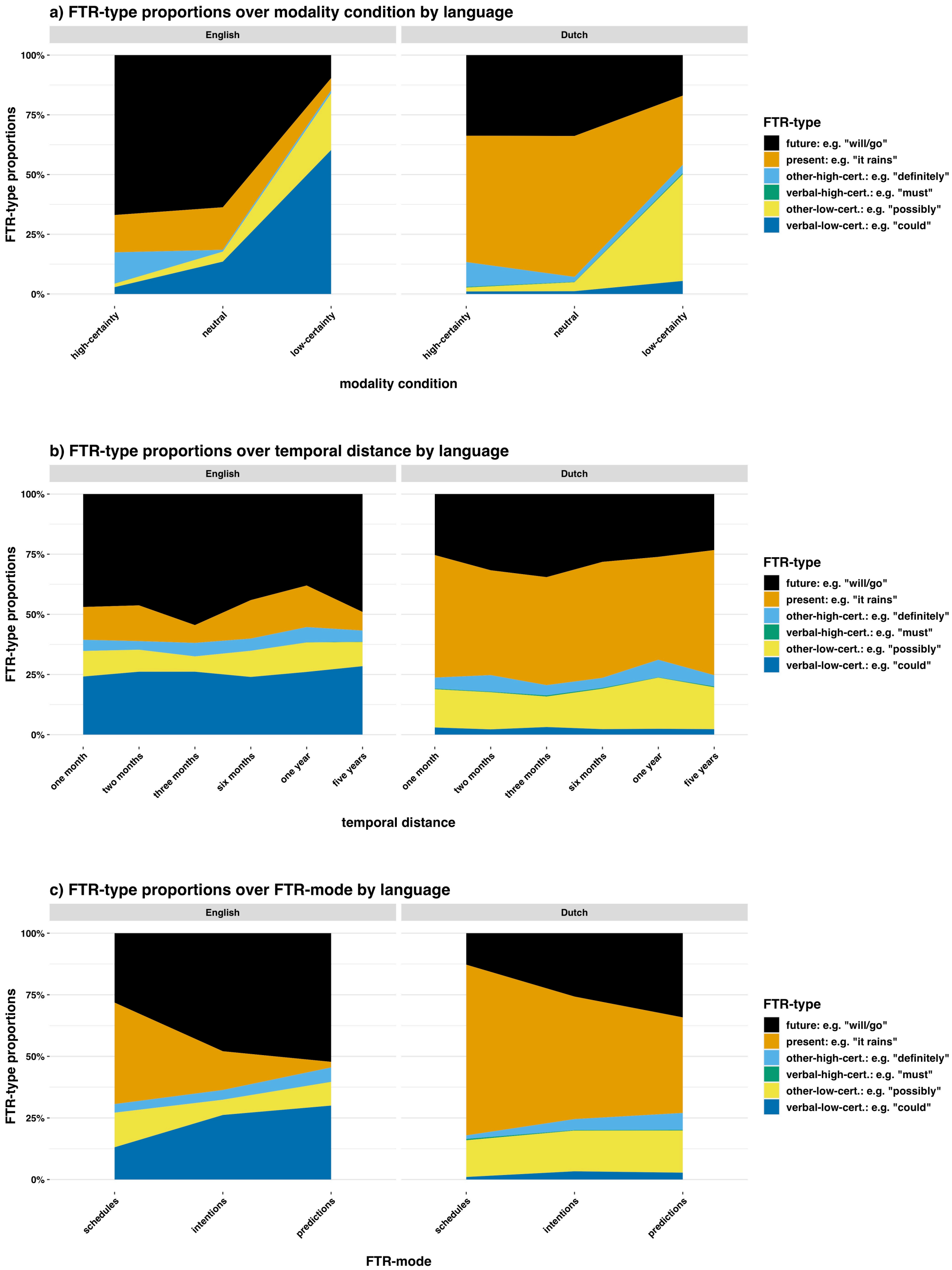


Figure 2.3: FTR-type proportions over modality condition, temporal distance, and FTR-mode (Study 2). As in Study 1, Dutch speakers used more present tense and fewer future tense constructions. English speakers use more low-certainty modal verbs. Dutch speakers made up for this to some degree through the use of other low-certainty constructions.

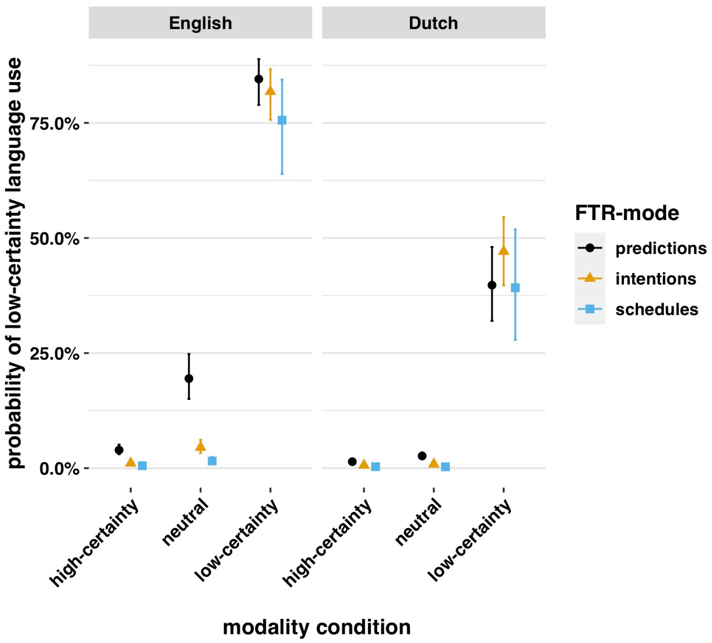


Figure 2.4: Low-certainty language use over modality condition by FTR-mode (Study 2). Confidence intervals here and for *Figs. 2.5* and *2.6* are calculated using the *R* package *ggpredict* by matrix-multiplying a predictor X by the parameter vector B to get the predictions, then extracting the variance-covariance matrix V of the parameters and computing XVX' to get the variance-covariance matrix of the predictions. The square-root of the diagonal of this matrix represents the standard errors of the predictions, which are then multiplied by ±1.96 for the confidence intervals (Lüdecke, 2019). English speakers used more low-certainty constructions, particularly when making predictions in the neutral condition, and in the low-certainty condition overall.

language use in the certain and neutral conditions. Indeed, relative to Dutch speakers, English speakers making predictions used significantly more low-certainty language, $e^{\beta} = 5.93$, $SE = 0.14$, $z = 13.08$, $p < .001$. A particularly striking effect is that English speakers used low-certainty language when they made predictions in the neutral condition, *Fig. 2.4*. This pattern is not evident in the Dutch data. These results support the uncertain predictions hypothesis. They suggest that the grammaticisation of FTR may involve increasing obligatorisation of the encoding of low certainty when making predictions.

2.3.2.2 The low-certainty-sensitivity hypothesis

Next we wanted to understand effects of modality condition. We had predicted that English speakers would be more sensitive to modality condition, using more low-certainty language in the low-certainty condition. As predicted, we found that English speakers were more sensitive to our certainty manipulation. Averaged across FTR-mode, English speakers in the low-certainty condition used significantly more low-certainty language than Dutch speakers did, $e^{\beta} = 5.87,\ SE = 0.15,\ z = 11.86,\ p < .001$, see *Fig. 2.4*. This indicates that in addition to using more low-certainty language generally, English speakers used more low-certainty language as a function of the low-certainty condition. They were more sensitive to our manipulation of certainty.

2.3.2.3 The English cross-modal-mapping hypothesis

Next we wanted to understand how temporal distance impacted low-certainty language use. We had predicted that English—but not Dutch—speakers would use more low-certainty language as a function of temporal distance.

We found that the effect of temporal distance was positive in English, and approached significance, *Table 2.3*. This suggests that English speakers may have been more likely to use low-certainty language as temporal distance increased, averaged across levels of FTR-mode and modality condition. A relevant question is how modality condition affected this slope. We found that the slope for temporal distance was significant in the low-certainty, $e^{\beta} = 1.15$, $SE = 0.07,\ z = 1.97,\ p = .049$, and neutral conditions, $e^{\beta} = 1.19,\ SE = 0.08,\ z = 2.29$, $p = .022$, see *Fig. 2.5*. A significant high-certainty condition by temporal distance interaction, *Table 2.3*, indicates that the slope in the high-certainty condition was not significant, see *Fig. 2.5*.

Was the pattern in Dutch different? It was. In Dutch, the slope for low-certainty language over temporal distance was not significant when averaged across all modality conditions, $e^{\beta} = 1.07,\ SE = 0.07,\ z = 0.99,\ p = .323$. Neither was it significant in any modality condition, $p > .05$.

These results support the English cross-modal-mapping hypothesis. English speakers used more low-certainty language in the neutral and low-certainty conditions as temporal distance increased. Dutch speakers did not.

As a robustness test, we wanted to check whether this was a general effect across all levels of temporal distance, or whether there were specific step changes at specific temporal distances. We therefore re-estimated the model but with temporal distance operationalised as an ordered factor. This allowed us to compare each factor level with the previous one. The only significant step change was for one year, $e^{\beta} = 1.38,\ SE = 0.15,\ z = 2.1,\ p = .036$. This suggests that temporal distance affects English use of low-certainty language, particularly as temporal distances reach and exceed one year (i.e. one year and five years, in our data). Further studies would be needed to identify further step changes as temporal distances increase beyond five years.

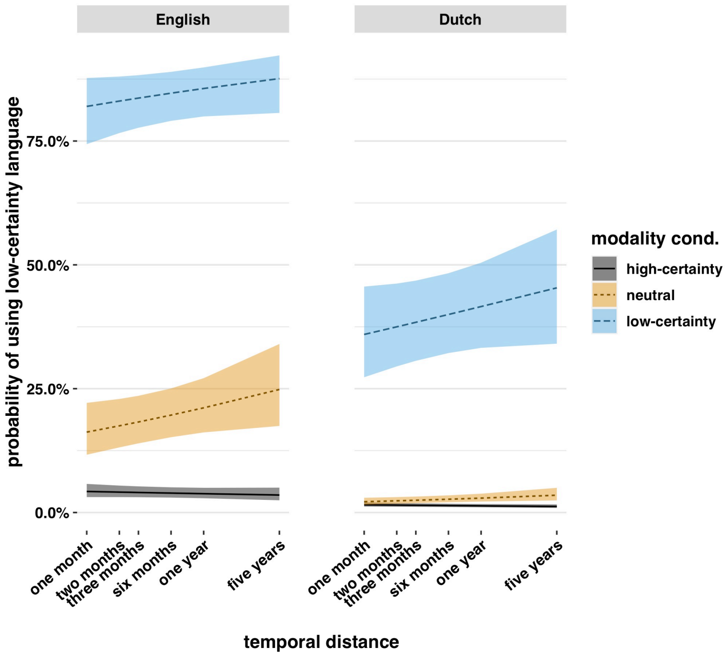


Figure 2.5: Low-certainty language use over temporal distance by modality condition (Study 2). In the low-certainty and neutral conditions, English speakers used more low-certainty language as temporal distance increased. Dutch speakers did not. The x-axis is log-scaled.

2.3.2.4 The linguistic-savings-distance hypotheses

On the basis of the linguistic savings hypothesis, we had predicted: (a) that participants would rate the future tense frame, "Ellie *will* arrive later on", as more temporally distal than the present tense frame, "John *is arriving* later on"; and (b) that Dutch participants would rate the future as more temporally proximal than English participants. To test these predictions we regressed subjective distance ratings over language and tense condition and the interaction between them. We used a multilevel linear regression with random intercepts for participant (these were significant, $\chi^2(1) = 388.44,\ p < .001$).

Was the future tense frame construed as more distant? It was not. Tense frame had no significant effect in either English, $\beta = 0.01,\ SE = 0.04,\ t(648) = 0.29,\ p = .773$, or Dutch, $\beta = -0.01,\ SE = 0.04,\ t(648) = -0.21,\ p = .833$. This indicates that participants were not construing the future tense statement as more temporally distal.

Did Dutch speakers construe the future more temporally proximal? They did not. In fact, relative to English speakers, Dutch speakers rated the future as more distal, $\beta = 0.68$, $SE = 0.12,\ t(648) = 5.88,\ p < .001$. This effect might have been limited to the temporal distance of the two "future/present" items (i.e. "later on"). To test whether it was, we re-estimated the model but using the objective distances in the distractor tasks, ranging from "later today" to "five years". We again found that Dutch speakers rated the future as more distal, $\beta = 0.61,\ SE = 0.13,\ t(648) = 4.84,\ p < .001$. This is the opposite to the direction predicted by the linguistic savings hypothesis.

Together, these results fail to support the hypothesis that tense framing impacts construals of temporal distance, and that therefore Dutch speakers construe the future as closer in time (cf. Chen, 2013).

2.3.2.5 Exploratory analyses: The subjective probability task

To explore the results of the subjective probability task, we regressed certainty ratings over an unordered factor which indexed each item. Because the items were not strictly comparable in between English and Dutch, we did this separately for each language. We included random intercepts for participant. This was significant in both languages, $p < .001$. We present the results in *Fig. 2.6.* We were particularly interested in the future tenses, given the conflicting accounts that they either encode modal strengthening or modal weakening. Interestingly, future tenses in both languages were rated as high certainty. This undermines accounts which suggest future tense marking encodes modal weakening.

However, the languages differed in subtle ways. In discussing these results, we will use the term "modal polarity" to refer to the one-dimensional scale between high and low certainty. English appeared to break modal polarity into finer gradations, with clearer differences between low certainty (*could/may/might*) and intermediate certainty (*should/probably/I think*). This suggests that English may oblige speakers to express greater degrees of precision about the likelihood of future outcomes.

Additionally, we found that relative to the present tense, Dutch speakers rated the future tense as significantly less certain, $\beta = -3.9,\ SE = 1.43,\ t(4770) = -2.72,\ p = .007$, and English speakers rated it as significantly more certain, $\beta = 3.5,\ SE = 1.43,\ t(3960) = 2.44,$ $p = .015$. This suggests that paradigmatic analyses may be a productive way of analysing the role of contrasting FTR constructions. We pointed out above that in English the highest certainty FTR structure available for prediction-based FTR is the future tense. In Dutch, present tense statements of fact are possible. Our results are compatible with the conclusion that this difference may have produced the differences in relative encoded certainty between the future and present tense in these languages. Moreover, the result suggests there are cross-linguistic differences in the modal strength of future tenses. This means obliging the use of the future tense could affect psychological discounting differently in different languages.

Finally, Dutch ratings of certainty for the future tense appeared to have greater variance. Treating ratings for both *zullen* 'will' and *gaan* 'be going to' as independent, variance in Dutch was higher ($M_{eng} = 79.5,\ SD_{eng} = 29.4;\ M_{dut} = 71.8,\ SD_{dut} = 29.4$). To test whether this was significant, we first calculated mean ratings per participant across ratings for both *zullen*

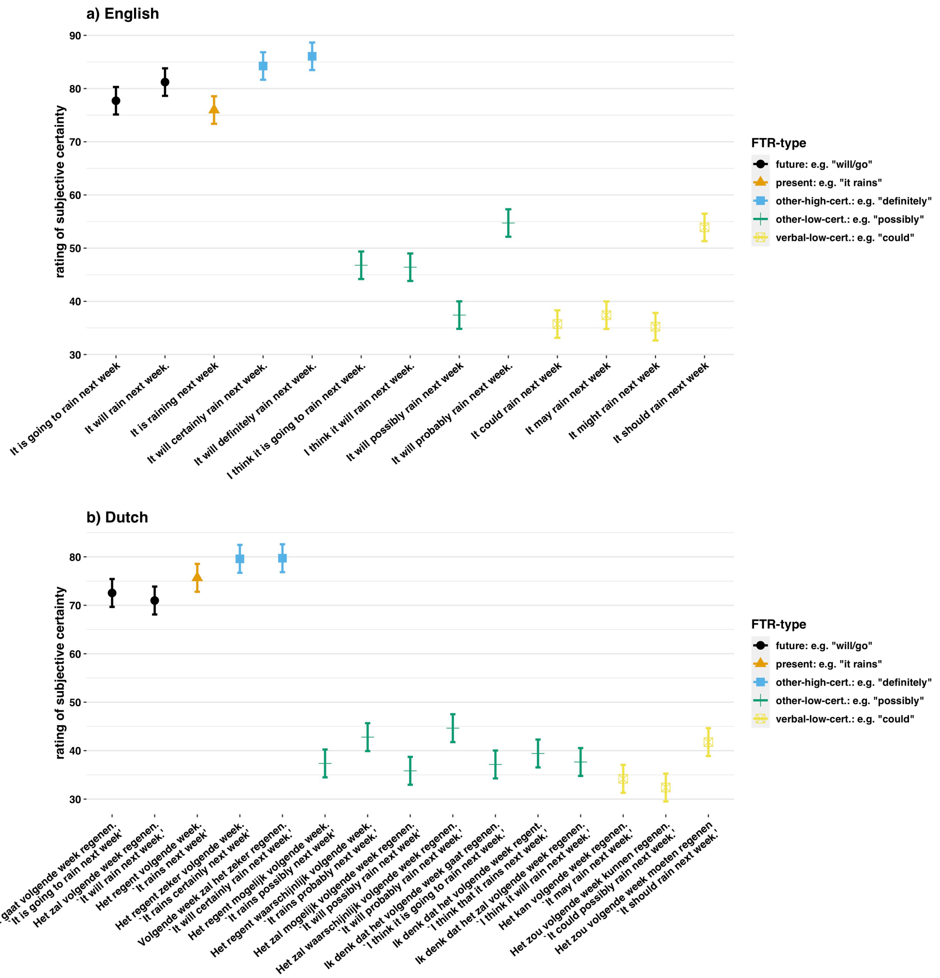


Figure 2.6: Subjective ratings of certainty by item and FTR type in English and Dutch (Study 2). In both languages, future and present tense appear to encode high certainty, while modal and other-low-certainty constructions encode low certainty. English speakers appeared to break modal polarity into finer gradations than Dutch speakers, with clearer differences between low-certainty (*could/may/might*) and intermediate-certainty (*should/probably/I think*) modal expressions.

'will' and *gaan* 'be going to'. This ensured that there was one observation per participant, so that assumptions of independence were met. Using an F test for equality of variances, we found that Dutch ratings for future tense certainty had higher variance $F_{317,329} = 1.33$, $p = .011$. This suggests there are greater individual differences in Dutch in terms of how future tense semantics are construed. This is also compatible with a paradigmatic analysis. If the Dutch future tense plays a less unique role relative to other possible FTR structures, this relative semantic indeterminacy may allow for greater individual differences in the way it is used.

2.3.3 Discussion

Study 2 supports the hypothesis that the encoding of modality is be implicated in FTR grammaticisation processes. We found that English speakers were more likely to mark their predictions using an low-certainty construction. English speakers also used more low-certainty language as a function of temporal distance. This suggests that English speakers may construe temporally distant events as increasingly uncertain. Additionally, English speakers were more sensitive to modality condition. They used more low-certainty language in the low-certainty condition. All these results suggest that, relative to Dutch speakers, English speakers are more likely to encode low-certainty notions when they talk about the future. The fact that it was mostly modal verbs which drove this effect, *Fig. 2.3*, suggest grammatical constraints are responsible.

In exploratory analyses of the subjective probability task, we found that both the future and present tenses were rated as high certainty. This suggests obliging their use would cause less not more discounting (cf. Chen, 2013). Additionally, differences in the relative modal polarity of the present and future tenses in English and Dutch suggest that FTR status may be a relevant determinant of modal future tense semantics.

Finally, we found no difference in subjective-temporal-distance ratings as a function of whether a future-referring statement was framed using the future or present tense. This suggests that the future tense does not encode temporal distance (cf. Chen, 2013). We also found that Dutch speakers rated future events as more distal than English speakers (cf. Chen, 2013). In combination, these findings suggest that temporal mechanisms hypothesised to be underpinning the relationship between FTR grammaticisation and temporal discounting should be reconsidered.

2.4 General Discussion

In general, the findings of both studies suggest that modality is an important aspect of FTR systems, and may be implicated in the process of FTR grammaticisation. This suggests that FTR systems should be studied holistically rather than focusing exclusively on tense, aspect, or modality. It also suggests that researchers investigating potential impacts of cross-linguistic differences in FTR grammaticisation on psychological discounting processes should seek to understand the role of modal language and probability discounting. Below, we focus on implications for the linguistic savings hypothesis.

2.4.1 FTR status: The weak/strong dichotomy

Do our results corroborate or undermine the FTR status dichotomy? English speakers used more low-certainty language. Low-certainty language use in English was also more sensitive to FTR-mode, probability condition, and temporal distance. This was mostly driven by modal verbs, which means that grammatical features of English may be involved in producing higher encoding of low certainty, i.e. the obligatory modal verb system.

This suggests that obligatory future tenses and stricter encoding of modality may arise from a unified underlying process. Dahl (1985) delineated a "futureless area" comprising European languages which do not oblige the future tense for prediction-based FTR. Obligatory tense marking in prediction-based FTR is suggested to be reasonable proxy for FTR grammaticisation in general. Our results suggest this includes the obligatorisation of modal verbs as well as future tenses. As such, in one sense, the FTR status dichotomy was supported. FTR appeared more grammaticised in English, with the noted caveat that modal FTR structures were implicated in this difference.

However, our results suggest a more nuanced typological approach is needed to better capture the cross-linguistic differences which may affect psychological discounting. Particularly, we found significant differences between Dutch and German speakers in terms of use of the future and present tenses. In many ways, German FTR grammar seemed to be intermediate between Dutch and English. Relative to Dutch speakers, Germans used more future tenses, fewer present tenses, and more low-certainty language. Relative to English speakers, they used less. This suggests that researchers should dispense with the notion that cross-linguistic differences can be captured using a single variable in large-scale statistical analyses.

An important point is that our results have no implications for whether it is *possible* to form nuanced linguistic FTR utterances in these languages. A fundamental idea which can be traced to Franz Boas (see Leavitt, 2011, p. 127), is that languages differ not so much in what *may* be said, but in what *must* be said. In pointing out that English speakers use more low-certainty language, we do not imply that Dutch and German FTR is deficient or simple. Our results rather suggest that English grammatical constraints nudge English speakers towards encoding more low-certainty modality.

2.4.2 Causal mechanisms: A modal account of observed findings

How might differences in the obligation to encode low-certainty FTR impact psychological discounting? In a recent risky intertemporal-choice task, Vanderveldt et al. (2015) found that a function of the following form best described empirical valuations of risky future rewards: $V = A/[(1+kD)^{sd} \times (1+h\theta)^{sp}]$. In this instance, V is monetary value, D is temporal distance, θ is odds against, k and h are parameters affecting the discounting rates, and sd and sp are scaling factors which have been found to best describe experimental evidence. This means that psychological discounting is better described by a discounting plane, than a discounting curve. Subjective value is a function of both the odds against and the time until the receipt of a future reward.

How might cross-linguistic differences in FTR grammaticisation impact such psychological discounting processes? Firstly, the mechanisms proposed by the linguistic savings hypothesis might still be in effect. However, they could just as easily apply to probabilistic discounting, i.e. we might predict speakers of languages which more strictly grammaticise FTR to both have relatively more precise beliefs about, and relatively lower estimates of, the probability of future events. We would thereby predict them to probabilistically discount more heavily. However, as the probability of a future reward decreases, temporal distance has an increasingly negligible effect on subjective value; in contrast probability discounting is relatively unaffected by temporal distance (Vanderveldt et al., 2015). We therefore suggest that differences in the grammaticisation of probability (i.e. modality) may be the more important factor in driving observed cross-cultural differences in discounting-related behaviour (cf. Chen, 2013). Real-world (risky) intertemporal decisions could be impacted by such probability discounting differences. If the English case generalises, this suggests that a "modal" account could plausibly explain many reported results (Chen, 2013; S. Chen et al., 2017; Chi et al., 2018; Figlio et al., 2016; Galor et al., 2016; Guin, 2017; Hübner & Vannoorenberghe, 2015b, 2015a; Liang et al., 2018; Lien & Zhang, 2020; Mavisakalyan et al., 2018; Pérez & Tavits, 2017; Roberts et al., 2015; Sutter et al., 2015; Thoma & Tytus, 2018).

2.4.3 Causal mechanisms: Temporal distance and precision

The results of the subjective-temporal-distance task did not support the linguistic savings hypothesis. English speakers rated future events as closer in time than Dutch speakers. This is the opposite to the direction expected if the English future tense encoded temporal distance. Additionally, we found that tense framing (future vs. present) had no effect on distance ratings. It is possible that this null result is an artefact of the single phrase we used: "_____ *is arriving/will arrive* later on." This constitutes a limitation. A difference might emerge with more distant FTR statements, or other phrases. Future research could take this up. However, our findings are consistent with findings that tense framing does not affect intertemporal decisions. Banerjee and Urminsky (2017) conducted a series of 6 experiments investigating this. They had participants make intertemporal choices, which were framed in either the present or future tense, i.e. "you get $10 in a week" vs. "you *will* get $10 in a week." In a series of several experiments which used a range of distance, such manipulations had no effect on participants' time preferences (a similar result is reported in Thoma & Tytus, 2018). This suggests that future tenses do not encode temporal distance, regardless of the temporal distances involved. Our findings corroborate this conclusion.

What do tenses encode? We found the present and future tenses were rated as high certainty in English and Dutch. This suggests obligatory future tense use would cause speakers to discount less, not more. In fact, the ratio of high-certainty (present + future + certain) vs. low-certainty language is the only linguistic feature we identified which might plausibly affect psychological discounting in the observed direction. This lends support to our general argument that FTR grammaticisation impacts psychological discounting because it affects speakers beliefs about future risk rather than their construals of future temporal distance and/or precision.

2.4.4 Contributions to work on temporal-distance representations

Dutch speakers rated the future as farther away. This contributes to a nascent body of literature which has begun to investigate how subjective ratings of future distance impact discounting (see Bradford et al., 2019; B. K. Kim & Zauberman, 2009; Zauberman et al., 2009). For instance, Thorstad and Wolff (2018) found that people whose tweets reference increasingly distant future times were more likely to invest in the future and less likely to undertake risky behaviour. Ireland, Schwartz, Chen, Ungar, and Albarracín (2015) found that US counties with higher rates of FTR tweets had lower rates of HIV. In this context, HIV exposure is expected to be impacted by time preferences because risky behaviours (e.g. intravenous drug use, unprotected intercourse) incur long terms costs (risk of contracting HIV) but confer short term benefits. Finally, using a measure similar to ours, Thorstad, Nie, and Wolff (2015) found that people who construed the future as farther away were more present oriented. Together, these results support Chen's (2013) proposal that subjective representational distance is a significant predictor of time preferences. However, we found that Dutch speakers represented the future as farther away. As far as we can tell, this is the first study to use time slider type tasks to identify cross-cultural differences of this nature. If this is related to cross-linguistic differences in FTR grammaticisation, this suggests that higher obligation to mark future statements may cause future events to be more salient to strong-FTR speakers. It is conceivable that this might cause them to perceive future outcomes as more proximal. However, if this were the case, it would cause strong-FTR speakers to be more future-oriented not less—as is hypothesised and observed. This entails that differences in construals of future distance are not likely to be causally implicated in the relationship between FTR grammaticisation and psychological discounting.

2.4.5 Conclusions

In general, we found that FTR status indexes cross-linguistic differences in the encoding of future modality. English speakers tended to encode low-certainty modality more than Dutch or German speakers. This was mostly driven by a more highly grammaticised modal verb system. These results undermine the notion that FTR grammaticisation is primarily about time, and call into question the validity of the causal mechanisms suggested in Chen (2013). If tense and modal FTR grammaticisation are generally correlated, it may be the case that observed cross-cultural differences in discounting-related behaviour actually involve probabilistic discounting driven by stricter encoding of modal notions in strong-FTR languages.

Economists continuing to work on this question might begin exploring the complex potential relationships between FTR grammaticisation and discounting. These processes are worth understanding: Psychological discounting processes are an important determinant of a wide range of behaviours, including health outcomes (Ireland et al., 2015; Vuchinich & Simpson, 1998), drug use (McKerchar & Renda, 2012), climate change attitudes (Mavisakalyan et al., 2018), educational performance (Figlio et al., 2016), pathological gambling (Hodgins & Engel, 2002), and investment in savings (Liu & Aaker, 2007). If the precise nature of the relationship between FTR grammaticisation and discounting is better understood, researchers may better understand how cross-linguistic differences impact behaviour. Detailed experimental work which combines behavioural economic techniques with usage-based typological linguistics should be employed to explore the precise relationships between cross-linguistic differences in FTR grammaticisation and psychological discounting.

Chapter 3

Future tense use causes less not more temporal discounting

Abstract

Recently, economists have suggested that the future tense—e.g. *will, be going to*—encodes notions of temporal distance. Alternatively, debates among linguists suggest the future tense encodes "modal" notions of high certainty. People devalue delayed outcomes as the time until they will occur grows longer ("temporal discounting"). If the future tense encodes temporal distance, its obligatory use would cause distal temporal representations and increased temporal discounting. People also devalue risky delayed outcomes as risk grows higher. If the future tense encodes high certainty, its obligatory use would therefore cause high-certainty representations and decreased temporal discounting (since delayed outcomes may be construed as inherently risky). While speakers of languages with obligatory future tenses appear to exhibit increased temporal discounting, whether future tense use is driving outcomes remains unclear, since no studies (that we know of) have measured how individual-level future tense use relates to temporal discounting. To test this, we elicited language about the future and then measured temporal discounting in speakers of English (which obliges the future tense) and Dutch (which does not). We used mediation analysis to test whether higher use of the future tense in English resulted in more or less temporal discounting. We found it resulted in less (Study 1), which replicated (Study 2). This suggests the English future tense encodes high certainty, not temporal distance. To address the issue that cultural differences could have driven aspects of our findings, we directed Dutch speakers to talk about the future either using the present or future tense (Study 3). Again, the future condition resulted in less and not more temporal discounting. These results suggest that obligatory future tense use is not responsible for observed correlations between cross-linguistic differences in future time grammar and behaviours involving temporal discounting. Obligatory low-certainty modal grammar—e.g. *could, may*—is a plausible alternative.

3.1 Introduction

Linguistic relativity is the hypothesis that the peculiarities of the languages we speak can shape, augment, or otherwise alter our cognitive processes and thereby give rise to cognitive differences between populations (Whorf, 1956). While scholars continue to debate the depth and extent of linguistic penetration into cognitive processes (Gleitman & Papafragou, 2012; Montemayor, 2019), a growing body of evidence suggests that differences in online attentional demands during language use may grow into entrenched offline cognitive differences (see Lupyan et al., 2020; Majid, 2018; Wolff & Holmes, 2011). The hypothesis has recently gained popularity among economists, who have contributed to a body of research which has found that cross-linguistic differences in the grammar of Future Time Reference (FTR) predict a range of behaviours which appear to involve "intertemporal decision making", where people must balance present versus future costs and rewards. By FTR we mean any statement which involves talking about future events.

3.1.1 An overview of how FTR grammar is hypothesised to affect intertemporal decisions

Most research into this question has advanced under the banner of what we will refer to as the temporal hypothesis (Chen, 2013). This hypothesis is based on two observations. First, some languages oblige speakers to use the future tense for FTR, while others do not. For instance:

(3.1) a. English
It will rain tomorrow.

b. Dutch

Morgen regent het.
tomorrow rain:PRS[1] it

'Tomorrow it rains.'

English speakers must use *will*, *be going to*, or *shall* in example (3.1a), whereas Dutch speakers are free to use the present tense in example (3.1b). In this framework, languages like English are referred to as "strong-FTR", while languages like Dutch are referred to as "weak-FTR".[2] Specifically, weak-FTR languages are those languages which do not oblige the future tense

[1]PRS. The verb *regenen* 'to rain' is in the present tense *regent* 'it rains'. I use the Leipzig rules for interlinear glosses (Comrie et al., 2015).

[2]Dahl (2013) points out that these terms are problematic because it is possible in all languages to refer to future time (whether or not the future tense is obliged). Better might be "strong-FTR_G" to indicate that the difference is a matter of Grammatical marking. However the terms are in widespread use. We will not deviate from them, but wish to acknowledge Dahl's (2013) point.

in prediction-based contexts and strong-FTR languages are those languages which are not weak-FTR (Chen, 2013). In this context, prediction-based FTR is contrasted with schedule-based and intention-based FTR (Dahl, 2000a). This is relevant, because it is quite common for languages to permit the present tense when referring to schedules, and, in general, for FTR usage to be sensitive to whether predictions, schedules, or intentions are involved (Dahl, 2000a). English is a good example: When referring to a schedule, e.g. *The Bears play at 7pm*, the present tense is perfectly acceptable. This is not the case for predictions, see example (3.1a). As such, when we write of languages "obliging" the future tense, we follow Chen (2013) in only referring to prediction-based FTR. We refer Chen's (2013) dichotomous weak/strong distinction as "FTR status."

The second observation is that people tend to devalue (or "temporally discounted") outcomes as the time until they will occur grows longer (L. Green et al., 2014; Rachlin et al., 1991). For instance, given the choice of receiving \$10 now or \$10 in five years, most would choose the former because they would have temporally discounted the later as a function of the delay. However, ask a person whether they would prefer to receive \$10 now or \$20 in a month and some will take \$10 now, while others will choose to wait for \$20 (L. Green & Myerson, 2004). We will refer to these differences in terms of "time preferences". Temporal discounting is therefore usually measured by giving participants a series of binary choices like those above. Participants who temporal discount less tend to choose the "future oriented" options to wait for future rewards, because they perceive such rewards as more valuable than participants who discount more, who tend to make the "present oriented" choice for the immediate reward.

The temporal hypothesis predicts that weak-FTR speakers should temporally discount less than strong-FTR speakers, and therefore tend to be more future oriented. Two causal mechanisms are hypothesised. First, habitual use of the present tense for FTR in weak-FTR languages is suggested to "collapse" future time with present time, causing weak-FTR speakers to perceive future events as more temporally proximal than strong-FTR speakers. We refer to this as the "distance mechanism". Second, failure to grammatically disambiguate future from present time in weak-FTR languages is suggested to cause speakers to represent the future temporal locations less precisely—the "precision mechanism". Either of these temporal representations would cause increased temporal discounting in speakers of weak-FTR languages (*Fig. 2.1*, Chapter 2; and see Chen, 2013). In this discussion, we focus on the distance mechanism.

To support his formulation of the distance mechanism, Chen (2013) outlines the following simple decision-making model. Consider someone making a decision about whether to pay a cost, C, now or take a future reward, $R > C$. This translates directly into experimental paradigms which measure time preferences by having participants make binary decisions between present and delayed rewards (see Section 3.2.1.2). Specifically, in choosing between \$10 now, and \$20 in a month, C is simply the value forgone by opting for the delayed option, i.e. \$10. Consequently, a participant will choose the delayed \$20, R, if its discounted value is greater than \$10. Imagine she is uncertain about the precise time she will receive the reward and holds the beliefs about time t with a distribution $F(t)$. This simply implies that the subjective representation of time is distributional. When people hear “two weeks from now” they have a idea of *about* how temporally distal this feels, but this is different from objective measurements of time (Zauberman et al., 2009; B. K. Kim & Zauberman, 2009; Bradford et al., 2019). If she discounts delayed rewards with rate δ, she will prefer R if and only if:

$$C < \int e^{-\delta t} R dF(t) \tag{3.1}$$

which is to say she will only choose the delayed reward if its subjective value—δ integrated over $F(t)$—is greater than the cost of waiting.

If weak-FTR speaker construe future events more proximally, this would in effect shift beliefs about temporal location such that $F_{Weak}(t) < F_{Strong}(t)$. If this were the case, it would result in lower willingness to choose delayed rewards in strong-FTR speakers:

$$\begin{gathered} \text{if } \forall t,\ F_W(t) < F_S(t) \text{ then} \\ \int e^{-\delta t} R dF_W(t) > \int e^{-\delta t} R dF_S(t) \end{gathered} \tag{3.2}$$

which is to say that if for all t, weak-FTR speakers believe future events are more temporally proximal, then the discounted value of delayed rewards will on average be relatively higher for weak-FTR speakers. On average, they will therefore be relatively more future oriented: more likely to save money for the future, exercise, and eat healthily.

3.1.2 Empirical support for the temporal hypothesis

In the seminal contribution which first proposed and tested the temporal hypothesis, Chen (2013) used typological research by Dahl (1985, 2000a) and others to create the dichotomous FTR status variable which classes $N = 129$ languages as either weak- or strong-FTR. He then used FTR status to predict a set of behaviours in which time preferences are implicated. Among other outcomes, weak-FTR speakers were significantly more likely to have saved each year and retired with more assets; examples of foregoing present spending to save for the future. They were less likely to have smoked and more likely to practice safe sex; examples of taking consideration of the delayed deleterious effects of present behaviour. Additionally, they were healthier, as indexed by obesity, peak blood flow, grip strength, and physical exercise levels. Such behaviours are impacted by time preferences because health decisions often involve trade-offs between immediate (dis)comfort and future (ill)health (Tate et al., 2015).

Since then, a number of studies have extended this basic approach, using FTR status in regression analyses to predict a range of country-, firm-, and individual-level behaviour to which time preferences appear to be relevant. For instance, German (weak-FTR) speaking households were found to save more, smoke less, and resort to credit less than French (strong-FTR) speaking households in Switzerland (Guin, 2017). FTR status has also been found to impact educational outcomes (Figlio et al., 2016), patience (Falk et al., 2018), and suicide rates (Lien & Zhang, 2020). Additionally, firms headquartered in weak-FTR-speaking countries engaged in more future-oriented behaviour (S. Chen et al., 2017), a pattern of results which has replicated across a variety of contexts (see Chi et al., 2018; Fasan et al., 2016; J. Kim et al., 2017; Liang et al., 2018). Country-level economic outcomes have been impacted as well. Weak-FTR-speaking countries engaged in more future-oriented management of their economies (Hübner & Vannoorenberghe, 2015a, 2015b). Finally, Mandarin speakers (weak-FTR) had more pro-environmental attitudes than Korean speakers (strong-FTR) (S. Kim & Filimonau, 2017); bilinguals expressed higher support for a fictitious pro-environmental "green tax" when they were tested in Estonian (weak-FTR) than Russian (strong-FTR) (Pérez & Tavits, 2017); weak-FTR speakers (across a number of languages) were more likely to adopt environmentally responsible behaviours (Mavisakalyan & Weber, 2018); and weak-FTR-speaking countries enacted more pro-environmental policies and exhibited stronger commitment to global environmental cooperation (Mavisakalyan & Weber, 2018). Such decisions can be construed as intertemporal choices, because climate change mitigation efforts tend to incur short-term costs but confer long-term benefits.

If the correlation identified in these studies is not spurious, it would constitute profound new ground for the linguistic relativity hypothesis, which has mostly been characterised by experimental research into processing and attentional biases (Wolff & Holmes, 2011), and has therefore been criticised as only having identified linguistic effects in studiously artificial experimental settings (McWhorter, 2014).

3.1.3 Problems with the temporal hypothesis

However, support for the temporal mechanisms outlined by Chen (2013) may not be as strong as findings suggest. First, the direction of the effect of FTR status been mixed. Countries with a higher percentage of strong-FTR speakers were found to have higher percentages of their populations who perceived climate change as serious, as well as have better performance on climate change mitigation action, as indicated by lower CO_2 and energy production per capita (Zhu et al., 2020). These effects run in the opposite direction to other analyses involving FTR and climate change, which generally find weak-FTR speakers tend to have greater support for pro-environmental policies (S. Kim & Filimonau, 2017; Mavisakalyan & Weber, 2018; Pérez & Tavits, 2017). Additionally, German (weak-FTR) speaking school children were more future-oriented than Italian (strong-FTR) speaking children in an experimental intertemporal-choice task. On the other hand, an experiment involving speakers of 5 languages spanning the FTR-status dichotomy found weak-FTR speakers were less future-oriented (Thoma & Tytus, 2018). Such mixed results suggest more research is needed to determine how FTR grammar may relate to temporal discounting.

Critically, at least eight independent experiments conducted in both weak- and strong-FTR languages have failed to find significant differences as a function of whether an intertemporal choice is framed in the present or future tense (Banerjee & Urminsky, 2017; J. I. Chen, He, & Riyanto, 2019; Thoma & Tytus, 2018). In general, these experiments test whether framing a future reward in the present ("you get $10 in a week") or the future tense ("you *will* get $10 in a week") impacts how likely participants are to choose the delayed option. The temporal hypothesis predicts future tense framing should cause participants to construe a future reward as more temporally distal, and therefore make them less likely to wait. However, no such pattern emerges. For instance, in an impressive set of six variants on this general design, Banerjee and Urminsky (2017) only found an effect of tense frame when no other temporal information was included, i.e. no temporal adverbials such as "in a week". The authors suggest this is weak evidence: It is not possible to rule out that participants construed the present tense frame as immediate and the future tense frame as delayed. This suggests the future tense is not construed as more temporally distal.

Our previous research supports this conclusion (Study 2, Chapter 2). When participants used a slider to estimate the temporal distance encoded by a simple expression, no difference emerged as a function of whether the expression was framed in the future or present tense, e.g. "John *arrives*" versus "*will arrive* later on". In addition to this, we found that Dutch (weak-FTR) speakers rated future outcomes as more distal than English (strong-FTR) speakers. This reverses the direction predicted by the temporal hypothesis (Study 2, Chapter 2). Taken together, these results suggest that the future tense does not encode temporal distance (cf. Chen, 2013), and that something else is going on with future tense semantics. Economists working on the temporal hypothesis might therefore benefit from consulting contemporary debates among linguists about future tense semantics. We do this in the next section.

3.1.4 An alternative modal account of future tense semantics

We propose an alternative, modal, account of future tense semantics, which we base on review of the linguistic literature. The expression of future time is very complex and often involves the expression of "modal" notions of ability, desire, (un)certainty, probability, volition, intention, and obligation (see Palmer, 2001; Bybee & Dahl, 1989; Bybee et al., 1994; Fries, 1956). Future tense semantics are hotly debated (see Broekhuis & Verkuyl, 2014; Cariani & Santorio, 2018; Dahl, 2000a; Enç, 1996; Fries, 1956; Giannakidou & Mari, 2018; Huddleston, 1995; Klecha, 2014; Salkie, 2010; Sarkar, 1998). There is an obvious epistemic—potentially ontological—imbalance between the past and the future: Whereas the past can be known, in our memories, the future cannot. This leads fairly transparently to modal complexities when it comes to the expression of FTR. Tenses are usually thought of as deictic expressions which move the temporal frame of reference relative to the time of speech, i.e. a future tense simply served to indicate that the references event is posterior to the time of speech (Klein, 1995). On the other hand, modal operators involves quantifying what is likely–unlikely, or possible–necessary, relative to various modal "bases". For instance, in *epistemic* modality, speakers express what what they believe to be (un)likely relative to what they know about the world (Palmer, 2001), e.g. *He will definitely win tonight!* The question becomes whether future "tenses" operate like past tenses—simply shifting referred-to time relative to time of speech—or whether their semantics involve modal notions. In fact, some accounts suggest that a definitional feature of future tense markers is that they are characterised by division of labour between temporal and modal semantics (Dahl, 2000a).

This indeterminacy may have contributed to ongoing arguments about the theoretical status of future "tenses". While some think of them as tenses *sensu stricto* (Kissine, 2008), others argue they should be subsumed within the category of modality (Giannakidou & Mari, 2018; also see Dahl, 2000a). For instance, some linguists have suggested *will* encodes modal weakening because it is used to mark predictions (Dahl, 2000a; Enç, 1996; Fries, 1956; Huddleston, 1995; Klecha, 2014). Others have suggested it encodes high certainty, like *must* (Giannakidou & Mari, 2018) or is closer to a prototypical future tense (Salkie, 2010). Critically, according to these contrasting accounts, future tense use should relate to future orientation in different ways. It is beyond the scope of this chapter to wade into these debates. We rather focus on two issues: (a) the idea that future tenses in different languages have different semantics; and (b) the idea that future tenses encode modal notions of high certainty.

A critical distinction here is between language-specific grammatical categories, and cross-linguistic typological categories (Dahl, 1985) (we use SMALL CAPS to refer to the latter). In this context the FUTURE TENSE may be thought of as a fuzzy category comprising a collection of features which tend to be cross-linguistically common of language-specific future tenses (Dahl, 1985). Some language-specific future tenses may prototypically represent the FUTURE TENSE, but none fully exemplify it (Dahl, 1985). Chen's (2013) argument makes a tacit claim about the semantics of the FUTURE TENSE, i.e. temporal distance, precision. This may be problematic given expected cross-linguistic variation the semantics of future tenses (see Bybee & Dahl, 1989; Bybee et al., 1994; Dahl, 2000a). This is particularly true because FTR status may be a relevant factor in determining future tense semantics.

A paradigmatic analysis is a useful framework for addressing these issues. Paradigmatic analyses attempt to understand the semantic contribution of a linguistic unit by contrasting it with the other units by which it may be replaced in a given context (Lyons, 1968). The context we focus on here is prediction-based FTR in English and Dutch, since this forms the basis of the FTR status dichotomy (Chen, 2013), and those are the languages in our sample. The future tense is formed in similar ways in English and Dutch. Both languages use either a modal auxiliary (English *will/shall*; Dutch *zullen* 'will') or a *de andative*, "go-based", construction

(English *be going to/gonna*; Dutch *gaan* 'be going to'). Both *zullen* and *will/shall* are part of a larger modal system comprising other modals which—at least among those which are permitted in prediction-based FTR—generally encode modal weakening relative to the future tense (English *could, may, might, should, would*; Dutch *kunnen* 'may') (Nuyts, 2000). Here, the relevant contrast is Dutch grammar also affords the use of the unmarked present tense for prediction-based FTR, as in example (3.1b). A difference emerges: The English future tense is the highest-certainty construction type available. This suggests the English future tense encodes high certainty, which echoes accounts that it operates like a high-strength marker of modal necessity similar to *must* (Giannakidou, 2017; Klecha, 2014). On the other hand, the Dutch future tense may be contrasted with unmarked present tense statements. Such statements are usually considered to be stronger than any modal quantification, i.e. because they are simply statements of fact (Giannakidou & Mari, 2018; Enç, 1996). This suggests that the Dutch future tense is not the highest-certainty construction type available. It may therefore play a less clear role when contrasted with other possible FTR constructions. This echoes accounts of *zullen* as encoding modally weakened notions of expectation or prediction (Broekhuis & Verkuyl, 2014). On the other hand, relative to *kunnen* 'may', *zullen* and *gaan* likely encode relative modal strength. This suggests *zullen* and *gaan* also encode high certainty.

In fact, in previous studies, this account was supported by ratings of future tense semantics in English and Dutch (Chapter 2). Participants used a slider to rate future- and present-tense FTR statements between high (100) and low certainty (0). In English, the future tense was rated as significantly higher certainty than the present tense. In Dutch, the present tense was rated as significantly higher. Both were in the upper ranges of the scale (> 70). This suggests that the English and Dutch future tenses both encode high-certainty modality, but that paradigmatic contrasts may play a role in determining relative modal semantics. We additionally found that there was significantly greater variation in Dutch rating of the future tense. This is compatible with our account that the Dutch future tense plays a less-unique paradigmatic role within Dutch FTR. These findings suggest that future tenses tend to encode high certainty, and also that cross-linguistic differences may be expected in such semantics. Additionally, FTR status may be a determinant of such differences. This undermines Chen's (2013) tacit argument about the semantics of the FUTURE TENSE. In the next section we outline how our modal account of the semantics of future tenses leads to different predictions than Chen's (2013) temporal account.

3.1.5 Divergent predictions of the modal and temporal accounts

If future tenses tend to encode notions of high certainty, their use might be expected to impact risk preferences. Obligatory use of high-certainty constructions might cause speakers to construe future outcomes as less risky. This would make them less risk adverse, e.g. more likely to choose a 50% chance of $20 over a guaranteed $10. Identical causal reasoning to Chen's simple decision-making model also applies to decisions involving risk. Imagine a decision maker is uncertain about the precise probability p she will receive a reward and holds beliefs about that probability with a distribution $F(p)$. As with temporal beliefs, this simply captures approximate representations of risk when the true probability of some future outcome in unknown, as is often the case. If she discounts risky rewards with rate λ, she will prefer R if and only if:

$$C < \int e^{-\lambda p} R dF(p) \tag{3.3}$$

which is to say that she will only prefer the discounted reward R if it is greater than the cost of gambling. For instance, given the choice between \$10 and a 50% chance of \$20, she will only choose the \$20 if its subjective value as a function of λ and $F(p)$ is greater than \$10. Strong-FTR languages oblige the use of a future tense for prediction-based FTR. If future tenses tend to encode high certainty, this might lead speakers of strong-FTR languages to construe future events as less risky, i.e. hold beliefs about probability $F_W(p) < F_S(p)$. This would translate directly into higher willingness to choose risky rewards:

$$\begin{gathered}\text{if } \forall p,\ F_W(p) < F_S(p) \text{ then} \\ \int e^{-\lambda p} R dF_W(p) < \int e^{-\lambda p} R dF_S(p)\end{gathered} \tag{3.4}$$

which is to say that for all p, if strong-FTR speakers believe future outcomes are relatively less risky, the value of the risky reward for strong-FTR speakers will also be higher.

There is evidence that people factor in risk when making what seem to be "purely" intertemporal decisions. For instance, previous experience of high inflation rates have been shown to impact putative (strictly) intertemporal decision-making outcomes in Polish (Ostaszewski et al., 1998) and Argentine (Macchia et al., 2018) participants. Considerations of this nature have led a growing number of researchers to focus on discounting when rewards are both delayed and risky (Vanderveldt et al., 2015, 2017; Yi et al., 2006). Such "risky intertemporal" decisions are measured by having participants choose between immediate, certain rewards, and delayed, uncertain ones, e.g. \$10 now or a 50% chance of \$40 in three months. Growing evidence attests to the utility of such approaches. For instance, probability and delay have been found to multiplicatively interact in risky intertemporal decision making (Vanderveldt et al., 2015). This indicates that delay discounting is contingent on the probability associated with an outcome and *vice versa*. If this is true, beliefs about future risk would also be expected to impact time preferences (contingent on the assumption that unmeasured risk may impact what seem to be purely intertemporal decisions).

This means Chen's (2013) temporal account and our modal account have different predictions about how obligatory future tense use should impact temporal discounting. The temporal account construes the *future tense* as encoding temporal distance (or precision), and therefore predicts that obliging the *future tense* should cause increased temporal discounting (Chen, 2013). On the other hand, the modal account understands future tenses as tending to encode high certainty (Broekhuis & Verkuyl, 2014; Salkie, 2010). Future tense use should therefore cause decreased temporal discounting. We refer to this as the modal hypothesis. Additionally, the temporal account (Chen, 2013) makes an argument about the semantics of the FUTURE TENSE. We have argued that future tenses may exhibit different semantics. The temporal account therefore predicts that future tense use will relate to temporal discounting in the same way in all languages; our account predicts cross-linguistic differences in this relationship.

We therefore conducted a series of studies designed to test these predictions. Our general empirical approach was to elicit free-text FTR data and temporal discounting in speakers of English and Dutch. We used mediation analyses to estimate the impact of speaking English on temporal discounting *via* elicited future tense use (Studies 1 and 2). To address possible cultural impacts on our outcomes, we then directed Dutch speakers to use the future tense and estimated whether this intervention impacted temporal discounting (Study 3).

3.2 Study 1

Our goal in Study 1 was to ascertain whether increased future tense use in English caused higher or lower temporal temporal discounting. Higher discounting supports the temporal account of tense semantics (distance/precision). Lower discounting support the modal account (high certainty).

3.2.1 Methods

3.2.1.1 Participants

An initial sample of $N = 240$ participants completed the study. However, one participant was excluded because their answers to the FTR-elicitation task were given in bad-faith and four were excluded because of missing demographic data. This left a final sample of $n = 235$ ($n = 113$ in English [$n = 62$ females, $n = 49$ males, $n = 2$ other], and $n = 122$ in Dutch [$n = 106$ females, $n = 16$ males]). English speakers were recruited from the University of Oxford and from Prolific Academic, and Dutch speakers were recruited from the Max Planck Institute for Psycholinguistics (MPI). All participants were screened such that they spoke the respective language of the survey as their first language, and were at least 18 years old. Data were collected in January and February 2017. Ethical approval for the study was granted by the University of Oxford internal review board, ref. no. R39324/RE001, and by the MPI for the data collection that occurred there. All participants were remunerated.

3.2.1.2 Materials

Study 1 comprised two tasks: an FTR-elicitation task designed to establish individual future-referring language and an intertemporal-choice task designed to measure temporal discounting.

The FTR-elicitation task: To establish FTR language we created a new FTR-elicitation task which we based on the EUROTYP FTR questionnaire (Dahl, 1985, 2000a), on which FTR status is predominantly based (Chen, 2013). Our version contained many new questions, included quantitative independent variables (which was not a feature in the original), and was able to be used with large samples in an online survey-based paradigm (the original was designed to be administered by bilingual trained linguists so was difficult to administer to a large sample) (see Robertson, 2019a). In this task, participants were given a context and a target sentence. The main verb in the target sentence was unconjugated, and participants were asked to render the target sentence given the context. All contexts referred to future events. Before starting the FTR-elicitation task, participants were given three past time examples. Participants were advised that there "were no correct answers," and that they should complete the questionnaire sentences, "as though they were speaking to a close friend." See Appendix I for items.

FTR elicitation is slow. To keep the length of the survey reasonable, we used a subset of the full (48 item) FTR questionnaire. In selecting this subset, we were concerned that the grammatical constraints of the items did not overwhelmingly drive results. In other words, we reasoned there might be certain items which did not permit any variation in possible responses. This would mean that speaker-level differences could not influence elicited FTR data, which would not have been ideal since we wanted to correlate speaker-level elicited FTR language with temporal discounting. We therefore conducted a pilot study with the full FTR-questionnaire and used this data to chose our subset. We regressed future tense use over an intercept allowed to randomly vary by both item and participant. We selected those items for which "item effects" explained the least variance. In other words, these were the language-matched sample of questions for which answers were least constrained by context (see Appendix D). This process resulted in a sample of $N = 29$ items. However, data from three items needed to be excluded for two reasons. Firstly, inclusion of tense or modal words in the contexts (*go, think*, etc.) made it impossible to identify whether participants had used the words themselves or had copied the prompt. Secondly, for one item, it was impossible to tell whether participants interpreted it as referring to present or future time.[3]

[3] The item was: "[Q: Do you think your dad will go to sleep?] A: Yes, he {BE} tired." (context is given in [brackets]). Participants overwhelmingly responded with e.g. *Yes he is tired*. It is impossible to tell whether this means that he is tired *now* or that he will be tired *later*.

There were 3 within-subjects factors: (1) distance from present (later today, tomorrow, one week, six months, one year, two years, ten years, and 25+ years, indeterminate, and ongoing predictions[4]); (2) modality condition (high-certainty, low-certainty, neutral); and (3) FTR-mode (whether the context involved a prediction or an intention).[5]

Following Dahl (2000a), we defined predictions as contexts in which the future outcome is not under the control of the speaker, requiring the speakers to estimate the likelihood of the event, e.g. "It {RAIN} tomorrow." We defined intentions as contexts which generally are under the control of the speaker (when people know what they or others intend to do), e.g. "I/you {DRIVE} to the store." Temporal distance indexed the temporal distance from the time of speech. We operationalised temporal distance by using temporal adverbials in the context ("one month", "two weeks", etc.) or by referring to well-known events which would occur a specific length of time after data collection, e.g. the U.S. election. We operationalised modality condition using parenthetical verbal "certainty information", which indicated how certain the participant was supposed to be:

neutral	**Context**: It's no use trying to swim in the lake tomorrow... **Target**: ...The water {BE} cold (then).
low-certainty	**Context**: Q: What is the weather forecast for next week? **Target**: A: It {RAIN} (a 50% chance).
high-certainty	**Context**: Don't bother investing in the tech industry... **Target**: ...Silicon Valley {CRASH} within two years (I'm very sure).

In the high-certainty condition, participants were told to imagine they were highly certain, in the low-certainty condition they were told to imagine they were highly uncertain, and in the neutral condition they were not given any certainty information. To avoid participants simply reproducing the certainty information in their responses, above each item was the directive to "omit material within parentheses."

Free text responses were classified using the FTR-type classifier (Robertson, 2019b), which is a deterministic closed-vocabulary keyword-based classification program we wrote in Python (Python Software Foundation, 2017). Simply scoring responses which were formally future tensed as "future tense" would have resulted in confounded results because it is possible in both English and Dutch to encode modal notions in addition to using either the future or present tense, e.g. *it will possibly rain tomorrow* (future + modal), *rain tomorrow is possible* (present + modal). Reasoning that semantic notions were more likely to reflect latent participant beliefs about the future (and therefore correlate with temporal discounting), we decided that

[4] The last two were categories in the original EUROTYP questionnaire; for example "[A: I have a headache. B: Take this medicine...] ...It {MAKE} you feel better." (ongoing prediction), and "[Don't bother investing in real estate...] ...the housing market {CRASH} (it's possible)." (indeterminate).

[5] After exclusions, no schedule-based FTR items remained.

the FTR-type classifier should classify according to semantics of the target sentence and not the tense structure. We therefore implemented the following class schema.

Future tense: Responses were classed as future tense if they used commonly accepted "future" auxiliaries or explicit temporal adverbials (English *will, shall, be going to, about to*; Dutch *zullen* 'will', *gaan* 'be going to', *staat op* 'about to'). Any response exhibiting these words, without additional modal words (high or low-certainty) was counted as future tense. Responses which met these criteria were scored with (future tense = 1), otherwise (0).

Present tense: Responses were classed as present tense if they conjugated the main verb in the target sentence using the present tense, were not classed as future tense, and did not use additional modal words (as for the future tense). Responses which met these criteria were scored with (present tense = 1), otherwise (0).

The FTR-type classifier's predictions using these schema were verified against a test set of $N = 1350$ ($n = 675$ English, $n = 675$ Dutch) responses from Studies 1–3, which were classified by trained human raters, and all accuracy metrics were $> .97$ (see Appendix D). Duplication of high-frequency responses means that the test set had high coverage, i.e. approximately 69% of responses in English and 73% in Dutch.[6].

In order to be able to compare elicited FTR language and discounting at the participant-level, we used elicited FTR language means per participant as our dependent variables. These were equal to $FTR_j = \sum_{n}^{i=1} \frac{FTR_{ij}}{n_j}$, where FTR are scored future/present responses for participant j, and n is the total number of responses by that participant; in other words, the elicited FTR proportions.

The intertemporal-choice task: In this task, participants made a series of binary choices between a small reward given now (the "smaller-sooner" reward), and a larger reward given after a delay (the "larger-later" reward). For example, "Would you rather have £7.50 now or £10 in two months?" (Dutch participants were given amounts in euros.) The amount of the larger-later reward was always £10. Prior to starting the temporal discounting task participants were told to "try to answer quickly and honestly, without thinking about it too much" and completed a training example with a larger-later reward of £10,000. The amounts of the smaller-sooner reward were £4.00-£9.50 by increments of £0.50; the delays of the larger-later reward were one day, two days, one week, two weeks, one month, six months, two years. Amounts and delays were fully crossed to produce a battery of test items ($12_{amounts} \times 7_{delays} = 84_{items}$).

[6]There were no negated, or "mixed modal" in Study 1, so these did not need to be excluded, see Section 1.4.1.4

To calculate participant-level discounting we followed the procedure given in Kirby et al. (1999). Research has shown that the following hyperbolic function fits real and hypothetical delay-discounted value well (Mazur, 1987; Kirby et al., 1999):

$$V = \frac{A}{1 + kD} \tag{3.5}$$

where V is the subjective value of a delayed reward A at a given delay D, and k is a scaling parameter which captures individual differences in discounting. Higher values of k imply more discounting, i.e. lower subjective value of the delayed larger-later reward, V.

To derive participant-level k, we followed Kirby et al. (1999). This involved calculating hypothetical values of k and retaining for each participant some k which best predicted empirical choices. Specifically, we calculated all k_{1-n} at <u>indifference</u> between LLR and SSR, i.e. all ks such that $SSR = LLR/(1 + k_i D)$ for all $n = 84$ values of SSR by D under the study. We then created hypothetical intertemporal choices for each k_{1-n} by using equation 3.5, i.e. if $SSR > \frac{LLR}{(1+kD)}$, we predicted $choice = SSR$, otherwise $choice = LLR$. We then retained k_i for each participant p_j which had the highest proportion of matches against p_j empirical choices. When more than one k had an equal number of matches, we took the geometric mean of all (Kirby et al., 1999). This method appeared to accurately capture participant-level discounting, correctly predicting 93.04% of all intertemporal choices. Since k tends to be approximately exponentially distributed, we used $log_e(k)$ as our dependent variable in order to better approximate regression assumptions about error normality (Kirby et al., 1999).

3.2.1.3 Procedure

The survey was hosted on Qualtrics, and all participants completed it online. English first-language speaking participants completed the survey in English, and Dutch first-language speaking participants completed the survey in Dutch. Before commencing, participants were asked to confirm their first language matched the survey language. If they did not, they were immediately ejected from the survey. Participants then answered some demographic questions (age, sex, income, education, marital status, and employment status). We never found that any of these significantly impacted outcomes, so they are henceforth disregarded (see Appendix D). All participants then completed the FTR-elicitation task and then the intertemporal choice task. In both tasks, item order was fully randomised and one item was displayed per page.

3.2.2 Results

We report descriptive statistics and correlations for Study 1 in *Table 3.1* and *Table 3.2*. We found English speakers used more future and fewer present tense constructions, which supports the classification of English as strong- and Dutch as weak-FTR. Future tense use

Table 3.1

Descriptive statistics for Studies 1–3

study	variable	language	condition	$\bar{X}$	SD	$\tilde{X}$	MAD^*	min	max	$Q_{0.25}$†	$Q_{0.75}$†
Study 1	future tense	English	-	0.59	0.13	0.62	0.11	0.19	0.85	0.52	0.69
		Dutch	-	0.33	0.16	0.31	0.14	0.00	0.68	0.23	0.42
	present tense	English	-	0.06	0.08	0.04	0.06	0.00	0.42	0.00	0.08
		Dutch	-	0.40	0.18	0.38	0.13	0.04	1.00	0.31	0.5
	$log_e(k)$	English	-	-3.15	2.11	-2.87	2.28	-7.20	1.54	-4.90	-1.80
		Dutch	-	-3.17	1.92	-2.89	2.00	-7.20	0.13	-4.57	-1.64
Study 2	future tense	English	-	0.59	0.16	0.62	0.12	0.04	0.92	0.5	0.7
		Dutch	-	0.20	0.18	0.17	0.19	0.00	0.88	0.04	0.29
	present tense	English	-	0.07	0.09	0.04	0.06	0.00	0.75	0.00	0.12
		Dutch	-	0.55	0.24	0.54	0.25	0.00	1.00	0.38	0.75
	$log_e(k)$	English	-	-2.97	1.32	-2.53	1.09	-5.49	-1.79	-3.82	-1.79
		Dutch	-	-3.77	1.38	-3.64	2.09	-5.49	-1.79	-5.49	-2.67
Study 3	future-high-certainty$_{w1}$	Dutch	-	0.44	0.22	0.50	0.25	0.00	0.92	0.25	0.58
	future tense$_{w1}$	Dutch	-	0.48	0.21	0.49	0.22	0.00	0.90	0.33	0.62
	future tense$_{w2}$	Dutch	present	0.07	0.14	0.01	0.02	0.00	0.86	0.00	0.05
		Dutch	future	0.88	0.22	0.97	0.02	0.01	0.99	0.92	0.99
	$log_e(k)change_{w2-w1}$	Dutch	present	-0.13	0.66	0.00	0.32	-2.62	1.58	-0.45	0.14
		Dutch	future	-0.17	0.80	0.00	0.40	-4.04	2.26	-0.46	0.08

*Median absolute deviation.
†Quartiles.

was negatively correlated with discounting in English but not Dutch. This suggests the English future tense encodes high certainty rather than temporal distance.

3.2.2.1 Mediation analysis

To test whether higher future tense use in English resulted in less discounting, we conducted a mediation analysis. In mediation analyses, the predictor is referred to as X, the mediator is referred to as M, and the outcome is referred to as Y (see *Fig. 3.1*). The $X \rightarrow M$ path is referred to as α, the $M \rightarrow Y$ path is referred to as β, and the $X \rightarrow Y$ path is referred to as τ', or the "direct effect". The $X \rightarrow M \rightarrow Y$ path is referred to as the "indirect effect"; it can be estimated as the product of the paths involved ($\alpha\beta$). It captures the effect of X on Y via M while controlling for the direct effect (τ'). The direct effect captures the effect of X on Y while controlling for the indirect effect ($\alpha\beta$). The "total effect" is the sum of $\alpha\beta$ and τ', and captures the total effect of X on Y, as in a normal regression (Yuan & MacKinnon, 2009). To allow our model to capture language-wise differences in temporal discounting over future tense, we allowed β to be conditional on language by including an interaction term for future tense by language, see *Fig. 3.1*.

Table 3.2

Pearson correlations for variables in Studies 1 and 2, by language

study	language	variable	future tense	present tense
Study 1	English	future tense		
		present tense	-0.57***	
		$log_e(k)$	-0.20*	0.09
	Dutch	future tense		
		present tense	-0.83***	
		$log_e(k)$	0.10	-0.10
Study 2	English	future tense		
		present tense	-0.59***	
		$log_e(k)$	-0.15**	0.12*
	Dutch	future tense		
		present tense	-0.73***	
		$log_e(k)$	-0.04	0.01

$^{***}p < .001$; $^{**}p < .01$; $^{*}p < .05$; $\cdot p < .1$

We used $log_e(k)$ as our outcome variable (Y), language (English = 1, Dutch = 0) as our predictor (X), and participant-level mean future tense use as our mediator (M). The model therefore took the following form:

$$\begin{aligned} fut_i &= \lambda_1 + \alpha lang_i + e_{1i} \\ log_e(k)_i &= \lambda_2 + \tau' lang_i + \beta_1 fut_i + \beta_2 fut_i lang_i + e_{2i} \end{aligned} \tag{3.6}$$

where $\lambda_{1,2}$ are intercepts, and α, τ', and β_1 are slope coefficients as described. Because β_1 is moderated by language, the indirect effect is calculated conditional on β_2, i.e. $ind.eff. = \alpha(\beta_1 + \beta_2 language)$. For Dutch ($language = 0$), this simplifies to the normal indirect effect as described, but for English ($language = 1$) the interaction term β_2 is added. For parity, we report the indirect effect conditioned on both α and β by language, where α_{eng} is α, and α_{dut} is the future tense intercept λ_1. This allows us to estimate the effect of observed English and Dutch levels of future tense use conditional on observed English and Dutch relationships between future tense use and discounting, see *Fig. 3.1*.

Bayesian statistics are well-suited to mediation analyses. As well as making no assumptions about the normality of sampling statistics, they allow for straight-forward inferences about any transformation of model parameters (i.e. path products) through carrying out the desired operation on posterior probability distributions (Vehtari, Gelman, Simpson, Carpenter, & Bürkner, 2019). We therefore used the *brms* package (Bürkner, 2017) in *R* (R Core Team, 2020) to estimate model parameters using a no U-Turn Hamiltonian Monte Carlo sampling procedure (Hoffman & Gelman, 2014; Stan Development Team, 2020) with uninfor-

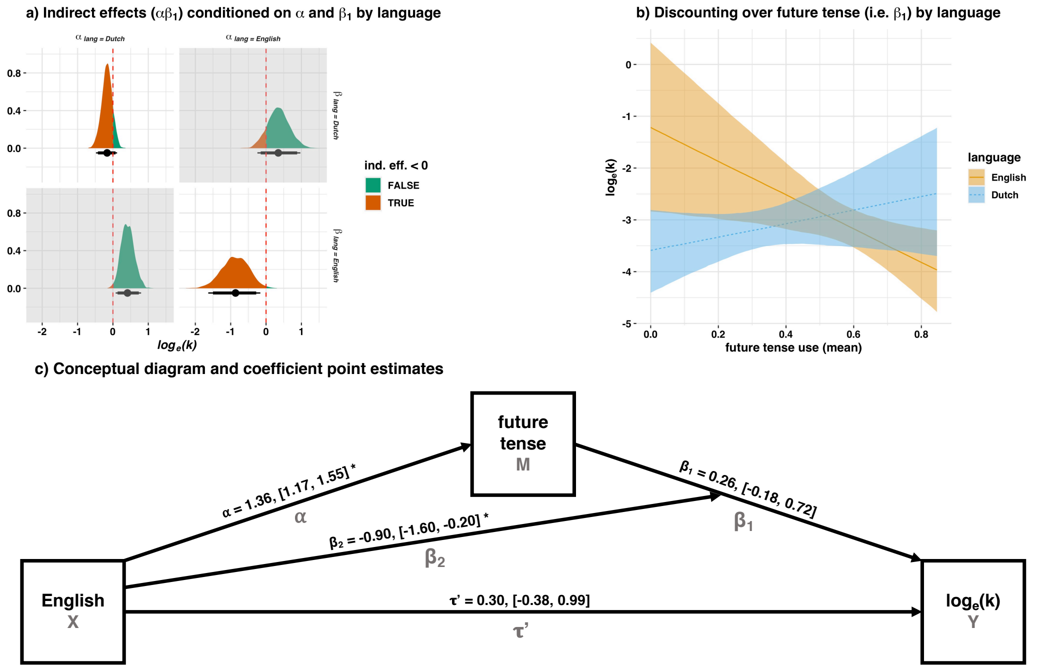


Figure 3.1: Study 1 mediation analysis, with conditional indirect effects of language on temporal discounting via future tense use. Effects are moderated by language, i.e. they are the indirect effects with both α and β estimated for each language. a) Effects are given by eff $= (\alpha|lang)(\beta_1 + \beta_2 lang)$, where $\alpha_{lang=dutch}$ is the future tense intercept. For Dutch ($language = 0$) this simplifies to a classical path product ($\alpha\beta$); for English ($language = 1$) the interaction term is added. Credibility intervals are plotted at 90% (thick bars) and 95% (thin bars). For two-tailed hypotheses, thin bars represent cut-off criteria; for one-tailed hypotheses, thick bars do (providing sign is as predicted). Counterfactual parameters are in grey. b) Marginal effects of participant-level $log_e(k)$ values over mean future tense use conditioned on language (i.e. β for each language) with 95% credibility intervals. Lower values indicate less discounting. c) Conceptual model diagram; interaction term is represented by paths intersecting. Full parameters are given in *Table 3.3*, keyed by path name. To ease interpretation, figures throughout use raw predictor values, not the mean-centred z-scaled variables used in regressions.
*95% CI does not contain zero.

mative priors (i.e. $Unif(-\infty\text{–}\infty)$), 4 chains, 2,000 iterations, and a burn in discard of 1,000 (the defaults). Inspection of caterpillar plots and $\hat{R}$s indicated that the estimation procedure had converged (Vehtari et al., 2019).

The predictors explained a large proportion of future tense variance and a small proportion of discounting variance, $R^2_{future} = 0.46$, $CI_{95\%} = [0.38, 0.52]$, $R^2_{log_e(k)} = 0.04$, $CI_{95\%} = [0.01, 0.09]$. These are a Bayesian adaptation of R^2, including spread statistics for values at the 2.5% and 97.5% percentiles (Gelman, Goodrich, Gabry, & Vehtari, 2018).

Fig. 3.1a reports indirect effects conditioned on language-wise differences in α and β, as explained. Counterfactual thinking is necessary to understand these parameters (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014). For instance, *Fig. 3.1a, bottom left* reports the (counterfactual, in grey) effect of the observed Dutch levels of future tense use ($\alpha|lang = Dutch$)

combined with the observed English relationship between future tense use and discounting ($\beta|lang = English$). The significant effect indicates that low use of the future tense in Dutch would cause increased discounting, were discounting over future tense use the same in Dutch as in English (which it is not).

The quantities of interest are the non-counterfactual parameter estimates (in white). In particular, we are interested in the (true) effect of English use of the future tense, i.e. the effect given α_{eng} & β_{eng}. This effect is negative, *Fig. 3.1a, bottom right*, and significant, $Est. = -0.87$, $CI_{90\%} = [-1.5, -0.28]$, $pp = .991$. In these parameter summaries, $Est.$ is the mean of the posterior probability distribution, CI is the x% credibility interval (90% in this case because the hypothesis is one-tailed, i.e. the modal hypothesis $\alpha\beta < 0$), and "*pp*" reports the probability that the posterior estimate of the effect matches the direction of the hypothesis. In contrast to frequentist *p* values, $pp > .95$ indicates evidence that the parameter matched the prediction exceeds 95% (Bürkner, 2017). In summary, English speakers used significantly more future tense constructions, i.e. α in *Fig. 3.1c*. In English, the relationship between future tense use and discounting was negative, *Fig. 3.1b*. As a consequence, the indirect effect was negative (as the modal account predicts), not positive (as the temporal account predicts).

3.2.2.2 Language-wise differences in discounting over future tense use

Future tense use had differential effects on temporal discounting in English and Dutch. In English, the relationship was negative and significant, i.e. β_2 in *Table 3.3*. In Dutch it was not significant, i.e. β_1 in *Table 3.3* (also see *Fig. 3.1b*). This undermines accounts which treat the FUTURE TENSE as exhibiting universal relationships with discounting (cf. Chen, 2013). Rather, it suggests future tenses relate to discounting differently in different languages, and that FTR status is a determinant of such relationships—as was suggested by our paradigmatic analysis of future tense semantics in English and Dutch.

3.2.2.3 Exploratory analyses: Population-level temporal discounting differences

A relevant question is whether English speakers discounted more in total. The temporal hypothesis predicts this should be the case (Chen, 2013). To test this, we calculated the total effect conditioned on English i.e. $tot.\ eff. = \tau' + \alpha(\beta_1 + \beta_2)$. We tested the temporal hypothesis prediction that $tot.\ eff. > 0$. It was not supported, $Est. = -0.57$, $CI_{90\%} = [-1.17, 0.02]$, $pp = .058$. In other words, we found no difference between these populations in temporal discounting. This was likely driven by the slightly positive, though non-significant (see τ' in *Fig. 3.1c*), direct effect being offset by the negative indirect effect. This analysis therefore fails to support the predictions of the temporal hypothesis that English speakers should temporally discount more than Dutch speakers.

Table 3.3

Regression coefficients for mediation analyses Studies 1–3

study	path name	outcome	predictor	estimate	SE	low-95%CI	high-95%CI^{**}	
Study 1	λ_1 ($\alpha_{lang=dut}$)	future tense	Intercept	-0.65	0.07	-0.78	-0.52	*
	λ_2	$log_e(k)$	Intercept	-3.00	0.24	-3.46	-2.52	*
	α ($\alpha_{lang=eng}$)	future tense	English	1.36	0.10	1.17	1.55	*
	β_1	$log_e(k)$	future tense	0.26	0.23	-0.18	0.72	
	τ'	$log_e(k)$	English	0.30	0.35	-0.37	0.99	
	β_2	$log_e(k)$	future tense * English	-0.90	0.36	-1.60	-0.20	*
Study 2	λ_1 ($\alpha_{lang=dut}$)	future tense	Intercept	-0.75	0.04	-0.82	-0.67	*
	λ_2	$log_e(k)$	Intercept	-3.82	0.12	-4.04	-3.6	*
	α ($\alpha_{lang=eng}$)	future tense	English	1.51	0.05	1.40	1.62	*
	β_1	$log_e(k)$	future tense	-0.07	0.11	-0.29	0.15	
	τ'	$log_e(k)$	English	1.10	0.17	0.77	1.44	*
	β_2	$log_e(k)$	future tense * English	-0.26	0.17	-0.59	0.07	
Study 3: Model 1	λ_1	future tense$_{w2}$	Intercept	-0.78	0.05	-0.88	-0.68	*
W = future tense$_{w1}$	λ_2	$log_e(k)change_{w2-w1}$	Intercept	-0.39	0.15	-0.69	-0.08	*
	α	future tense$_{w2}$	future condition	1.83	0.07	1.69	1.97	*
	β_1	$log_e(k)change_{w2-w1}$	future tense$_{w2}$	-0.34	0.16	-0.67	-0.03	*
	β_2	$log_e(k)change_{w2-w1}$	future tense$_{w1}$	0.07	0.06	-0.05	0.19	
	τ'	$log_e(k)change_{w2-w1}$	future condition	0.56	0.32	-0.05	1.21	
	β_3	$log_e(k)change_{w2-w1}$	future tense$_{w2}$ * future tense$_{w1}$	-0.01	0.07	-0.13	0.13	
Study 3: Model 2	λ_1	future tense$_{w2}$	Intercept	-0.78	0.05	-0.88	-0.67	*
W = future-high-	λ_2	$log_e(k)change_{w2-w1}$	Intercept	-0.38	0.16	-0.69	-0.08	*
certainty$_{w1}$	α	future tense$_{w2}$	future condition	1.83	0.07	1.69	1.98	*
	β_1	$log_e(k)change_{w2-w1}$	future tense$_{w2}$	-0.34	0.16	-0.66	-0.02	*
	β_2	$log_e(k)change_{w2-w1}$	future-high-certainty$_{w1}$	0.04	0.07	-0.08	0.17	
	τ'	$log_e(k)change_{w2-w1}$	future condition	0.54	0.33	-0.10	1.18	
	β_3	$log_e(k)change_{w2-w1}$	future tense$_{w2}$ * future-high-certainty$_{w1}$	0.07	0.07	-0.06	0.19	

All predictors were mean-centered and z-scaled so coefficient point estimates represent predicted changes in outcomes as a function of changes in 1 standard deviation in the predictors. Point estimates are the mean of the posterior probability distribution. Bayesian R^2 equivelents are reported in the main text.
*Bayesian 95% Credibility Interval (CI) does not contain zero.
**Credability intervals are reported at 90% in the main text because they test one-tailed hypotheses.

3.2.3 Discussion

The principal result of Study 1 is that obligatory future tense use in our strong-FTR language (English) resulted in less rather than more discounting. This suggests the English future tense encodes high certainty, not temporal distance or precision (cf. Chen, 2013). This supports the modal and not the temporal account. Additionally, we found no difference in temporal discounting between English and Dutch speakers, which fails to support the temporal account (cf. Chen, 2013). Finally, future tense use (β) had differential effects on discounting in English and Dutch (cf. Chen, 2013). This was expected given our paradigmatic analysis. The English future tense is the highest-certainty FTR construction allowed for prediction-based FTR, whereas the Dutch future tense may be contrasted with the unmarked present tense. Our findings are compatible with the conclusion that this is a factor in the non-significant relationship between future tense use and temporal discounting in Dutch.

3.3 Study 2

The Study 1 results undermine the widely-cited temporal account which has been used to motivate and interpret a great deal of work (Bernhofer, Costantini, & Kovacic, 2019; Chen, 2013; S. Chen et al., 2017; Chi et al., 2018; Fasan et al., 2016; Figlio et al., 2016; Falk et al., 2018; Gotti et al., 2020; Guin, 2017; Hübner & Vannoorenberghe, 2015b, 2015a; J. Kim et al., 2017; S. Kim & Filimonau, 2017; Liang et al., 2018; Lien & Zhang, 2020; Mavisakalyan et al., 2018; Pérez & Tavits, 2017; Schellen, 2019; Sutter et al., 2015; Thoma & Tytus, 2018). Rather than high future tense use in English resulting in more temporal discounting, it resulted in less. If our findings generalise to other languages, this would mean that something other than obligatory future tenses were driving observed effects of FTR status on intertemporal behaviour. Given the extent of the findings this result undermines, a replication seemed appropriate. Additionally, we reasoned that Study 1 could have been underpowered. In Study 1, the relationship between future tense use and discounting in Dutch was actually positive, though not significantly different from zero. If this parameter *were* significant, the direction of this relationship would have supported the temporal hypothesis. It is plausible that the Dutch future tense could encode temporal distance/precision while the English future tense encodes high certainty. If Study 1 had been underpowered, we might have made a type II error. In Study 2, we therefore replicated Study 1, but with a larger sample. We also adjusted that amounts and delays in the intertemporal-choice task; our aim was to increase the resolution of the task to better capture correlations between future tense use and discounting (see Appendix D). We were interested in whether the updated amounts, delays, and sample size would result in the Dutch indirect effect reaching significance, and also whether the negative indirect effect in English would replicate.

3.3.1 Methods

3.3.1.1 Participants

An initial sample of $N = 625$ participants completed Study 2 (this number was determined by a power analysis, see Appendix D). However, $n = 7$ participants were excluded because inspection indicated their FTR-elicitation responses were given in bad faith and $n = 12$ participants were excluded because a survey error resulted in excessive missing data. After these exclusions, a final sample of $n = 606$ participants completed Study 2 ($n = 301$ in English [$n = 134$ males, $n = 163$ females, and $n = 4$ other], and $n = 305$ in Dutch [$n = 127$ males, $n = 177$ females, and $n = 1$ other]). English speakers were recruited from Prolific Academic and Dutch speakers from Qualtrics. All participants were screened such that they spoke the respective language of the survey as their first language, and were 18 or over. Data were collected in October and November 2018 (Dutch) and May 2020 (English). Ethical approval for the study was granted by the University of Oxford internal review board, ref. no. R39324/RE001. All participants were remunerated.

3.3.1.2 Materials

Study 2 again comprised two tasks: the same FTR-elicitation task as in Study 1, and an intertemporal-choice task with updated amounts and temporal delays.

The FTR-elicitation task: The FTR-elicitation task was identical to Study 1, and the dependent variables were the same. However, in addition to the three items which needed to be excluded from Study 1, an additional two items were excluded from Study 2, leaving a final sample of $N = 24$ items. This was because an oversight meant that a specific temporal adverb ("by 2018") was in the past at the time of data collection (see Appendix D).

The intertemporal-choice task: The amounts of the smaller-sooner rewards in Study 2 were £7.50-£9.25 by increments of £0.25, and the delays of the larger-later reward were 2 weeks, 1 month, 1½ months, 2 months, 2½ months, 3 months, 4 months, and 4½ months.[7] Amounts and delays were fully crossed to produce a battery of test items ($8_{amounts} \times 8_{delays} = 64_{items}$). The dependent variable was again $log_e(k)$, as in Study 1. The procedure for estimating participant-level k again appeared to accurately capture participant-level discounting, correctly predicting 94.29% of responses.

3.3.1.3 Procedure

Procedure was identical to Study 1. English-speaking participants completed the survey in English and Dutch-speaking participants completed the survey in Dutch. Task order, presentation, and demographic covariates were exactly replicated.

3.3.2 Results

Replicating Study 2, future tense use was negatively correlated with discounting in English but not in Dutch, *Table 3.2*. This suggests that higher future tense use in English would again result in less discounting, as the modal hypothesis predicts, and that the Dutch indirect effect would again be non-significant.

3.3.2.1 Mediation analysis

To test this, we re-estimated the same model with the same estimation parameters as in Study 1 (equation 3.6), but with the Study 2 data. This time, the model explained a similar proportion of future tense use variance and a slightly higher proportion of discounting variance, $R^2_{future} = 0.57,\ CI_{95\%} = [0.53, 0.6]$, $R^2_{log_e(k)} = 0.09,\ CI_{95\%} = [0.05, 0.14]$.

Once again, we found that discounting over future tense use was negative in English (*Fig. 3.2b*) and that the indirect effect conditioned on English was negative (*Fig. 3.2a, bottom right*), as the modal hypothesis predicts, $Est. = -0.5,\ CI_{90\%} = [-0.83, -0.19],\ pp = .995$. This replicated our main Study 1 finding that higher future tense use in English resulted in decreased—rather than increased—discounting.

[7] Three and a half months was accidentally not included due to an oversight.

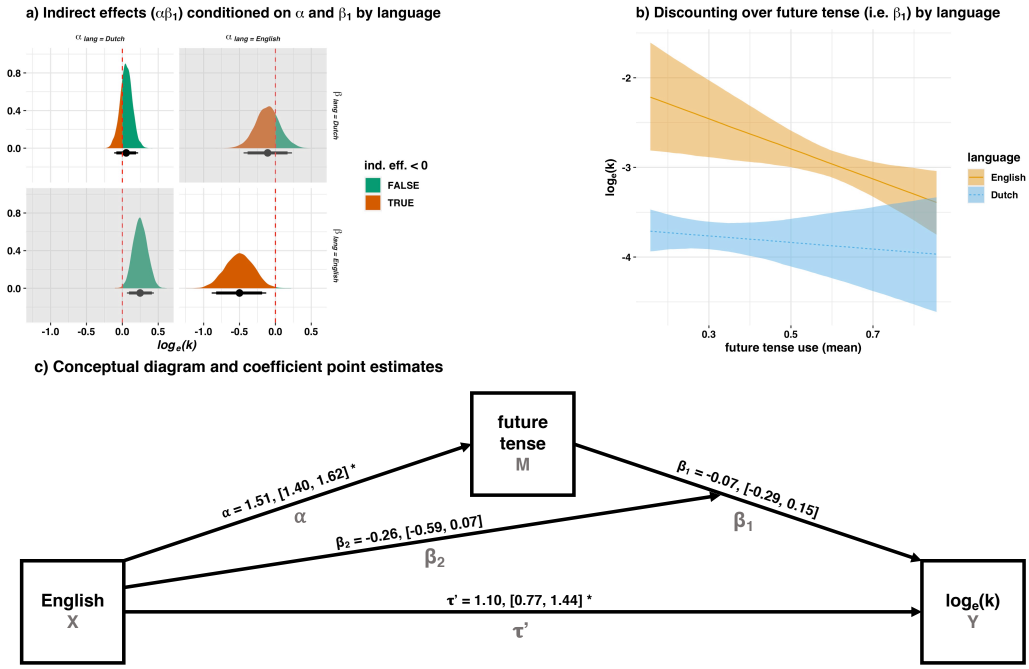


Figure 3.2: Study 2 mediation analysis, with conditional indirect effects of language on temporal discounting via future tense use, moderated by language. a) Estimates of conditional indirect effects; see *Fig. 3.1* for notes on interpretation. b) Marginal effects of participant-level $log_e(k)$ over mean future tense use conditioned on language (95% credibility intervals). Lower values indicate less discounting, so higher estimations of delayed future rewards. c) Conceptual model diagram; full parameters are given in *Table 3.3*, keyed by path name. Study 1 replicated: The indirect effect in English was negative and did not contain zero (*a, bottom right*). The indirect effect in Dutch was non-significant (*a, top left*).
*95% *CI* does not contain zero.

Again, discounting over future tense use in Dutch was not significant, i.e. β_1 for Study 2 in *Table 3.3* (see *Fig. 3.2b*). Again, the indirect effect conditioned on Dutch was not significant, $Est. = 0.05,\ CI_{90\%} = [-0.08, 0.19],\ pp = .252$ (*Fig. 3.2a, top left*). In other words, even after increasing the sample size, the indirect effect in Dutch was again not significant. Our Study 1 null finding replicated. This again suggests that future tense use has differential effects in weak- and strong-FTR languages.

3.3.2.2 Exploratory analyses: Population-level discounting differences

This time the direct effect of language on temporal discounting was positive and significant, *Fig. 3.2c*. As a consequence, the total effect conditioned on English was also positive, as predicted by the temporal hypothesis, $Est. = 0.6,\ CI_{90\%} = [0.31, 0.88],\ pp > .999$. This indicates there was a small but significant difference in temporal discounting between English and Dutch speakers. Exponentiating parameter point estimates so they represent changes in k

and plugging these into equation 3.5 indicates that the average subjective value of £/€10 at a four-week delay was approximately €9.19 for Dutch speakers and £8.62 for English speakers.

3.3.3 Discussion

In Study 2—as in Study 1—higher use of the future tense in English resulted in less, not more, temporal discounting. The non-significant effect in Dutch also replicated, despite the improved power of Study 2. The substantive difference between the two studies is that in Study 2 the direct effect of English on temporal discounting was positive and significant. As a consequence, the total effect was also positive and significant. In Study 1, it was not. Our adjustment of intertemporal delays and amounts may have led to this difference. Study 2 had relatively more proximal delays (Study 2 max: 4½ months; Study 1 max: 2 years), and relatively more valuable smaller-sooner reward amounts (Study 2 min: £/€7.50; Study 1 min: £/€4). These adjustments may have increased the resolution of the task to identify smaller differences in time preferences. The temporal hypothesis predicts English speakers should temporally discount more than Dutch speakers, so this result might be viewed as supporting the temporal account. However, the indirect effect of English on temporal discounting via future tense use was negative. This suggests that higher English future tense use is not causally implicated in higher English discounting, as is hypothesised by the temporal account. This entails that other factors are responsible. A plausible account is that time preferences are affected by constraints in English that oblige low-certainty modal FTR—e.g. *could, may*, or *might*. We address this idea in the General Discussion (Section 3.5.1).

3.4 Study 3

3.4.1 Motivations

Study 3 had two principal goals. First, we wanted to address potential cultural confounding. Second, we wanted to explore what was happening with the relationship between the Dutch future tense and temporal discounting. We outline these issues below.

3.4.1.1 Cultural confounding

The principal finding in Studies 1 and 2 is the negative indirect effect of future tense. However, these studies do not address the possibility that cultural, rather than grammatical, differences might have been responsible for higher English future tense use. Linguistic relativity research involves a minimum of three related steps: (a) Cross-linguistic differences need to be identified; (b) correlated differences in non-linguistic cognition need to be established; and (c) causal evidence linking (a) to (b) must be developed. Studies 1 and 2 go some way to addressing all three of these steps. English speakers used more future tense constructions (a), this was shown to impact temporal discounting (b), and mediation analyses established that language use was involved in our outcomes (c). However, Studies 1 and 2 were still essentially correlational. Cultural, not grammatical, factors could have driven higher English use of the future tense. For instance, English speakers—for whatever reason—might construe future outcomes as higher certainty. This might be driving higher future tense use, rather than grammatical constraints. We consider this unlikely because it is well known that the English future tense is more obligatorised than the Dutch future tense (Dahl, 1985, 2000a). It is nonetheless possible.

3.4.1.2 The Dutch future tense

We were perplexed by repeated non-significant relationships between future tense use and temporal discounting in Dutch. In an earlier study, Dutch speakers rated the future tense as encoding high certainty (*Fig. 2.6*, Chapter 2). Why did it not relate to temporal discounting? We speculated that non-significant findings could have resulted from large individual differences in modal future-tense semantics between Dutch speakers. Unlike English, where the future tense is the highest-certainty construction type which is grammatically acceptable for future predictions (Dahl, 2000a), Dutch speakers are free to use either the future or present tense. Both of these are suggested to encode strong modality by most linguists (Broekhuis & Verkuyl, 2014; Giannakidou & Mari, 2018, 2018; Salkie, 2010). The paradigmatic role of the Dutch future tense in the context of the broader FTR system may therefore be less clear. This is compatible with an earlier finding that Dutch ratings of future tense certainty had exhibited greater variance than English ratings (Section 2.3.2.5, Chapter 2). Further, some participants rated the future tense as relatively more certain, while others rated the present tense as relatively more certain (Appendix D). This raises the possibility that temporal discounting over future tense use is positive for some participants and negative for others. This would depend on individual differences in how these structures are used to encode modal notions of certainty. This might explain apparently flat relationships between future tense use and temporal discounting in Studies 1 and 2.

3.4.2 Overview and hypotheses

We designed Study 3 to address these issues. It was pre-registered with AsPredicted (Appendix B) and conducted entirely with Dutch participants. It used a mixed, repeated-measures design. In "Wave 1", participants initially completed an FTR-elicitation and then intertemporal-choice task, as in Studies 1 and 2. Then, several weeks later, in "Wave 2", the same participants completed a similar survey. The only difference was that in Wave 2, before starting the FTR task, participants were randomly directed to use either the present or the future tense (we refer to this as "tense condition"). By intervening directly in participants' language production, this allows us to rule out cultural factors which might have driven language use in Studies 1 and 2.

As in Studies 1 and 2, the FTR-elicitation task included a modality manipulation. Some contexts (neutral condition) included no specific certainty information, while others (high-certainty condition) included a direction for participants to imagine they were highly certain. This made several important changes to extant priming paradigms which have tested for effects of framing intertemporal choices in the present or future tense, e.g. "you *will get* $____ in a month" vs. "you *get* $____ in a month" (Banerjee & Urminsky, 2017; J. I. Chen et al., 2019; Thoma & Tytus, 2018). Firstly, we reasoned a production- rather than comprehension-based paradigm might increase "strength" of the prime. Secondly, measuring FTR language beforehand meant we knew the extent to which participants tended to use the future tense, both in general and across the modality conditions of the FTR task—and we had a baseline measure of discounting.

First, we made a prediction about how Wave 1 future tense use would relate to discounting. We reasoned that future tense use differences across modality condition should reflect participants' understanding of the semantics of the future tense. Participants who construed the future tense as encoding high certainty would be more likely to use the future tense in the high-certainty condition, and participants who construed it as encoding low certainty would be less likely to. Therefore, discounting over future tense use should be negative for participants who use the future tense in the high-certainty condition, and positive for participants who do not. We refer to this prediction as the modal use of tense hypothesis.

Second, we made competing predictions from the perspectives of the temporal and modal hypotheses about how Wave 2 tense condition would impact discounting. Similarly to Studies 1 and 2, if directing participants to use the future tense caused increased discounting, this would support the temporal hypothesis. If it caused decreased discounting, it would support the modal hypothesis.

Third, we made predictions from the contrasting perspectives of the modal and temporal accounts about how tense condition would impact discounting conditional on individual differences in previously elicited FTR data. It is important to note strong negative correlations between future and present tense use in Dutch participants, *Table 3.2*. This indicates that some Dutch speakers exhibited strong-FTR-like results (high future tense use) and others exhibited weak-FTR-like results (high present tense use). With this in mind, according to the temporal hypothesis, directing low future-tense-using participants to use the future tense should make them construe future events more distally and therefore to discount more. Directing high future-tense-using participants to use the future tense should have no effect (as they presumably already construe future events distally). The temporal hypothesis therefore predicts an interaction between Wave 1 future tense use and Wave 2 future tense use as a function of tense condition. Conversely, we have made an argument that semantics of the Dutch future tense are modal and exhibit significant individual differences. Therefore, directing participants to use the future tense should only affect discounting conditional on participant-level understandings of the modal semantics of the future tense. For example, directing a *future tense* ≈ *higher-certainty* participant to use the future tense should make them construe future events as high certainty and therefore to discount less. Directing a *future tense* ≈ *lower-certainty* participant to use the future tense should make them construe future events as low certainty and therefore to discount more. The modal hypothesis therefore predicts that there should be an interaction between Wave 1 use of the future tense in the high-certainty condition and Wave 2 future tense use as a function of tense condition. We refer to these contrasting hypotheses as the tense condition interaction hypotheses.

3.4.3 Methods

3.4.3.1 Participants

In Study 3, $N = 147$ ($n = 53$ males, $n = 92$ females, and $n = 2$ other) native-Dutch-speaking participants completed Wave 1; and after attrition $N = 127$ ($n = 45$ males, $n = 80$ females, and $n = 2$ other) of these completed Wave 2 approximately 2 weeks later (inter-wave period $M = 16.05$ $SD = 12.73$ days). All participants confirmed that their first language was Dutch, otherwise they were ejected from the survey immediately. Due to a survey error, there was missing age data for $n = 39$ participants. Participants were recruited from Radboud University, the CREED centre at the University of Amsterdam, and from Prolific Academic. Data were collected between March–June 2018. Ethical approval was granted by the University of Oxford internal review board, ref. no. R39324/RE001. All participants were remunerated.

3.4.3.2 Materials

Study 3 again comprised two tasks: an updated FTR-elicitation task and an intertemporal-choice task.

The FTR-elicitation task: The FTR-elicitation task was presented identically to Studies 1 and 2. The factors of the FTR-elicitation task were: modality condition (high-certainty, neutral)[8]; FTR-mode (intention, prediction, scheduling); and temporal distance (today, tomorrow, one week, one month, six months, one year, two years, ten years, twenty-five + years, indeterminate, and ongoing predictions). However, in Study 3 it comprised 100 items. We reasoned this would maximise the likelihood of tense condition having an impact on discounting (see below). For the same reasons as in Studies 1 and 2, data from 20 items needed to be excluded from analyses (see Appendix D for exclusions and Appendix I for the items in

[8]The low-certainty condition was removed to make the contexts compatible with tense condition.

the FTR-elicitation task). This meant conditions were not balanced. The dependent variables were participant-level mean future tense use and, additionally, mean future tense use factored by modality condition. This is equal to $FUT_{jk} = \sum_{n}^{i=1} \frac{FUT_{ij}}{n_{jk}}$, where FUT is (1) for a response using the future tense else (0), and n is the total number of responses by participant j within modality condition k. We refer to the resultant condition-wise proportions as "future-neutral", "future-high-certainty", etc.

The intertemporal-choice task: The intertemporal-choice task was identical to Studies 1 and 2, except for it used different amounts and delays. The amounts of the smaller-sooner reward were: €7.00-€9.75 by increments of €0.25. The delays of the larger-later reward were: one month, two months, three months, four months, five months, six months, nine months, and one year. The amount of the larger-later reward was constant at €10.00, and a test battery was created by fully crossing amounts and delays, $12_{amounts} \times 9_{delays} = 108_{items}$. The dependent variable was again $log_e(k)$. However, we now had baseline discounting in Wave 1 (see below), so we used change relative to baseline as our dependent variable, as specified in our pre-registration, i.e. $log_e(k)change = log_e(k)_{Wave\ 2} - log_e(k)_{Wave\ 1}$. Positive values indicate increased discounting, and negative values indicate decreased discounting, as with $log_e(k)$.

3.4.3.3 Procedure

Participants initially completed a baseline survey comprising the FTR-elicitation and intertemporal-choice tasks (Wave 1). This was essentially a replication of Studies 1 and 2: Participants answered the same demographic questions as in Studies 1 and 2, and then completed the FTR-elicitation task, followed by the intertemporal-choice task. Training and instructions were the same as Studies 1 and 2. Several weeks later, the same participants completed a second survey (Wave 2). Wave 2 was identical to Wave 1, except demographic questions were not collected a second time, and tense condition was added.

We operationalised tense condition in the following way: Before beginning the FTR-elicitation task, participants were randomly told to either give their responses strictly using the present tense or the future tense. In both conditions, participants were instructed that "all sentences are about future events." In the present condition, $n = 63$ participants were instructed to "only use the present tense", while in the future condition, $n = 64$ participants were instructed to "only use the future tense." All participants were told to "still try to answer naturally, as if you were talking to a friend." All participants were given an example question involving a woman buying a car. They were presented with four example responses, two in the present tense (one in the simple present *...koopt ze een auto...* '...she buys a car..', and the other in the present progressive *...aan het kopen...* '...she is buying...'), and two in the future tense (one using *zullen* 'will', the other *gaan* 'be going to'). In the present condition,

Table 3.4

Pearson correlations for variables in Study 3

	future tense$_{w1}$	future-high-cert$_{w1}$	future tense$_{w2}$
future-high-cert$_{w1}$	0.72***		
future tense$_{w2}$	0.15	0.15·	
$log_e(k)change_{w2-w1}$	0.08	0.05	-0.10

***$p < .001$; **$p < .01$; *$p < .05$; ·$p < .1$

participants were told the correct answers were the present tense examples, and in the future condition, participants were told the correct answers were the future tense examples. After this, all participants were given a second training example, but this time with a multiple choice question again with two present tense and two future tense choices. Participants were not allowed to advance until they had correctly identified the two choices which matched their assigned tense condition.

Study 3 was hosted on Qualtrics, and all participants completed it online. For both tasks, item order was randomised and one item was displayed per page.

3.4.4 Results

The tense condition manipulation was successful. In the present tense condition, Wave 2 present tense use was high, and future tense use was low. In the future condition, Wave 2 future tense use was high and present tense use was low, *Table 3.1*. See *Table 3.4* for Study 3 correlations.

3.4.4.1 The modal use of tense hypothesis

We first established whether temporal discounting over future tense was conditioned on use of the future tense in high-certainty contexts, as the modal account predicts. To test this, we estimated the following ordinary linear regression using the full $N = 147$ Wave 1 data:

$$log_e(k)_i = \beta_0 + \beta_1 fut.n_i + \beta_2 fut.hc_i + \beta_3 fut.n_i fut.hc_i + e_i \tag{3.7}$$

where $fut.n$ is mean future tense use in the neutral condition (i.e. "future-neutral") and $fut.hc$ is mean future tense use in the high-certainty condition (i.e. "future-high-certainty"). We found that this model did not explain a significant proportion of the variance, $F(3, 143) = 1.86$, $p = .139$, $R^2 = 0.04$. However, the future-neutral by future-high-certainty interaction was significant, *Table 3.5*. For participants who used the future tense to encode high certainty, temporal discounting over future tense use was negative, *Fig. 3.3*. For participants who did

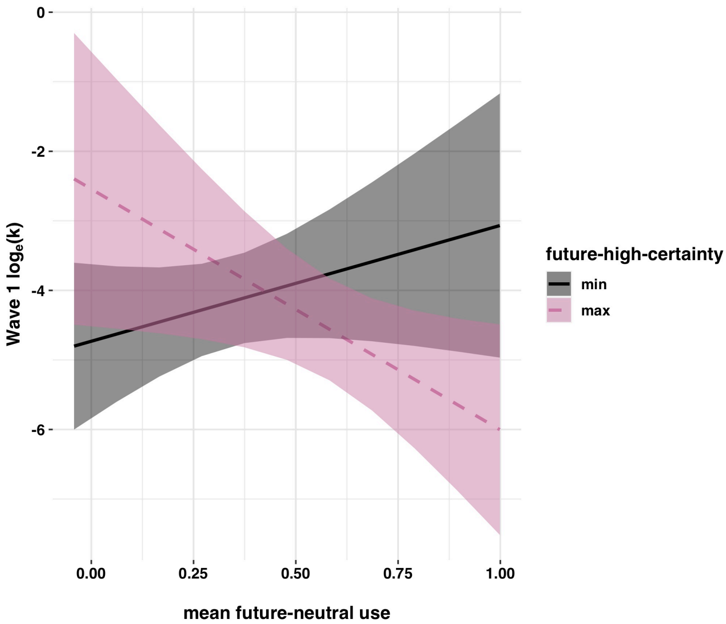


Figure 3.3: Discounting over future-neutral by future-high-certainty in Dutch in Study 3 (Wave 1). Plotted values are the marginal effects, ± 1.96 SE (Lüdecke, 2019). Lower values indicate less discounting, so higher estimations of delayed future rewards. Effects conditioned on future-high-certainty are depicted in black and purple to ease visual differentiation from conditioning on language, depicted in blue and beige in other figures.

not use the future tense to encode high certainty, temporal discounting over future tense use was positive, *Fig. 3.3.* This result supports the modal use of tense hypothesis.

3.4.4.2 The tense condition interaction hypotheses

Next we tested the contrasting predictions involving tense condition, using a moderated mediation analysis. This time, our predictor (X) was tense condition, our mediator (M) was Wave 2 future tense use, and our outcome (Y) was $log_e(k)change_{w2-w1}$. In other words, tense condition was our treatment effect, but since we had a quantitative measure of compliance (actual future tense use at Wave 2), we estimated the effect of condition on discounting via Wave 2 future tense use. We estimated one model to test each hypothesis. To test the predictions of the temporal hypothesis, Wave 1 future tense use was the moderator ("W"). We refer to this model as "Model 1". To test the predictions of modal hypothesis W was Wave 1 future-high-certainty use ("Model 2"):

$$\begin{aligned} fut.w2_i &= \lambda_1 + \alpha\ cond_i + e_{1i} \\ log_e(k)change_i &= \lambda_2 + \tau' cond_i + \beta_1 fut.w2_i + \beta_2 fut.w1_i + \beta_3 fut.w2_i fut.w1_i + e_{2i} \end{aligned} \tag{3.8}$$

Table 3.5

Regression coefficients for future-neutral by future-high-certainty interaction in Study 3

outcome	predictor	β	SE	t-value	p
$log_e(k)_{w1}$	(Intercept)	-4.06	0.15	-26.91	<.001
	future-neutral	-0.16	0.17	-0.96	.34
	future-high-certainty	-0.06	0.17	-0.36	.717
	future-neutral * future-high-certainty	-0.25	0.12	-2.04	.043

All predictors were mean-centered and z-scaled so βs represent predicted changes in outcomes as a function of changes in 1 standard deviation in the predictors. R^2 and F statistics are reported in the main text.

where $fut.w1$ (W), is either Wave 1 future tense use (Model 1) or Wave 1 future-high-certainty use (Model 2), depending on which hypothesis we were testing. We estimated these models with the same parameters as Studies 1 and 2, using a matched sample of the $N = 127$ participants who had completed both waves. See *Fig. 3.4c* for a conceptual diagram and *Table 3.3* for full parameter estimates.

Model 1 explained a large proportion of future tense use and a moderate proportion of discounting change scores, $R^2_{future_{w2}} = 0.84$, $CI_{95\%} = [0.81, 0.85]$, $R^2_{log_e(k)change} = 0.07$, $CI_{95\%} = [0.01, 0.15]$. We found the indirect effect with Wave 1 future tense use (W) held at mean (0) was negative, $Est. = -0.63$, $CI_{90\%} = [-1.11, -0.15]$, $pp = .982$. This indicates that directing participants to use the future tense resulted in less temporal discounting—as the modal hypothesis predicts—not more—as the temporal hypothesis predicts, *Fig. 3.4a*. Additionally, the indirect effect did not change conditional on Wave 1 future tense, *Fig. 3.4a*. This indicates that the effect of condition on discounting via Wave 2 future tense use was not moderated by Wave 1 future tense use. Again, the predictions of the temporal hypothesis were not supported.

Model 2 explained a large proportion of future tense use and a moderate proportion of discounting change scores, $R^2_{future_{w2}} = 0.84$, $CI_{95\%} = [0.81, 0.85]$, $R^2_{log_e(k)change} = 0.07$, $CI_{95\%} = [0.01, 0.16]$. The indirect effect with Wave 1 future-high-certainty (W) held at mean (0) was again negative, $Est. = -0.62$, $CI_{90\%} = [-1.13, -0.13]$, $pp = .98$ (*Fig. 3.4b*). However, this time the indirect effect was moderated by Wave 1 future-high-certainty use, *Fig. 3.4b*. This indicates that tense condition had a differential effect depending on the extent to which participants used the future tense to encode high certainty at Wave 1. At the highest levels of Wave 1 future-high-certainty use (the 75$^{\text{th}}$ and 100$^{\text{th}}$ percentiles in *Fig. 3.4b*), the indirect effect was non-significant. As Wave 1 future-high-certainty use decreased, the indirect effect grew increasingly negative. This indicates that low future-high-certainty using participants discounted less when directed to use the future tense. High future-high-certainty using participants were not affected.

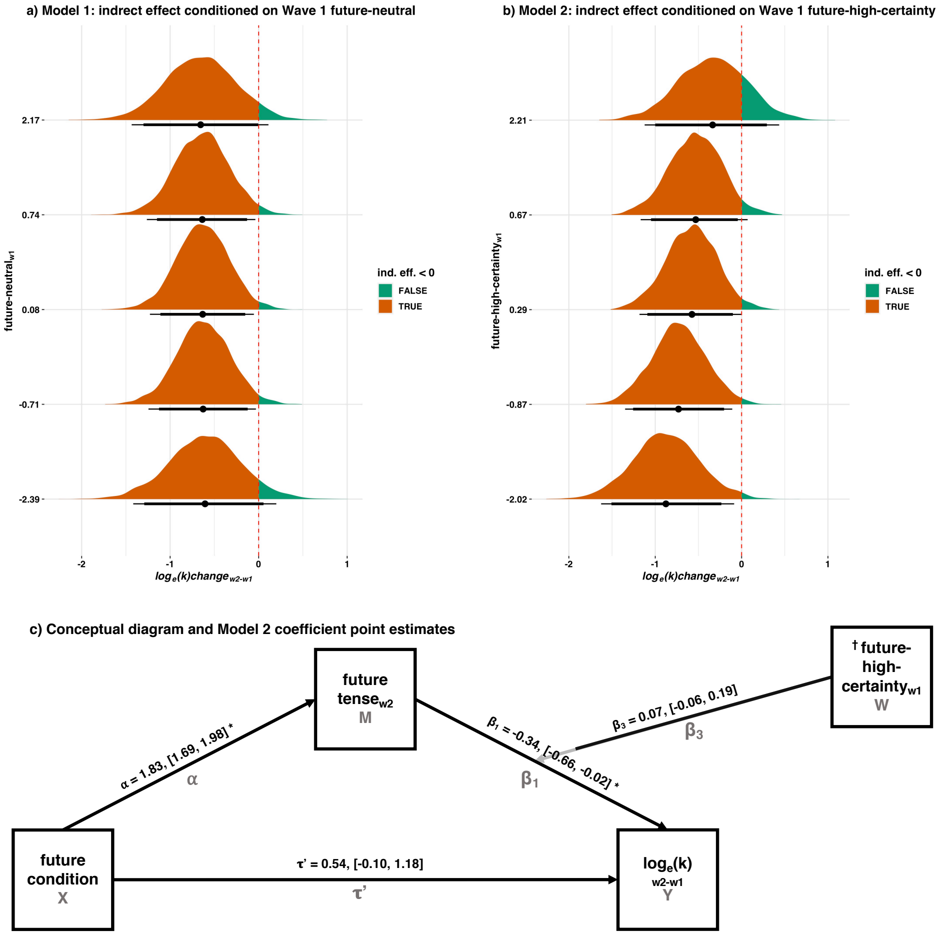


Figure 3.4: Study 3 moderated mediation analysis with conditional indirect effects of tense condition on $log_e(k)_{w2-w1}$ change scores via Wave 2 future tense use. a) Model 1: indirect effect conditioned on Wave 1 future tense use. Effect does not change across W. b) Model 2: indirect effect conditioned on Wave 1 future-high-certainty use. Effect changes across W. In (a) and (b), effects are calculated as $eff = \alpha(\beta_1 + \beta_3 W)$, where W is Wave 1 future tense (Model 1) and Wave 1 future-high-certainty (Model 2) at the 0%, 25%, 50%, 75%, and 100% percentiles. Lower values indicate less discounting, so higher estimations of delayed future rewards. c) conceptual diagram with parameter estimates for Model 2. Reference level of tense condition is the present condition, so parameters represent the effect of the future condition. See *Table 3.3* for full parameter estimates.
*95% *CI* does not contain zero.
†For Model 1, W = future tense$_{w1}$, as explained.

3.4.5 Discussion

Study 3 involved three findings. First, when we tested the modal use of tense hypothesis, we found that temporal discounting over future tense use was moderated by use of the future tense to encode notions of high certainty—as we predicted. This suggests that rather than the Dutch future tense encoding temporal distance, it encodes modal notions of certainty and uncertainty. It also goes some way to explaining the flat relationships obtained in Studies 1 and 2: We may have simply estimated a linear relationship through an interaction. Finally, it suggests that there are significant individual differences in the notions Dutch speakers use the future tense to encode, and that temporal discounting is differentially impacted contingent on these differences.

Second, we found that directing participants to use the future tense resulted in less temporal discounting, not more. This echoes the negative effect of higher future tense use in English, and suggests that the Dutch future tense encodes notions of high certainty, not temporal distance or precision (cf. Chen, 2013). Additionally, the exogenous manipulation of language production rules out the possibility that cultural differences drove Dutch use of the future tense.

Third, we found that the indirect effect of tense condition on temporal discounting via Wave 2 future tense use was moderated by Wave 1 future-high-certainty use, but not Wave 1 future tense use. Again, this suggests that the Dutch future tense is involved in the encoding of modal notions and not temporal distance or precision. However, this effect goes in the opposite direction to what we predicted. We had postulated that low future-high-certainty use meant participants construed the future tense as encoding lower certainty. Therefore, directing such participants to use the future tense should make them discount more. Accordingly, our prediction assumed that high future-high-certainty use meant participants construed the future tense as encoding high certainty. Therefore, directing such participants to use the future tense should make them discount less. However, we found the opposite: no effect for high future-high-certainty users, and a negative effect for low future-high-certainty users, *Fig. 3.4b*. This suggests some kind of ceiling effect where high future-high-certainty users were not effected. Whatever the case, the findings suggest that the Dutch future tense encodes high certainty—even for speakers who did not use it in high-certainty contexts.

3.5 General discussion

Most contemporary research into FTR grammaticisation and temporal discounting has been theoretically framed using the temporal account proposed by Chen (2013). According to this account, the FUTURE TENSE encodes either temporal distance or precision relative to the PRESENT TENSE. Either of these are hypothesised to cause strong-FTR speakers to temporally discount more. Conversely, we gave a plausible account that future tenses tend to encode high-certainty modality. Contingent on the idea that construals of risk can impact intertemporal decisions, obligatory future tense use in strong-FTR languages would cause speakers to discount less. We referred to this as the modal hypothesis.

The present work therefore contributes to this body of research by investigating how future tense use correlates with temporal discounting at the speaker-level. The results are clear: It is difficult to construe a single way in which they support the temporal hypothesis. Most importantly, we found that future tense use was negatively related to discounting in English, and the effect of high future tense in English resulted in decreased discounting—the direction predicted by the modal but not the temporal hypothesis (Studies 1 and 2). In so far as this may generalise, this undermines the temporal hypothesis that obligatory future tense use in strong-FTR languages is driving observed increased discounting. This suggests that much of the work interpreted to support the temporal hypothesis should be reconsidered in this light.

The relationship between future tense use and discounting in Dutch was more complicated. We found no effect of future tense use on temporal discounting in Dutch in Studies 1 and 2. In Study 3, we found a interaction between future tense and future-high-certainty use. This suggests that a modal framework best explains the semantics of the future tense in Dutch, but the Dutch future tense may be characterised by a greater individual differences in what people use the future tense to encode.

However, in Study 3 we found that directing Dutch participants to use the future tense caused them to discount relatively less, the same direction as the mediated effect of English in Studies 1 and 2. As far as we know, this is the only experiment to date in which a tense prime has impacted temporal discounting. It is unclear how persistent such priming effects may be (less than 1 hour in our experiment). However, to the extent that cross-linguistic differences in FTR grammaticisation exert similar population-level influences on FTR usage, the results are consistent with the hypothesis that such differences might give rise to differences in temporal discounting between populations. As with Studies 1 and 2, the effect was in the opposite direction to that predicted by the temporal hypothesis, and supported our modal account. Additionally, we found that this indirect effect was moderated by previous future-high-certainty use but not future tense use. This suggests that modal and not temporal semantics are encoded by the Dutch future tense. Again, this adds support to the modal but not the temporal account.

3.5.1 What is responsible for apparently higher discounting in strong-FTR speakers?

Consistent negative indirect effects of future tense use undermine the temporal mechanisms hypothesised by Chen (2013). What, then, is to be made of findings that Chen's (2013) FTR status variable is a consistent predictor of intertemporal behaviour (Bernhofer et al., 2019; Chen, 2013; S. Chen et al., 2017; Chi et al., 2018; Fasan et al., 2016; Figlio et al., 2016; Falk et al., 2018; Gotti et al., 2020; Guin, 2017; Hübner & Vannoorenberghe, 2015b, 2015a; J. Kim et al., 2017; S. Kim & Filimonau, 2017; Liang et al., 2018; Lien & Zhang, 2020; Mavisakalyan et al., 2018; Pérez & Tavits, 2017; Schellen, 2019; Sutter et al., 2015;

Thoma & Tytus, 2018)? On one hand, correlational research involving linguistic predictors is notorious for producing spurious relationships due to non-independence between languages resulting from historical processes of cultural inheritance which see causally unrelated bundles of traits being passed on from antecedent to descendant cultures (Roberts & Winters, 2013). When linguistic history controls are added, results have been mixed (Roberts et al., 2015), or have failed to replicate (Gotti et al., 2020). Reported correlations may be spurious. On the other hand, FTR grammaticisation could be impacting temporal discounting, just not via the tense-based temporal mechanisms hypothesised by Chen (2013). In this case, new theoretical accounts would need to be developed to explain the observed effects of FTR status.

One possibility is that strong-FTR languages oblige the use of low-certainty modal constructions for FTR. English is a good example. When English speakers make predictions, it is not obligatory to use the future tense (cf. Chen, 2013). What *is* obligatory is a modal verb. For instance, in example (3.1), it is acceptable to say any of *It can/could/may/might/shall/should/will/would/is going to rain tomorrow.* With the exception of the future tense constructions (*will*, *be going to*, and *shall*), the other modals generally encode modally weak notions of low certainty. In a previous study using a similar FTR-elicitation task, we found English speakers used more low-certainty FTR constructions than Dutch speakers (Studies 1 and 2, Chapter 2). This suggests strong-FTR grammar obliges speakers to use low-certainty modals to talk about the future. Could this cause them to construe future events as more risky?

If this were the case, it could explain effects of FTR status. A crucial point is that risk may be an overarching confounding factor in many outcomes involving real-world behaviour assumed to be impacted by <u>time</u> preferences. Most real-world intertemporal decisions involve some degree of risk. For instance, someone deciding whether to smoke a cigarette has to balance the present pleasure against future costs as a function of delay and probability, i.e. *when* and *if* she will develop smoking-related health issues. People devalue ("probability discount") outcomes as their odds of occurring grow lower (L. Green et al., 2014; Rachlin et al., 1991). As with temporal discounting, probability discounting can differ in extent. Ask a person whether he would prefer to receive \$10 or a 50% chance of \$20, and some will take the \$10 while others will take the risky \$20 (L. Green & Myerson, 2004). We will refer to these differences in terms of "risk preferences". Many of the outcomes predicted by FTR status involve real-world intertemporal decision making involving risk. If strong-FTR speakers construe future outcomes as riskier, effects of FTR status might result from a difference in <u>risk</u> preferences.

Contemporary research has found that models which multiplicatively combine probability and temporal discounting functions best predict subjective future value (Vanderveldt et al., 2015, 2017). In other words, the true unknown discounting function is approximated by a multiplicative relationship between temporal and probability discounting. In a multiplicative model, a decision maker will prefer R if and only if:

$$C < \int e^{-\delta t} R dF(t) \times \int e^{-\lambda p} R dF(p) \tag{3.9}$$

which is simply saying that he will prefer a future reward R if its subjective value discounted as a function of delay, risk, and their conditional relationship is greater than the cost of waiting. Under this model, if strong-FTR speakers construe future events as more proximal (the temporal hypothesis), or more risky (the modal hypothesis), this would translate into lower willingness to wait for risky future rewards:

$$\begin{aligned}
&\text{if } \forall t,\ F_W(t) < F_S(t),\ \textit{or} \\
&\quad \forall p,\ F_W(p) > F_S(p), \text{ then} \\
&\quad \int e^{-\delta t} R dF_W(t) \times \int e^{-\lambda p} R dF_W(p) > \int e^{-\delta t} R dF_S(t) \times \int e^{-\lambda p} R dF_S(p)
\end{aligned} \tag{3.10}$$

where $F(p)$ represents beliefs about risk, and $F(t)$ represents beliefs about temporal distance. This simply says that if strong-FTR speakers construe future events as either relatively more temporally distal or risky, they will tend to undervalue risky future outcomes compared to weak-FTR speakers (all else equal). On one hand, the temporal hypothesis simply predicts that FTR status results in beliefs $F_W(t) < F_S(t)$ and thereby impacts time preferences. On the other, higher low-certainty modal verb use could result in beliefs $F_W(p_{FUT}) > F_S(p_{FUT})$ and thereby impact time preferences when outcomes are also risky. The subscript p_{FUT} is used to indicate that it is specifically future probabilities which are represented as riskier. There is no obligation in English to use weak modals when talking about present-time probabilities. Therefore, only beliefs about future risky outcomes are hypothesised to be impacted. We might refer to this as a (broader) modal hypothesis. Either this broader modal account or the temporal account could give rise to observed impacts of FTR status on real-world risky intertemporal outcomes.

However, our findings suggest the (broader) modal hypothesis is more likely. The modal account would explain a number of findings that strong-FTR speakers are more risk-averse. For instance, in a globally balanced sample, strong-FTR speakers were found to be more risk-averse (Falk et al., 2018). In a sample of first-generation immigrants in Europe, strong-FTR speakers tended to hold fewer risky assets (Bernhofer et al., 2019). Additionally, strong-FTR-speaking clients and strong-FTR-speaking financial managers tended to hold less volatile portfolios in a sample of very high net worth individuals from Europe, the Middle East, Asia, and Latin America (Schellen, 2019). The idea that FTR status impacts risk preferences is a more parsimonious account of such findings than the idea that FTR status causes strong-FTR speakers to (temporally) discount delayed losses to a greater extent.

3.5.2 Temporal discounting and the (broader) modal hypothesis

A relevant point is that our results present a challenge to the modal account given above. Crucially, the temporal hypothesis predicts time preferences should be impacted, but the modal hypothesis does not. Our measures were of time preferences. While there was no difference in Study 1, English speakers temporally discounted more in Study 2. This is predicted by the temporal but not modal hypothesis. It is important to differentiate real-world (risky) intertemporal outcomes from experimentally measured intertemporal-choice tasks. FTR status is a consistent predictor of the former, but not the latter. In particular, Sutter et al. (2015) found higher future orientation in weak-FTR speakers, while Thoma and Tytus (2018) found lower future orientation in weak-FTR speakers. Our findings echo such inconsistent effects of FTR status on experimentally measured intertemporal outcomes. It could be the case that these mixed results represent random fluctuation. Apparently significant results could be, in reality, type I errors. On the other hand, it could be the case that time preferences are impacted by differential obligation to use low-certainty modal terms.

We need only ask whether it is possible to conclusively rule out the possibility that—even in "pure" intertemporal decisions—participants do not believe some degree of risk is involved. There is evidence that even pure intertemporal decisions are a special case of risky intertemporal decisions made under conditions of uniform high certainty, i.e. where $p_{FUT} = 1$ (Luckman et al., 2018). Given this, it seems self-evident that some people might construe delayed high-certainty outcomes as being risky to some degree. Since even moderate delay may imply some degree of uncertainty, this seems plausible. In this case, beliefs about future risk could impact intertemporal decision making. Put concretely, if $F(p_{FUT} = 1) < 1$, then $F(p_{FUT})$ would likely impact intertemporal decision tasks made under conditions of certainty. If speakers of strong-FTR languages tend to hold beliefs $F_S(p_{FUT}) < F_W(p_{FUT})$, then it seems reasonable to hypothesise they would be more likely to hold beliefs $F(p_{FUT} = 1) < 1$. This would directly translate into lower willingness to wait for future rewards, even in intertemporal decision-making tasks.

Is there any evidence that strong-FTR speakers tend to hold beliefs $F(p_{FUT} = 1) < 1$? Possibly. In a previous study, English but not Dutch speakers used an increasing number of low-certainty terms as temporal distance increased from the time of speech (Study 2, Chapter 2). This suggests that temporally distant outcomes are construed by strong-FTR speakers as increasingly risky. This lends support to the idea that strong-FTR speakers might hold beliefs $F(p_{FUT} = 1) < 1$. Additionally, the consistent negative relationship between future tense use and temporal discounting suggests that obligatory use of high-certainty terminology drives decreased temporal discounting. This is not a trivial result since temporal discounting tasks are supposed to involve intertemporal decision making under uniform 100% certainty. This suggests low-certainty terms would have similar effects but in the other direction. Future research should test this by directly measuring beliefs about time and probability, as well as use of low-certainty FTR language, and test how these impact (risky) intertemporal decision making in weak and strong-FTR speakers.

3.5.3 The semantics of the Dutch future tense

Our research contributes to contemporary debate on the semantics of *zullen*. Broekhuis and Verkuyl (2014) point out that the Dutch present tense can be used for FTR, and *zullen* can be used to mark present and future predictions. They argue the contribution of *zullen* must be *only* modal. Our results tend to support such lines of argumentation, and make a contribution where formal analyses have struggled to precisely characterise the modal semantics of *zullen*. For instance, Broekhuis and Verkuyl (2014) suggest that *zullen* quantifies over sufficient worlds to make the outcome a "confident expectation." Conversely, our results cast doubt on any universal semantics for *zullen*. The fact that the relationship between use of *zullen* and discounting switched signs conditional on the extent to which participants used *zullen* to encode certainty (future-high-certainty) suggests, rather strangely, that *zullen* encodes contrasting modal notions depending on who is using it.

However, it seems difficult to reconcile the Study 3 mediation results with the interaction in Study 3. On one hand, the interactions between future-neutral and future-high-certainty suggest that the Dutch future tense encodes different (high-certainty, low-certainty) modal notions for different speakers. On the other, the future condition caused lower relative discounting, which suggests that the Dutch future tense encodes high certainty for most speakers. Additionally, this indirect effect grew increasingly negative as future-high-certainty use decreased. This suggests that even low future-high-certainty using participants still construed the future tense as encoding high certainty. This undermines the idea that use of the future tense in modal contexts indexes the modal semantics of the future tense.

In fact, this seeming contradiction is compatible with previous findings that Dutch participants rated *zullen/gaan* as encoding relatively lower certainty then the present tense, but still encoding high certainty. Specifically, the future-neutral by future-high-certainty interaction might reflect participants understanding *zullen* and/or *gaan* to encode relatively lower certainty than the present tense. Greater variance was also identified in Dutch ratings of future tense semantics, which might also contribute to this interaction. Nonetheless, mean certainty for the future tense was high $\bar{X}_{dut} = 75.70, SD_{dut} = 29.50$. This might explain the negative mediated effect in Study 3.

A major caveat to the Study 3 findings is that modality condition in the FTR-elicitation task was not balanced across the other factors. Findings involving interactions with these factors must be interpreted extremely cautiously. A balanced design to investigate these issues would be needed to draw firmer conclusions. The negative indirect effect in Study 3, and the effects in Studies 1 and 2, should not have been affected by these design issues (i.e. since elicited FTR language is aggregated across all the factors of the FTR-elicitation task).

3.5.4 Conclusions

Psychological discounting processes are an important determinant of a wide range of behaviours, including health outcomes (Vuchinich & Simpson, 1998), drug use (McKerchar & Renda, 2012), pathological gambling (Hodgins & Engel, 2002), and investment in savings (Liu & Aaker, 2007). If the correlation between FTR grammaticisation and (risky) intertemporal preferences is not spurious, the breadth of reported findings attest to the potential impact cross-linguistic differences may have on psychological discounting processes which underpin these behaviours. Understanding the causal mechanisms involved will be critical for research to generate usable policy applications or therapeutic interventions. Our research indicates that the causal model espoused in much of this research should be reconsidered. We consistently found that use of the future tense was associated with less not more discounting. We therefore conclude that obligatory use of low-certainty modal FTR terms is more likely to be responsible for observed linguistically-driven differences in behaviours involving (risky) intertemporal decision making.

Chapter 4

Low-certainty modal constructions not future tenses cause increased psychological discounting

Abstract

Speaking a language that obliges the future tense for linguistic Future Time Reference (FTR) is hypothesised to cause speakers to perceive delayed outcomes as temporally distal and therefore less valuable ("temporal discounting"). Conversely, we hypothesise that the obligation to use low-certainty modal verbs, e.g. *may*, causes speakers to construe risky *and* delayed outcomes as riskier and therefore less valuable ("combined discounting"). We experimentally elicited free-text FTR responses and measured construals of future temporal distance and discounting of (risky) delayed rewards in speakers of Dutch (which does not oblige FTR marking) and English (which does). We used statistical mediation models, which parse out an *indirect effect* of a predictor on an outcome via a mediating variable, a *direct effect* of a predictor on an outcome, and a *total effect*—the other two combined. English speakers construed future events as temporally proximal, which in turn caused less (rather than more) temporal discounting (*indirect effect*). The *direct effect* went in the opposite direction, meaning there was no population-level difference in temporal discounting (no *total effect*, Study 1). When we measured combined discounting, English speakers discounted more (a *total effect*, Study 2). Further, English speakers used more low-certainty language and this caused increased combined discounting (an *indirect effect*). Future tense use did not impact construals of temporal distance, meaning future tense use did not mediate an effect on combined discounting (Study 2). We addressed possible cultural confounds by testing Dutch-English bilinguals (Study 3). When bilinguals were tested in English, they used more low-certainty modal verbs, which in turn caused increased combined discounting (an *indirect effect*). This was moderated by self-reported English proficiency, indicating language, not culture, is responsible for effects. Results suggest effects of FTR grammar are not driven by obligatory future tenses affecting temporal discounting, but by obligatory low-certainty modals affecting combined discounting.

4.1 Introduction

Cross-linguistic differences in Future Time Reference (FTR) grammar may influence the way people make "intertemporal decisions" (Chen, 2013; for review, see Mavisakalyan & Weber, 2018). Future time reference refers to any linguistic utterance about future events. Intertemporal decisions are decisions that involve balancing present outcomes against future ones, e.g. deciding whether to forgo present spending to invest in savings. In making such decisions, people devalue ("temporally discount") delayed outcomes as delays grow longer. For instance, given the choice of receiving \$10 now or \$10 in five years, most would choose the former because they would have temporally discounted the latter as a function of the delay (L. Green et al., 2014). Temporal discounting is therefore usually measured by offering participants a series of binary choices between a smaller reward given now (the "smaller-sooner reward", SSR) and a larger reward given later (the "larger-later reward" LLR). How much people temporally discount is subject to individual differences. Ask one person whether they would prefer to receive \$10 now or \$20 in a month and some will take the \$10, while others will wait for the \$20 (L. Green & Myerson, 2004). People who temporally discount less are more likely to make "future oriented" decisions to wait for future rewards which they construe as more valuable. People who temporally discount more tend to make "present oriented" decisions to forgo delayed gains in favour of present ones. Such differences are referred to in terms of "time preferences".

4.1.1 An overview of the temporal hypothesis

Most research into cross-linguistic affects on temporal discounting has investigated what we call the "temporal hypothesis" (Chen, 2013). According to this account, temporal semantics are involved in the relationship between FTR grammar and temporal discounting (Chen, 2013). The hypothesis is based on two observations. First, some languages oblige speakers to use the future tense when they make FTR, while other languages do not:

(4.1) a. English
It will rain tomorrow.

b. Dutch

Morgen regent het.
tomorrow rain:PRS[1] it

'Tomorrow it rains.'

[1]PRS. The verb *regenen* 'to rain' is in the present tense *regent* 'it rains'. I use the Leipzig glossing rules for interlinear glosses (Comrie et al., 2015).

In English, it is not grammatical to use the present tense when speakers make *predictions* about the future, as in example (4.1a). In Dutch, it is (4.1b). We refer to languages like English as "strong-FTR" and languages like Dutch as a "weak-FTR" (Chen, 2013). Above, we emphasised *predictions* because English *does* permit present-tense FTR for talking about schedules, e.g. *My show is at 6pm.* This is cross-linguistically common enough (Dahl, 2000a) that strong-FTR is defined according to whether the future tense is obliged for prediction-based FTR (see Chen, 2013). We refer to the dichotomous weak/strong distinction as "FTR status".

The temporal hypothesis predicts that speakers of strong-FTR languages temporally discount more than speakers of weak-FTR languages (Chen, 2013). Two reasons for this are proposed (see Chen, 2013). Here, we focus on one. In this account, present-tense FTR conveys a sense of immediacy; obligatory use of the future tense therefore delineates future from present time, causing strong-FTR speakers to perceive future outcomes as more temporally distal than weak-FTR speakers (Chen, 2013).

This would lead directly to increased temporal discounting in strong-FTR speakers (Chen, 2013). Consider someone making a decision about whether to pay a cost, C, now or take a future reward, $R > C$. This simple decision translates directly into experimental paradigms that establish time preferences. For instance, in choosing between and SSR of \$10 now and an LLR of \$20 in a month, C is the value forgone by opting for the delayed option, i.e. the value of the SSR \$10. Imagine she is uncertain about the precise time she will receive the reward and holds the beliefs about time t with a distribution $F(t)$. If she discounts delayed rewards with rate δ, she will prefer R if and only if (Chen, 2013):

$$C < \int e^{-\delta t} R dF(t) \tag{4.1}$$

which is to say she will only choose the delayed reward if its subjective value is greater than the cost of waiting. Consequently, a participant will choose the delayed \$20, if and only if its discounted value given δ integrated over $F(t)$ is greater than \$10. If strong-FTR speakers construe future outcomes as more distal, this would, in effect, shift temporal beliefs about delayed outcomes such that $F_{Strong}(t) > F_{Weak}(t)$. If this were the case, on average, strong-FTR speakers would exhibit lower willingness to choose delayed rewards:

$$\begin{gathered} \text{if } \forall t,\ F_S(t) > F_W(t), \text{ then} \\ \int e^{-\delta t} R dF_S(t) < \int e^{-\delta t} R dF_W(t) \end{gathered} \tag{4.2}$$

This is to say: If, for all t, strong-FTR speakers believe future events are more temporally distal, then their discounted value of delayed rewards will be relatively lower than for weak-FTR speakers. For instance, it is predicted that speakers of strong-FTR languages are more likely to chose an immediate \$10 over a delayed \$20 (because their subjective valuation of the \$20 is lower).

4.1.2 Support for (and issues with) the temporal hypothesis

Most work testing the temporal hypothesis has utilised correlational paradigms where FTR status is used as a predictor of naturalistic intertemporal outcomes. For example, Chen (2013) found weak-FTR speakers were 30% more likely to have saved money in any given year and had accumulated 39% more wealth by retirement. Other work has applied Chen's (2013) approach to a variety of outcomes (Bernhofer et al., 2019; S. Chen et al., 2017; Chi et al., 2018; Figlio et al., 2016; Falk et al., 2018; Guin, 2017; Hübner & Vannoorenberghe, 2015b, 2015a; J. Kim et al., 2017; S. Kim & Filimonau, 2017; Liang et al., 2018; Lien & Zhang, 2020; Mavisakalyan et al., 2018; Schellen, 2019; Thoma & Tytus, 2018). Strong-FTR speakers appear to be less future-oriented, as predicted by the temporal account. However, some results have been mixed (Gotti et al., 2020; Roberts et al., 2015), particularly in experimentally-controlled intertemporal tasks (Sutter et al., 2015; Thoma & Tytus, 2018). This suggests there is a difference between experimental outcomes which measure temporal discounting and natural intertemporal behaviour. Additionally, effects of framing intertemporal choices in the future versus present tense (i.e. "you <u>get</u> \$10 in a week" versus "you <u>will</u> get \$10 in a week") have repeatedly been null (Banerjee & Urminsky, 2017; J. I. Chen et al., 2019; Thoma & Tytus, 2018). According to the temporal hypothesis "will get" should be construed as more distal and therefore more discounted. Crucially, in previous studies we found that compared with Dutch (weak-FTR) speakers, future tense use in English (strong-FTR) speakers caused less not more temporal discounting (Chapter 3). At the same time, English speakers construed future outcomes as more temporally proximal (Chapter 2). Both results go in the opposite direction to the predictions of the temporal account. Together, these findings suggest: (a) Strong-FTR speakers are less future oriented than weak-FTR speakers in real-world data; and (b) the temporal account used to explain this outcome should be reconsidered. In the next section, we outline an alternative account.

4.1.3 The modal hypothesis

We hypothesise that low-certainty modal constructions are a factor in how FTR grammar impacts (risky) intertemporal decisions. A critical point is that FTR structures are often poorly described by the notion of "tense", which involves shifting the referenced time frame relative to the time of speech (Lyons, 1968; Mezhevich, 2008; Te Winkel, 1866; Broekhuis & Verkuyl, 2014). Because the future cannot be known at the time of speech, FTR grammar often involves the obligation to encode notions of possibility, intention, certainty, obligation, desire, necessity, conditionality, chance, and probability (Bybee et al., 1994; Palmer, 2001; Lassiter, 2015). The term for such notions is modality. English grammar is a good example. In English, it is not obligatory to use the future tense in example (4.1a) (cf. Chen, 2013). Rather, English speakers are obliged to use a modal verb. Apart from future tense constructions, any of *can, could, may, might, should* or *would* are acceptable. These modal verbs imply future-shifted time reference. Critically, relative to the future tense constructions—*will, shall* and *be going to*—they encode weakened certainty. In a series of previous studies, we found that English speakers were more likely than Dutch speakers to use low-certainty modal FTR constructions, particularly when they made predictions about the future (Chapter 2). This was mostly driven by use of obligatory English modal verbs, though Dutch speakers partially made up the difference through more frequent use of other low-certainty modal construction types, e.g. *mogelijk* 'possibly', *waarschijnlijk* 'probably', and *Ik denk dat...* 'I think that'.

A critical point is that people devalue ("probability discount") risky rewards as their odds of occurring becomes lower (L. Green et al., 2014; Rachlin et al., 1991). As with temporal discounting, there are individual differences: Ask one person whether they would prefer to receive $10 or a 50% chance of $20, and some take the $10 while others chose the risky $20. "Psychological discounting" is a general term used for discounting of outcomes which are either risky or delayed (on other factors, see Białaszek et al., 2019). Probability and delay interact in the estimation of subjective value. For instance, previous experience of high inflation rates impacted time preferences in Polish (Ostaszewski et al., 1998) and Argentine (Macchia et al., 2018) participants, and evidence suggests present orientation in intertemporal decisions is driven by a simple preference for certainty when one outcome is immediate (Andreoni & Sprenger, 2012). Considerations of this nature have led to interest in discounting when rewards are both delayed and risky (Vanderveldt et al., 2015, 2017; Yi et al., 2006). Such "risky intertemporal" decisions are measured by having participants choose between immediate, certain rewards, and delayed, risky ones, e.g. $10 now or a 50% chance of $40 in three months. Wen take the view that such paradigms better represent real-world intertemporal decisions than outcomes which are (just) delayed or (just) risky. This is because most real-world intertemporal decisions involve some degree of risk and/or uncertainty (Vanderveldt et al., 2015). For instance, someone deciding to smoke a cigarette must balance the immediate hedonic rewards with the future costs as a function of both *when* and *if* he believes he may develop smoking-related health problems. We use the term "combined discounting" to refer to outcomes which are both delayed and risky.

The obligation to use low-certainty modal FTR constructions might cause strong-FTR speakers to construe future events as more risky. Much work on modal semantics has used ideas derived from formal modal logic. In this view, modal operators quantify over a set of possible worlds (Kratzer, 2012). For instance, a weak modal operator like *may* serves as a marker of possibility, quantifying over at least one possible world. Conversely, recent accounts treat modal expressions as scalar operators which map transparently onto notions of probability. Rather than Boolean quantification, modal semantics involve encoding the likelihood of events on a one-dimensional scale between high ($p = 1$) and low ($p = .5$) certainty[2] (Lassiter, 2015; Moss, 2015; Santorio & Romoli, 2017). Evidence suggests scalar accounts capture modal semantics better than notions of Boolean quantification, since the latter yields incorrect predictions in some linguistic contexts (Lassiter, 2015). With this in mind, it is uncontroversial that modal constructions encode weakened certainty relative to the future tense (Palmer, 2001; Enç, 1996; Huddleston & Pullum, 2002). For instance, in order of encoded certainty, *The Bears... could/might/may < should < shall/will/are going to... win tonight*. In Dutch, *De Bears... kunnen < zullen/gaan... winnen vanavond* 'the bears may < will/are going to... win tonight'. This ordering is supported by a previous study in which English and Dutch participants used a slider to rate FTR constructions between high and low certainty (Chapter 2). If such operators map onto scalar notions of probability, the obligation to use "low-probability" modal constructions could cause strong-FTR speakers to construe risky future outcomes as having a lower probability of occurring. This would, in effect, shift beliefs such that $F_{Strong}(p_{FUT}) < F_{Weak}(p_{FUT})$ where p_{FUT} indicates future probabilities specifically. We limit our hypothesis to future probabilities because English grammar only obliges speakers to use low-certainty terms for future predictions. It is unclear why this would affect beliefs about present-time probabilities.

Such beliefs would translate directly into increased discounting of outcomes which are both delayed and risky (combined discounting). Models which multiplicatively combine probability and temporal discounting functions best predict subjective future value (Vanderveldt et al., 2015, 2017). A realistic account of real-world intertemporal decision making should therefore incorporate beliefs about future probability $F(p_{FUT})$, time $F(t)$, and their conditional relationship. Consider someone making a decision about whether to pay a cost, C, now or take an risky and delayed reward, $R > C$. Imagine she is uncertain about the precise probability p_{FUT} and time t she will receive R. She holds beliefs about p_{FUT} with a distribution $F(p_{FUT})$ and beliefs about t with a distribution $F(t)$. She discounts over p_{FUT} with rate λ, and discounts over t with rate δ. She will prefer R if and only if:

$$C < \int e^{-\lambda p_{FUT}} R dF(p_{FUT}) \times \int e^{-\delta t} R dF(t) \tag{4.3}$$

which is simply saying that she will prefer a risky future reward R if its discounted value integrated over her beliefs about delay and risk, and given their conditional relationship, is greater than the cost C of waiting. Under this model, if strong-FTR speakers hold beliefs,

[2]Negated high certainty ($p = 0$) might be added, for cases where a speaker is highly certain something is NOT the case. However, linguistic negation is highly complex and the FTR-type classifier cannot accurately classify negations. Such cases are avoided by eliciting language under only $p >= .5$ (see Section 4.3.1.2).

$F_S(p_{FUT}) < F_W(p_{FUT})$, this would translate into increased combined discounting of rewards which were both delayed and risky:

$$\text{if } \forall p_{FUT},\ F_S(p_{FUT}) < F_W(p_{FUT}), \text{ then}$$
$$\int e^{-\lambda p_{FUT}} R dF_S(p_{FUT}) \times \int e^{-\delta t} R dF_S(t) < \int e^{-\lambda p_{FUT}} R dF_W(p_{FUT}) \times \int e^{-\delta t} R dF_W(t) \tag{4.4}$$

which is to say if strong-FTR speakers construe risky future events as having a lower probability of occurring, they will undervalue risky future outcomes relative to weak-FTR speakers (all else equal). This is the modal hypothesis.

4.1.4 Support for the modal hypothesis

The modal account would make sense of seemingly contradictory results. In particular, if strong-FTR speakers temporally discount future losses more than weak-FTR speakers, they should be more willing to invest in risky assets (Schellen, 2019). However, one study found that strong-FTR investors as well as financial managers held less risky portfolios. There was also an interaction: When strong-FTR investors made investment decisions in collaboration with strong-FTR financial managers, portfolio risk was further reduced (Schellen, 2019; see also Falk et al., 2018). This is compatible with the interpretation that strong-FTR speakers discount risky future gains more than weak-FTR speakers.

4.1.5 Overview of the predictions of the two accounts

We present the temporal and modal accounts in *Fig. 4.1.* Here, the difference between speaker-level usage and language-level grammatical constraints is critical. While it may be common to refer to strong-FTR languages "obliging" speakers to use the future tense (Chen, 2013), typological linguists tend to talk about clines of obligatorisation (Dahl, 2000a). Similarly, usage-based linguists think of grammar not as a set of "hard" rules but as a set of population-level tendencies which emerge from speaker-level usage via processes of entrenchment as speakers use and re-use linguistic constructions, and which are understood in differing ways by individual speakers (Barlow, 2013; Bybee, 2006). While *a* modal verb is obligatory in English prediction-based FTR, speakers are free to chose one which may reflect their beliefs. If FTR language use reflects the beliefs which underpin psychological discounting—e.g. about future risk and/or delay—relationships between speaker-level FTR usage and psychological discounting would be correlational. Future time reference usage could both *affect* and *reflect* beliefs. In contrast, language-level grammatical norms do not *reflect* individual speakers' beliefs: Effects of FTR grammar on language use involve one-directional causality. While not

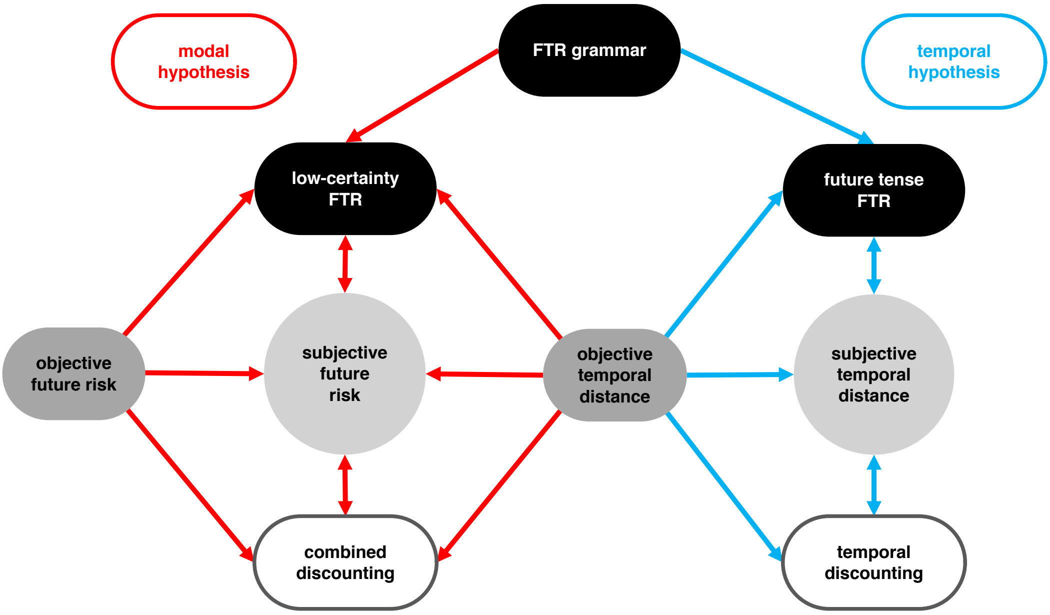


Figure 4.1: Conceptual diagram using directed causal graphs to represent the modal and temporal hypotheses. Relationships assumed to involve one-directional causality are depicted with single-pointed arrows. Relationships assumed to involve bi-directional causality (correlation) are depicted with double-pointed arrows. Expected effects of objective future certainty and distance are also depicted.

deterministic, FTR grammar sets the bounds of acceptability and thereby influences speaker-level beliefs by constraining the language it is acceptable to use.

A critical point is that under equation 4.4, the temporal hypothesis makes the same language-level predictions as the modal hypothesis. If strong-FTR speakers hold beliefs $F_W(t) < F_S(t)$, it also leads to more combined discounting in strong-FTR speakers. This means both accounts predict less future orientation in strong-FTR speakers in real-world intertemporal decision making. This is problematic because most results interpreted as supporting the temporal hypothesis involve naturalistic delayed outcomes which likely involve risk. While, Chen (2013) found weak-FTR speakers were more likely to have saved money, the outcome did not specify whether savings involved investment in risky assets such as stocks (World Values Survey Association, 2014). Other naturalistic outcomes are likewise not granular enough to rule out unmeasured risk (S. Chen et al., 2017; Chi et al., 2018; Falk et al., 2018; Fasan et al., 2016; Figlio et al., 2016; Guin, 2017; Hübner & Vannoorenberghe, 2015b, 2015a; J. Kim et al., 2017; S. Kim & Filimonau, 2017; Liang et al., 2018; Lien & Zhang, 2020; Schellen, 2019; Zhu et al., 2020). As such, these findings could be interpreted as supporting either account.

Therefore, a critical difference is that in "pure" intertemporal decisions under conditions of 100% certainty, the temporal hypothesis predicts an effect of FTR status whereas the modal hypothesis does not. A second difference is the modal hypothesis does not make any predictions about temporal beliefs, while the temporal account predicts weak-FTR speakers should construe future outcomes as more proximal. Finally, the accounts make different predictions about which FTR structures mediate the relationships between FTR status and combined discounting. Mediation involves a predictor impacting an outcome *via* one or more intermediary variables (Hayes, 2013). The temporal account predicts that future tense use should be associated with distal temporal construals and consequently increased combined discounting, *Fig. 4.1.* The modal account predicts low-certainty modal constructions should be associated with riskier representations of outcomes and consequently increased combined discounting, *Fig. 4.1.*

To test these predictions we therefore conducted a series of studies which measured FTR language use, subjective ratings of temporal distance, temporal discounting, and combined discounting in speakers of English (strong-FTR) and Dutch (weak-FTR). We used mediation analyses to test the two accounts.

4.2 Study 1

The goal of Study 1 was to test the predictions of the temporal hypothesis, in particular, that Dutch speakers should construe future events as more proximal than English speakers and that this should consequently cause increased temporal discounting.

4.2.1 Methods

4.2.1.1 Participants

A sample of $N = 234$ participants ($n = 113$ in English [$n = 62$ females, $n = 49$ males, and $n = 2$ other], and $n = 121$ in Dutch [$n = 105$ females and $n = 16$ males]) completed Study 1. Data were collected in January and February 2017. Native (British) English speakers were recruited from the University of Oxford and from Prolific Academic. Native Dutch speakers were recruited from the Max Planck Institute for Psycholinguistics (MPI). Ethical approval for the study was granted by the University of Oxford internal review board, ref. no. R39324/RE001, and by the MPI for the data collection that occurred there. All participants provided informed consent and were remunerated.

4.2.1.2 Materials

Study 1 consisted of two tasks: an intertemporal-choice task designed to measure temporal discounting and a task designed to measure construals of future temporal distance.

The intertemporal-choice task: In this task, participants made a series of binary choices between an SSR, which was varied, and an LLR, which was fixed at £10. For example, "Would you rather have £7.50 now or £10 in two months?" (In this and the other studies, Dutch participants were given amounts in euros.) There were two within-subjects factors: (1) the amounts of the SSR (£4.00–£9.50 by increments of £0.50); and (2) the delays of the LLR (one day, two days, one week, two weeks, one month, six months, two years). Amounts and delays were crossed to produce a battery of $12_{amounts} \times 7_{delays} = 84$ items. Prior to starting, participants were told to "try to answer quickly and honestly, without thinking about it too much" and completed a training example with an LLR of £10,000.

The dependent variable was calculated using an adaptation of the procedure given in Kirby et al. (1999). Research has shown that the following hyperbolic function fits real and hypothetical delay-discounted value well (Mazur, 1987; Kirby et al., 1999):

$$V = \frac{A}{1 + kD} \tag{4.5}$$

where V is the subjective value of a delayed reward A at a given delay D, and k is a scaling parameter which captures individual differences in discounting. We first derived participant-level ks (Kirby et al., 1999). This involved calculating hypothetical values of k and retaining for each participant some k which best predicted empirical choices. Specifically, we calculated all k_{1-n} at indifference between LLR and SSR, i.e. all ks such that $SSR = LLR/(1+k_iD)$ for all $N = 84$ values of SSR by D under the study. We then created hypothetical intertemporal choices for each k_{1-n} by using equation 4.5, i.e. if $SSR > \frac{LLR}{(1+k_iD)}$, we predicted $choice = SSR$, otherwise $choice = LLR$. We then retained k_i for each participant p_j which had the highest proportion of matches against p_j empirical choices. When more than one k had an equal number of matches, we took the geometric mean of all equally-performing ks (Kirby et al., 1999).

To keep results comparable with later studies, we then calculated indifference points for each participant at each delay. The indifference point is the value of the SSR at which a participant is indifferent in the choice between the SSR and LLR, i.e. the approximate point at which the cost C of forgoing the present reward is equal to the discounted reward R in equations 4.1 and 4.3. The indifference point approximates the subjective value of the delayed reward (L. Green & Myerson, 2004).

To calculate indifference points, we plugged participant-level ks derived from the Kirby et al. (1999) method, as well item-level delays, into equation 4.5. This resulted in a set of identical indifference points (Vs) for each participant within each delay under the study, i.e.

for each delay, indifference points were duplicated across of SSE amount. These were adjusted in the following way:

$$f(c) = \begin{cases} V_p, & \text{if } c_p = c_a \\ V_a, & \text{otherwise} \end{cases} \tag{4.6}$$

where V_p is Kirby et al. (1999) indifference point, V_a is the actual value of the participant's choice (i.e. the SSR amount if the participant chose the SSR, otherwise the LLR £10), c_p is the predicted choice, and c_a is the actual choice. To calculate a single indifference point for each participant and distance, we took the mean of the resultant Vs, i.e. $V_{ij} = \sum_{n}^{i=1} \frac{V_{ij}}{n_j}$, where V_{ij} are the values resulting from equation 4.6 and n is the total number of choices made by participant i in delay j. This was the most effective method of the several we tested (see Appendix E). The resultant indifference points were our dependent variable. They predicted participants' choices better than those derived simply from equation 4.5, correctly predicting actual choices in 93.48% of observations.

The subjective-temporal-distance task: This task was designed to measure how proximally or distally participants construed future events. Following Zauberman et al. (2009) and B. K. Kim and Zauberman (2009), participants used a slider to rate whether they construed a given temporal distance as "very close to now" (0) or "very far from now" (10). For each item, participants were directed to "indicate with the slider how far away from NOW the given time feels to you." There was one within-subjects factor: the objective distances, which were the same as in the intertemporal-choice task. Numbered slider intervals were not displayed. Prior to starting, participants were given a training example involving distant future time (40 years) and were told to "try to answer quickly and honestly, without thinking about it too much." Participant ratings (0-10) were the dependent variable.

4.2.1.3 Procedure

Study 1 was hosted on Qualtrics, and all participants completed it online. It had a mixed-design. The within-subjects factors for each task are described above. The between-subjects factor was language. There were two (otherwise identical) versions of the survey. Native English speakers completed the English survey and native Dutch speakers completed the Dutch survey. Participants confirmed their native language, or else they were immediately ejected. Participants completed the intertemporal-choice task, followed by the subjective-temporal-distance task. In both tasks, item order was randomised and one item was displayed per page.

Table 4.1

Descriptive statistics for Studies 1–3

study	variable	language	condition	$\bar{X}$	SD	$\tilde{X}$	MAD^*	min	max	$Q_{0.25}{}^\dagger$	$Q_{0.75}{}^\dagger$
Study 1	subjective value	English	-	7.29	3.21	8.9	1.58	0.02	10	5.29	9.82
		Dutch	-	7.34	3.18	8.96	1.48	0.08	10	5.59	9.81
	subjective temporal distance	English	-	1.74	2.15	1	1.48	0	10	0	3
		Dutch	-	2.9	3.2	1	1.48	0	10	0	5
Study 2	low-certainty language	English	-	0.5	0.5	1	0	0	1	0	1
		Dutch	-	0.28	0.45	0	0	0	1	0	1
	future tense	English	-	0.33	0.47	0	0	0	1	0	1
		Dutch	-	0.29	0.46	0	0	0	1	0	1
	subjective temporal distance	English	-	386.07	312.71	318.69	343.33	1	1000	108.17	601.93
		Dutch	-	390.49	313.25	325.72	350.41	1	1000	113.75	611.5
	subjective value	English	-	18.1	9.92	16.95	12.97	0.55	34.45	10.39	26.8
		Dutch	-	19.31	9.91	18.05	12.97	0.55	34.45	11.48	27.89
Study 3	English proficiency	-	-	2.96	0.75	3	0	1	4	3	3
	verbal-low-certainty	-	English	0.23	0.42	0	0	0	1	0	0
		-	Dutch	0.1	0.3	0	0	0	1	0	0
	subjective value	-	English	18.71	9.66	16.95	11.35	0.55	34.45	10.39	26.8
		-	Dutch	19.28	9.78	18.05	12.97	0.55	34.45	11.48	27.89

*Median absolute deviation.
†Quartiles.

4.2.2 Results

See *Table 4.1* for descriptive statistics. The temporal hypothesis predicts English speakers should rate future outcomes as more distal and therefore temporally discount more. We tested this using a simple mediation model. In mediation analyses, a predictor "X" (language) is assumed to effect an outcome "Y" (indifference points) via a mediating variable "M" (subjective temporal distance). In mediation terminology, the $X_{lang} \rightarrow M_{subj.\ dist.}$ path is referred to as α, the $M_{subj.\ dist.} \rightarrow Y_{indif.}$ path is referred to as β, and the $X_{lang} \rightarrow Y_{indif}$ path is referred to as τ', or the "direct effect". The direct effect (τ') captures the impact of X on Y while controlling for the "indirect effect". Similarly, the indirect effect $X_{lang} \rightarrow M_{subj.\ dist} \rightarrow Y_{indif.}$ captures the effect of X on Y *via* M while controlling for the direct effect τ'. The indirect effect can be calculated as the product of the paths involved, $\alpha\beta$ (Hayes, 2013). The "total effect" is the sum of $\alpha\beta$ and τ', and captures the total effect of X on Y, as in a normal regression. To test our hypothesis, we implemented the following mediation model:

$$\begin{aligned} subj.dist._{ij} &= \lambda_1 + \alpha lang_i + \kappa_{1q} X_{1qi} + u_{1j} + e_{1i} \\ subj.val._{ij} &= \lambda_2 + \tau' lang_i + \beta subj.dist_i + \kappa_{2q} X_{2qi} + u_{2j} + e_{2i} \end{aligned} \tag{4.7}$$

where α, τ', and β are slope coefficients for parameters of interest, as described; κ is a vector coefficients for a vector of $X_{1,2q}$ control variables; $\lambda_{1,2}$ are intercepts; and $e_{1,2i}$ are error terms. The outcomes are subjective distance ratings *subj.dist* and subjective values *subj.val.* (indifference points) for observation i and participant j. The random terms, $u_{1,2j}$, are assumed

Table 4.2

Regression coefficients for mediation analyses Studies 1–3

study	path name**	outcome	predictor	name	Est. = OR	estimate	SE	low-95%CI	high-95%CI†	
Study 1	λ_1	subjective temporal distance	Intercept	-		0.19	0.04	0.12	0.27	*
	$\boldsymbol{\alpha}$		English	X		-0.41	0.06	-0.52	-0.3	*
	κ_1		temporal distance	-		0.68	0.02	0.65	0.71	*
	λ_2	subjective value	Intercept	-		7.57	0.17	7.25	7.9	*
	$\boldsymbol{\tau'}$		English	X		-0.49	0.24	-0.97	-0.02	*
	$\boldsymbol{\beta}$		subjective temporal distance	M		-1.21	0.07	-1.35	-1.07	*
	κ_2		temporal distance	-		-1.25	0.06	-1.37	-1.12	*
Study 2	λ_1	low-certainty language	Intercept	-	yes	0.17	0.17	-2.11	-1.44	*
	$\boldsymbol{\alpha_1}$		English	X	yes	5.37	0.17	1.36	2.01	*
	κ_1		odds against	-	yes	3.32	0.12	0.97	1.45	*
	κ_2		temporal distance	-	yes	1.2	0.12	-0.06	0.42	
	κ_3		item order	-	yes	1.12	0.02	0.07	0.16	*
	κ_4		odds against * temporal distance	-	yes	0.94	0.12	-0.3	0.18	
	λ_2	future tense	Intercept	-	yes	0.3	0.11	-1.44	-1	*
	$\boldsymbol{\alpha_2}$		English	X	yes	1.28	0.13	0	0.5	
	κ_5		odds against	-	yes	0.54	0.08	-0.77	-0.47	*
	κ_6		temporal distance	-	yes	1	0.08	-0.14	0.15	
	κ_7		item order	-	yes	1.01	0.02	-0.03	0.05	
	κ_8		odds against * temporal distance	-	yes	1	0.08	-0.16	0.14	
	λ_3	subjective temporal distance	Intercept	-		0	0.02	-0.04	0.04	
	$\boldsymbol{\delta}$		future tense	M_2		0.01	0.01	-0.02	0.04	
	κ_9		odds against	-		0.01	0.01	0	0.03	
	κ_{10}		temporal distance	-		0.64	0.01	0.62	0.65	*
	κ_{11}		item order	-		-0.04	0.01	-0.05	-0.03	*
	κ_{12}		odds against * temporal distance	-		0	0.01	-0.01	0.02	
	λ_4	subjective value	Intercept	-		19.61	0.47	18.68	20.55	*
	$\boldsymbol{\tau'}$		English	X		-0.95	0.66	-2.22	0.35	
	$\boldsymbol{\beta_1}$		low-certainty language	M_1		-0.94	0.13	-1.19	-0.69	*
	$\boldsymbol{\beta_2}$		subjective temporal distance	M_3		-0.2	0.08	-0.35	-0.05	*
	κ_{13}		odds against	-		-4.08	0.05	-4.18	-3.97	*
	κ_{14}		temporal distance	-		-1.47	0.07	-1.61	-1.34	*
	κ_{15}		item order	-		0.03	0.05	-0.07	0.13	
	κ_{16}		odds against * temporal distance	-		0.6	0.05	0.5	0.69	*
Study 3	λ_1	verbal-low-certainty	Intercept	-	yes	0.02	0.18	-4.21	-3.5	*
	$\boldsymbol{\alpha_1}$		English	X	yes	4.53	0.06	1.38	1.63	*
	$\boldsymbol{\alpha_2}$		English proficiency	W	yes	1.19	0.14	-0.09	0.44	
	κ_1		odds against	-	yes	4.95	0.12	1.37	1.84	*
	κ_2		temporal distance	-	yes	0.99	0.12	-0.24	0.22	
	κ_3		item order	-	yes	1.3	0.03	0.2	0.32	*
	κ_4		condition order	-	yes	1.02	0.13	-0.22	0.27	
	α_3		English * English proficiency	-	yes	1.21	0.06	0.07	0.31	*
	κ_5		odds against * temporal distance	-	yes	0.97	0.12	-0.26	0.2	
	λ_2	subjective value	Intercept	-		19.28	0.42	18.46	20.07	*
	$\boldsymbol{\tau'}$		English	X		-0.5	0.11	-0.72	-0.29	*
	$\boldsymbol{\beta}$		verbal-low-certainty	M		-0.41	0.17	-0.73	-0.07	*
	κ_6		odds against	-		-4.28	0.06	-4.39	-4.17	*
	κ_7		temporal distance	-		-1.23	0.05	-1.33	-1.13	*
	κ_8		condition order	-		0.63	0.43	-0.19	1.49	
	κ_9		item order	-		0.1	0.05	0	0.21	
	κ_{10}		odds against * temporal distance	-		0.39	0.05	0.29	0.49	*

All non-dichotomous predictors were mean-centered and z-scaled so coefficient point estimates represent predicted changes in outcomes as a function of changes in 1 standard deviation in the predictors. For dichotomous outcomes (future tense, low-certainty language, verbal-low-certainty) estimates are exponentiated logits (i.e. $e^{log_e(\frac{\pi}{1-\pi})}$, where *pi* is the probability that the *outcome* = 1. This simplifies to $\frac{\pi}{1-\pi}$, the odds ratio. Otherwise, estimates represent change in outcomes as a function of change of 1 in predictors. Point estimates are the mean of the posterior probability distribution. Probabilities are converted to odds against, $OA = (1-p)/p$, where p is the certainty condition over 100. Temporal distance is measured in days. See Appendix E for discussion of these modelling decisions. Bayesian R^2 equivelents are reported in *Table 4.3*.
*Bayesian 95% Credibility Interval (CI) does not contain zero.
Paths which correspond to markup in *Figs. 4.2, 4.4,* and *4.6* are in **bold and red.
†Credibility intervals are reported in the main text at 90% because they test one-tailed hypotheses.

to be drawn from the normal distributions with means of zero and standard deviations drawn from the sample, $u_{1,2j} \sim N(0, \sigma_u^2)$. These allow intercepts to vary clustered by participant to account for expected inflation of standard errors which would result from correlation of multiple responses made by each participant. See *Tables 4.2* and *Table 4.3* for parameter estimates and *Fig. 4.2b* for a conceptual diagram.

Bayesian statistics are well-suited to mediation analyses. As well as making no assumptions about the normality of sampling statistics, they allow for straight-forward inferences about any transformation of model parameters (i.e. $\alpha\beta$ path products) through carrying out the desired operation on posterior probability distributions (Vehtari et al., 2019). We therefore

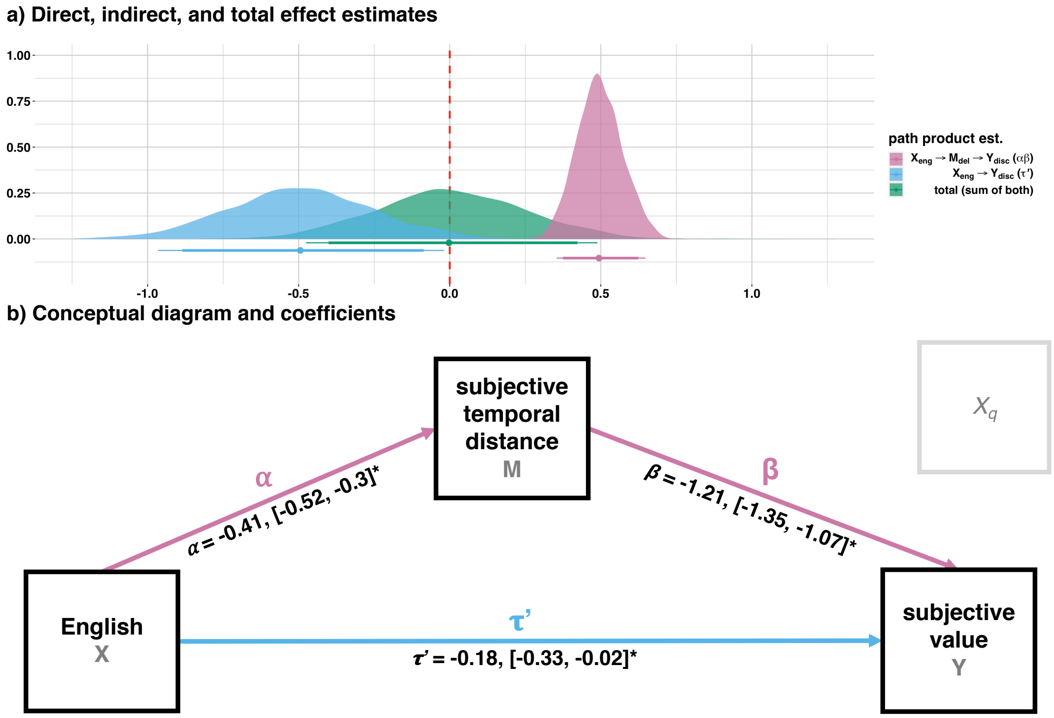


Figure 4.2: Study 1 results. a) Posterior estimates for paths and path products for the effect of speaking English on subjective value (indifference points) via ratings of subjective temporal distance in a temporal discounting task. In Bayesian analysis, coefficient Credibility Intervals (CIs) represent the probability of the coefficient falling within the specified range (see Van Dongen, 2006). Above and in *Figs. 4.4* and *4.6*, CIs are plotted at 90% (thick bars) and 95% (thin bars). For two-tailed hypotheses, thin bars represent cut-off criteria. For one-tailed hypotheses, thick bars do (providing sign is as predicted). b) Conceptual diagram with coefficient estimates and 95% CIs [lower, upper]. Path names correspond to *Table 4.2*. Above and in *Figs. 4.4* and *4.6*, non-dichotomous variables were mean-centred and z-transformed, i.e. $x = \frac{x_i - \bar{X}}{\sigma}$, so $\bar{X} = 0$, $\sigma = 1$. Coefficients therefore represent changes in standard deviations. See *Table 4.2* for other X_q control variables.
*Coefficients for which the 95% CI does not contain zero are significantly different from zero.

used the *brms* package (Bürkner, 2017) in R (R Core Team, 2020) to estimate model parameters using a no U-Turn Hamiltonian Monte Carlo sampling procedure (Hoffman & Gelman, 2014; Stan Development Team, 2020). We specified uninformative priors $Unif(-\infty - \infty)$, 4 chains, $2,000$ iterations, and a burn in discard of 400. Inspection of caterpillar plots and $\hat{R}$s indicated that parameter estimates were stable (Vehtari et al., 2019).

The indirect effect: Subjective-temporal-distance ratings were negatively related to subjective future value (β in *Fig. 4.2b*). However, contrary to the prediction of the temporal hypothesis, Dutch speakers rated future events as relatively more distal (α in *Fig. 4.2b*). As a consequence, the indirect effect ($\alpha\beta$) was positive rather than negative, *Fig. 4.2a*. Correspondingly, evidence for $\alpha\beta < 0$ was weak, $Est. = 0.5$, $CI_{90\%} = [0.37, 0.62]$, $pp < .001$,

Table 4.3

Bayesian R^2 equivalents for regressions in Studies 1–3

study	outcome	name	R^2	SE	$Q_{2.5}$	$Q_{97.5}$
Study 1	subjective temporal distance	M	0.62	0.01	0.6	0.64
	subjective value	Y	0.74	0.01	0.72	0.75
Study 2	low-certainty language	M_1	0.44	0.00	0.43	0.44
	future tense	M_2	0.25	0.01	0.24	0.26
	subjective temporal distance	M_3	0.57	0.00	0.57	0.58
	subjective value	Y	0.64	0.00	0.63	0.65
Study 3	verbal-low-certainty	M	0.39	0.01	0.38	0.40
	subjective value	Y	0.59	0.00	0.58	0.59

These are a Bayesian adaptation of R^2 following (Gelman et al., 2018), including spread statistics for values at the 2.5% and 97.5% percentiles.

and evidence for $\alpha\beta > 0$ was strong, $Est. = 0.5$, $CI_{90\%} = [0.37, 0.62]$, $pp > .999$. In these summaries, "*pp*" reports the probability that the posterior estimate of the effect matches the predicted direction. In contrast to frequentist p values, $pp > .95$ indicates evidence exceeds 95% (Bürkner, 2017). The positive indirect effect indicates English speakers construed future events more proximally than Dutch speakers, and therefore had higher indifference points (less discounting). This is the opposite direction to the prediction of the temporal hypothesis.

The total effect: A further relevant question is how English and Dutch speakers differed overall in terms of temporal discounting. The temporal hypothesis predicts English speakers should discount more, whereas the modal hypothesis only predicts a difference when outcomes are both risky and delayed. To test this, we calculated the total effect $(\alpha\beta) + \tau'$. We found this was not different from zero, $Est. = 0$, $CI_{90\%} = [-0.4, 0.42]$, $pp = .504$, see *Fig. 4.2a*. This was driven by the positive indirect effect being cancelled out by a negative direct effect, *Fig. 4.2a*.

4.2.3 Discussion

These results fail to support the predictions of the temporal hypothesis. Ratings of subjective temporal distance *were* negatively associated with subjective future value. However, English speakers rated future events more proximally than Dutch speakers, and this caused the indirect effect to go in the opposite to predicted direction. Additionally, we found no temporal discounting differences between English and Dutch speakers, where one is predicted by the temporal hypothesis. This suggests that factors other than temporal construals are driving observed effects of FTR status on naturalistic outcomes (cf. Chen, 2013).

4.3 Study 2

The Study 1 results had undermined the predictions of the temporal account, but important questions remained. Principally, Study 1 did not test the modal hypothesis. This was our goal with Study 2. We designed it to answer two questions. First, Study 1 found no difference in temporal discounting, but would a difference emerge when rewards were both risky and delayed?

Since this is predicted by both accounts, we also conceived Study 2 as contrasting the modal and temporal predictions about *how* such an effect would arise. Would higher use of low-certainty language in English cause increased combined discounting, as the modal hypothesis predicts, or would higher use of the future tense in English cause distal temporal representations, and therefore increased combined discounting, as the temporal hypothesis predicts? The temporal account seemed unlikely, given Study 1. We therefore hypothesised (pre-registered with AsPredicted, see Appendix B) that the modal prediction (*language* $\rightarrow$ *low-certainty* $\rightarrow$ *increased combined discounting*) would be supported, and that the temporal prediction (*language* $\rightarrow$ *future tense* $\rightarrow$ *temporal distance* $\rightarrow$ *increased combined discounting*) would not.

4.3.1 Methods

4.3.1.1 Participants

An initial sample of $N = 441$ participants passed attention checks (see below) to complete Study 2 ($n = 219$ in English and $n = 222$ in Dutch). English participants were native English speakers residing in the UK, and Dutch participants were native Dutch speakers residing in the Netherlands. From the initial sample, $n = 16$ participants were excluded for giving bad faith responses, e.g. repeatedly entering the same string of meaningless characters in the FTR-elicitation task or providing invariant responses in the subjective-temporal-distance task. These exclusion criteria were part of a pre-registered experimental plan (Appendix B). After exclusions, a final sample remained of $N = 425$ participants ($n = 209$ in English [$n = 110$ females, $n = 96$ males, $n = 3$ other], $n = 216$ in Dutch [$n = 87$ females, $n = 128$ males, $n = 1$ other]). Data were collected between March and April 2021. Participants were recruited via a Qualtrics participant panel. Ethical approval was granted by the Cardiff University internal review board, ref. no. ENCAP/Roberts/15-02-2021. All participants provided informed consent and were remunerated.

4.3.1.2 Materials

Study 2 comprised three tasks: an FTR-elicitation task designed to establish free-text FTR language use, a risky-intertemporal-choice task designed to measure combined discounting of outcomes which were both delayed *and* risky, and a subjective-temporal-distance task similar to Study 1.

The FTR-elicitation task: In this task, participants were given a context and a target sentence with the main verb in the infinitive. Their task was to type in the target sentence with the main verb conjugated. Since FTR status is defined according to prediction-based FTR (Dahl, 2000a; Chen, 2013), all items were predictions. Participants were advised there "were no correct answers" and that they should complete the questionnaire sentences "as though they were speaking to a close friend". Prior to starting, they were given one item with example responses, and one training item where they provided a response. These were in the past tense in order to avoid biasing participants. See Appendix I for items. There was one attention check: At a random point, participants were instructed to enter the word "dance" (Dutch "dans"). If they failed to do this, there were ejected from the survey immediately.

There were two within-subjects factors: temporal distance from time of speech (later today, one week, one month, three months, six months, and one year) and certainty (50%, 60%, 70%, 80% 90%, and 100%). Temporal distance was operationalised using temporal adverbials in the contexts, e.g. "in one week", "in a month", "a year from now", etc. Certainty was operationalised by providing some "certainty information" for each item. Certainty information was given in numerical percentages, for example:

Context: Chris's brother {SEND} him some money next month. You never know with him... When he get it...

Certainty: 50% certain.

Target: ...he {SPEND} it at the bar.

A typical response would be "He might spend it at the bar". Prior to starting, participants were told "there will be some 'certainty information' included in the context". They were informed, "this indicates how certain you are about what you are saying". They were then directed, "please imagine you are this certain and write down what you would say". Following this, they were given a past tense example. Certainty and temporal distance were crossed. The $n = 1$ item where temporal distance was "later today" and certainty was 100% was excluded (see below). This resulted in a test battery of $6_{t.dist.} \times 6_{cert.} - 1 = 35_{items}$.

The $N = 14,875$ resultant free-text responses were numerically scored using the the FTR-type classifier (Robertson, 2019b). The FTR-type classifier is a deterministic, closed-vocabulary, keyword-based computer program written in Python (Python Software Foundation, 2017). It comprises a set of keyword lists and rules which together classify text data. It categorises responses in terms of four exclusive semantic categories: future tense, present tense, low-certainty, and high-certainty. The latter two are further divided into two non-exclusive categories based on whether a modal verb or some other construction type is used (see below). Each category is coded (1) to indicate a response is a positive example, otherwise (0). These categories comprise the dependent variables for this task.

Relativity effects—including those presently investigated—often involve the extent to which grammatical constraints spotlight aspects of experience (Wolff & Holmes, 2011). The question is not which formal grammatical elements are involved, but which aspects of experience are spotlighted. As such, it was appropriate to have semantics take priority over formal criteria in the FTR-type classification scheme. In cases of so-called "modal concord", it is possible in English and Dutch to use a lexical modal periphrastic modifier in combination with a modal verb, e.g. *It could possibly rain tomorrow* or *It will possibly rain tomorrow* (Huitink, 2012). When the future tenses *will* and *zullen* are used in this way, their semantics are "dominated" by the modal modifier. For example, *It will...* expresses high certainty, whereas *It will possibly...* expresses low certainty. The same is true for present-tense FTR constructions (see Section 1.4.1.4). The FTR-type classifier reflects this: Responses which used both a tense and a modal modifier were classed as the modal category but not also the tense category. For instance *It will possibly rain tomorrow* would be classed as other-low-certainty but not also the future tense. We describe the categories below (see Appendix H for full keyword lists).

Present tense: Responses were classed as present tense if they conjugated the main verb using the present tense and failed to be classed as any of the other categories.

Future tense: Responses were classed as future tense if they used commonly accepted future auxiliaries or explicit temporal adverbials (e.g. English *will, shall, be going to, gonna, about to*, Dutch *zullen* 'will', *gaan* 'be going to', *staat op* 'about to'). Any response which used these words, without additional modal words, was counted as future tense.

Verbal-low-certainty: Responses which used low-certainty modal verbs were classed as verbal-low-certainty (e.g. English *can, could, may, might, would, ought, should*; Dutch *kunnen* 'may', *zou/zouden* 'would').

Verbal-high-certainty: Responses which used modal verbs which encode high certainty were classed as verbal-high-certainty (e.g. English *must*; Dutch *moeten* 'must').

Other-low-certainty: Responses which used modal expressions indicating low certainty (apart from modal verbs) were classed as other-low-certainty. This includes what we term "modal modifiers" (e.g. English *possibly, probably, potentially*; Dutch *misschien* 'perhaps', *mogelijk* 'possibly', *waarschijnlijk* 'probably', *wellicht* 'maybe'). It also includes low-certainty mental state predicates (e.g. English *think, believe, reckon*; Dutch *denken* 'think', *annehm* 'assume', *veronderstellen* 'suppose'), as well as low-certainty epistemic modal particles (e.g. Dutch *wel eens*, approximately 'well be').

Other-high-certainty: Responses which used modal expressions which encode high certainty (apart from modal verbs) were classed as other-high-certainty. This category comprises modal modifiers exclusively (e.g. English *certainly, definitely, absolutely*; Dutch *zecker* 'certainly', *definitief* 'definitely').

Responses which met none of these criteria were not given a category. To check whether the FTR-type classifier was accurate, linguistically-trained annotators hand coded a subset of $N = 379$ responses ($n = 179$ in English, and $n = 200$ in Dutch). Test data were chosen by randomly selecting $n = 25$ responses from each FTR-type category, $n = 25$ responses which were given no category, and the $n = 25$ most common responses from each language. The English sample is smaller because there were just $n = 4$ verbal-high-certainty responses in English. Duplicated responses means the test data covered $n = 2,821$ unique responses (21.47% of the English data, and 17.27% of the Dutch data). Against human annotations, $F1 > .95$ for all categories, where $F1 = (2rp)/(r + p)$; r is recall (true positive predictions tp over tp plus the true negative tn predictions, $r = tp/(tp + tn)$); and p is precision (tp over tp plus false positive fp predictions, $p = tp/(tp + fp)$). This means $F1$ balances the type I and type II error rates. High scores indicate the FTR-type classifier's predictions were reliable.

The FTR-type classifier cannot accurately classify negations or responses which use words from two conflicting class-criteria keyword lists. We refer to these as "mixed modal" responses. In the first instance, modal keywords switch polarity in the presence of negations. For instance, *I'm not certain it will rain tomorrow* expresses low certainty. However, the high-certainty class-criterion keyword *certain* means it would be classed as high-certainty. In the second, *Rain tomorrow is certainly possible* expresses moderate certainty, but would be classed as both verbal-high-certainty and verbal-low-certainty because of the class-criterion keywords *certainly* and *possible*. We therefore detected the presence of negations using an averaged perceptron tagger following Collins (2002) but with Brown cluster features as described by Koo et al. (2008) and using greedy decoding (implemented in the Python library *spaCy* (Explosion AI, 2020)). We excluded negations and mixed modal response ($n = 279$). These exclusion criteria were part of our pre-registered experimental plan (Appendix B). This left a final sample of $N = 14,596$ observations.

The risky-intertemporal-choice task: In this task, participants made a series of binary choices between an immediate smaller-sooner and a delayed and risky larger-later reward. For example, "Would you rather receive £17.50 now or a 60% chance of £35 in two months?" Prior to starting, participants were told to "try to answer quickly and honestly, without thinking about it too much" and completed a training example with an LLR of £10,000. There were two within-subjects factors: delay and probability. These were the same as in the FTR-elicitation task, and were crossed to produce a test battery of $N = 35$ delay by probability pairs (the today by 100% item was excluded because the subjective value of two rewards under identical delays and probabilities would be the same). The amount of the LLR was £35 (this was changed from £10 in order to make the nominal value, $LLR \times probability$, more difficult for participants to mentally calculate). There were two attention checks: At two random points, participants were given a choice between two non-risky, immediate rewards of different values. If they chose the smaller one, they were ejected from the survey immediately.

Following Vanderveldt et al. (2015), participants completed 5 trials per probability by delay pair, meaning participants completed a total of $35 \times 5 = 175$ trials. The value of the SSR was adjusted each trial depending on participants' responses. On the first trial of each probability by delay pair, the value of the SSR was $LLR/2 = 17.5$. This was adjusted upwards or downwards by an amount x. On the first trial of each delay by probability pair, this was set to $x_1 = (LLR - SSR)/2 = (35 - 17.5)/2 = 8.75$. The value of x was halved each iteration, $x_{n+1} = \frac{x_n}{2}$. For example, if a participant chose the LLR on the first trial n, then $SSR_{n+1} = SSR_n + x_1 = 17.5 + 8.75 = 26.25$; if, following this, they chose the SSR, then $SSR_{n+2} = SSR_{n+1} - \frac{x_1}{2} = 26.25 - 4.38 = 21.87$. This procedure "homes in" on the point of indifference between the delayed, risky LLR and the SSR: the indifference point. The dependent variable was the value the SSR would have taken on the 6th trial (which was not conducted). This captures the approximate subjective value of the discounted LLR under each delay by probability pair (Vanderveldt et al., 2015).

The subjective-temporal-distance task: In this task, participants were presented with a phrase and used a slider to rate whether they construed it as "close to now" (1) or "far from now" (1000). The scale was expanded to facilitate more granular measurements. Accuracy was recorded to two decimal places, and numbered slider intervals were not shown to participants. Participants were directed to "indicate with the slider how far away from NOW the given event feels to you." The phrase was "Manchester United has a ____% chance of winning ____". There were two within-subjects factors: delay and probability. These matched the other tasks and were crossed to produce $N = 35$ trials. Items were constructed by imputing delays and probabilities into the base phrase, e.g. "Manchester United has a 40% chance of winning in one year". Prior to starting, participants were given a training example involving a past event, and were told to "try to answer quickly and honestly, without thinking about it too much." Participant ratings (1.00–1000.99) were the dependent variable. There was one attention check: At a random point, participants were told to move the slider to the far right. If they did not, they were ejected immediately.

4.3.1.3 Procedure

Study 2 was hosted on the Qualtrics survey platform and was conducted online. It had a mixed-design. The within-subjects factors for each task are described above. There was one between-subjects factor: survey language. At the beginning of the survey, participants confirmed their first language and current residence. English speakers confirmed they were native English speakers currently residing in the United Kingdom and Dutch speakers confirmed they were native Dutch speakers currently residing in the Netherlands. If they did not, they were immediately ejected from the survey. Additionally, Qualtics panel targeting systems were used to advertise to participants who matched these criteria. If a participant's IP address location did not match their language condition, they were blocked from accessing the survey. Participants completed the FTR-elicitation task, then the risky-intertemporal-choice task, then the subjective-temporal-distance task. Within the tasks, item order was randomised and one item was displayed per page.

4.3.2 Results

Descriptive statistics for Study 2 are given in *Table 4.1*, and FTR-elicitation results are presented in *Fig. 4.3.* Elicited FTR-types were more sensitive to certainty than temporal distance, which suggests FTR structures encode modal rather than temporal notions. English speakers used slightly more future tense constructions, *Table 4.1.* However, English speakers also used more low-certainty constructions, and this was mostly driven by low-certainty modal verbs, *Fig. 4.3.*

Which, if any, of these FTR differences affected combined discounting? Both hypotheses predict English speakers should discount more, but the modal account predicts low-certainty language should drive this and the temporal account predicts future tense use should drive this. As mentioned above, we found in a previous study that—while English speakers used more low-certainty modal verbs—Dutch speakers used more other-low-certainty constructions (Chapter 2). This pattern was again evinced in the present results: Mean use of other-low-certainty language was slightly higher in Dutch ($\bar{X}_{dut} = 0.23$, $SE = 0.01$; $\bar{X}_{eng} = 0.20$, $SE = 0.01$). To capture these cross-linguistic differences, we created a new variable which was (1) when either verbal-low-certainty or other-low-certainty constructions were used, otherwise (0). We refer to this as "low-certainty language." We implemented the following mediation model:

$$\begin{aligned}
log_e(\frac{\pi_{LC_{ijk}}}{1-\pi_{LC_{ijk}}}) &= \lambda_1 + \alpha_1 lang_i + \kappa_{1q} X_{1qi} + u_{1j} + h_{1k} \\
log_e(\frac{\pi_{FUT_{ijk}}}{1-\pi_{FUT_{ijk}}}) &= \lambda_2 + \alpha_2 lang_i + \kappa_{2q} X_{2qi} + u_{2j} + h_{2k} \\
subj.dist._{ij} &= \lambda_3 + \delta FUT_i + \kappa_{3q} X_{3qi} + u_{3j} + e_{3i} \\
subj.val._{ij} &= \lambda_4 + \tau' lang_i + \beta_1 LC_i + \beta_2 subj.dist_i + \kappa_{4q} X_{4qi} + u_{4j} + e_{4i}
\end{aligned} \tag{4.8}$$

where π_{ijk} is the probability that $y = 1$ for low-certainty language (LC) and the future tense (FUT) for observation i, participant j, and item k. For the linguistic outcomes, intercepts are randomly varied by item h_k as well as participant u_j. This is to accommodate expected idiosyncratic effects of FTR-elicitation items k. As with u_j random intercepts for participant,

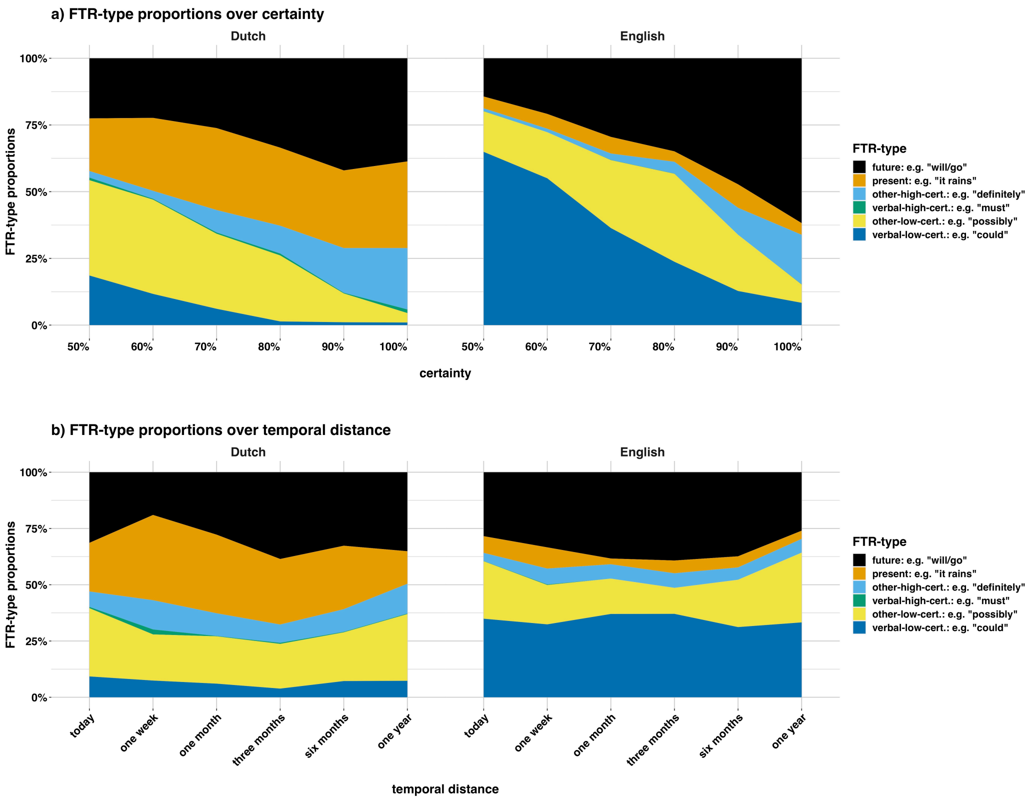


Figure 4.3: Study 2 elicited FTR-type proportions by certainty condition, temporal distance, and language. English speakers used more low-certainty constructions, and this was mostly driven by modal verbs. In English and Dutch, certainty condition had a greater effect on FTR type than did temporal distance. Reported values above and in *Fig. 4.5* are the proportions per language and FTR-task factor, i.e. means per level normalised to sum to 1, $(x = x_i / \sum_n x)$.

these are assumed to be drawn from the normal distribution with mean of zero and standard deviation drawn from the sample, $h_{1,2k} \sim N(0, \sigma_h^2)$. Other parameters are the same as in Study 1. See *Table 4.2* and *Table 4.3* for parameter estimates and *Fig. 4.4b* for a conceptual diagram. We estimated this model in *brms*, using the same estimation settings as in Study 1. Inspection of caterpillar plots and $\hat{R}$s indicated that parameter estimates were stable (Vehtari et al., 2019).

Indirect effects: We first tested the prediction that English speakers would use more low-certainty language and this would result in increased combined discounting. As predicted, English speakers used more low-certainty language (α_1 in *Fig. 4.4b*), and this was negatively associated with subjective future value (β_1 in *Fig. 4.4b*). As a consequence, the indirect effect ($\alpha_1\beta_1$) was negative, as predicted, *Fig. 4.4a2*. Evidence for $\alpha_1\beta_1 < 0$ was strong, $Est. = -1.59$, $CI_{90\%} = [-2.04, -1.17]$, $pp > .999$. This supports the predictions of the modal

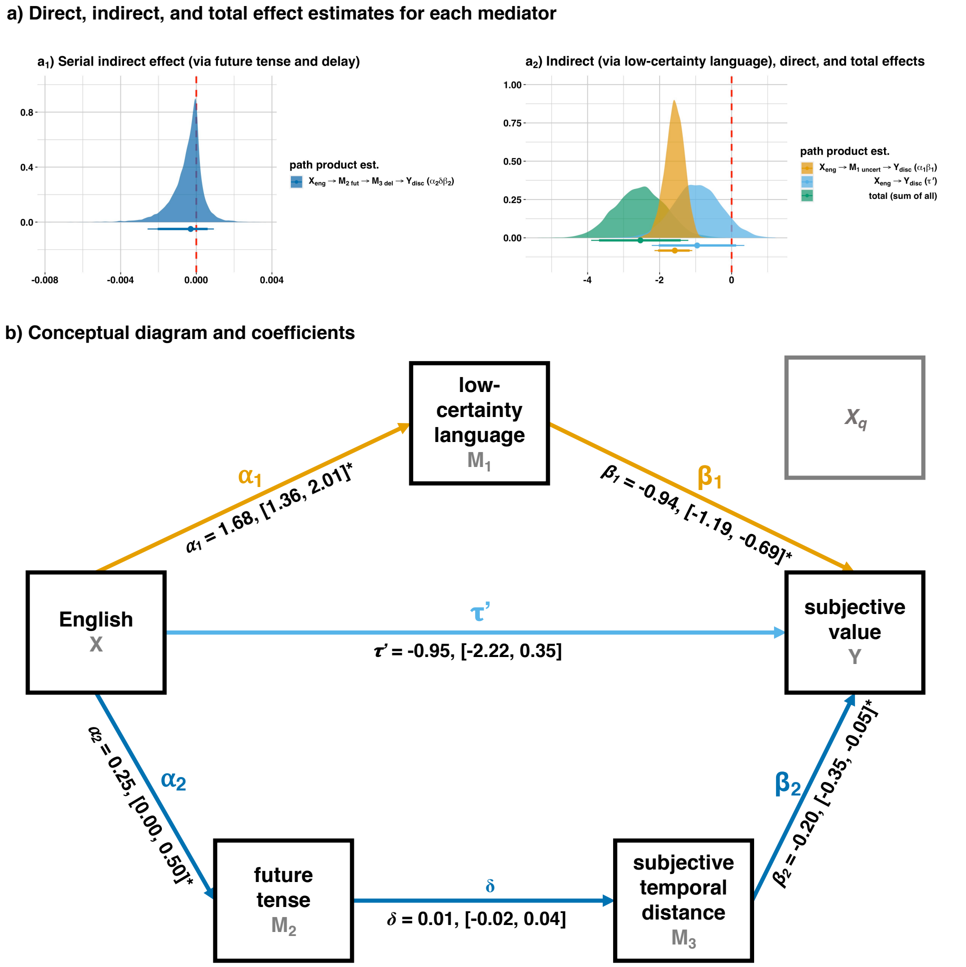


Figure 4.4: Study 2 results. a) Posterior estimates for paths and path products for the effect of speaking English on subjective value (indifference points) via low-certainty language (the modal hypothesis), future tense, and subjective temporal distance (the temporal hypothesis). a$_1$) Future tense use did not impact subjective temporal distance so had no effect on combined discounting. a$_2$) English speakers used more low-certainty language, and as a consequence exhibited increased combined discounting (the modal hypothesis). Above, and in *Fig. 4.6*, operations are carried out on the unstandardised coefficients. This makes path products comparable with τ', but means interpreting the extent of the impact on Y is difficult, since α is measured in log-odds and β is not (see Muthén, 2011). The direction and significance of all hypothesis tests were unchanged using standardised coefficients, Appendix E.
b) Conceptual diagram with coefficient estimates. Path names correspond to *Table 4.2*. See *Table 4.2* for other X_q control variables.
**Coefficients for which the 95% CI does not contain zero are significantly different from zero.

hypothesis. English speakers used more low-certainty language and as a consequence had lower indifference points (more combined discounting).

Next, we tested the prediction that English speakers would use more future tense constructions, and this would result in more distal subjective distance ratings, which would in turn result in more combined discounting. We had predicted this would not be the case. English speakers did use more future tense constructions (α_2 in *Fig. 4.4b*), and subjective distance ratings were negatively associated with subjective value (β_2 in *Fig. 4.4b*). However, future tense use did not impact subjective distance ratings (δ in *Fig. 4.4b*). As a consequence, the serial indirect effect ($\alpha_2\delta\beta_2$) contained zero, *Fig. 4.4a*$_1$. Evidence for $\alpha_2\delta\beta_2 < 0$ was weak, $Est. = 0$, $CI_{90\%} = [0, 0]$, $pp = .743$. This fails to support the temporal hypothesis. Future tense use did not impact subjective-temporal-distance ratings. As a consequence, future tense use had no impact on combined discounting.

The total effect: Did English speakers discount more, now that future outcomes were both delayed and risky? To test this, we calculated the total effect by summing the products of the paths ($\alpha_1\beta_1 + \alpha_2\delta\beta_2 + \tau'$). This was negative and did not contain zero, *Fig. 4.4a*$_2$. Evidence for this was strong, $Est. = -2.54$, $CI_{90\%} = [-3.68, -1.41]$, $pp > .999$. This indicates that English speakers had lower indifference points than Dutch speakers (more discounting). The direct effect (τ') contained zero, *Fig. 4.4a*$_2$. This indicates the negative total effect was driven by the negative indirect effect via low-certainty language use, which is what is predicted by the modal hypothesis.

4.3.3 Discussion

The modal hypothesis was supported and the temporal hypothesis was not. English speakers used more low-certainty modal constructions and this was in turn associated with lower indifference points (more combined discounting). On the other hand, future tense use was not related to subjective-temporal-distance ratings. As a consequence, the serial indirect effect contained zero. Additionally, in contrast to Study 1—which found no difference in temporal discounting—Study 2 found English speakers exhibited higher combined discounting of rewards that were both delayed and risky. This supports the modal hypothesis, since the serial mediation effect was not significant and the indirect effect via low-certainty language was.

4.4 Study 3

Study 2 found English speakers used more low-certainty language and this resulted in increased combined discounting. This supports the central prediction of the modal hypothesis. However, it left two important questions unanswered.

First, cultural rather than grammatical differences could have driven increased low-certainty language use in English. For example, non-linguistic norms or institutions in Anglophone culture could cause English speakers to view future outcomes as risky and therefore to use more low-certainty modal language. This is unlikely, since Study 2 reflects well-known features of English grammar, i.e. that English requires a modal verb for prediction-based

FTR. Nonetheless, Study 2 was essentially correlational. To address the concern that cultural differences could have driven increased use of low-certainty language in English speakers, we replicated Study 2 with Dutch-English bilinguals. Since effects would result from testing the same individuals at different times in different languages, this goes some way to isolating the effects of language from culture.

Second, the mediator in Study 2 was any low-certainty language use. In Study 3, we tested the specific prediction that low-certainty modal verb use affected combined discounting, since English's obligatory modal verb system is the relevant cross-linguistic contrast with Dutch.

We predicted that bilinguals in the English condition would use more low-certainty modal verbs and this would result in increased combined discounting (this was pre-registered with AsPredicted, see Appendix B). Additionally, we included a measure of self-reported English proficiency. If high-English-proficiency bilinguals used more low-certainty modal verbs and this led to differential impacts on combined discounting, this suggests linguistic rather than cultural differences drove effects.

4.4.1 Methods

4.4.1.1 Participants

There were two rounds of data collection in Study 3. An initial sample of $N = 320$ participants passed attention checks to complete the first round; after attrition $N = 201$ of these completed the second. Drop outs were excluded from analyses, leaving a final sample of $N = 201$ participants ($n = 85$ females, $n = 115$ males, and $n = 1$ other). Participants were native Dutch speakers currently residing in the Netherlands who self-rated as being able to at least "can have basic conversations on familiar topics" in English. Data were collected between March and May 2020. Participants were recruited via a Qualtrics participant panel. Ethical approval for the study was granted by the Oxford internal review board, ref. no. R39324/RE001. All participants provided informed consent and were remunerated.

4.4.1.2 Materials

There were two tasks in Study 3: an FTR-elicitation task and a risky intertemporal choice task. These were identical in every way to the tasks in Study 2 (including attention checks and survey-ejection criteria).

The FTR-type classifier was again used to classify free-text responses. Responses exhibiting mixed modality and negations were excluded ($n = 232$), leaving a final sample of $n = 13,838$ observations from an original sample of $N = 35_{items} \times 201_{participants} \times 2_{lang.\ conds}$ $= 14,070$. Linguistically trained annotators hand coded a subset of $n = 464$ responses ($n = 225$ in English, and $n = 239$ in Dutch). Duplicated responses means the test data covered $n = 3,482$ responses (29.16% of the English data, and 21.18% of the Dutch data).

Against human annotations, *F1* $> .97$ for all categories, indicating the FTR-type classifier's predictions were reliable.

4.4.1.3 Procedure

The study was hosted on the Qualtrics survey platform and was conducted online. It had a within-subjects, repeated-measures design. Two otherwise-identical versions of the survey were created: one in English, the other in Dutch. Participants completed the survey initially in one language ("Wave 1"), and then approximately one month later ($\bar{X} = 29.9$, $SD = 4.43$ days) the same participants completed it in the other language ("Wave 2"). We refer to the language manipulation as "language condition." Order was randomly assigned. In Wave 1 $n = 93$ completed the survey in Dutch and $n = 108$ did so in English. Before starting, participants confirmed they were native Dutch speakers currently residing in the Netherlands, otherwise they were immediately ejected from the survey. Qualtics' panel targeting systems were used to advertise to participants who matched these criteria. If a participant's IP address location was not the Netherlands, access was blocked. At the beginning of the survey, participants self-reported their English proficiency using the following scale: (1) "ask directions and answer simple questions"; (2) "can have basic conversations on familiar topics"; (3) "functional in most contexts (e.g. could tell a story or fill out a form), but not fluent"; (4) "fluent with occasional mistakes, clearly a foreigner"; and (5) "very fluent, can use the language as well as a native language". Participants who answered (1) were immediately ejected from the survey. Mean proficiency was $\bar{X} = 3.96$, $SD = 0.75$. After this, participants completed the FTR-elicitation task followed by the risky-intertemporal-choice task. Within both tasks item order was randomised and one item was displayed per page.

4.4.2 Results

Bilingual FTR elicitation differences echo the Study 2 findings, *Table 4.1*, *Fig. 4.5.* Use of the future tense as well as low-certainty language was higher in the English condition, and this was mostly driven by use of low-certainty modal verbs. This appeared to be moderated by self-reported English proficiency, with high-proficiency speakers using more low-certainty modal verbs, *Fig. 4.5.*

Did higher low-certainty modal verb use lead to increased discounting, and was this moderated by English proficiency? To test this, we implemented a moderated mediation model. In moderated mediation, the moderator is referred to as "W" (in this case English proficiency). Moderated mediation is essentially mediation with interaction term(s) included (Hayes, 2013). We were interested in how English proficiency interacted with (or "moderated")

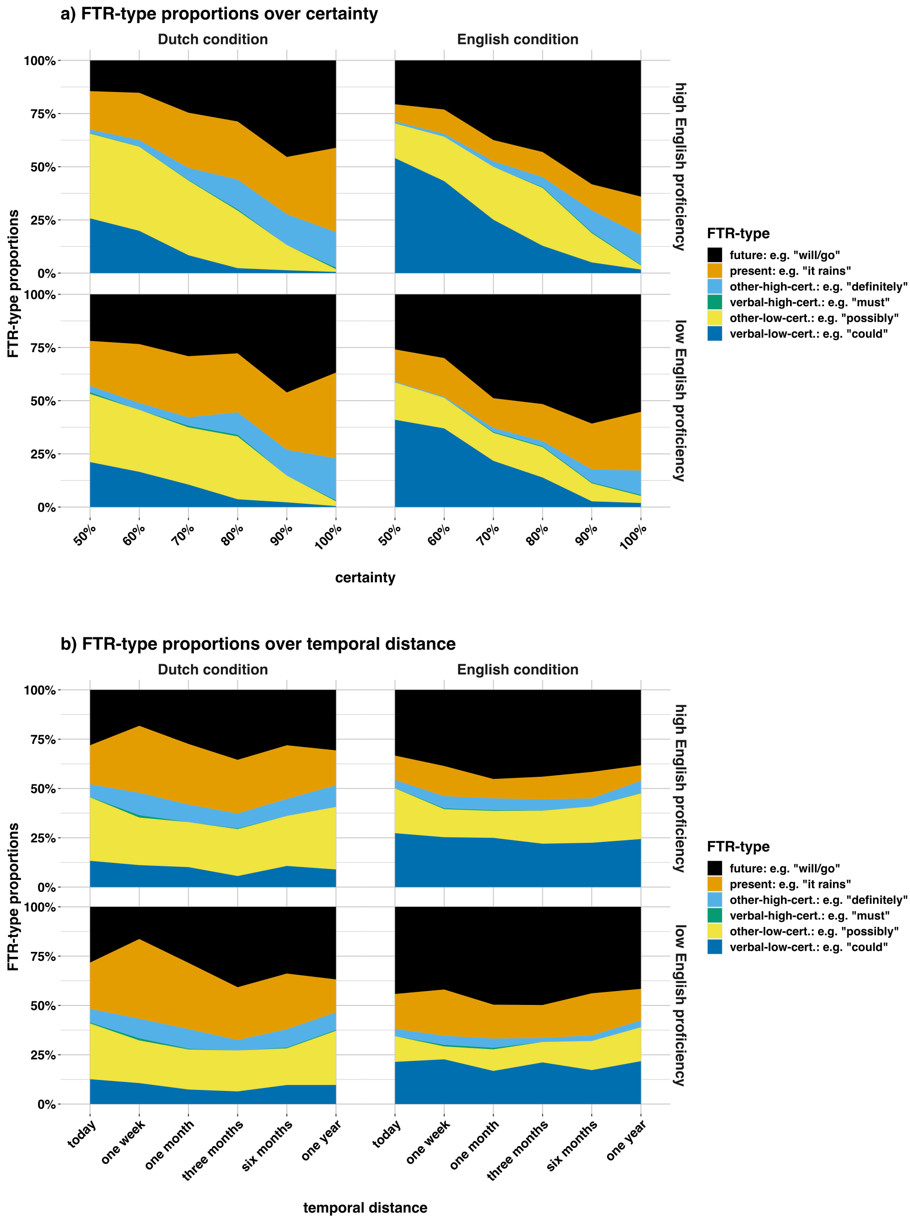


Figure 4.5: Study 3 elicited FTR-type proportions by certainty condition, temporal distance, and language, median-split on self-reported English proficiency. High-proficiency bilinguals used more low-certainty modal verbs, reflecting results in Study 2.

the indirect effect of language condition on combined discounting via low-certainty modal verb use. To test this, we implemented the following model:

$$
\begin{aligned}
log_e\left(\frac{\pi_{VLC_{ijk}}}{1-\pi_{VLC_{ijk}}}\right) = \; & \lambda_1 + \alpha_1 cond._i + \alpha_2 E.prof._i + \alpha_3 cond._i E.prof._i + \\
& \kappa_{1q} X_{1qi} + u_{1j} + h_{1k} \\
subj.val._{ij} = \; & \lambda_2 + \tau' cond._i + \beta VLC_i + \kappa_{2q} X_{2qi} + u_{2j} + e_{2i}
\end{aligned}
\tag{4.9}
$$

where $\pi_{VLC_{ijk}}$ is the probability a low-certainty modal verb (VLC) is used for observation i, participant j, and item k. Other terms are analogous to Studies 1 and 2. The difference is α_3, which captures the differential impact of language condition on verbal-low-certainty use conditioned on English proficiency ($E.prof.$ or W). See *Table 4.2* for full parameter estimates and *Fig. 4.6c* for a conceptual diagram. We estimated this model in *brms*, using the same settings as in Studies 1 and 2. Inspection of caterpillar plots and $\hat{R}$s indicated that parameter estimates were stable (Vehtari et al., 2019).

The indirect effect: We first tested whether the indirect effect of language condition on discounting via low-certainty modal verb use was negative at the mean of English proficiency. As with any interaction, calculating a marginal effect involves summing the parameters in which a predictor of interest is involved and multiplying these by the interacting term. To calculate the conditional indirect effect, this is then multiplied by β, i.e. $(\alpha_1 + \alpha_3 W)\beta$ (Hayes, 2013). English proficiency was mean-centred such that $\bar{X}_{fluency} = 0$, so at mean proficiency this simplifies to $\alpha_1\beta$. At other values of proficiency, the interaction term is added. As we had predicted, the indirect effect ($\alpha_1\beta$) was negative and did not contain zero, *Fig. 4.6a*. Evidence for this was strong, $Est. = -0.04,\ CI_{90\%} = [-0.06, -0.03],\ pp > .999$. This indicates that bilinguals used more low-certainty modal verbs in the English condition and this resulted in lower subjective values (more combined discounting). This is the core prediction of the modal hypothesis.

The moderated indirect effect: Was the indirect effect moderated by English proficiency? The interaction term between language condition and English proficiency was positive and did not contain zero, i.e. α_3 in *Fig. 4.6c*. This indicates that high-proficiency participants used more low-certainty modal verbs in the English condition than did low-proficiency participants, see *Fig. 4.5*. Did this lead to differential impacts on combined discounting? It did. As proficiency increased, the indirect effect grew increasingly negative, *Fig. 4.6b*. This indicates that high-proficiency participants used more low-certainty modal verbs as a function of language condition. As a consequence, they discounted more than low-proficiency speakers. This suggests that language exposure is implicated in outcomes, which supports the modal hypothesis.

The total effect: Did the English condition cause bilinguals to discount more? To test this we calculated the total effect at mean proficiency ($\alpha_1\beta + \tau'$). We found it did not contain zero, *Fig. 4.6a*. Evidence for this was strong, $Est. = -0.18,\ CI_{90\%} = [-0.22, -0.15],\ pp > .999$. This indicates that bilinguals in the English condition discounted more than in the Dutch condition. It echoes the Study 2 findings that English speakers discounted more when outcomes were both delayed and risky. This also supports the modal hypothesis.

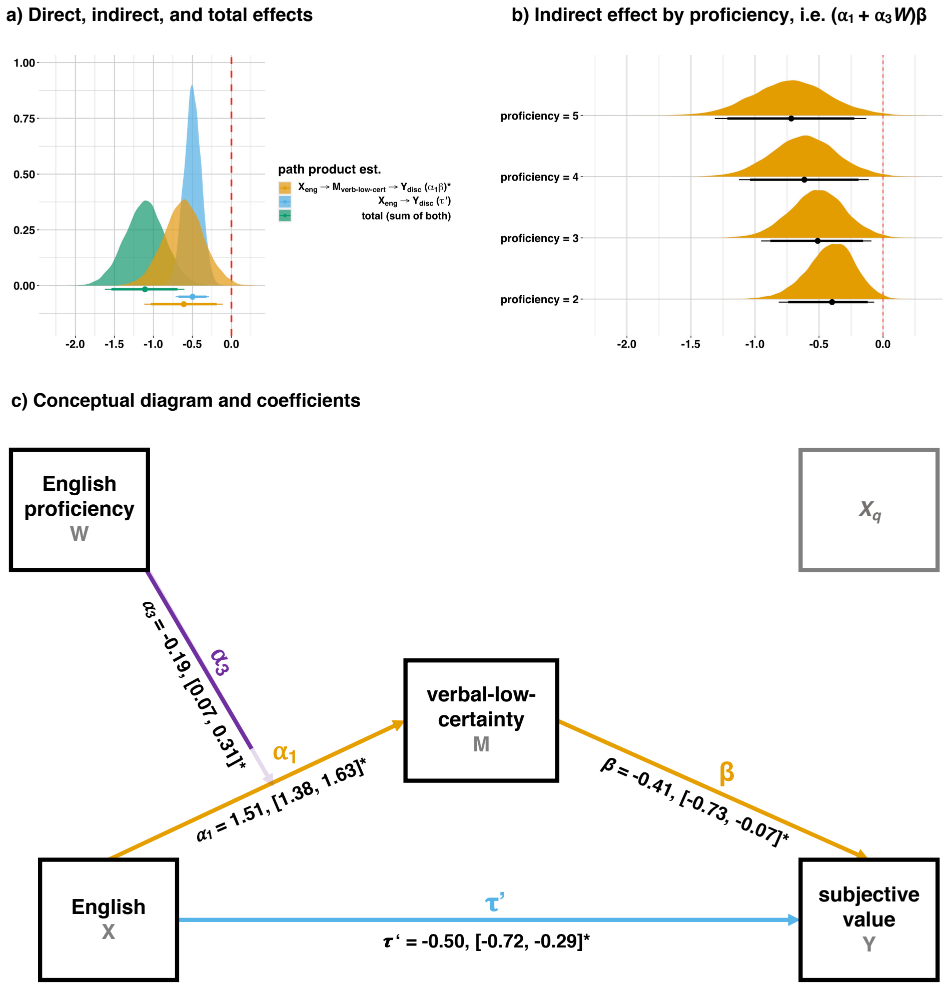


Figure 4.6: Study 3 results. a) Posterior estimates for paths and path products for the effect of the English condition on subjective value (indifference points) for Dutch-English bilinguals. Self-reported English proficiency was mean centred so (a) depicts the indirect effect at mean English proficiency. b) The indirect effect of language condition on subjective value via low-certainty modal verbs, conditioned on English proficiency (W). More proficient speakers used a increasing number of low-certainty modal verbs, which caused the indirect effect to grow increasingly negative—a moderated mediation effect. c) Conceptual diagram with coefficient estimates. Path names correspond to *Table 4.2*. See *Table 4.2* for other X_q control variables.
*Coefficients for which the 95% CI does not contain zero are significantly different from zero.

4.4.3 Discussion

Study 3 established that English grammatical constraints which require speakers to use low-certainty modal verbs when they make predictions about the future resulted in increased combined discounting in bilingual individuals. The fact that higher low-certainty modal verb use mediated the effect suggests English grammatical constraints on FTR are involved, since modal verbs are obligatory for English prediction-based FTR. The treatment and control groups in Study 3—arguably—shared the same culture, since they are the same people measured at different times. This therefore strengthens the evidence that cross-linguistic differences in FTR grammar are a causal factor in increased English combined discounting of risky and delayed outcomes. The fact that self-reported English proficiency moderated this effect supports this conclusion. Study 3 therefore strengthens causal support for the modal hypothesis and for the general hypothesis that grammatical constraints on FTR affect psychological discounting.

4.5 General discussion

Independently of contrasting tests between the temporal and modal accounts, the results support the general hypothesis that cross-linguistic differences in FTR grammar cause differences in psychological discounting. Study 2 found increased combined discounting in English speakers and this effect was entirely mediated by low-certainty language use. Because this left open the possibility that cultural differences drove English speakers to use more low-certainty language, we replicated the result in a population of Dutch-English bilinguals (Study 3). Testing bilinguals in English resulted in increased low-certainty modal verb use and consequently increased combined discounting. This was moderated by self-reported English proficiency. This suggests cross-linguistic differences in FTR grammar cause differences in psychological discounting between populations. Below, we outline support for the temporal and modal hypotheses and discuss wider implications.

4.5.1 The temporal account was not supported

Study 1 was conceived to specifically test the predictions of the temporal account. English speakers rated future events as more temporally proximal—the opposite direction to that predicted by the temporal account. As a consequence, English speakers had higher—not lower—indifference points (less temporal discounting). Additionally, Study 2 found that use of the future tense had no impact on subjective temporal distance, as predicted by the temporal account. Consequently, higher future tense use in English had no impact on combined discounting. Finally, Study 1 found no temporal discounting difference between English and Dutch speakers. This undermines the temporal but not the modal hypothesis, because the temporal hypothesis predicts a difference in temporal discounting while the modal hypothesis does not, see *Fig. 4.1.* These results constitute strong evidence against the temporal hypothesis in general, and the temporal-distance mechanism in particular. More research would be needed to explore the other hypothesised temporal mechanism, which involves the idea that strong-FTR speakers have more precise representations of future temporal events (see Chen, 2013).

4.5.2 The modal account was supported

Study 2 found English speakers used more low-certainty language and this resulted in increased combined discounting—the central prediction of the modal hypothesis. To address the concern that cultural differences could have driven increased use of low-certainty language in English speakers, we replicated this finding with Dutch-English bilinguals. Since effects resulted from testing the same individuals at different times in different languages, this goes some way to isolating the effects of language from culture. When tested in English, bilinguals used more low-certainty modal verbs, and this resulted in increased combined discounting—again, the central prediction of the modal hypothesis.

A relevant critical perspective on these findings is that switching between languages may activate cultural norms and schemas in bilingual participants (see S. X. Chen, 2015). The idea is that language learning is usually associated with a specific cultural context. When bilinguals switch between languages, this may activate the values, beliefs, attitudes, and construals which are associated with that cultural context (see Grosjean, 1982; Ervin, 1964). In this perspective, manipulations of language are to some extent also manipulations of culture. If Anglophone culture involves more present-oriented attitudes than Dutch culture, could switching between languages have activated such attitudes in Study 3? We view this as unlikely for three reasons. First, Study 3 involved the elicitation of language. Low-certainty modal verb use affected combined discounting. This suggests cross-linguistic grammatical differences are a causal factor, since English obliges a modal verb for future predictions. Second, self-reported English proficiency moderated the indirect effect of language condition on combined discounting via low-certainty modal verb use. This suggests that linguistic exposure was involved in producing the effects identified in Study 3. Third, it seems incoherent to give a language-independent account of the idea that bilinguals are bicultural. How could language switching activate cultural values and beliefs if language were not implicated in the maintenance and transmission of such values? Such an account would therefore rely on what some have described as *the* central idea of linguistic relativity: "the idea that culture, *through* language, affects the way we think" (Gumperz & Levinson, 1996, p. 1, emphasis original). For these reasons, we interpret Study 3 as providing strong evidence that the obligation to use low-certainty modal verbs for predictions about the future causes increased combined discounting in English speakers—the modal hypothesis.

4.5.3 Bilingual findings

Study 3 complements similar findings with bilingual participants. In a between-subjects design, Pérez and Tavits (2017) found bilingual participants who completed a survey in Estonian (weak-FTR) were more likely to support a fictitious future-orientated environmental tax policy than when other bilingual participants completed the same survey in Russian (strong-

FTR). This suggests that bilinguals discounted less when completing a weak-FTR survey. Additionally, Sutter et al. (2015) found that German (weak-FTR) speakers were more future-oriented than Italian (strong-FTR) speakers in a temporal discounting task conducted on a sample of school children in bilingual South Tyrol. Together, these findings suggest that strong-FTR speakers discount more than weak-FTR speakers, even when they share similar cultural environments. Our work complements these findings by implementing (as far as we know) the first within-subjects test of FTR status. Echoing previous work, we found that bilinguals discounted more when directed to use a strong-FTR language (English). An important point is that our results undermine the temporal account which previous work with bilinguals has been interpreted as supporting (Pérez & Tavits, 2017; Sutter et al., 2015), since our study specifically teased apart, for the first time, the temporal and modal hypotheses.

4.5.4 Subjective temporal distance and psychological discounting

The subjective-temporal-distance results contribute to ongoing research into the relationship between subjective construals of future time and intertemporal decision making (B. K. Kim & Zauberman, 2009; Zauberman et al., 2009). They echo previous findings that distal temporal representations are associated with increased temporal discounting (Thorstad et al., 2015). For instance, Thorstad and Wolff (2018) found that people whose tweets reference distant future times were more likely to invest in future outcomes. This suggests that people who think about far-distance future events are more future-oriented. As far as we can tell, our results are the first to use time-slider methods to establish cross-cultural differences in representations of future temporal distance. However, they suggest this variation is unrelated to FTR grammar in English and Dutch. In fact, English and Dutch do not use the future tense to encode temporal distance. Rather, temporal distances are lexically expressed using temporal adverbials (“one week”, “next month”, “tomorrow”, etc.). Future research might explore a limited version of the temporal hypothesis applied to languages which do use future tenses to encode notions of temporal distance. For instance, the Northeastern Bantu language Gĩkũyũ in Kenya has different future tense morphemes for near and remote FTR (Cable, 2013).

4.5.5 Mixed temporal discounting results

The total effect in Study 1 contained zero. In other words, there was no total difference between English and Dutch speakers in a temporal discounting task. In previous studies, we found mixed results, only sometimes identifying a difference in temporal discounting between Dutch and English speakers (Chapter 3). This echoes other mixed results in the few studies involving FTR status and experimentally-elicited temporal discounting (i.e. rather than naturalistic outcomes). In particular, effects identified by Sutter et al. (2015) in school children in the South Tyrol supported the predictions of the temporal account (more discounting in strong-FTR-speaking children). On the other hand, (Thoma & Tytus, 2018) found strong-FTR speakers discounted less in an experimental intertemporal-choice task conducted with speakers of five languages. Our findings add to these mixed results. They suggest that FTR status is more likely to impact combined discounting than (just) temporal discounting.

Assuming real-world intertemporal decision making is usually characterised by some degree of risk, this suggests that differences in naturalistic intertemporal outcomes are driven by the modal effects identified here. One of the reasons that researchers continue to use non-experimental regression-based approaches to test the temporal hypothesis is that Chen (2013) provides a handy dichotomous independent variable (FTR status) which can be used to predict natural outcomes. While this approach has yielded fascinating insights, we suggest more painstaking methods involving smaller language samples, closer attention to cross-linguistic differences, and more tightly controlled risky-intertemporal-choice paradigms will deepen our understanding of this question in the future.

4.5.6 Causal order and cultural confounds

Above, we raised the possibility that if English speakers psychologically discount more, this could be driven by culturally-informed low-certainty schemas for future outcomes. This might lead to higher use of low-certainty language, i.e. *English* → *discounting* → *low-certainty language*. By specifying combined discounting as the mediator, Study 2 allows this alternative account to be tested. When we did this, the indirect effect was null (Appendix E). This supports the conclusion that English grammar impacts combined discounting not the other way around.

4.5.7 Conclusions

Psychological discounting processes are an important determinant of a wide range of behaviour. This includes health decisions (Ireland et al., 2015; Vuchinich & Simpson, 1998), drug use (McKerchar & Renda, 2012), pathological gambling (Hodgins & Engel, 2002), and investment in savings (Liu & Aaker, 2007). FTR status has been found to predict a similarly diverse range of behaviour including health outcomes such as obesity, peak blood flow, grip strength, and physical exercise levels (Chen, 2013); educational outcomes such as math and reading scores, graduation, absenteeism, retention, and disciplinary incidents (Figlio et al., 2016); firm-level investment in research and development (Chi et al., 2018); country-level economic growth as indicated by output per worker, patents, innovation performance, and capital stock (Hübner & Vannoorenberghe, 2015b); country-level likelihood of having taken climate change mitigation action (Mavisakalyan & Weber, 2018); and even country-level suicide rates and support for euthanasia (Lien & Zhang, 2020).

If the correlation identified in these studies is not spurious, it would constitute new ground for the linguistic relativity hypothesis, which has been characterised by experimental work on processing and attentional biases (Wolff & Holmes, 2011). Whereas such outcomes have been criticised as having few applications to real-world behaviour (McWhorter, 2014), psychological discounting is a consistent predictor of a range of important outcomes outside of the laboratory.

The reported real-world effects of FTR status suggest cross-linguistic differences in FTR grammar affect (risky) intertemporal behaviours. However, they do not provide strong causal evidence—as we have done—and they do not untangle the causal mechanisms (temporal, modal) which may be involved. If policy and public communications strategies are to be informed by research, it is critical to understand which language differences impact behaviour. Our results provide evidence that FTR grammaticisation impacts psychological discounting processes. They indicate that the obligatory use of low-certainty modal verbs is more likely than the obligatory use of future tenses to be responsible for observed effects of FTR status.

Chapter 5

A big-data and experimental analysis of the relationship between time perspective, anxiety, and depression

Abstract

Anxiety and depression are mood disorders which negatively impact many people. Psychology studies suggest depression is associated with future time horizons, or how "far" into the future people tend to think, and anxiety is associated with temporal discounting, or how much people devalue delayed future rewards. Separate studies from linguistics and economics have shown that individual differences in how people refer to future time predict temporal discounting. Yet no one—that we know of—has investigated whether future time reference habits are a marker of anxiety and/or depression. In Study 1, we analysed data from the social media website Reddit to test for relationships between these variables. Users who had previously posted popular contributions to forums dedicated to anxiety and depression referenced the future and the past more often than controls, had more proximal future and past time horizons, and significantly differed in their linguistic future time reference patterns. They used fewer future tense constructions (e.g. *will*), fewer high-certainty constructions (e.g. *certainly*), more low-certainty constructions (e.g. *could*), more bouletic modal constructions (e.g. *hope*), and more deontic modal constructions (e.g. *must*). This motivated Study 2, a survey-based mediation analysis. Self-reported anxious participants represented future events as more temporally distal and therefore temporally discounted to a greater degree. The same was not true of depression. We contextualise the results using two theoretical frameworks: construal level theory and functionalist approaches to clinical psychology. We conclude that methods which combine big-data with experimental paradigms can help identify novel markers of mental illness, which can aid in the development of new therapies and diagnostic criteria.

5.1 Introduction

Millions of people suffer from mental illness world wide. For example, in North America the prevalence of mental health disorders in adults may be as high as 20% (National Institute of Mental Health, 2019). In addition to the impacts on well-being—important in their own right—the estimated economic costs of mental illness in North America are as much as US$83.1bn annually (Dewa & McDaid, 2011). At the same time, medical science has struggled to predict the onset of mental illness from physiological markers which has made preventative medicine difficult to implement (Kapur, Phillips, & Insel, 2012). A potentially promising avenue is to identify linguistic markers of mental illness. One way to do this is using naturalistic linguistic data from social-media websites. This has lead to work which uses social media to predict various aspects of personality, including the onset and diagnosis of mental illness (see Guntuku, Yaden, Kern, Ungar, & Eichstaedt, 2017; Goldstone & Lupyan, 2016).

For instance, the language used in previous posts on the social-media site Reddit was successfully used to predict whether a user would later post to a forum dedicated to mental illness (Thorstad & Wolff, 2019a). Tweets by Twitter users who self-diagnosed as suffering from depression or post traumatic stress disorder were successfully classified above chance levels (Coppersmith, Dredze, Harman, Hollingshead, & Mitchell, 2015). Additionally, posts to 11 different Reddit forums dedicated to mental illness were successfully differentiated with an average of 72% precision (Gkotsis et al., 2017).

Many predictive studies have implemented "black-box" machine learning algorithms which focus on predicting mental health outcomes from prior social media data. While predictive approaches are undoubtedly valuable, model parameters are not always interpretable by human practitioners (Rai, 2020). Additionally, black-box approaches may risk amplifying markers of more general structural inequality that can be associated with mental illness (see Eubanks, 2018; Obermeyer, Powers, Vogeli, & Mullainathan, 2019). Another strand of research therefore seeks to complement predictive approaches with descriptive ones that attempt to generate psychological insights from the analysis of social media data. For instance, Thorstad and Wolff (2019a) used per-word model weights to understand which words were most predictive of a user later posting to a Reddit forum focused on depression. Similarly, De Choudhury, Counts, Horvitz, and Hoff (2014) found language use differences on Facebook posts as a function of whether survey respondents had been clinically diagnosed with postpartum depression.

In the present study we therefore implemented a hybrid methodology, using a combination of machine-learning and descriptive natural language processing to explore the relationships between future time horizons, linguistic future time reference patterns, and anxiety and depression. The choice of these variables was motivated by previous research which suggests that how people construe future events—their future time perspective—may be a key marker of anxious and/or depressive tendencies (Dilling & Rabin, 1967; Pulcu et al., 2014; Steinglass et al., 2017).

Future time perspective encompasses how people think about and relate to future events. It includes questions such as: Do people tend to save for the future or spend resources now, do they conceive of future events as proximal or distal, and do they tend to imagine temporally proximal or distal future events? Measures of future time perspective have been found to be significant markers of anxiety (Papastamatelou, Unger, Giotakos, & Athanasiadou, 2015), major depressive disorder (Pulcu et al., 2014) and social anxiety disorder (Steinglass et al., 2017).

Particularly, people suffering from depression have been found to have contracted “future time horizons” relative to controls (Dilling & Rabin, 1967). This is usually measured using the Wallace task (Wallace, 1956). In this task, participants are asked to think of 10 events that are likely to occur in their future lives, and what age they expect to be when the events will occur. Their future time horizon is then calculated by subtracting their current age from the ages they give. This approximates how far into the future participants tend to imagine themselves.

Pathological gamblers (Hodgins & Engel, 2002), alcoholics (Smart, 1968), and heroin addicts (Petry, Bickel, & Arnett, 1998) were all found to have shortened future time horizons compared with healthy controls. While these studies do not explicitly study depression, substance abuse disorders and depression co-occur often enough to warrant potential connections between these results and underlying depressive tendencies (Barrault & Varescon, 2013; Blaszczynski & McConaghy, 1989; Felton et al., 2020). Since anxiety tends to involve excessive “anticipation of future threat” (American Psychiatric Association, 2015, p. 1), the subjective experience of future events seems almost inherently salient to the experience of anxiety. However, the relationship between anxiety and future time perspective seems less well understood then for depression.

Most research into anxiety and time perspective has used temporal discounting paradigms. Temporal discounting captures the extent to which people devalue delayed future outcomes as a function of the length of time until they will occur. In general, people tend to devalue delayed rewards as a function of wait time, but the extent of such devaluation differs between individuals (L. Green & Myerson, 2004). These differences are usually measured by giving participants a series of binary choices between immediate and delayed rewards, for instance \$10 now vs. \$20 in a month. “Present-oriented” people tend to devalue future outcomes and would therefore opt for the immediate \$10, while “future-oriented” people tend to devalue less steeply and would therefore opt to wait for the \$20. The term “time preferences” refers to individual differences in tasks like this. These measures are not unrelated to time horizons. Shorter time horizons tend to be associated with increased discounting (Thorstad & Wolff, 2018; Thorstad et al., 2015), and researchers increasingly understand temporal discounting processes in terms of subjective representations of future time (B. K. Kim & Zauberman, 2009; Zauberman et al., 2009), i.e. as opposed to discounting rates over objective time.

While people suffering from major depressive disorder have been found to temporally discount more than healthy individuals (Felton et al., 2020; Pulcu et al., 2014), the results around anxiety have been mixed. Some findings suggest people suffering from social anxiety discount more (Rounds et al., 2007), while others report the effect goes in the opposite direction (Steinglass et al., 2017), or that there is no significant relationship (Jenks & Lawyer, 2015). This suggests more research is needed to untangle how future time perspective may relate to the experience of anxiety.

An understudied topic that may help to resolve these issues is the linguistics of how people talk about the future. A growing body of research indicates that linguistic Future Time Reference (FTR) may be an important factor in shaping future time perspective. FTR is a catch-all term used for any linguistic statement which refers to future events, whether or not it uses the future tense (Dahl, 1985, 2000a). For instance, *It could rain tomorrow*, *It will rain tomorrow*, and *I hope it rains tomorrow* are all FTR statements. Recent work suggests that cross-linguistic differences in FTR grammar impact speakers' time preferences. Chen (2013) provides a framework for how this might occur. He hypothesised that languages which oblige speakers to use the future tense for FTR might cause speakers to perceive the future as farther away and therefore discount more. For example, in English, speakers are obliged to use *will* in the sentence *It <u>will</u> rain tomorrow.* On the other hand, speakers of Dutch are free to use the present tense, *Het regent morgen*, 'It rains tomorrow.' Chen (2013) theorised that consistently using the present tense for FTR would collapse future into present time, effectively causing speakers to construe objective future dates as subjectively more proximal and therefore temporally discount less (i.e. be more willing to wait for subjectively more proximal future rewards). In support of this hypothesis, he found that speakers of languages like Dutch were more likely to have saved each year and retired with more assets. Since then, numerous other studies have found that speakers of languages like Dutch tend to behave as though they discount less (e.g. Chi et al., 2018; Falk et al., 2018; Figlio et al., 2016; Liang et al., 2018; Mavisakalyan et al., 2018; Pérez & Tavits, 2017; Zhu et al., 2020).

However, in addition to constraints on future tense use, cross-linguistic differences in FTR grammar can involve constraints on the encoding "modal" notions of possibility, probability, and certainty (Chapter 2). For instance, it is not actually obligatory to use *will* in the sentence *It will rain tomorrow.* Rather, English grammar obliges speakers to use one of *It could/may/might/should/shall/would/is going to rain tomorrow.* These "modal" verbs generally encode notions of weakened certainty relative to the future tense constructions (*shall, will,* and *be going to*). Obligatory use of low-certainty modals for FTR in English drove differences in time preferences involving risky outcomes between English and Dutch speakers (Chapter

4). This suggests that FTR usage habits reflect latent beliefs about the risk and/or temporal distance of future events, and are related to time perspective.

However, the relationship between FTR habits and anxiety and depression has not been studied. We therefore undertook an exploratory study which used big-data methodology to understand whether people who had posted to Reddit forums dedicated to anxiety and depression differed from a random sample in terms of their FTR language use and future time horizons.

5.2 Study 1

The basis of Study 1 is an analysis of data downloaded from Reddit. Reddit is a social media platform with over 430m average monthly users (Reddit Inc., 2020). The site is similar to a message board or traditional forum. Users ("redditors") communicate in forums ("subreddits") dedicated to specific topics. There are no character limits to user contributions ("posts"), so users are free to comment in as much detail as they wish. Subreddits including topic focus and rules on acceptable behaviour are created and self-monitored by moderator teams comprising leading subreddit members. Comments are "upvoted" and "downvoted", and can be sorted by various criteria, for instance most upvoted of all time. The names of subreddits usually reflect their topical focus and are prefaced with *r/*, for instance, *r/news* is dedicated to sharing US and international news stories.

The largest subreddits dedicated to anxiety and depression are *r/Anxiety* and *r/depression*. In December, 2020, *r/Anxiety* had 417k members, and *r/depression* had 712k members (Reddit Inc, 2020). Both are focused on providing support and advice for people dealing with issues related to anxiety and depression (respectively) and both allow posts by sufferers or those close to them.

5.2.1 Methods

This study was approved by the internal review board of Brunel University London (Ref. 7863-A-Jan/2018-10690-1).

5.2.1.1 Data acquisition

We were concerned that language used when posting to *r/Anxiety* and *r/depression* might differ from a random sample not due to the habits of speech of the posters, but due to rules of discourse imposed by moderators. We therefore implemented a "snowball" sampling approach, first constructing a "seed" sample of redditors who had posted to *r/Anxiety* and *r/depression*, and then a final sample comprising these same redditors' top posts in any subreddit.

Specifically, we used the Python Reddit API Wrapper (Boe et al., 2020), to download the $1,000$ "hottest" comments in each of *r/Anxiety* and *r/depression.* Here "hotness" is defined by a Reddit proprietary algorithm for ranking content. It is calculated by taking the log_{10} of the net number of upvotes and adding it to the number of 12.5 hour periods that have passed since the first ever Reddit post. This means that, for each successive 12.5 hour period, a post must get ten times the number of upvotes than it had in the previous 12.5 hour period to retain the "hot" score it had when it was 12.5 hours younger. Its value is that it captures topical content specific to a given subreddit, where other measures like "top", which counts total number of upvotes, are disproportionately captured by "stickied" content that is permanently visible.

This resulted in a sample of $n = 910$ posts from *r/Anxiety* and $n = 902$ posts from *r/depression.* (As scores for user-deleted submissions are still returned by the API, the number of usable posts was less than 1,000.) We then downloaded up to the top $1,000$ most recent posts made to any subreddit by each unique redditor in these "seed" data. These data were downloaded 30 October 2019. This resulted in $n = 224,009$ posts made by redditors who had once posted popular content in *r/Anxiety* (the "anxiety" condition), and $n = 182,555$ posts made by redditors who had once posted popular content in *r/depression* (the "depression" condition). However, we found no differences between these conditions on any measure in this study, and so combined these into a single "mental health" condition. We constructed a control sample of $n = 213,132$ posts by repeating an identical process with the *r/All* subreddit, which is a compendium of all content generated on Reddit. These data were downloaded on 1 February 2020. We refer to the contrast between the mental health and control conditions as "data-source condition".

This meant our final sample comprised $N = 619,696$ posts, drawn from $N = 13,651$ unique subreddits ($n = 6,978$ [mental health], $n = 2,781$ [control], and $n = 3,856$ crossover subreddits which had been posted to in both the mental health and control conditions), and $N = 4,205$ unique redditors ($n = 1,810$ [mental health], $n = 2,395$ [control], no crossovers). It should be noted that this process resulted in some posts from the original seed subreddits still being retained in the final sample ($n = 10,145$ posts from *r/Anxiety*, and $n = 17,319$ from *r/depression*). Excluding these data did not substantively change any results, i.e. cause significance to cross critical thresholds or coefficients to change sign, so we did not exclude them.

5.2.1.2 Data preprocessing

Inspection of the data indicated that, due to their unrestricted length, a number of posts contained references to both past and future time. We therefore split posts into constituent sentences using a non-monotonic transition system dependency parsing algorithm (Honnibal & Johnson, 2015) implemented in Python (Python Software Foundation, 2017) in the open-source natural language processing package *spaCy* (Explosion AI, 2020). This resulted in $N = 2,014,181$ sentences ($n = 1,549,023$ [mental health], $n = 465,158$ [control]).

5.2.1.3 Temporal horizons

To estimate temporal horizons, we used the Stanford University Time (SUTime) temporal tagging system (Chang & Manning, 2012). SUTime is a rule-based deterministic natural language tagging system which uses regular expressions which combine both keywords (e.g. "yesterday") and rules (e.g. "[DATE] at [TIME]") to identify temporal expressions and convert them to numerical data. It processes absolute dates (e.g. November 7, 2012), relative dates (e.g. "this Friday"), and combinations of these (e.g. "tomorrow at 3pm"). For relative dates, we used time of posting (available from the Reddit API) as the reference date. For instance, if a redditor referenced "tomorrow" on 10-06-2020, SUTime would return 11-06-2020. To calculate time horizons (days), we subtracted time of posting from the SUTime reference values. In cases where there was more than one time reference in one sentence, we took the mean, i.e. $H(s) = \sum_r \frac{D(r)}{n}$; where $H(s)$ is the time horizon for a given s sentence, r is a temporal reference, $D(r)$ is the number of days from time of posting, and n is the number of temporal references in the sentence (Thorstad & Wolff, 2018). Such methods with data from Twitter have been found to positively and significantly correlate with participants' future time horizons as measured by the Wallace task, as well as predict both collective and individual intertemporal decision making (i.e. temporal discounting) (Thorstad & Wolff, 2018).

5.2.1.4 Time reference classification

To estimate whether sentences were FTR or Past Time Reference (PTR), we used an active learning paradigm implemented in Prodigy (Montani & Honnibal, 2020) to train a machine learning "time reference" classifier. Active learning is a way of reducing the number of annotated examples machine learning models need to generate reliable predictions in supervised learning paradigms. In supervised learning, human raters annotate data and machine learning models are then trained to predict annotations from input data (in this case raw text). Active learning speeds up this process by keeping a machine learning model "in the loop" during the annotation process, and selecting the un-annotated data about which the model is most uncertain. This increases the number of annotations which are near the boundary criteria of the unknown function the model is attempting to approximate, and reduces the number of annotations the model is already able to confidently predict, thereby reducing the total number annotations needed.

The authors used Prodigy to annotate text examples for both FTR ($n = 4002$ [$n_{accept} = 1403$, $n_{reject} = 2406$, $n_{ignore} = 193$[1]]), and PTR ($n = 2382$ [$n_{accept} = 660$, $n_{reject} = 1704$,

[1]For indeterminable cases.

Table 5.1

Time reference and FTR-type classifier performance metrics in Study 1

classifier	category	sub-cat.	accuracy	precision	recall	*F1*
time reference	FTR	-	.91	.93	.73	.82
	PTR	-	.81	.83	.35	.49
FTR type	future tense	-	.84	.81	.92	.86
	present tense	-	.93	.94	.99	.96
	low-certainty	verbal	.99	$>.99$	.99	.99
		other	.96	$>.99$	.97	.98
	high-certainty	verbal	.98	$>.99$	.98	.99
		other	.98	.99	.98	.99
	bouletic modality	-	.94	.99	.94	.97
	irrealis	-	.89	.99	.89	.94

Accuracy is defined as $a = (tp + tn)/(tp + fp + fn + tn)$ where tp is the number of true positives, tn is the number of true negatives, fp is the number of false positives and fn is the number of false negatives. Accuracy captures the classifier's performance without prioritising either positive or negative examples. Precision is $p = tp/(tp + fp)$; it captures the model's likelihood of being correct if it makes a positive prediction and is therefore sensitive to the model's type I error rate. Conversely, denominator in recall is the false negatives, $r = tp/(tp + fn)$; it therefore captures whether the model tends to miss true examples, and is sensitive to the model's type II error rate. *F1* is the harmonic mean of recall and precision, $F1 = (2rp)/(r + p)$, and attempts to balance the two.

$n_{ignore} = 18$]). Because the active learning process resulted in overlaps, there were $N = 4,947$ unique text examples. For PTR, all annotations were drawn from our Reddit data; for FTR, annotations were drawn from a combination of a dataset of Tweets and Reddit data (see Appendix F). Because we wanted to cast as wide a net as possible to better understand linguistic usage within the broad notional domains of future and past time reference, our annotation criteria were semantic rather then formal. Any statement which made FTR/PTR was classed positively, regardless of formal tensing. For instance, "they could win", "it will be ok", "it rains tomorrow", "I go in three weeks", "my train is arriving soon", and (in November) "Christmas is soon" would all be classed as FTR. Similarly, "it has rained", "I saw her", "it was tiresome", "I'm walking along the other day", and "I'm tired from my run this morning" would all be classed as PTR.

Using Prodigy, we trained a model on these data using 80%/20% train/evaluate split, 10 iterations over the data, a dropout rate of 0.2, and exclusive categories. The Prodigy text classification system wraps a *spaCy* ensemble model comprising a bag of words model and a 1-dimensional Convolution Neutral Net (CNN) where token vectors are calculated using a CNN, mean pooled and used as features in a feed-forward network (Explosion AI, 2020). Our trained model can be loaded as a *spaCy* module in Python and will be available as part of the FTR-type classifier (Robertson, 2019b). Because not all the FTR annotations had been drawn from the Reddit data and we wanted to make sure the model's out of sample performance was acceptable, we annotated two additional "test" datasets drawn exclusively from the Reddit data ($N_{FTR} = 477$, and $N_{PTR} = 489$). The model outputs scalar probabilities that an item of text data is FTR or PTR. We converted these to categorical data by splitting on $p = 0.5$:

$$f(p_{tr}) = \begin{cases} 1, & \text{if } p_{_{TR}} >= 0.5 \\ 0, & \text{otherwise} \end{cases} \tag{5.1}$$

where $p_{_{TR}}$ is the probability p that an item of text data is either FTR or PTR. This allowed us to test accuracy against categorical human annotations. Model performance on these withheld data was generally good, *Table 5.1*. Low recall for PTR classifications indicates that the model tends to miss some true PTR statements. This is because we prioritised precision; high precision indicates that when the model does predict PTR it is correct 83% of the time. We reasoned that having reliable positive predictions was more important than identifying a high percentage of true positives. Why FTR performance is better is unclear; perhaps more data would be useful.

5.2.1.5 FTR-type classification

To establish FTR-usage patterns, we used the FTR-type classifier (Robertson, 2019b). It is closed-vocabulary deterministic keyword-based classification program written in Python (Python Software Foundation, 2017). It classifies short future-referring text documents into 7 exclusive categories based on the presence of keywords which encode semantic domains most salient to FTR. One of these categories is further broken down according to the word classes of the keywords involved. See Chapter 2 for a description and justification of these categories for English FTR. Each category is dichotomous, and coded (1) for positive examples, otherwise coded (0).

Two categories are tense-based and the other 5 are based on modal semantics. Modality generally involves quantifying what is possible and/or necessary relative to any of several common modal "bases" (Palmer, 2001). For instance, epistemic modality involves speakers expressing what they think is probable or unlikely relative to an epistemic "base", i.e. what they know and believe about the world (Nuyts, 2000); bouletic modality involves speakers quantifying what they think is possible or desirable relative to a desiderative base, i.e. what they hope or want to occur (Palmer, 2001); and deontic modality involves speakers quantifying what they think is obligatory or necessary relative to a normative base, i.e. what they think *must* or *should* occur (Palmer, 2001). The 7 categories are as follows.

The future tense: Sentences which use the future tense without any other modal words are classed in this category, e.g. *Tomorrow it will/is going to/shall rain* would be classed as future tense because it uses *will, is going to*, or *shall* and not also any modal class-criterion keywords (see below).

The present tense: Sentences which are FTR but fail to be classed in any other category are classed as present tense, e.g. *tomorrow it rains* would be classed as present tense because it does not use the future tense or any other modal class-criterion keywords (see below).

Low-certainty epistemic modality: Sentences which express low certainty about future events are classed in this category. This is further broken down depending on whether a modal verb or some other keyword class is used, e.g. *It could/may/might/should rain tomorrow* would all be classed as verbal-low-certainty because they use the low-certainty modal verbs *could, may, might* or *should*, while *It will possibly/probably/potentially rain tomorrow* and *I think it will rain tomorrow* would be classed as other-low-certainty because they express low certainty without using a modal verb. The reason for this subdivision is that our previous research, as well as review of the literature (Nuyts, 2000), suggests that modal verb use is more constrained by features of English grammar—modal verbs are obligatory whereas other-low-certainty constructions are not. We reasoned that individual differences in FTR usage would be more likely to manifest in the linguistic units which are freer to reflect speakers' beliefs and intentions.

High-certainty epistemic modality: Sentences which express high certainty are classed in this category, e.g. *It will definitely/certainty/absolutely rain tomorrow* would be classed as other-high-certainty. In English FTR, this domain is typically encoded using high-certainty epistemic adjectival or adverbial modifiers.

Deontic modality: Sentences which express necessity and/or obligation are classed in this category, e.g. *You must come tomorrow* and *I have to/need to pick up more groceries tomorrow* would be classed as deontic modality because they use the deontic modal verb *must*, and deontic a *have to/need to* construction. These are typical ways of expressing deontic notions in English FTR (Nuyts, 2000).

Bouletic modality: Sentences which express desires and hopes for the future are classed in this category, e.g. *I hope it rains tomorrow*, *I want it to rain tomorrow*, and *I wish it would rain tomorrow* would be classed as bouletic modality because they use the bouletic modal verbs *hope*, *want*, and *wish*.

Irrealis: Sentences which express counterfactual realities are classed in this category, e.g. *If it rains, I'll go out*, *If he would only go out with me, I would be happy* would be classed as irrealis because they express non-factual realities using the conditional *if*, and *would*.

Since the FTR-type classifier is intended to classify based on semantics, responses in the present tense which expressed modal semantics (e.g. *Tomorrow it could rain*) were classed according to the modal notions, and not formal tense categories. Similarly, responses in the future tense but which expressed modal notions (e.g. *Tomorrow it will possibly rain*) were classed according to the modal notions and not the tense category. This is because *It will possibly rain* and *It will rain* express different semantics and this is reflected in the classification scheme.

To check FTR-type classification accuracy on our Reddit data, we had trained assistants hand-annotate an evaluation dataset. To select data, we randomly selected from sentences for which p_{FTR} >= 50%, n = 50 from each FTR-type category, plus n = 50 uncoded sentences, and n = 50 randomly selected sentences (no overlaps). After exclusions for negations this resulted in N = 447 sentences. We checked predictions against this evaluation dataset, and updated the FTR-type classification code to fix any discrepancies. In order to check this had not resulted in over-fitting to the evaluation dataset, we repeated this process with a second test dataset, this time randomly selecting n = 30 sentences as described above (N = 299 total sentences). Accuracy metrics against this test dataset are presented in *Table 5.1*.

5.2.2 Results

Descriptive statistics for Study 1 are presented in *Table 5.2*.

5.2.2.1 Statistical approach

The data contained multiple sentences from individual posts, multiple posts from individual redditors, and multiple posts and redditors across various subreddits. Using ordinary regressions which assume observations are independent might therefore have risked underestimating actual standard errors and increasing Type I error rates. Our statistical solution to this issue was to calculate "design effect" (*deff*) for each potential source of within-group variation (redditors, posts, subreddits) for each dependent variable in the study. We used *deff* to ascertain whether it was necessary to employ mixed regression techniques to address complex data structures. Kish (1965) defines the *deff* of some sample statistic as "the ratio of the actual variance of a sample of the variance of a simple random sample of the same number of elements" (p. 258). In mixed regression modelling, *deff* can be calculated using the Intraclass Correlation Coefficient (ICC) and the average group size, c, such that (Muthén & Satorra, 1995):

$$deff = 1 + (c - 1) \times ICC \tag{5.2}$$

Table 5.2

Descriptive statistics for Studies 1 and 2

study	variable	condition	n	$\bar{X}$	SD	min	max
Study 1	FTR	mental health	1,549,023	0.07	0.26	0	1
		control	465,158	0.06	0.25	0	1
	PTR	mental health	1,549,023	0.13	0.33	0	1
		control	465,158	0.09	0.29	0	1
	$log_e(\|H_{days}\|)$	mental health	28,249	2.26	2.68	-8.97	11.39
		control	6,999	2.77	2.91	-7.84	11.33
	$\|H_{days}\|$	mental health	28,249	332	2,134.62	0	88,850.2
		control	6,999	923.93	4,629.42	0	83,386.52
	future tense	mental health	95,429	0.26	0.44	0	1
		control	25,302	0.27	0.45	0	1
	present tense	mental health	95,429	0.16	0.36	0	1
		control	25,302	0.17	0.37	0	1
	other-low-certainty	mental health	95,429	0.12	0.33	0	1
		control	25,302	0.1	0.3	0	1
	verbal-low-certainty	mental health	95,429	0.1	0.3	0	1
		control	25,302	0.09	0.28	0	1
	other-high-certainty	mental health	95,429	0.04	0.19	0	1
		control	25,302	0.04	0.2	0	1
	verbal-high-certainty	mental health	95,429	0.04	0.19	0	1
		control	25,302	0.03	0.18	0	1
	bouletic modality	mental health	95,429	0.17	0.38	0	1
		control	25,302	0.12	0.33	0	1
	irrealis	mental health	95,429	0.14	0.35	0	1
		control	25,302	0.14	0.34	0	1
Study 2	depression	-	202	6.86	9.66	0	40
	anxiety	-	202	4.4	7.07	0	34
	subjective temporal distance	-	2222	31.77	32.67	0	100
	$log_e(k)$	-	202	-3.06	2.26	-9.06	1.74

where ICC is the ratio of within-group to total variance. *Deff* is sensitive to either high $ICCs$ or high group sizes. We calculated *deff* for each grouping factor and each dependent variable, using the entire Study 1 data. For all dependent variables, *deff* scores for posts were low, $1.01 \leq deff_{posts} \leq 1.12$, as were *deff* scores for subreddits, $1.03 \leq deff_{subred} \leq 1.27$. However, *deff* scores for redditors tended to be higher, $1.1 \leq deff_{red} \leq 4.48$. We were therefore justified in ignoring group-wise correlation within posts and subreddits, but estimated multilevel regressions with random intercepts clustered by redditor. This approach allowed us to avoid over-inflating standard errors to within acceptable criteria (i.e. $\approx< 10\%$ according to *Table 3* in Lai & Kwok, 2015).

5.2.2.2 Frequency of time reference

The first thing we established was how frequently redditors made non-present time reference. To address this, we estimated two regressions, one for past time reference and the other for future time reference. We used mixed regressions with a logistic link function and random intercepts clustered by redditor to regress the binary FTR and PTR variables resulting from equation 5.1 over *m.heal*, a dummy for control (0) versus mental health (1) condition:

$$log_e(\frac{\pi_{ij}}{1-\pi_{ij}}) = \beta_0 + \beta_1 m.heal_i + u_j \tag{5.3}$$

where π_{ij} is the probability that $y_{FTR} = 1$ for the future time regression, and $y_{PTR} = 1$ for the past time regression, for observation i and redditor j. The random term u_j allows intercepts to vary by redditor and is assumed to be drawn from the normal distribution with a mean of zero and a standard deviation drawn from the data, i.e. $u_j \sim N(0, \sigma_u^2)$.

For mixed effects models like these, standard calculation of R^2 is complicated by the fact that there are multiple variance components explained by the model (i.e. fixed $\beta_0+\beta_1 m.heal_{ij}$ and random u_j). We therefore used the *MuMIn* package (Bartoń, 2020) to calculate pseudo $R^2_{marginal}$ and $R^2_{conditional}$ (Nakagawa, Johnson, & Schielzeth, 2017); $R^2_{marginal}$ represents the variance explained by the fixed effects, while $R^2_{conditional}$ represents the variance explained by the random effects + the fixed effects. For the FTR model, $R^2_m = 0.001,\ R^2_c = 0.019$ (where more than one method to calculate these quantities was available, we report the mean). This indicates that approximately 0.1% of the variance was explained by data-source condition. For the PTR model $R^2_m = 0.007,\ R^2_c = 0.05$, indicating approximately 0.7% of variance was explained by data-source condition.

We found that redditors in the mental health condition were significantly more likely to make FTR, $e^\beta = 1.23,\ SE = 0.02,\ z = 10.41,\ p < .001$, and PTR, $e^\beta = 1.58,\ SE = 0.03,\ z = 17.11,\ p < .001$ (βs are exponentiated so represent changes in odds ratios). This indicates that redditors who had posted popular contributions to *r/Anxiety* and *r/depression* had approximately 23% higher odds of referencing the future and approximately 58% higher odds of referencing the past, see *Fig. 5.1.*

5.2.2.3 Future and past time horizons

Next we investigated whether there were differences in temporal horizons derived from SUTime. We excluded observations with no temporal expressions according to SUTime. Analyses were then conducted with the $n = 35,248$ observations for which SUTime had identified a temporal expression. Recall, time horizons were calculated by subtracting the Reddit post date from the SUTime reference date, so negative values indicate PTR and positive values indicate FTR. We created a dichotomous "time reference" dummy. For positive temporal horizons (i.e. FTR) this was 0; for negative temporal horizons (PTR) this was 1. We then transformed time horizons into a single, positive measure by taking their absolute value. The resultant temporal horizons were highly right skewed ($skew = 13.44$), and truncated at zero. We therefore use the natural log of the absolute values of time horizon as our dependent

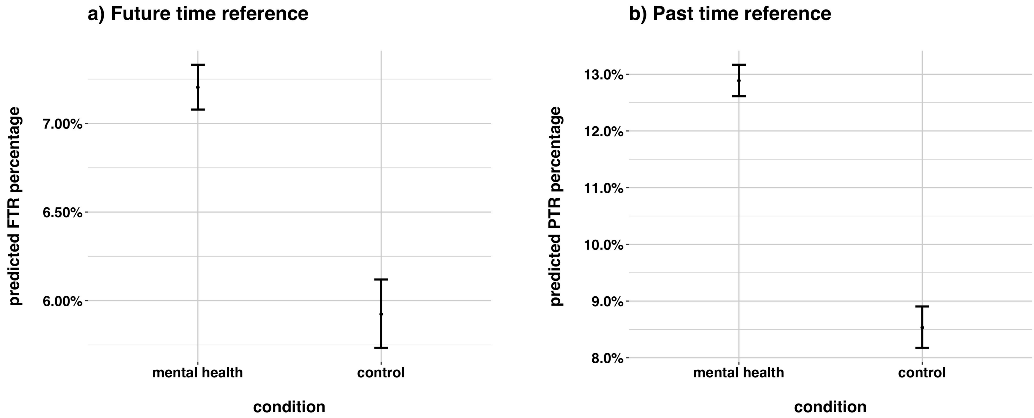


Figure 5.1: Differences in probability of making: (a) FTR and (b) PTR as a function of data-source condition in Study 1. Redditors in the mental health condition exhibited higher probability of referencing the past and future relative to control. Above and in *Figs. 5.2* and *5.4*, confidence intervals account for random effect variance by matrix-multiplying a predictor X by the parameter vector B to get the predictions, then extracting the variance-covariance matrix V of the parameters and computing XVX' to get the variance-covariance matrix of the predictions. The square-root of the diagonal of this matrix represents the standard errors of the predictions, which are then multiplied by ±1.96 (Lüdecke, 2019).

variable. We regressed this over condition, the time reference dummy, and the interaction between them (in case effects of condition differed for future and past time horizons):

$$log_e(|H_{ij}|) = \beta_0 + \beta_1 m.heal_i + \beta_2 t.ref_i + \beta_3 m.heal_i t.ref_i + u_j + e_i \tag{5.4}$$

where $t.ref$ is the time reference dummy; u_j are random intercepts for participant, as above; and e_i is an error term. Inspection of regression residuals indicated the log transformation resulted in a better approximation of regression assumptions of error normality than untransformed time horizons, so we proceeded. The fixed components of the model explained approximately 8% of the variance, $R^2_m = 0.08,\ R^2_c = 0.16$. We found redditors in the mental health condition had significantly shorter time horizons, $(e^\beta - 1) \times 100 = -24.42$, $SE = 0.08,\ t(2791.38) = -3.56,\ p < .001$ (βs are transformed as indicated so represent percentage change in y). This indicates that time horizons were approximately 24% shorter in the mental health condition, *Fig. 5.2*. There was also a significant effect of time reference. Compared with FTR, time horizons for PTR were significantly longer, $(e^\beta - 1) \times 100 = 400.28$, $SE = 0.07,\ t(35247.58) = 24.69,\ p < .001$. This indicates past time horizons were approximately 4 times longer than future ones, *Fig. 5.2*. The interaction term was not significant, $(e^\beta - 1) \times 100 = -1.98,\ SE = 0.07,\ t(35242.89) = -0.27,\ p = .786$. This indicates that there was no difference in the effect of data-source condition between PTR and FTR.

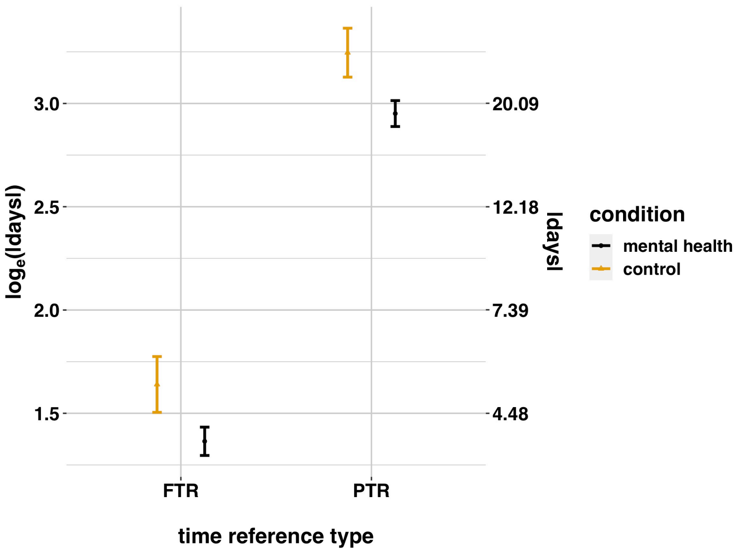


Figure 5.2: Differences in temporal horizons as a function of time reference type and data-source condition in Study 1. Redditors in the mental health condition exhibited contracted time horizons relative to control. Left y-axis is measured in the natural log of the absolute value of temporal distances (the dependent variable); right y-axis is the exponential transformation of the left y-axis, to yield the absolute value of time horizons (days).

5.2.2.4 FTR-type analysis

We next tested whether there were differences in how redditors referred to the future, i.e. in FTR type. We excluded all observations where $p_{FTR} \leq 50\%$ according to the time reference classifier. This left $n = 145,130$ observations. However, because of the keyword methods it employs, the FTR-type classifier cannot handle negations (e.g. *Rain tomorrow is not possible* expresses high certainty but would be classed as low-certainty because of the presence of the low-certainty keyword *possible*). We therefore detected the presence of negations using an averaged perceptron tagger following Collins (2002) but with Brown cluster features as described by Koo et al. (2008) and using greedy decoding (implemented in *spaCy* (Explosion AI, 2020)). We excluded any observations which used negations ($n = 24,399$), leaving a final sample of $n = 120,731$.

To give an overview, we present FTR-type usage proportions across the whole sample, *Fig. 5.3*. We found that, rather than the future tense marking the preponderance of FTR statements, FTR marking in English tends to be fairly equitably distributed across the different semantic categories estimated by the FTR-type classifier.

To investigate whether there were differences between the control and mental health conditions, we regressed each FTR types over data-source condition with a logit link function and random intercepts clustered by redditor, i.e. equation 5.3 for each FTR-type variable.

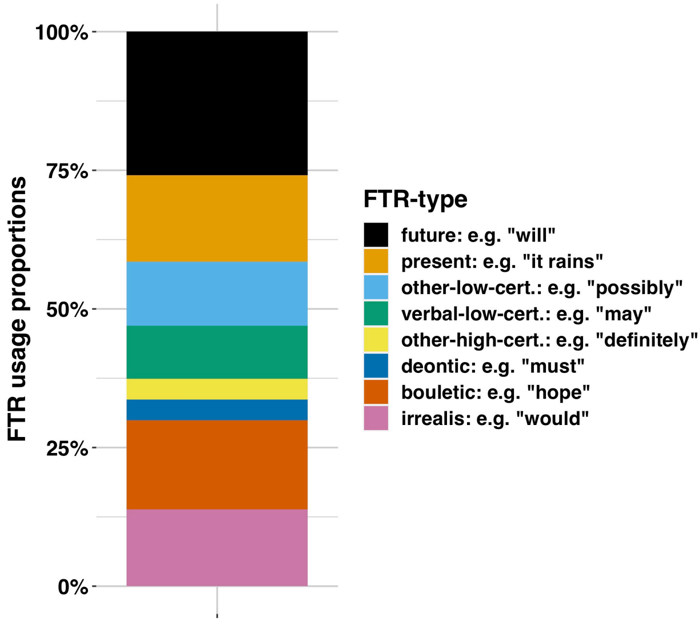


Figure 5.3: Plotted values are the proportional breakdown of FTR type in Study 1, i.e. the proportions per type normalised to sum to 1, $x_k = ((\sum_{i=1}^{n_k} x_{ik})/n_k)/\sum_{i=1}^{k} x_k$ for k FTR types. This suggests FTR is spread across multiple semantic domains, not just tense (results are similar even when modality does not dominate tense (see Appendix F).

Statistical results are reported in *Table 5.3* and depicted in *Fig. 5.4*. We found that redditors in the mental health condition were less likely to use the future tense, *Fig. 5.4a*, other-high-certainty constructions, *Fig. 5.4e*, and the present tense (for the present tense, this only approached significance, $p = .058$)[2], *Fig. 5.4b*. The semantic domain which all these FTR types encode is high certainty (see Chapter 2 and Giannakidou & Mari, 2018; Salkie, 2010). This suggests that a bias against speaking about the future in high-certainty terms among sufferers of anxiety and depression may be driving this effect. This conclusion is supported by the fact that we found redditors in the mental health condition to be less likely to use low-certainty terminology (whether other or verbal, *Fig. 5.4c–d*). They also exhibited higher use of deontic modal terms, which encode future obligations and necessities, *Fig. 5.4f*. Additionally, redditors in the mental health condition used more bouletic modal language, which relates to hopes, plans, and intentions for the future, *Fig. 5.4g*. We found no difference in use of irrealis terms, *Fig. 5.4h*.

[2] We note that setting α_{crit} to .05 is arbitrary and join researchers who take the view that there is not much substantive difference between p-values which closely approach $\alpha_{crit} = .05$ from either side (e.g. Thiese, Ronna, & Ott, 2016).

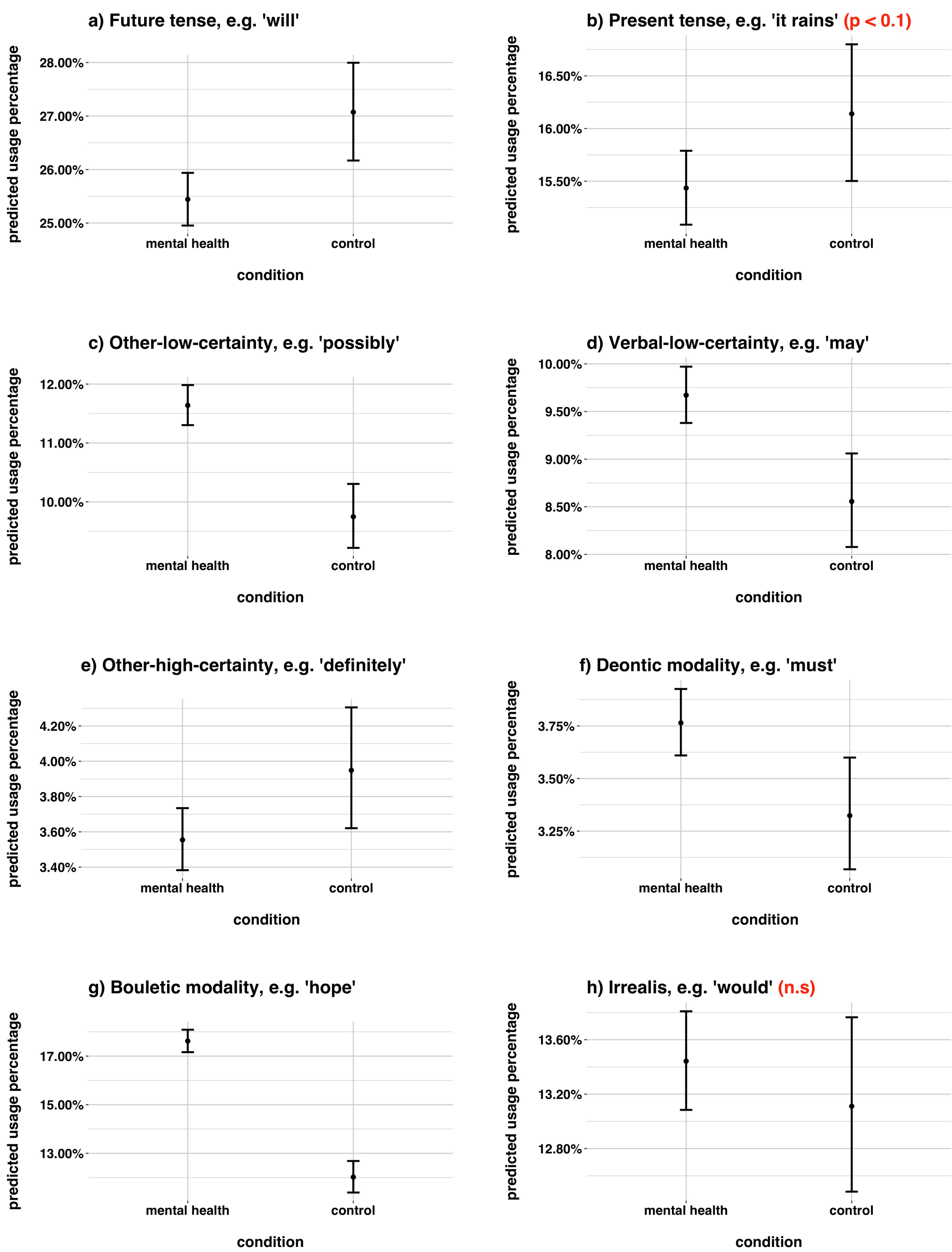


Figure 5.4: Differences in probability of using different FTR types when referring to the future as a function of data-source condition in Study 1. Redditors in the mental health condition exhibited higher probability of using low-certainty language, bouletic language, and verbal deontic language, and lower probability of using the future tense and other-high-certainty language.

Table 5.3

FTR type by condition on Reddit (Study 1)

outcome	predictor	e^{β}	SE	z-score	
future tense	(Intercept)	0.37	0.02	-41.96	***
$R^2_m < .001,\ R^2_c = .023$	mental health condition	0.92	0.03	-3.11	**
present tense	(Intercept)	0.19	0.02	-67.42	***
$R^2_m < .001,\ R^2_c = .016$	mental health condition	0.95	0.03	-1.89	·
other-low-certainty	(Intercept)	0.11	0.03	-70.7	***
$R^2_m = .001,\ R^2_c = .028$	mental health condition	1.22	0.04	5.56	***
verbal-low-certainty	(Intercept)	0.09	0.03	-74.01	***
$R^2_m < .001,\ R^2_c = .025$	mental health condition	1.14	0.04	3.71	***
other-high-certainty	(Intercept)	0.04	0.05	-69.4	***
$R^2_m < .001,\ R^2_c = .042$	mental health condition	0.9	0.05	-2.07	*
deontic	(Intercept)	0.03	0.04	-80.02	***
$R^2_m < .001,\ R^2_c = .021$	mental health condition	1.14	0.05	2.71	**
bouletic modality	(Intercept)	0.14	0.03	-63.61	***
$R^2_m = .007,\ R^2_c = .041$	mental health condition	1.57	0.04	12.72	***
irrealis	(Intercept)	0.15	0.03	-65.91	***
$R^2_m < .001,\ R^2_c = .025$	mental health condition	1.03	0.03	0.88	

Coefficients represent the change in odds ratio of each predictor compared to the intercept.
$^{***}p < .001$; $^{**}p < .01$; $^{*}p < .05$; $\cdot p < .1$

5.2.3 Discussion

Taken together, the Study 1 results present an interesting pattern. We found that when making FTR, redditors in the mental health condition were more likely to use low-certainty terminology, bouletic terminology, and deontic terminology, and less likely to use the future tense, the present tense (p = .0.58), and high-certainty terminology. The future tense (Giannakidou & Mari, 2018; Salkie, 2010), the present tense (Giannakidou & Mari, 2018; Salkie, 2010), and other-high-certainty terms all encode notions of modal high certainty. The finding that redditors in the mental health condition used more of these high-certainty FTR types, and fewer low-certainty FTR types, suggests that the experience of anxiety and/or depression may be characterised by increased uncertainty about future events. Additionally, higher use of bouletic terminology suggests increased desire for future change.

We found redditors in the mental health condition were more likely to make past and future time reference and their time horizons were shortened relative to control. This suggests that anxiety and depression may be characterised by heightened salience of future and past events, but within a shorter temporal window.

We found no differences between the anxiety and depression conditions. We reasoned that self-selecting to post in *r/Anxiety* and *r/depression* was a weak diagnostic criteria, especially given high co-morbidity rates between the two mood disorders (Groen et al., 2020). We therefore undertook to differentiate the relationships between anxiety, depression, and temporal horizons in a survey paradigm over which we could exert a greater degree of control.

5.3 Study 2

Our goals for Study 2 were twofold. First, we wanted to understand more about representations of temporal distance and how this might relate to temporal discounting in sufferers of anxiety and depression. We hypothesised that shortened temporal horizons in sufferers of anxiety and/or depression might be driven by temporally distal representations future events. If sufferers of anxiety and depression tend to talk about temporally proximal time frames on Reddit, this might be because they represent such dates as subjectively more distal. Distal representations of subjective temporal distance is a predictor of increased temporal discounting (Thorstad et al., 2015; B. K. Kim & Zauberman, 2009; Zauberman et al., 2009). This means such perceptual biases might be driving increased temporal discounting in sufferers of anxiety. We therefore hypothesised that sufferers of anxiety and/or depression would represent future events more distally than healthy individuals, which would in turn give rise to increased temporal discounting. Second, we aimed to disambiguate whether such cognitive biases were characteristic of anxiety, depression, or both.

5.3.1 Methods

5.3.1.1 Participants

A sample of $N = 202$ participants ($n = 103$ females, $n = 98$ males, and $n = 1$ other) passed attention checks to complete the study. Data were collected in March 2020. All participants were English speakers residing in the United Kingdom and were at least 18 years old. Participants were recruited from Amazon Mechanical Turk Prime and completed the survey online. The study was approved by the internal review board of Brunel University London (Ref. 22348-MHR-Jan/2020- 24407-1). Participants were paid the UK living wage on a pro-rata basis.

5.3.1.2 Materials

Study 2 consisted of 3 tasks: (1) an intertemporal-choice task designed to establish temporal discounting; (2) a "time-slider" task designed to establish subjective representations of future temporal distance; and (3) the Depression Anxiety and Stress Scales 21 item measure (DASS-21) which measures depression, anxiety, and stress levels (Lovibond & Lovibond, 1995).

The intertemporal-choice task: In this task, participants made repeated binary decisions between a larger reward (the "larger-later reward") offered after a delay and a smaller reward (the "smaller-sooner reward") offered immediately; for instance, "Would you prefer £50 now of £100 in six months?" The smaller-sooner reward was always smaller than the larger-later reward, which was fixed at £100. The factors of this task were the amounts of the smaller-sooner reward (£50–£95 by increments of £5), and the delays of the larger-later reward (later today, tomorrow, one week, one month, two months, three months, six months, one year, two years, five years, and ten years). If participants chose the larger-later reward, this was scored with (1), otherwise (0). Amounts and delays were fully crossed to produce a test battery of $10_{amounts} \times 11_{delays} = 110_{items}$. Prior to starting, participants were told to "Try to answer quickly and intuitively, without thinking about it too much." Item order was randomised and one item was displayed per page.

The following attention check was implemented. At two random points during the task, participants were given a choice between two present-time rewards of different values. If they chose the smaller reward, they were ejected from the survey immediately.

To be able to model linear relationships between discounting (which is non-linear over time (L. Green & Myerson, 2004)) and participant-level variables (anxiety, depression), we calculated a participant-level measure of discounting. Research has shown that the following hyperbolic function fits real and hypothetical delay discounted value well (Mazur, 1987; Kirby et al., 1999):

$$V = \frac{A}{1 + kD} \tag{5.5}$$

where V is the subjective value of a delayed reward A at a given delay D, and k is a scaling parameter which captures individual differences in discounting. Higher values of k imply lower V, i.e. more discounting. To derive participant-level ks, we followed Kirby et al. (1999). This involved calculating hypothetical values of k and retaining for each participant some k which best predicted empirical choices. Specifically, we calculated all k_{1-n} at indifference between LLR and SSR for each D and SSR, i.e. all ks such that $SSR = LLR/(1 + k_i D)$ for all values of SSR and D under the study. We then predicted hypothetical intertemporal choices for each k_{1-n} and retained k_i for each participant p_j which had the highest proportion of matches against p_j empirical choices. When more than one k had an equal number of matches, we took the geometric mean (Kirby et al., 1999). The resultant k values correctly predicted participants empirical choices for 93.5% of observations. Since k tends to be non-normally distributed, we used $log_e(k)$ as our participant-level measure of discounting (Kirby et al., 1999). This is the dependent variable for this task.

The time-slider task: In this task, participants used a slider to rate whether they construed a given objective temporal distance between "close to now" (0) and "far from now" (100). To increase a sense of subjectivity, numbered slider intervals were not displayed. This was based on similar methods in Zauberman et al. (2009) and B. K. Kim and Zauberman (2009). Rather than time horizons—how far into the future participants tend to think—the time-slider task established how subjectively distant participants felt given temporal distances to be. The only factor was the objective temporal distances. These matched the intertemporal-choice task. For each item, participants were directed to "indicate with the slider how far away from NOW the given time feels to you." Prior to starting, participants were given a training example involving a past time reference (9 months ago), and were told to "try to answer quickly and intuitively, without thinking about it too much." Item order was randomised and one item was displayed per page.

The Depression, Anxiety, and Stress Scale: To establish depression and anxiety levels, participants completed DASS-21 (Lovibond & Lovibond, 1995). In the DASS-21, participants self-reported using a Likert scale what extent a given statement applied to them. The Likert scale options were "did not apply to me at all" (0), "applied to me to some degree, or some of the time" (1), "applied to me to a considerable degree, or a good part of time" (2), and "applied to me very much or most of the time" (3). For each item, participants were instructed to "Please read each statement and use the scale provided to indicate how much the statement applied to you over THE PAST WEEK. There are no right or wrong answers. Do not spend too much time on any statement." The DASS-21 is broken down into 3 dimensions, each comprising 7 items, designed to separately measure depression, anxiety, and stress. Example items from each dimension are as follows: "I felt I wasn't worth much as a person" (depression), "I felt I was close to panic" (anxiety), and "I found it difficult to relax" (stress). See Appendix F for items. The DASS-21 is not a diagnostic tool *vis-à-vis* the discrete diagnostic categories of, for instance, the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2013). Rather, it is based on research which suggests that the differences in depression, anxiety, and stress experienced by healthy and clinical populations are a matter of degree (Lovibond & Lovibond, 1995). It is therefore designed to span the boundaries between clinical and healthy populations. Dependent variables are the sum of scores for each dimension multiplied by 2, and therefore theoretically range between (0) and (42).

5.3.1.3 Procedure

The survey was hosted on Qualtrics. It had a within-subjects correlational design, meaning there were no independent variables or factors other than the within-subjects factors described above for each task. Participants provided informed consent; indicated if they were male, female, or other; confirmed they were 18 or over; and confirmed they were native English

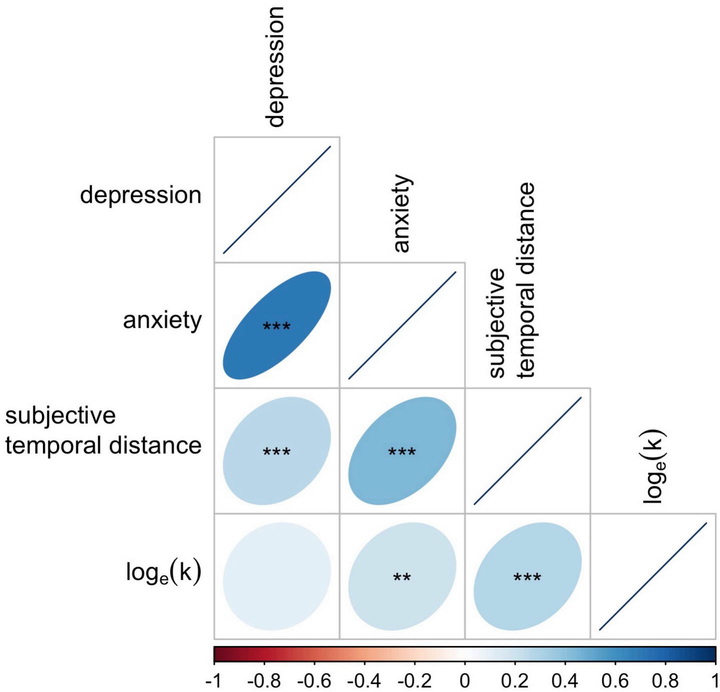


Figure 5.5: Bivariate Pearson correlations for Study 2.
$^{***}p < .001$; $^{**}p < .01$; $^{*}p < .05$; $\cdot p < .1$

speakers. Participants who indicated they were under 18 or not native English speakers were ejected from the survey immediately. Participants completed the intertemporal-choice task, the time-slider task, and then the DASS-21.

5.3.2 Results

Descriptive statistics for Study 2 are presented in *Table 5.2*. Correlations are presented in *Fig. 5.5*. Anxiety and depression scores were positively correlated, which may reflect widely noted co-morbidity between the two mood disorders (Groen et al., 2020). Both anxiety and depression scores were positively correlated with subjective-temporal-distance ratings. This suggests sufferers of anxiety and depression represent future events more distally. Subjective delay ratings were positively correlated with $log_e(k)$ scores. This indicates that participants who represented future outcomes as more distal also exhibited increased temporal discounting.

To understand the conditional relationship of anxiety and depression on discounting via temporal-distance ratings, we conducted a mediation analysis. In mediation analyses, a predictor X is assumed to effect an outcome Y via a mediating variable M. For instance, vitamin C levels mediate the relationship between eating citrus fruit and the development of scurvy (Pearl & Mackenzie, 2018). Mediation analyses can be conducted by combining two regressions into a single model: an $X \rightarrow M$ regression and as $X + M \rightarrow Y$ regression. In mediation terminology, the $X \rightarrow M$ path is refereed to as α, the $M \rightarrow Y$ path is referred to as β, and the $X \rightarrow Y$ path is referred to as τ'. The "indirect effect", $X \rightarrow M \rightarrow Y$, can be estimated as the product of the paths involved, $\alpha\beta$, and captures the effect of X on Y via M while controlling for the "direct effect" τ'. Similarly, the direct effect captures the effect of X on Y while controlling for the indirect effect. The "total effect" is the sum of $\alpha\beta$ and τ', and captures the total effect of X on Y, as in a normal regression (Yuan & MacKinnon, 2009).

We used $log_e(k)$ as our outcome variable (Y), anxiety and depression scores from the DASS-21 as our predictor variables ($X_{1,2}$), and participant-level mean subjective temporal-distance scores as our mediating variable (M). The model therefore took the following form:

$$\begin{aligned} subj.dist_i &= \lambda_1 + \alpha_1 anx_i + \alpha_2 dep_i + e_{1i} \\ log_e(k)_i &= \lambda_2 + \tau_1' anx_i + \tau_2' dep_i + \beta subj.dist_i + e_{2i} \end{aligned} \tag{5.6}$$

where $\lambda_{1,2}$ are intercepts, $\alpha_{1,2}$ are slope coefficients for the effects of anxiety and depression on mean subjective temporal distance, $\tau_{1,2}'$ and β are slope coefficients for the effects of anxiety, depression, and mean subjective distance on $log_e(k)$ temporal discounting, and $e_{1,2}$ are error terms. See *Fig. 5.6b* for conceptual diagram.

Bayesian statistics are well-suited to mediation analyses. As well as making no assumptions about sampling statistic normality, they allow for straight-forward inferences about any transformation of model parameters (i.e. path products) through simply carrying out the desired operation on posterior probability distributions (Vehtari et al., 2019). We therefore used the *brms* package (Bürkner, 2017) to derive posterior probability distributions for model parameters using a no U-Turn Hamiltonian Monte Carlo sampling procedure (Hoffman & Gelman, 2014; Stan Development Team, 2020) with uninformative priors (i.e. $Unif(-\infty\text{-}\infty)$), 4 chains, $2,000$ iterations, and a burn in discard of $1,000$ (the defaults). Inspection of caterpillar plots and $\hat{R}$s indicated that the estimation procedure had converged (Vehtari et al., 2019).

The predictors explained moderate proportions of the variance, $R^2_{subj.temp.dist} = 0.34$, $CI_{95\%} = [0.26, 0.41]$, $R^2_{log_e(k)} = 0.29$, $CI_{95\%} = [0.21, 0.36]$. These are a Bayesian adaptation of R^2 following (Gelman et al., 2018), including spread statistics. They indicate that anxiety and depression explained approximately 20% of the variance in subjective temporal distance; and anxiety, depression, and subjective temporal distance explained approximately 10% of the variance in temporal discounting ($log_e(k)$ scores).

Results are reported in *Fig. 5.6.* We had made the directional hypotheses that anxiety and/or depression would be associated with more distal subjective distances and therefore increased discounting—higher $log_e(k)$. We tested the one-tailed hypotheses that these effects would be positive. As we had predicted for anxiety, we found a significant positive indirect,

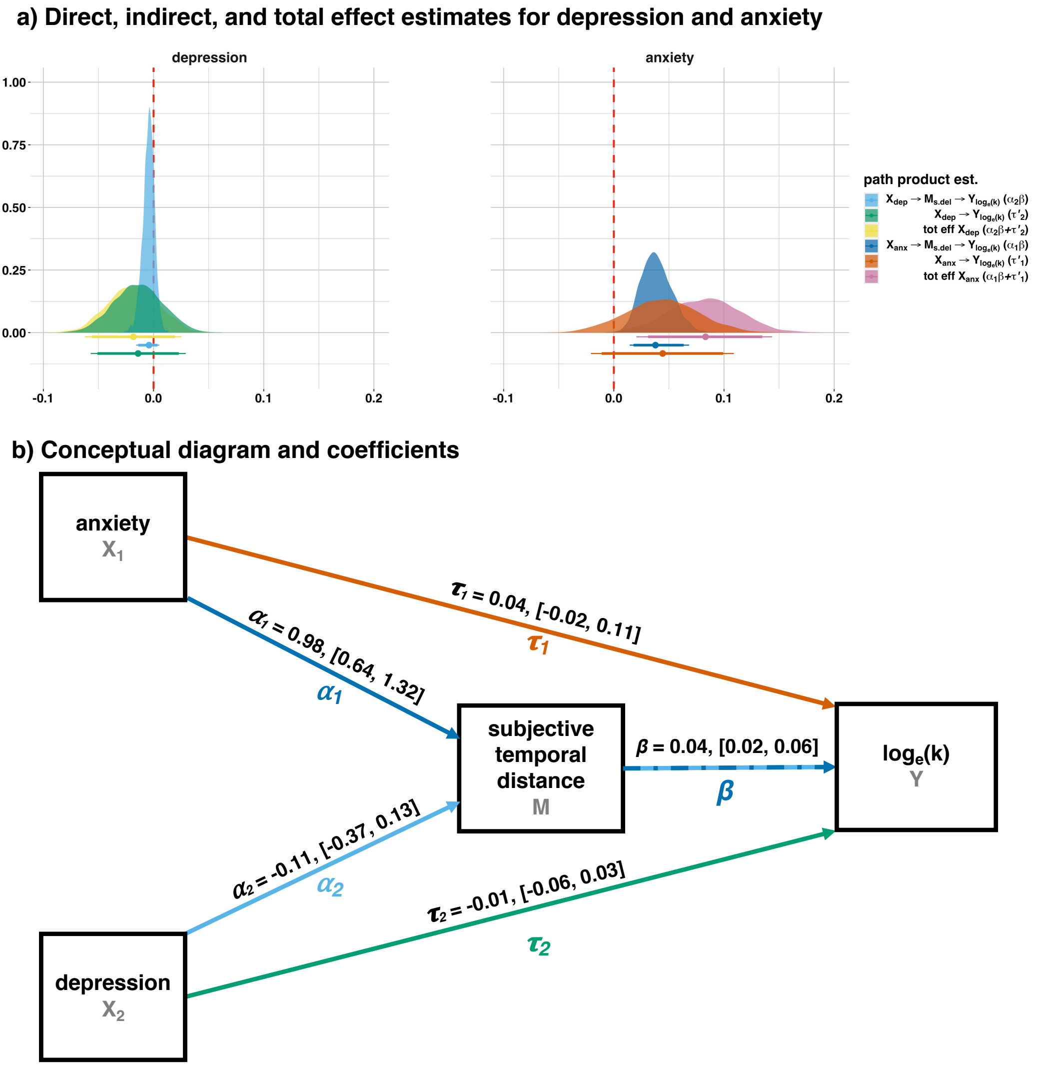


Figure 5.6: Study 2 mediation results. a) Posterior estimates for paths and path products for the effects of anxiety and depression on discounting ($log_e(k)$) via participant-level mean ratings of subjective future temporal distance. Bayesian Credibility Intervals (CI) are plotted at 90% (thick bars) and 95% (thin bars). For two-tailed hypotheses, thin bars represent cut-off criteria. For one-tailed hypotheses, thick bars do (providing sign is as predicted). b) Conceptual diagram, with parameter point estimates and 95% CIs. If the CI does not contain zero, parameters are significantly different from zero. The values for the intercepts are, $\lambda_{1:subj.dist.} = 28.25,\ [26.10, 30.42]$, and $\lambda_{2:log_e(k)} = -4.41, [-5.16, -3.64]$.

$Est. = 0.04,\ CI_{90\%} = [0.02, 0.07],\ pp > .999$[3], and total effect, $Est. = 0.11,\ CI_{90\%} = [0.06, 0.17],\ pp > .999$, see *Fig. 5.6a*. The direct effect of anxiety was not significant, $Est. = 0.07,\ CI_{90\%} = [0.02, 0.13],\ pp = .982$. This indicates that participants who self-reported as

[3]In these parameter summaries, "*pp*" reports the probability that the posterior estimate of the effect matches the direction of the hypothesis. In contrast to frequentist *p* values, $pp > .95$ indicates evidence for the parameter matching the prediction exceeds 95% (Bürkner, 2017).

high-anxiety also had more distal representations of future events. As a consequence, they devalued future outcomes to a greater extent. The non-significant direct effect indicates that the entirety of the effect of anxiety on temporal discounting was transmitted via construals of temporal distance, *Fig. 5.6*.

On the other hand, we found no significant effects for depression (indirect: $Est. = -0.01$, $CI_{90\%} = [-0.02, 0]$, $pp = .158$, direct: $Est. = -0.02$, $CI_{90\%} = [-0.06, 0.01]$, $pp = .146$, total: $Est. = -0.03$, $CI_{90\%} = [-0.07, 0.01]$, $pp = .103$, see *Fig. 5.6a*). This indicates that, when controlling for anxiety, depression scores had no impact on temporal distance or temporal discounting, *Fig. 5.6*, despite bi-variate correlations between depression and temporal-distance ratings, *Fig. 5.5*.

5.3.3 Discussion

These findings suggest that anxious participants tended to construe future events more distally, and that this in turn caused them to temporally discount to a greater degree. The fact that the direct effect of anxiety was not significant indicates the entirety of the effect of anxiety on temporal discounting was mediated by subjective-temporal-distance ratings. This suggests reported associations between anxiety and temporal discounting are mediated by construals of future temporal distance. On the other hand, we found no significant relationships between depression and subjective temporal distance or temporal discounting.

5.4 General discussion

In Study 1, we found that language used by redditors who had posted to *r/Anxiety* and *r/depression* differed significantly from control. Firstly, we found they were more likely to make non-present time reference, both past and future. This suggests that sufferers of anxiety and/or depression may spend more time thinking about future and past events rather than present ones. It may be important to note that we did not build a tri-class past/present/future time reference classifier, so the reference category in these analyses is not present time. Rather it is anything that is not future or past time reference, for instance statements of ability *It can rain a lot here*; statements of fact, *The sun is hot*; or present time reference, *I am hot*. In light of this, it would be incorrect to infer that redditors in the mental health condition spent less time making present time reference. Rather, they spent more time making past and future time reference.

While relationships between future time perspective and anxiety and depression have been established, such research does not illuminate the extent to which the future is salient to sufferers from these mood disorders. As far as we know, this is the first study to find that the experience of anxiety and depression may be characterised by an increased amount of time spent talking about the future and the past.

We also found significant (though small) differences in FTR type. This suggests that habits of thinking about the future may differ between healthy populations and sufferers of anxiety and/or depression. The largest differences seemed to be driven by lower use of the future tense and higher use of bouletic language relative to control, *Fig. 5.4*. The semantic domain which bouletic modal language encodes is desiderative, having to do with hopes, desires, and plans for change. This suggests that future thinking in sufferers of anxiety and/or depression is characterised by desire for future change rather than simply statements of fact.

We found higher use of low-certainty language relative to control. This suggests that sufferers of anxiety and/or depression may construe the future more uncertainly then healthy populations. This raises the possibility that probability discounting may be a relevant marker of anxiety and/or depression. Probability discounting is similar to delay discounting, but involves the devaluation of risky rather than delayed rewards (L. Green & Myerson, 2004). It is measured in a similar way, by giving participants binary choices, but rather than immediate and delayed, the choice is between certain and risky rewards; for instance a 100% chance of receiving $10 vs. a 50% chance of receiving $20. People who probabilistically discount more would tend to choose high-certainty options because they devalue low-certainty ones to a greater degree. Recent research involving rewards which are both probabilistic and delayed, e.g. a 100% chance of receiving $10 vs. a 50% chance of receiving $20 in three months, has found that these factors interactively predict discounting rates (Vanderveldt et al., 2015). This suggests that real-world intertemporal decision making (in which people must balance present vs. future options as in delay discounting paradigms) which usually involves some degree of uncertainty about the future may be driven by both probabilistic and temporal discounting rates.

The relationship between probability discounting and anxiety and/or depression is understudied relative to temporal discounting. While one study found probability discounting was not a significant marker of major depressive disorder (Hart et al., 2019), our results suggest that anxiety and/or depression may be characterised by low-certainty construals of future events. This may warrant further investigation into potential relationships between anxiety, depression, and low-certainty future construals and their role in risky intertemporal decision-making processes.

Additionally, we found higher use of deontic modal language relative to control. The primary deontic modal verb in English is *must*, and the other criterion keywords in the FTR type classifier in this category are *have to*, *need to*, and variants thereof. This suggests that redditors in the mental health condition were more likely to encode *obligation* and *necessity*. This echoes the finding that people who experience mental health distress overuse “absolute words”. Specifically, words with totalising predication, for instance *absolutely*, *always*, *completely*, *constant*, *definitely*, *entire*, *every*, *everyone*, *everything*, *full*, *must*, *never*, *nothing*, *totally* and *whole* (Al-Mosaiwi & Johnstone, 2018). Though these words encode more than deontic modal language, there is some overlap between the predicate function of these lexical expressions of absoluteness, and the totalising nature of strong modal expressions of obligation (*need to, have to, must*) and conditional necessity (*If X, must Y*).

It is unclear whether these patterns are causal or descriptive. In other words, habitually construing future events in low-certainty, highly-constrained, bouletic terms may contribute to the development and maintenance of anxiety and/or depression. On the other hand, such FTR usage differences may simply be symptomatic of anxiety and/or depression. In either case, we appear to have isolated new markers of these mood disorders vis-à-vis FTR.

Study 1 found contracted time horizons in the mental health condition relative to control. In other words, redditors in the mental health condition referenced more proximal future and past times. In Study 2, we used a different measure of subjective temporal distance. Grasping this difference is crucial. The time horizon measured in Study 1 approximated how far into the future people tended to talk about. The subjective-temporal-distance measure in Study 2 gave participants objective future dates and asked them how far away they construed those dates to be. We found that anxious participants tended to rate future events as more temporally distal. This suggests that subjective representations of temporal distance (i.e. how far away DATE feels to an individual), and time horizons (i.e. how far into the future an individuals tends to think) may be negatively correlated. More research using time slider tasks and the Wallace task could explore this further.

In Study 2, we found subjective-temporal-distance ratings mediated a relationship between anxiety and temporal discounting. This suggests that subjective representations of future time may be partially driving reported links between anxiety and temporal discounting (Rounds et al., 2007), and adds evidence to the inference that the relationship is significant and positive (contra Jenks & Lawyer, 2015; Steinglass et al., 2017). On the other hand, we found no relationship between our measures of discounting and depression, despite the fact that depression has been previously linked to contracted future time horizons (Dilling & Rabin, 1967), and increased temporal discounting (Felton et al., 2020; Pulcu et al., 2014). This may be driven by symptom severity and/or reward size effects. Reward size has a well-established effect on discounting rates; larger future rewards are discounted less, i.e. people are more willing to wait when future rewards are large even when present rewards are proportionally identical (L. Green & Myerson, 2004). Different reward sizes may therefore be responsible for differences in reported findings. Further, there is evidence that reward size affects temporal discounting in depressive individuals differently to healthy ones. Pulcu et al. (2014) compared discounting between clinically diagnosed present sufferers of Major Depressive Disorder (MDD), remitted MMD (rMDD), and healthy controls, with small (£30), medium (£55) and large (£80) rewards. They found no significant differences between rMDD participants and controls, and only found significant differences between MDD and controls for large rewards. Since we used a reward size comparable to their large condition (£100), this suggests that depressive symptom severity in our sample was not high enough to detect an effect. It is an important point that depression was significantly correlated with subjective-temporal-distance scores, *Fig. 5.5*. This suggests that—had we not also measured anxiety—depression would have had an impact on temporal discounting via subjective temporal distance. However, when anxiety was also included as a predictor in the mediation analysis, there was no significant effect of depression *Fig. 5.6*. Since the two mood disorders are often co-morbid (Groen et al., 2020), our results suggest including measures of both is important when attempting to understand the relationship between either and time perspective.

5.4.1 A functionalist perspective on the results

Functionalist approaches to mental health and wellness offer a theoretical framework which may help make sense of our findings. A functionalist perspective simply asks "what is the purpose of a mental state?" The relevant mental states here are those associated with anxiety and depression. Purpose is understood in terms of the relationship between an organism and its environment. In this view, mental states and their suites of behavioural correlates are seen to help organisms survive in stochastic but statistically predictable environments (see Brunswik, 1952, 1956). A functionalist approach to the understanding and treatment of mental health therefore assumes there is a utility (a function) to common patterns of thought and feeling—even those which seem exaggerated in people who suffer from mental health problems.

This is contrasted with the descriptive approach taken in, for instance, the DSM-5, which seeks to describe the commonly-occurring attributes of the various mental illnesses (American Psychiatric Association, 2013). From the taxonomic perspective of the DSM-5, anxiety and depression look profoundly different. In anxious states, individuals experience heightened autonomic arousal (Rozenman, Sturm, McCracken, & Piacentini, 2017). Systolic blood pressure may rise (Gerra et al., 2000), and heart rate may rise (Rozenman et al., 2017), or exhibit suppressed variability (Rozenman et al., 2017), and breathing rate may be more variable (Wilhelm, Trabert, & Roth, 2001). Patients tend to describe recursive loops of rapid negative thoughts, or constant ideation of potential catastrophes. Emotions may be dysregulated, with powerful experiences of fright and frustration often occurring. The mind and body are poised for action, and are vigilant for potential dangers. The phenomenology of depressive states is markedly different. Individuals experience reduced autonomic arousal (Volkers et al., 2003). Endogenous opioids (endorphins) rise (Al-Fadhel, Al-Hakeim, Al-Dujaili, & Maes, 2019), and motor activity is reduced (Volkers et al., 2003). Rather than a racing mind, patients report their thinking becomes slow and foggy. Emotions feel dulled, if not completely numbed. Physical energy is low, and patients describe feeling as though they are in a state of lassitude and torpor.

Despite these marked differences, a functionalist approach takes the view that anxiety and depression are similar because they are both strategies which sufferers employ to deal with perceived "threats" to their safety and/or well-being. A useful way of understanding what is meant by "threat" is the concept of homeostatic salience. This refers to elements of the environment which are experienced as relevant to immediate well-being (Suvak & Barrett, 2011). For instance, ensuring the provision of one's basic physical needs has clear homeostatic relevance. Even in a simple physical example, there are individual differences in what may be construed as a threat. Someone living in an air-conditioned building connected to a robust power grid would be unaffected by an approaching heat wave, but the same phenomenon would have high homeostatic salience to someone living without such amenities. However, modern approaches to clinical psychology understand more complicated feelings of belonging and attachment can also have high homeostatic salience (at least since Bowlby, 1969). In this view, there are a vast array of ways people feel they might be harmed. Clinicians working from a functionalist perspective are therefore inclined to see anxiety and depression as useful, if different, strategies for dealing with threats to homeostatic regulation. Put another way, functional clinicians try to understand the, possibly idiosyncratic, ways in which sufferers of anxiety and depression construe their environment as threatening. This implies that in most cases of anxiety and depression, clinicians can expect to find something which the sufferer perceives as having homeostatic salience.

This line of thinking clarifies the present results in several ways. First, redditors in the mental health condition used more bouletic FTR language, which involves the expression of desire and intention. This suggests a neutral, objective, or uninvested viewpoint may be inaccessible. If anxiety and depression are coping strategies to deal with homeostatic threats, it makes sense that sufferers would be more invested in future outcomes. They would therefore use more language which expresses desire for specific eventualities. Secondly, redditors in the mental health condition exhibited contracted time horizons relative to controls. In threatening environments, distal time frames lose salience. To someone who is hypothermic, nothing is important relative to the immediate provision of warmth. The need to get warm constricts which possibilities matter, as well as the timeline on which they matter. In navigating an environment which is construed as threatening, sufferers of anxiety and depression may disregard distal events which seem irrelevant compared with more proximal ones. Third, redditors in the mental health condition used more low-certainty FTR constructions. Uncertainty and threat are closely related. Mild hypothermia may be disregarded if a warm house is minutes away. On the other hand, uncertainty about environmental affordances seems inherently threatening. Increased low-certainty language use may therefore either: (a) reflect threatening construals about which sufferers are uncertain; or (b) contribute to low-certainty construals which heighten feelings of threat.

The experience of agency may also be related to modal notional categories important to the expression of future time reference. Patients suffering from anxiety and depression often report their sense of agency is constricted by self doubt, feelings of helplessness, or resignation to the inefficacy of action. This may interact with English grammar to produce higher proportions of low-certainty FTR construction in sufferers of anxiety and depression. There are lots of ways people can talk about the future, and different reasons for doing so. A widely-used framework is to break down these reasons into schedule-based, prediction-based, and intention-based FTR (Dahl, 1985, 2000a). In schedule-based FTR people talk about well-known scheduled events, e.g. *The train arrives at 6*; in intention-based FTR people talk about their own intentions and plans for future action, e.g. *I'm coming over later*; and in prediction-based FTR people make predictions about less knowable events, e.g. *The Cubs can probably win tomorrow.*[4] These categories index modal notions of certainty and uncertainty. For instance, most intention-based FTR is higher certainty than prediction-based FTR. English FTR grammar is sensitive to these distinctions. Critically, it is obligatory in English to use either a future tense construction or a modal verb for prediction-based FTR, e.g. *That coin could/may/might/should/would/shall/will/is going to land on heads.* With the exception of the future tense constructions, these generally express low-certainty modality. In contrast to this, modal/tense marking may be eschewed in schedule-based, and some intention-based FTR statements. For instance, *The show is at 5* (schedule), or *I'm leaving her after Easter* (intention) are acceptable, while *?The coin is landing/?lands on heads tomorrow* (prediction) is not.[5]

The point here is that the tendency to form linguistic utterances about one's own intentions may be bound up with feelings of agency. If sufferers of anxiety and depression feel a lack of agency, this may leave them more dependent on external loci of control. They may therefore feel at the mercy of unpredictable, capricious, and uncertain forces, and be more likely to form prediction-based FTR statements. This could interact with English grammar and result in higher use of low-certainty language. Conversely, if sufferers of anxiety and depression experience reduced agency, this may imply low-certainty representations of future events and lead directly to increased use of low-certainty modal terms.

5.4.2 Construal level theory: Psychological distance, abstraction, and FTR

A second theoretical perspective might also be brought to bear on the results. Construal Level Theory (CLT) posits that "distant" events are represented in similar ways across different modalities, including space, time, probability, and social familiarity. The general term for cross-modal distance is "psychological distance". CLT claims there is a relationship between psychological distance, and concreteness and abstraction. In this context, "concreteness" refers to things that can be physically perceived (touched, seen, etc.), and "abstraction" refers

[4] On the basis of the Study 2 results, we might add desire-based FTR to this schema, given its high proportion of total FTR statements.

[5] Unacceptable language usage is marked with a preceding question mark.

to things which cannot. CLT claims that psychologically distal events are represented more abstractly, and psychologically proximal events are represented more concretely. For instance, socially distal (an alien compared with a brother), spatially distal (Mars compared with London), probabilistically distal (winning the lottery compared with eating breakfast tomorrow), and temporally distal (the Stone Age compared with yesterday) events are represented less in terms of concrete sense modalities and more in terms of abstract schemas (Trope & Liberman, 2010). Reciprocally, it is hypothesised that abstraction cues expectations of psychological distance, and concreteness expectations of psychological proximity (Trope & Liberman, 2010; for critical perspectives, see Trautmann, 2019). The value of CLT is that its claims provide an empirically supported link between psychological distance and areas of cognition as diverse as moral orientation (Eyal & Liberman, 2012), self-perception (Liberman & Förster, 2009), emotionality (Woltin et al., 2011), tolerance (Luguri, Napier, & Dovidio, 2012), conformity (Ledgerwood & Callahan, 2012), creativity (Ledgerwood & Callahan, 2012), and aesthetic response (Durkin, Hartnett, Shohamy, & Kandel, 2020).

Crucially, many outcomes in the present study involved representations hypothesised by CLT to be represented in terms of psychological distance. Specifically, modal notions of high and low certainty may map onto probability. For instance, strong modals such as *must* and *will* indicate high-probability outcomes ($p \to 1$), while weak modals such as *could, may* and *might* indicate indeterminate probability outcomes ($p \approx .5$) (see Chapter 2; and on connections between modality and scalar notions of probability, see Lassiter, 2015; Moss, 2015; Santorio & Romoli, 2017). Additionally, future and past time reference are generally temporally distant compared with present time reference, and time horizons are a concrete measure of temporal distance.

There is evidence that depression, especially, is associated with overly-abstract representations (Gjelsvik, Lovric, & Williams, 2018; Ratcliffe, 2015), and can be ameliorated by exposure to concrete sensory stimuli (Watkins et al., 2012; Werner-Seidler & Moulds, 2012; Takano & Tanno, 2010). Comparable claims have also been made for anxiety (White & Wild, 2016; Goldwin & Behar, 2012). If so, individuals with anxiety and depression may experience the world in a manner that makes it difficult to resolve mental representations into concrete forms, or may give disproportionate attention to abstract domains. If there is a relationship between abstraction and psychological distance, this might explain our results.

Redditors in mental health condition were more likely to make non-present time reference compared with controls. If sufferers of anxiety and depression represent experience more abstractly, CLT predicts this will result in more frequent non-present time reference, because non-present time reference is predicted to involve more abstraction. In fact, this claim is supported by the finding that references to the past and future involved more abstract language than references to the present in a sample of blog posts (Thorstad & Wolff, 2019b).

Additionally, redditors in the mental health condition exhibited shortened time horizons relative to controls. This seems out of keeping with the CLT prediction that high abstraction should result in reference to more distal time frames. However, a crucial point is that we found higher use of low-certainty language in the mental health condition. Both anxiety and depression are associated with the experience of uncertainty, i.e. all possible outcomes may be represented as equally likely and are thus unpredictable (Hirsh, Mar, & Peterson, 2012; Davis, Neta, Kim, Moran, & Whalen, 2016; Grillon, Baas, Lissek, Smith, & Milstein, 2004; Grupe & Nitschke, 2013; Ouimet, Gawronski, & Dozois, 2009; Saulnier, Allan, Raines, & Schmidt, 2019; Carleton et al., 2012). In our data, sufferers of anxiety and/or depression exhibited decreased time horizons, but increased low-certainty language. A possible explanation is that the future events which preoccupy sufferers of anxiety and depression may be not so proximal as to be easily predictable or so distal as to be irrelevant. In other words, they may be in the "Goldilocks" zone where worries are relevant.

A further point is CLT maintains that low probability events are associated with more abstract representations. The idea is that people use more abstract language to talk about low-certainty events, and *vice-versa.* For instance, compare *I am going to replace a piston* with *I am going to replace a component.* The latter is more abstract and evinces less certainty. Similarly, compare *I live in a small brown stone walk-up in Queens with a green front door and my tabby cat, Jefferson* with *If I win the lottery, I'll buy a big mansion.* The latter is lower probability and more abstract. If individuals with anxiety and depression construe experience more abstractly (Gjelsvik et al., 2018; Ratcliffe, 2015; White & Wild, 2016; Goldwin & Behar, 2012), this might itself drive low-certainty FTR language.

5.4.3 Therapeutic applications

Therapeutic interventions might be developed based on the findings. Correlations between language use and mental unwellness provide deepened evidence of the power of discursive constructionism. Discursive constructionism is a theoretical framework which has become widely-used by clinical practitioners over the last several decades. In this view, systems exist not in external reality, but within "language and communicative action" (Anderson & Goolishian, 1988, p. 373) and so these structures "are locally determined through dialogical exchange" (Anderson & Goolishian, 1988, p. 373). As John Austin puts it, words "do things" (1962). That is, the manner in which we talk about things can change how they are construed and experienced. In this perspective, people in anxious and depressed states discursively reshape the objects of their concern. The power of discursive constructionism is that by tuning into the subtleties of language use, therapists can exert influence in an intentional direction. From this perspective, the present results support discursive constructionist approaches by suggesting that illness experiences are embedded in language use patterns. This begs the question: Decoupled from their linguistic correlates, would the phenomenology of anxiety and depression still be the same? Therapeutic intervention could encourage people to use language which expresses higher agency, more certainty, or expanded temporal horizons. These simple interventions could help treat anxiety and depression.

5.4.4 Conclusions

Anxiety and depression continue to affect the lives of many (Dewa & McDaid, 2011). By implementing descriptive big-data analyses, we used the information available on social media to develop novel insights into the habits of speech and thought which may characterise anxiety and depression. We showed that redditors who had once posted to forums dedicated to the discussion of anxiety and depression made more frequent past and future time reference, and used a different distribution of FTR types. They used more bouletic, more low-certainty, more deontic, and fewer future tense constructions. While these effects were small, they might be further developed as markers of these disorders. Additionally, we found posters to mental health forums exhibited contracted time horizons relative to control. This motivated a mediation analysis in Study 2 which found that subjective representations of future temporal distance mediated a relationship between anxiety and temporal discounting. Anxious participants construed future events as more distant and were therefore less likely to wait for future rewards. We then outlined why these results might be expected from two theoretical perspectives: functionalist approaches to clinical psychology and construal level theory.

If progress is to be made in understanding and treating these disorders, new markers need to be explored which can help develop more accurate diagnostic criteria. Additionally, predictive approaches to online social data should be complemented by descriptive ones which allow insights from the analysis of social media data to be tested in controlled environments and applied in novel therapies. Our results suggest that such approaches may yield fruitful insights.

Chapter 6

General discussion

In this dissertation I have investigated the relationship between FTR status[1] and (risky) intertemporal decision making with a view to understanding whether cross-linguistic differences in FTR grammar impact the psychological discounting processes that underpin such decisions. In addition to investigating *whether* there is evidence for such effects, I focussed on identifying the mechanisms that may be responsible for mediating such effects. Specifically, I established whether time or probability was principally involved. For linguistic predictors, the question was whether low-certainty modal verbs or future tenses were involved in mediating effects. For (risky) intertemporal outcomes, the question was whether FTR status impacted delayed outcomes (temporal discounting) or risky delayed outcomes (combined discounting).

I contrasted Chen's (2013) temporal hypothesis with the modal hypothesis I develop. Although Chen's (2013) original predictions involve temporal discounting only, I applied his ideas about the effect of future tense use on temporal beliefs to combined discounting outcomes. I reasoned that both hypotheses predict strong-FTR speakers should exhibit increased combined discounting (but for different reasons). Whereas the temporal hypothesis posits that tenses impact temporal beliefs and therefore combined discounting, the modal hypothesis posits that modals impact beliefs about risk and therefore combined discounting. In general, findings support the modal account and fail to support the temporal one. This suggests that effects of FTR status involve the modal mechanisms I have proposed, rather than the temporal ones proposed by Chen (2013). Specifically, speakers of strong-FTR languages are suggested to discount risky future outcomes to a greater degree than speakers of weak-FTR languages because strong-FTR languages grammatically oblige the use of weak modals—*could, may, might, should*—to encode uncertainty about the future.

[1]In order to avoid repeatedly referring to the specific contrast between English and Dutch which formed the basis for my investigation, I use the term FTR status. I do not intend to make broader claims about effects of FTR status in other languages. FTR status is simply used to indicate the contrast as applied to the (limited) languages in my sample. See Section 6.6.2 for a discussion of generalisability.

In the first instance, none of the predictions of the temporal account were supported. First, future tense use caused decreased rather then increased temporal discounting (Study 1, Chapter 3), which replicated (Study 2, Chapter 3), and was re-produced in a forced-production paradigm when we directed Dutch speakers to use future tense constructions and then measured temporal discounting (Study 3, Chapter 3). This suggests that future tenses encode modal high certainty rather than temporal distance or precision (cf. Chen, 2013). Second, future tense use had no impact on ratings of subjective temporal distance (Study 2, Chapter 4), again suggesting that the English future tense does not encode temporal notions (cf. Chen, 2013). Third, subjective-temporal-distance ratings were not different when a future-referring statement was framed in the present vs. the future tense: Using a time-slider, English and Dutch participants rated "Ellie *will* arrive later on" and "John *is arriving* later on"[2] as being the same distance away in time (Study 2, Chapter 2). Chen's (2013) temporal account predicts *zal* 'will' constructions should be rated as more temporally distal. This further suggests that the future tenses in English and Dutch do not encode temporal distance (cf. Chen, 2013). Fourth, English speakers rated future outcomes as more proximal than Dutch speakers (Study 2, Chapter 2), a pattern which replicated (Study 1, Chapter 4). As a consequence, English speakers temporally discounted less rather than more (Study 1, Chapter 4), once again undermining the idea that the future tense encodes temporal distance or precision (cf. Chen, 2013). Fifth, population-level differences in temporal discounting were mixed. In one study, there was no difference between English and Dutch speakers (Studies 1 of Chapters 3 and 4[3]); in another, English speakers temporally discounted more (Study 2, Chapter 3). This provides unclear evidence that English speakers temporally discount more than Dutch speakers (cf. Chen, 2013). These findings provide strong convergent evidence against the temporal account developed by Chen (2013).

In the second instance, the modal account I develop was supported. First, language-elicitation tasks demonstrated that the grammaticisation of FTR does not only involve the increased obligation to use future tenses, but also to use low-certainty modal verbs (Chapter 2). Second, increased use of low-certainty constructions in English caused increased combined discounting (Study 2, Chapter 4). Third, this finding replicated with a sample of Dutch-English bilinguals, which helps address cultural confounding (Study 3, Chapter 4; see Section 6.6.3 for discussion). Forth, use of the obligatory English modal verb system drove the effect—the core prediction of the modal hypothesis—and this was moderated by self-reported English proficiency, which suggests language is involved in producing outcomes. Fifth, whereas temporal discounting results were mixed, three independent studies found that English speakers exhibited increased combined discounting of outcomes which were both risky and delayed

[2]Dutch "Ellie *zal* later aankomen" and "John *arriveert* later".

[3]Studies 1 of Chapters 3 and 4 reports different aspects of the same experiment so do not constitute independent evidence.

(Studies 2 and 3, Chapter 4; Appendix G). Sixth, the general importance of modality to how people talk about the future was supported by a corpus analysis of data from the social-media website Reddit (Study 1, Chapter 5). In this study, I downloaded a large sample of Reddit posts and analysed them statistically. Rather than the future tense predominating, a variety of modal construction type encoding notions of probability, obligation, and desire comprised the preponderance of English FTR. This suggests the obligation to encode low-certainty modality impacts a large proportion of naturally-occurring FTR statements, bolstering the plausibility of the hypothesis that speakers' (risky) intertemporal preferences could be affected (see Section 6.1.4.1). Together, these results provide strong convergent evidence for the modal account I develop.

Numerous findings involving real-world (risky) intertemporal behaviour have been interpreted as supporting Chen's (2013) temporal hypothesis (e.g. Chen, 2013; S. Chen et al., 2017; Chi et al., 2018; Falk et al., 2018; Figlio et al., 2016; Guin, 2017; Hübner & Vannoorenberghe, 2015b; J. Kim et al., 2017; S. Kim & Filimonau, 2017; Liang et al., 2018; Lien & Zhang, 2020; Mavisakalyan et al., 2018; Pérez & Tavits, 2017; Sutter et al., 2015; Thoma & Tytus, 2018). This interpretive framework should be reconsidered. The modal hypothesis may be more plausible.

In these conclusions, I revisit the issues discussed in the general introduction to this thesis (Chapter 1). First, I discuss the key questions which empirical work sought to answer (introduced in Section 1.5). I then discuss contributions to work on language typology and grammaticisation, with particular emphasis given to contrasts with Dahl's (1985, 2000a) approach (outlined in Section 1.4.1). Next, I revisit Chen's (2013) FTR status typology and argument about FUTURE TENSE semantics (reviewed in Section 1.3.2). Following this, I evaluate my contributions in the context of common contemporary criticisms of the linguistic relativity hypothesis (identified in Section 1.2). I then discuss whether my findings can be brought to bear on the Hopi time controversy (discussed in Section 1.1). Finally, I acknowledge some limitations of the current work and conclude.

6.1 Key questions answered

I have suggested that linguistic relativity investigations must, at a minimum:

(a) Establish relevant cross-linguistic differences
(b) Identify correlated non-linguistic cognitive (or behavioural) differences
(c) Provide causal evidence for the impact of (a) on (b)

I organised my 8 key questions using this framework. In this summary, I return to these key questions and evaluate my contributions to answering them. Questions are again organised using this ternary framework.

6.1.1 Establishing cross-linguistic differences

The first step was to identify cross-linguistic differences which might be relevant to (risky) intertemporal outcomes. This resolves into two questions. First, do English speakers use more low-certainty modal constructions? Second, what do future tenses "mean"? These two questions are addressed below.

6.1.1.1 English speakers used more low-certainty modal language

A key question is whether English (strong-FTR) speakers use more low-certainty modal constructions than Dutch or German (both weak-FTR) speakers. FTR status is not a very rich descriptor of FTR grammaticisation: It simply indexes whether the future tense is obliged for prediction-based main-clause FTR. What this broadly indicates about FTR is unclear. Recall, Dahl (2000a) suggests that obligatory tense marking of prediction-based FTR is a reasonable indicator for FTR grammaticisation. However, he is deliberately vague about what is indicated, noting only that obligatory use of *will* in an English weather forecast suggests FTR is grammaticised "in English in a way that it is not in Finnish" (2000a, p. 311). *What* this says about the differences between English and Finnish FTR is left unsaid. Despite this, Chen (2013) applied Dahl's criteria to a large (N = 129) language sample. The question is: Are low-certainty modals—as well as the future tense—higher frequency in English? Although the English modals are more grammaticised, Dutch and German speakers could have made up the difference by using other FTR construction types.

I answered this question by having English, Dutch, and German speakers complete the FTR-elicitation task. I processed the data with the FTR-type classifier, and statistically analysed the results (Study 1, Chapter 2). English speakers used relatively more low-certainty terms, and this was specifically driven by use of the obligatory English low-certainty modal verb system—*could, may, might,* and *should*. Dutch and German speakers made up for this to some extent through the use of periphrastic other-low-certainty constructions (e.g. *denken* 'think', *mogelijk* 'possibly', or *wel* approximately 'well'). However, encoding of low certainty was higher in English.

English speakers were also more sensitive to the factors of the FTR-elicitation task (Study 2, Chapter 2). First, they used more low-certainty constructions in the low-certainty condition (when they were directed to imagine they were uncertain). This suggests that the threshold at which English speakers begin to express weak modality is lower. Second, English speakers used more low-certainty terms when they talked about distal time frames. This suggests they construe temporally distant events as risky and/or uncertain. Third, they used more low-certainty terms when they made predictions. This suggests the grammar of English FTR obliges speakers to encode modal notions which characterise FTR-mode (*schedules* $\approx$ *high certainty*; *intentions* $\approx$ *moderate certainty*; and *predictions* $\approx$ *low certainty*). However, marking of FTR-mode is distributed between the various modals in the English system. Key features of English FTR may therefore be overlooked by focussing exclusively on the future tense. Particularly, it is misleading to say *will* is required for prediction-based FTR, because this is a general feature of the English modals. This suggests that the FTR system—as a whole—is more sensitive to modal notional categories of FTR-mode.

6.1.1.2 Future tenses encode modal notions and not temporal distance

A critical part of establishing cross-linguistic differences in FTR language use is understanding the future tense semantics. Chen (2013) argues that the FUTURE TENSE exhibits temporal semantics, whereas I have argued that future tenses tend to encode modal high certainty. In fact, I suggest that the semantics of tenses are inherently high certainty. Future tenses are usually defined in terms of shifting the temporal frame of reference to after the time of speech (Klein, 1995). They therefore imply a speaker is highly certain about the outcome—otherwise they would use a modal expression. Here, a paradigmatic analysis is useful. The English future tense is the only high-certainty expression possible in prediction-based FTR. This suggests its contrastive semantic contribution is to encode high certainty relative to the other weaker modals.

This should be seen in the broader context of debate about FUTURE TENSE semantics. Some linguists have argued that the tendency of the FUTURE TENSE to mark predictions implies it encodes epistemic weakening (Giannakidou & Mari, 2018). Evidence for this draws on future tenses as a markers of present-time predictions. When used in such a way, future tenses encode modal weakening relative to unmarked constructions. For instance, on hearing a knock on the door, an English speaker can say either:

(6.1) a. *That will be the postman.*
b. *That is the postman.*

Example (6.1a) encodes modal weakening compared with (6.1b). This is quite a common feature of the FUTURE TENSE: Similar uses have been noted for Dutch, German, Greek, Italian (Giannakidou & Mari, 2018), and French (Lyons, 1968). The question becomes whether the *future tense* also encodes modal weakening when used for FTR. Some argue this is the case (e.g. Giannakidou & Mari, 2018), while others see future-marking *will* as having different semantics to present-marking *will* (Salkie, 2010). Three accounts emerge: (a) The FUTURE TENSE encodes temporal distance (Chen, 2013); (b) future tenses tend to encode low-certainty modal weakening; and (c) future tenses tend to encode high-certainty modal strengthening.

This thesis brought several lines of evidence to bear on clarifying these positions. In the first instance, correlations between future tense use and (risky) intertemporal outcomes were used to make inferences about tense semantics (Chapter 3). Secondly, future tense use was analysed as a function of the factors of the FTR-elicitation task: FTR-mode, modality condition, and temporal distance (Chapter 2). Thirdly, participants used a slider to indicate their understanding of FTR semantics in terms of temporal distance (close vs. distant in time) and modal polarity (low- vs. high-certainty) (Study 2, Chapter 2). Below, I evaluate the three accounts of future tense semantics with reference these lines of evidence.

First, I found no evidence for the suggestion that the FUTURE TENSE encodes temporal distance. There was no effect on temporal-distance ratings when a simple FTR statement was framed in the present vs. future tense (Study 2, Chapter 2). Moreover, English speakers rated future events as closer not farther away (Study 2, Chapter 2; Study 1, Chapter 4). This is the opposite to the predicted direction if the English future tense encoded temporal distance. Null and opposite effects constitute strong evidence against the idea that the FUTURE TENSE encodes temporal distance. I did not directly test precision effects, but found that temporal discounting was negatively related to future tense use in English (Studies 1 and 2, Chapter 3), which would not be predicted if the future tense encoded temporal precision.

Secondly, accounts of the FUTURE TENSE as encoding modal weakening were not supported. In fact, English speakers rated *will* and *be going to* as higher certainty than present-tense FTR (Study 2, Chapter 2). Additionally, use of the English future tense was negatively correlated with temporal discounting (Studies 1–3, Chapter 3; Appendix G), which argues against accounts of *will* as encoding modal weakening. This leaves the possibility that the English future tense encodes high certainty. This makes it look more like a tense than a modal, but it appears to be a constituent of an FTR system which is—as a whole—preoccupied with encoding of modality.

A further consideration is whether there are cross-linguistic differences in future tenses semantics. Paradigmatic analyses suggest this, since the future tenses in Dutch and German contrast with present-tense FTR, making their oppositional roles less clear. Perhaps due to this, the Dutch case was more complicated. While Dutch speakers rated *zullen* ‘will’ and *gaan* ‘be going to’ as encoding high certainty (Study 2, Chapter 2), they rated these constructions as lower-certainty than the present tense. This suggests that modal weakening of the sort apparent in example (6.1) applies to FTR in weak-FTR languages (on this, see Broekhuis & Verkuyl, 2014). Additionally, in contrast with the English future tense, which exhibited consistent negative correlations with temporal discounting, the relationship between such measures and future tense use in Dutch was characterised by complicated interactions (Chapter 3). At the same time, ratings of certainty for the Dutch future tense exhibited greater interpersonal variation. This suggests that the Dutch future plays a less well-defined role in FTR than the English future, or at least that its use is subject to greater interpersonal variation. Taken together, the results suggest the semantics of the Dutch future are modal in character. However, exactly what they conclusively demonstrate remains murky. The critical findings involve early versions of the FTR-elicitation task which had imbalanced factors. This makes interpretation difficult. Future research might build on the techniques I developed to more fully investigate the semantics of *zullen* 'will' and *gaan* 'be going to'.

The semantics of language-specific future tenses are suggested to vary. This undermines accounts which treat FUTURE TENSE semantics as universal (i.e. Chen's [2013] argument). In fact, FTR status may be a relevant factor in determining the semantics of future tenses. More weakly grammaticised future tenses in weak-FTR languages may be characterised by greater semantic variation and indeterminate modal semantics. More strongly grammaticised future tenses in languages like English may be characterised by clearly-defined roles within FTR systems, i.e. the encoding of high certainty. It is an important observation that there is common ground and variation in the semantics of FTR structures in closely-related languages. As noted above, both the English and Dutch future tenses appeared to encode modal notions (Chapters 2 and 3). However, English future tense use was consistently related to discounting outcomes, while interactions characterised the relationship in Dutch (Chapter 3). This suggests there are overarching similarities between the English and Dutch future tenses, but also important differences. This echoes findings involving the lexical semantics of cutting and breaking events in Western Germanic languages. Specifically, English, Dutch, German, and Swedish speakers described cutting and breaking events in terms which made a distinction between whether the site of disconnection was predictable (Majid, Bowerman, Van Staden, & Boster, 2007), but there were differences between how verb semantics clustered the same events (Majid, Gullberg, Van Staden, & Bowerman, 2007).

6.1.2 Identifying correlated differences in non-linguistic cognition

The next step was to identify correlated population-level differences in psychological discounting. In the context of my research, the relevant question is whether temporal or combined discounting differences are associated with FTR status. The temporal account developed by Chen (2013) predicts a difference in temporal discounting (of just delayed rewards); the modal account I develop only predicts a difference in combined discounting (of risky and delayed rewards).

6.1.2.1 Combined discounting was affected, temporal discounting results were mixed

Chen (2013) hypothesised that FTR status should impact inter<u>temporal</u> decisions; I have referred to this idea as the temporal hypothesis. On the other hand, I argue that observed population-level behaviours might equally be explained by differences in temporal or combined discounting. I hypothesised that increased encoding of low certainty may be a wide-spread feature of strong-FTR languages. Strong-FTR speakers may only discount more than weak-FTR speakers when future outcomes are delayed *and* risky. I refer to this as the modal hypothesis. This issue is not addressed by wide-ranging evidence that FTR status is correlated with population-level behavioural differences. This is because observed behaviours might be attributed to differences in temporal or combined discounting, since real-world intertemporal decisions often involve risk. In fact, evidence that FTR status impacts temporal discounting in experimentally controlled settings has yielded mixed results (e.g. Thoma & Tytus, 2018; Sutter et al., 2015). A pertinent question is therefore whether temporal or combined discounting is associated with FTR status.

I answered this question by having participants complete intertemporal-choice tasks, as well as risky-intertemporal-choice tasks. The results support the modal account that FTR status impacts combined discounting but not necessarily (just) temporal discounting. I did find one population-level difference in time preferences between English and Dutch speakers (Study 2, Chapter 3). However, this did not replicate (Studies 1, Chapters 3 and 4). In other words, when I analysed strictly intertemporal decision making, results were mixed. In contrast to this, English speakers discounted more when outcomes were both risky and delayed in three independent studies (Studies 2 and 3, Chapter 4; Appendix G). This indicates FTR status is more consistently associated with psychological discounting when outcomes are both delayed and risky, as compared to when they are (just) delayed. This suggests behavioural differences in real-world (risky) intertemporal behaviour are attributable to the modal mechanisms I have identified.

6.1.3 Developing causal evidence

The third step was to develop causal evidence which links cross-linguistic differences to the relevant outcomes. I developed causal evidence by having English and Dutch participants complete the FTR-elicitation task and a (risky) intertemporal decision-making task. Mediation models were employed to estimate the impact of language on (risky) intertemporal outcomes via speaker-level use of various FTR construction types. To further clarify the contribution of language (compared with culture) to producing the observed outcomes, I also worked with bilinguals, and used directed language-production paradigms.

6.1.3.1 Higher use of the future tense in English caused decreased temporal discounting

Chen's (2013) account of FUTURE TENSE semantics implies increased use of the future tense should cause speakers to discount more. In fact, the effect consistently went in the opposite direction (Studies 1 and 2, Chapter 3; Appendix G). These results are inconsistent with the theoretical mechanisms proposed by Chen (about either distance or precision), widely cited in the economic literature. If the English future tense is a good model for other future tenses in strong-FTR languages (a big if, see Section 6.6.2), this would render Chen's (2013) temporal account implausible. By demonstrating this effect using mediation models, I am able to provide the clearest causal evidence to date that cross-linguistic differences impact psychological discounting. How is this the case? Causal models, while constituted from correlational data, provide causal evidence because they flexibly represented the kinds of causal processes which researchers postulate occur in the real world and give rise to experimental and observational data (Pearl & Mackenzie, 2018). The nature of such causal evidence can be difficult to put one's finger upon. In the context of this thesis, it can be limbed by imagining the evidence I would have presented had I merely regressed discounting outcomes over the contrast between English and Dutch. In fact, the *total effects* I report are such regression coefficients (Hayes, 2013). With only these to interpret, Study 1 Chapter 3 would find a null effect, and Study 2 of Chapter 3 would find English speakers discounted more. Stripped of the mediation results, this does not add up to very much. On other other hand, in the mediation effects, causal inferences emerge; The finding that higher future tense use in English caused less discounting strongly suggests Chen's (?) causal models are flawed.

6.1.3.2 Higher use of low-certainty modals in English caused increased combined discounting

The core prediction of the modal account I develop is that use of low-certainty FTR terms should cause increased combined discounting of risky and delayed outcomes. This prediction was supported. Higher modal verb use in English caused increased discounting in risky intertemporal decision-making outcomes (Studies 2 and 3, Chapter 4; Appendix G). Since the effect of future tense use went in the opposite direction (Studies 1 and 2, Chapter 3), or had no impact when the conditional effect *via* temporal-distance ratings was estimated (Study 2, Chapter 4), this suggests that the obligation to use low-certainty FTR structures is the principal cross-linguistic difference responsible for observed effects of FTR status. Once again, the path analyses I conducted begin to answer causal questions about the nature of cross-linguistic effects on (risky) intertemporal preferences. Here, they strongly support the causal processed postulated by my modal account.

6.1.3.3 Cross-linguistic differences may affect (risky) time preferences

Throughout this thesis I use mediation modelling and measurement of FTR language use to provide evidence that cross-linguistic differences between English and Dutch are involved in producing observed differences in (risky) intertemporal outcomes. Do such approaches rule out the possibility that non-linguistic cultural differences drove language use? Not necessarily. It could be the case that cultural and not grammatical differences were driving FTR language use differences. I take the view that such criticisms fail Occam's razor because they need to invoke some kind of non-linguistic system of cultural values which is correlated with—

but somehow independent from—known grammatical features of language. The elicited FTR results correspond to well-known grammatical features of English, i.e. the obligatory modal verb system. The simplest explanation is that such features play a causal role in affecting non-linguistic cognition, even if they also reflect and/or transmit cultural values (Gumperz & Levinson, 1996).

Nonetheless, to empirically address this issue, I used a within-subjects design with a sample of Dutch-English bilinguals. Participants completed an FTR-elicitation task and then a risky intertemporal decision-making task. They did so first in one language, and then, after approximately one month, in the other (Study 3, Chapter 4). Such methods address cultural confounding because the control and treatment samples are identical (they are the same individuals, tested at different times in different languages). These methods therefore go some way[4] to isolating language from culture. Bilinguals used more low-certainty modals in the English condition, which was in turn associated with increased combined discounting. This was moderated by self-reported English proficiency, so more proficient bilinguals used more low-certainty modals and this led to differential impacts on combined discounting. This suggests that language-level grammatical differences are responsible for observed effects of FTR status. It also strengthens the evidence for the modal hypothesis that the encoding of low-certainty modality is the mechanism by which this effect is transmitted.

A second method I employed to address this question was the use of a directed language-production task in a sample of Dutch speakers (Study 3, Chapter 3). I first established a baseline by having participants complete an undirected FTR-elicitation task and then a temporal-discounting task. Several weeks later the same participants completed a second survey which was identical except for one factor: In the FTR-elicitation task, half the participants were directed to use the future tense and half were directed to use the present tense. Relative to the previously-measured baseline, this manipulation impacted temporal discounting. However, by exogenously manipulating language use, the study rules out cultural factors which could have caused English speakers to use more future tense constructions in correlational cross-cultural experiments. This suggests that the grammatical obligation to use the future tense can impact time preferences. However, a critical point is that participants who were directed to use the future tense temporally discounted <u>less</u> not <u>more</u>. As with contrasts between English and Dutch (Studies 1 and 2, Chapter 3), this supports the idea that the Dutch future tense encode notions of high certainty rather than temporal distance or precision.

6.1.4 Further questions: Big-data and functional relativity

Below I shift to questions which reach beyond the relationship between cross-linguistic differences and cognition. Linguistic FTR and intertemporal decision making remain the

[4]They only go *some* way because it might be argued that bilingual people are also bicultural to some extent (S. X. Chen, 2015). See Section 6.6.3.

loci of enquiry, but new issues are taken up. In particular, I bring my results to bear on a description of naturally-occurring English FTR, with a view to making connections with relativity questions. Next, I discuss results which are pertinent to the relationship between FTR language use and anxiety and depression.

6.1.4.1 English FTR is characterised by a blend of modal and tense constructions

If some linguistic feature is hypothesised to impact cognition, the feature needs to be used with enough frequency to plausibly have an effect. It would make little sense to hypothesise that a grammatical distinction should cause a broad cognitive difference if its use is very low frequency. For instance, Dahl (2013) makes the case that Chen's (2013) prediction-based FTR criteria may not indicate much, if speakers mostly talk in intention-based or schedule-based FTR. In a similar way, the modal hypothesis would be undermined if modal constructions formed only a very minor proportion of FTR statements. An important question becomes what is the relative frequency of modal FTR.

I investigated this question using big-data corpus methods (Study 1, Chapter 5). Specifically, I downloaded a large sample of naturally-occurring text data from the social media website Reddit ($N = 619,696$ unique posts comprising $N = 2,029,956$ sentences). I then built a machine learning classifier which estimates the likelihood that a short text item refers to the future. Future-referring sentences ($n = 145,130$ sentences, or 7.15%) were then passed to the FTR-type classifier so as to understand the relative proportions of FTR type which characterise English future time reference. I found that modal statements formed a larger proportion of total English FTR than did tensed statements. This was still true even when I relaxed the dominance structure of the FTR-type classifier such that constructions such as *It will probably rain tomorrow* were counted as future tense *and* other-low-certainty rather than *just* other-low-certainty (Appendix F). This indicated that—in English—modal constructions of various forms are a highly frequent way of referring to the future. Therefore, the hypothesis that the differential obligation to encode low-certainty FTR impacts risky intertemporal decision-making processes is plausible.

6.1.4.2 Time perspective and linguistic FTR habits differ between healthy populations and people who suffer from anxiety and depression

This question may be seen as a departure from the preoccupations of much of this thesis, which has investigated *structural* relativity effects of cross-linguistic differences on cognition (Lucy, 1997). With this question, I shift to investigating *functional* relativity, which involves the relationship between language and thought in different speech communities within the same language (see Lucy, 1997; Gumperz & Levinson, 1996). In particular, I focus on the relationship between FTR language use and the modes of thinking which may characterise anxiety and depression.

To understand this relationship, I supplemented the Reddit corpus with posts which were made by people who had once posted a popular comment to forums dedicated to the discussion of anxiety and depression (the "mental health" condition) (Study 1, Chapter 5). Language used by these people was then compared to a sample drawn from all of Reddit (the control condition). Language use in the mental health condition differed significantly from the control condition. People in the mental health condition tended to make more non-present time reference (past and future) and exhibited contracted time horizons (past and future), i.e. they referenced more temporally proximal events. When they talked about the future, how they did so differed as well. People in the mental health condition used more low-certainty language, more deontic language, more bouletic language, and fewer future tense statements. These findings suggest that people who suffer from anxiety and depression spend more time thinking about non-present time, but reference more temporally proximal time frames. At the same time, they may simultaneously feel less certain and more constrained about future outcomes. Higher bouletic modal usage suggests they relate to future outcomes more in terms of their own desires, rather than simple future tense statements of fact.

There was no difference between people who had posted to forums about anxiety compared with depression. As such, it was unclear which of these two mood disorders was driving results. Even had a difference been identified, having once posted to an online forum is a poor diagnostic criterion. Therefore, I conducted an experimental study (Study 2, Chapter 5). Participants completed a temporal-discounting task, a time-slider task, and a measure of anxious and depressive tendencies. Anxious individuals rated future events as more distant and consequently temporally discounted to a greater degree. The same was not true of depressive participants.

In combination, these results are compatible with some form of *functional* relativity (Lucy, 1997). The results suggest habitual language use contributes to the development and maintenance of anxiety and depression. Here an important point is my previous studies investigated *structural* relativity effects. With *structural* relativity, there are good theoretical reasons to treat language as exerting exogenous effects. For instance, it is difficult to see how someone's (risky) intertemporal preferences might impact which language they first learn (see Section 6.6.4). This is not true of *functional* relativity effects: Habits of thought endemic to the experience of anxiety and/or depression could have driven language use, rather than language use contributing to mental ill health. Online forums dedicated to the discussion of anxiety and depression may be a prototypical speech community, and self-diagnosis as suffering from anxiety and/or depression is the criteria according to which members self-select to join. I therefore make no claims about the direction of the association between language use and mental ill health. It is easy to see that reciprocal causal effects might be involved. Language use could both contribute to the development and maintenance of anxiety and depression, and be symptomatic of modes of thinking which characterised these mood disorders.

Nonetheless, future research could leverage the insights I developed. Treatment programs could answer the question: Can concerted effort to speak about the future in prescribed ways lead to lessening of anxious and/or depressive symptoms? Future time reference was involved in approximately 7% of the sentences in the corpus, and modal statements formed a large proportion of these. This suggests such a program could have a positive impact. However, the differences in FTR-type proportions between the mental health and control conditions were minor, *Fig. 5.4*. The effect-sizes of treatments based on these markers might be correspondingly small.

6.2 Evaluating the approach taken by this thesis: The measurement of linguistic FTR

The innovative aspects of my methodological approach involve the creation of a new FTR-elicitation questionnaire and FTR-type classifier. These elements of this thesis are heavily indebted to Dahl (1985, 2000a), but also represent significant changes to his approach. I identified four key differences between my and Dahl's (1985, 2000a) approaches: (a) I focus more on individual differences; (b) this entails a (much) smaller sample of languages; (c) I emphasise the importance of modality to the investigation of FTR; and whereas (d) Dahl is concerned with the investigation of grammatical categories, I emphasise the investigation of semantic categories. I do not believe I need to further justify the languages in my sample or belabour the importance of modality to FTR. As such, I focus the comments below on: (a) individual differences; and (d) my emphasis on semantic categories.

6.2.1 Individual differences

A major point of difference between my and Dahl's (1985, 2000a) work is that I focus on establishing the extent of individual differences in elicited FTR language between speakers of the same language. Below, I outline the implications of my findings regarding within-language differences.

6.2.1.1 Usage-based grammar and linguistic relativity

The extent of the individual differences I found in elicited FTR language, particularly for Dutch, support recent calls for ongoing cross-pollination between research in linguistic relativity and usage-based cognitive linguistics (Günther et al., 2016). On one hand, much linguistic relativity research progresses by identifying prospective typological differences at the language-level, and then exploring potential non-linguistic cognitive effects of these (see discussion in Lucy, 1997, 2016). On the other, usage-based cognitive linguistics emphasises that no two speakers of a language share the same grammars (Barlow, 2013; Dabrowska, 2012; Günther et al., 2016). Such approaches understand language to be represented as a network of syntactic-semantic constructions which are subject to life-long change through frequency-based processes of entrenchment via language use (Bybee, 2006; Günther et al., 2016). There is good evidence that within-language differences exist (Barlow, 2013; Dabrowska, 2012), and that linguistic representation changes each time a speaker uses language (Bybee, 2006). In fact, the kind of cross-linguistic difference which might give rise to population-level cognitive differences is often evinced by two speakers of the same language, or even one speaker at different times (Kay, 1996). This is unlikely to be the case for every conceivable typological distinction, but it may be for some. I found that some Dutch speakers used the future tense as much as English speakers, and some hardly ever used it. In other words, I found FTR-status-like differences within Dutch, and complex differences in how future tense use related to temporal discounting between Dutch speakers. This finding meshes with theoretical accounts which combine linguistic relativity with usage-based notions. In fact, it has long been suggested that cross-linguistic differences can give rise to cognitive differences between speakers of different languages (*structural* relativity) at the same time as relativity effects can arise between speech communities within the same language (*functional* relativity) (Lucy, 1997). However, within-language accounts are usually troubled with issues of circular causality. Do speakers use constructions because of endemic cognitive differences or do they possess endemic cognitive differences because of the network of constructions which they happen to have arrived at? Because of these issues, most *functional* relativity research has focused on natural linguistic exposure effects which can be captured statistically, such as education (Lucy, 1997; Vygotsky, 1987; Croijmans et al., 2020).

Do the results in this dissertation clarify how *functional* and *structural* relativity accounts may interact? Possibly. A major goal of this work has been to provide a quantitative account of how language-level grammatical features impact cognition via speaker-level usage. In particular, I used mediation models to estimate the conditional impact of language on (risky) intertemporal outcomes via speaker-level use of FTR constructions. Both language-level and speaker-level differences are accommodated by such models. Moreover, since language use is shown to cause population-level outcomes, these analytic techniques provide better causal evidence than much experimental work which tests different populations. I therefore join contemporary approaches which emphasise the value of exploring how language use may relate to cognition both within- and between-languages (Günther et al., 2016).

Further cross-pollination between usage-based and relativity work will surely be productive. Future research might seek to explore how speaker-level usage relates to discounting in different speech contexts. For instance, language use may equally reflect speakers beliefs, or their communicative goals *vis-à-vis* other speakers: Does a speaker use *will* because she believes a future event to be high certainty, or because she wants her interlocutor to believe this? An interesting application of my research would be to expand my methods to investigate how language-level constraints might impact such conversational dynamics and thereby cognitive processes in different speech contexts. Lucy (1997) puts forward Roman Jakobson's (1960) well-known schema for the different semiotic functions of language as a potential starting point. Jakobson's typology is rooted in the dynamics of speech contexts. Among other things, it distinguishes between: (a) *referential* contexts; (b) *emotive* contexts; and (c) *conative* contexts. In the first instance, *referential* contexts involve speakers referencing and/or predicating something about the speech environment. *Referential* contexts tend therefore to be associated with third-person pronominal forms. In the second, *emotive* contexts involve speakers expressing their own sentiments and emotions about the speech context. *Emotive* contexts may therefore involve first-person pronominal forms. In the third, *conative* contexts involve the use of language to engage people to act. *Conative* contexts therefore tend to be expressed using second-person pronominal forms (see also Lucy, 1997). I have not worked out precise theoretical predictions for how—or whether—FTR language use should correlate with (risky) intertemporal outcomes across these different speech contexts. I merely point towards this as a potential framework with which an empirical starting point could be made. Jakobson's schema provides not only a useful ontology of speech contexts but also links them to simple grammatical features of English. These features could, for instance, be used to classify naturally-occurring language of the sort analysed in Chapter 5. Experimental paradigms drawn from behavioural economics which use economic games to study interactive processes might also be leveraged to control aspects of interactive dynamics.

6.2.1.2 Big data linguistics: Structured priming and linguistic acceptability judgments

While this is not intended as a criticism, Dahl (1985, 2000a) only used a handful of speakers from each sampled language. In a similar vein, acceptability judgments from single speakers (often the authors themselves) are still a wide-spread basis for making conclusions about language-level features (e.g. Salkie, 2010; Giannakidou & Mari, 2018; Sarkar, 1998). Acceptability judgments typically involve native speakers deciding whether a given utterance is grammatical, and linguistic research has historically relied on such judgments to form theories on the cognitive representation of language, and on linguistic typology (Branigan & Pickering, 2017).

The extensive individual differences I identify support contemporary arguments that such acceptability judgments based on a small number of informants may be a poor source of language-level data (Bard, Robertson, & Sorace, 1996; Branigan & Pickering, 2017; Dabrowska, 2010; Gibson & Fedorenko, 2013). My work joins modern data-driven approaches which rely on structured priming paradigms (Branigan & Pickering, 2017) and acceptability judgments from non-specialist speakers (Dabrowska, 2010) to investigate linguistic representation (see also Sorace & Keller, 2005). My results suggest typological studies should attempt to sample enough speakers to identify the extent of between-speaker variation (say $N \geq 200$, or as many as possible in languages with few speakers). Such methods will likely identify typological clines rather than strict categories. In this sense, the results echo work on manner of motion in verb-framed versus satellite-framed languages. In this area, initial categorical typological approaches (Talmy, 1985) have given way to accounts of cross-linguistic clines in the salience of the path and manner of motion (Ibarretxe-Antuñano, 2008; Slobin, 2006). Where I have used elicitation, researchers have proposed various sources of language data—including elicitation, naturally-occurring corpora, and language usage in creative fiction—to better understand typological clines (Slobin, 2006; Verkerk, 2013). In general, the results attest to within-language variation between speakers, which has obvious implications for both typological research and linguistic relativity. A consideration for relativity work is whether a typological contrast might simply not be realised in some individual speaker, or whether it might be realised inconsistently by one speaker at different times (Kay, 1996).

In conclusion, the emphasis on measuring individual differences proved fruitful in this thesis. I have quantitatively integrated theories from usage-based grammar with contemporary empirical work on linguistic relativity. I further hope to have deepened contemporary knowledge about the relationship between language-level grammatical features and speaker-level language use—at the least, I have added another example attesting to the extent of differences between speakers.

6.2.2 Semantic approaches to grammaticisation

Grammaticisation is typically defined as those diachronic processes where lexical linguistic items develop into grammatical ones (Dahl, 2000a). Much of the EUROTYP volume (Dahl, 2000b) is rooted in the study of grammaticisation. As such, a pertinent question not directly addressed above is what insights to grammaticisation my emphasis on semantic categories may yield. The approach taken in this dissertation is well-described by what Nuyts (2000) refers to as a cognitive-pragmatic approach. According to Nuyts, such approaches begin with a broad conceptual domain—FTR in this case—and ask how its expression is distributed across the linguistic units involved. A further distinction made by Nuyts (2000) is that cognitive-pragmatic approaches seek to understand language in terms its dynamic role in interacting with non-linguistic aspects of the human mind. This is contrasted with traditional linguistic research which may be characterised by an "inclination to limit the scope of attention to purely linguistic aspects of language use, and to the organisation of the linguistic system per se" (Nuyts, 2000, p. xv). The obvious example is Dahl's (2000a, 1985) work on grammaticisation, which seeks to understand the grammatical categories of tense and aspect by analysing (exclusively) linguistic data. My results suggest cognitive-pragmatic approaches may supplement traditional linguistic analyses. Rather than some specific morphosyntactic category, they emphasise investigating processes of grammaticisation across multiple grammatical and functional elements, and in terms of extra-linguistic phenomena (in this case [risky] intertemporal decision making).

Dahl (2000a) characterises grammaticisation as involving several related processes which can occur at different rates and times. These include, *obligatorisation*, i.e. the process whereby a given item becomes increasingly obligatory as it grammaticises; *semantic bleaching*, i.e. processes which see original meanings slowly lost; *loss of communicative motivation*, i.e. processes which see linguistic units used even when they contribute no information to a linguistic utterance. An example of *loss of communicative motivation* is one must use the past tense in *I walked to the store yesterday* even though the temporal adverb *yesterday* renders any tensing of *walk* unnecessary. The English modals, including *will/shall*, are usually suggested to be more grammaticised than the Dutch equivalents (Nuyts, 2000). By eliciting similar constructions at different stages of grammaticisation in two closely-related languages, my results may be brought to bear on some aspects of these grammaticisation processes, though care should be taken drawing conclusions about diachronic processes from synchronic data. Nonetheless, while not strictly comparable, English and Dutch may be similar enough to facilitate speculative insights.

Particularly for prediction-based FTR, I found higher relative use of the modals in English (including the future tense). This supports that idea that there is increased *obligatorisation* in the English modals. However, the results do not suggest the English future tense is more *semantically bleached*. If the English future were more semantically bleached, we would expect to see a weaker relationship between future tense FTR and intertemporal choices in English than Dutch. However, the opposite pattern was true. In English, there were consistent correlations between future tense use and intertemporal decisions. In Dutch this relationship was inconsistent (Chapter 3). This suggests that the English future tense is characterised by relatively less, not more, semantic bleaching. Rather, its unique semantic contribution seems to derive from its paradigmatic contrast with the other possible FTR constructions—a system of contrasts which is different in Dutch because the present-tense is possible in prediction-based FTR. Additionally, I found that the direction of the correlation between the Dutch future tense and intertemporal outcomes switched signs conditionally on an interaction (Study 3, Chapter 3). There was a negative relationship for people who tended to use the future tense in high-certainty contexts, and a positive relationship for people who did not. This suggests there are greater individual differences in the representation of Dutch future tense semantics than in English. In fact, when I had participants rate the future tense in terms of high to low certainty, the spread for Dutch ratings was greater (Study 2, Chapter 2). In conclusion, the (more grammaticised) English future appears to exhibit more consistent semantics. Rather than semantic bleaching, this suggests that grammaticisation may sometimes involve regularisation among members of a speech community of the relationship between linguistic structures and the notional domains they encode.

Traditional approaches to grammaticisation would not be surprised by this conclusion: *Will* may exhibit what Kemmer calls a “doughnut configuration” (Kemmer, 1993, p. 225). Specifically, *will* has mostly lost its original meaning from the proto-Germanic *willan* “to want”, but as *will* has grammaticised, its principal use has become encoding high certainty rather than marking desiderata. However, peripheral bouletic meanings still surround this newer primary function, like a doughnut. This can be seen in, for instance, *I will that the anonymous referee accept this manuscript for publication.* An important point is semantic bleaching is apparent when *will* is used with a modal modifier, e.g. *will* does not contribute much meaning to *It will possibly rain later.* The above analysis is therefore contingent on my coding scheme, which does not count *will* as “future tense” when it appears with a modal modifier.

An interesting finding is that the Dutch future tense was associated with decreased discounting in combined discounting (Appendix G) but not temporal discounting (Studies 1 and 2, Chapter 3). A relevant question is why the correlation with temporal discounting was significant in English: If the English future tense encodes high certainty, it is not obvious why it should correlate with a strictly intertemporal outcome. One possibility is that time and probability are more entangled for English speakers. Some evidence suggests this. Particularly, English but not Dutch speakers used more low-certainty terms as a function of temporal distance in the FTR-elicitation task (Study 2, Chapter 2). This suggests English but not Dutch speakers construe increased delay as increasingly uncertain. Additionally, assuming the future tense encodes notions of high certainty, the fact that future tense use correlated with intertemporal decisions in English but not Dutch suggests increased overlap between time and probability in English speakers (Studies 1 and 2, Chapter 3). One possibility is that the increased obligation English grammar places on speakers to encode future uncertainty results in relatively more cross-modal mapping between time and probability in English speakers. In terms of grammaticisation, such cross-modal mapping might be expected to contribute over long periods to grammatical *obligatorisation* processes. If English speakers believe the future to be uncertain, grammatical norms of English may come to reflect this over time. Again, rather than semantic bleaching, it appears that grammaticisation can lead to tighter links between grammar and notional domains.

The extent to which bouletic structures were common in English FTR is perhaps surprising. They represented a significant proportion of all FTR statements in the Reddit corpus (Study 1, Chapter 5). This suggests that Dahl’s ternary division of FTR-mode into intentions, predictions, and schedules (2000a) may need revision. The FUTURE TENSE may tend to crop up as a marker of intentions and predictions (but not desires). On the other hand, bouletic modal structures appear to be important to the expression of FTR. A quaternary system may be appropriate: intentions, predictions, schedules, and desires. This observation would not be possible if the inquiry were limited to the expression of the future tense.

In conclusion, this thesis focussed on semantic rather than grammatical categories. Data from (risky) intertemporal tasks were leveraged to make inferences about language. Such tactics yielded insights about grammaticisation which could not have been made from language data alone. Such approaches may complement traditional linguistic investigations.

6.3 Revisiting Chen's (2013) linguistic work: FTR status and future tense semantics

In the introduction to this thesis (Section 1.3.2) I presented Chen's (2013) approach to establishing cross-linguistic differences in two parts: first, the creation of FTR status based on whether a language permits the present tense for main-clause prediction-based FTR; second, his argument that the FUTURE TENSE encodes temporal distance or precision. I argued that—while oversimplified—Chen's (2013) creation of FTR status was well-grounded in the typological literature, with the critical exception that it disregards modal structures. In contrast, I took a critical stance to Chen's (2013) arguments involving FUTURE TENSE semantics.

Little more needs to be said on the latter point. Chen's (2013) arguments about FUTURE TENSE semantics were not supported. Moreover, differences in how the English and Dutch future tenses correlated with temporal discounting (Chapter 3) undermine arguments about the general semantics of the FUTURE TENSE. Similarly, I hope the importance of modal structures to FTR has by now been demonstrated.

An outstanding question is how well FTR status described elicited future tense use patterns in English, Dutch, and German. There were differences between Dutch and German, both weak-FTR languages. Relative to Germans, Dutch speakers used more present and fewer future tense constructions. This suggests FTR status is overly broad. It points future research towards more granular investigations of language and away from using FTR status as a predictor of further economic outcomes. In general, this conclusion echoes early receptions to Chen's (2013) project by linguists, who point to cases where the FTR system of some language is poorly-described by the weak/strong dichotomy (e.g. Dahl, 2013; McWhorter, 2014; Pullum, 2012; Pereltsvaig, 2011). On the other hand, English speakers used more future tense constructions than either Dutch or German speakers (Study 1, Chapter 2). This suggests that FTR status may still be a useful proxy for FTR grammaticisation, modal complexities notwithstanding.

In conclusion, the results support the proposition that obligatory future tense use in prediction-based main-clause FTR is a good proxy for FTR grammaticisation—English (the strong-FTR language) exhibited more grammaticised FTR than either Dutch or German.

However, the results strongly suggest modal structures are implicated, and point future research toward confronting this possibility in further languages. Conversely, while FTR status was supported with caveats, Chen's (2013) arguments about the FUTURE TENSE encoding temporal distance or precision were not supported.

6.4 Putting this thesis in context: Contemporary concerns about linguistic relativity

Several key issues surrounding the linguistic relativity hypothesis were identified in the introduction (Section 1.2). These critical perspectives were illustrated, where possible, with examples from relativity research involving time. The issues were: (a) the correlation problem, i.e. whether language *causes* observed effects; (b) issues of magnitude, i.e. to what extent relativity affects may be (un)important; and (c) concerns around scope, i.e. which mental processes may be subject to relativity affects. In this section, I revisit these critical perspectives with a view to characterising my contributions. I then touch briefly on the potential for cross-pollination between relativity work on intertemporal decision making and relativity work on space-time metaphors.

6.4.1 The correlation problem: Does the present work contribute to evidence that language *causes* observed effects?

Whether an observed correlation represents a causal relationship can be difficult to establish. This is just as true in relativity research as in any other area of science. To help think clearly about this issue, I introduced a theoretical framework for parsing how correlational evidence can strengthen grounds for causal inference: The Royal College Framework (Section 1.2.1). The framework was developed to parse the causal evidence provided by the association between smoking cigarettes and lung cancer (Pearl & Mackenzie, 2018). The framework includes notions such as: *specificity* (one cause should have one effect, not a litany); *strength* (a higher dose should have a stronger effect); *consistency* (multiple repeated findings in different populations); and *coherence* (the plausibility of the effect given relevant theory and knowledge).

The repeated reproduction of key findings throughout this dissertation helps build evidence for the *consistency* of the effects under investigation. Particularly, higher use of low-certainty modal language in English was shown to mediate increased discounting in three experiments (Studies 2 and 3, Chapter 4; Appendix G), and future tense use was repeatedly shown to mediate decreased discounting (Studies 1–3, Chapter 3; Appendix G). In terms of *coherence*, results are generally *coherent* with theoretical accounts which suggest that grammatical coding of some domain can spotlight notional aspects of experience (in this case uncertainty about the future) and thereby influence non-linguistic cognition (Wolff & Holmes, 2011) (in this case [risky] intertemporal decisions). These aspects of the work seem relatively straightforward, and as such I devote this section to the discussion of *specificity* and *strength*.

6.4.1.1 Specificity (one cause, one effect)

The notion of *specificity* is pertinent to the interpretation of several findings in this dissertation. For instance, in Chapter 4, I tested whether the modal or temporal hypothesis is a more plausible theoretical account for the impact of FTR grammaticisation on (risky) intertemporal decision making. Under the modal hypothesis, it is predicted that English speakers should discount more in combined outcomes involving risk and delay, but not necessarily in outcomes only involving delay (though see Section 3.5.2). Both predictions were generally supported. Effects of FTR status in time-preference-only decisions were mixed. In one experiment no effect emerged (Study 1, Chapter 3); in the other, English speakers temporally discounted more (Study 2, Chapter 3). A relevant point is that the significant finding involved a time-preference task in which delays and reward amounts were calibrated to identify small differences, and a large sample size (N = 606 participants) designed to have high power was used. Apart from this—arguably special—circumstance, other time-preference effects were null. On the other hand, I found consistent effects of FTR status on combined discounting. This suggests that, in general, the *specific* predictions of the modal hypothesis were supported.

A second example is that the modal hypothesis predicts that the *English* → *low-certainty* → *combined discounting* path will be positive (more low-certainty language, more discounting). On the other hand, it predicts that the *English* → *future tense* → *subjective temporal distance* → *combined discounting* path will be null. This is because under the modal account, the future tense is not hypothesised to encode notions of temporal distance or precision. The modal predictions were both true, another example of *specificity*. Taken together, these findings involve confirmatory and dis-confirmatory tests against specific positive and negative predictions to develop causal evidence. In general, this evidence supports the modal hypothesis.

6.4.1.2 Strength (a higher dose should have a larger effect)

The best example of dose response in the findings presented in this dissertation comes from Study 3 of Chapter 4, which involved testing for linguistic exposure effects (English vs. Dutch) on risky intertemporal outcomes in Dutch-English bilinguals. Critically, I found that English proficiency moderated the effect of language condition on discounting via low-certainty modal verb use. High-proficiency English speakers tended to use more low-certainty modal verbs—a pattern of language use which closely mirrored language-level differences between English and Dutch first-language speakers (see *Figs. 4.3* and *4.5*). As a function of this, high-proficiency bilinguals also tended to discount risky future outcomes more than low-proficiency bilinguals. As such, increased exposure to English grammatical norms (either in terms of their life experience with speaking English or in terms of language production during the experiment) resulted in a stronger effect. This is a prototypical example of *strength.* These findings therefore bolster the causal inference that increased use of low-certainty modal language in English FTR results in increased combined discounting.

An additional question is whether *strength* can be invoked to interpret cross-population results where English and Dutch speakers were tested. In terms of the population-level statistical effects of FTR status, there are not gradations of dose. In my statistical models, there is a single parameter which captures the average effect of English compared with Dutch. This means the concept of *strength* does not apply to the language-level mediation effects, i.e because the effect of language is constant. However, throughout the empirical work of this dissertation, I found various correlations between speaker-level use of FTR structures and (risky) intertemporal decision making. For instance, increased use of low-certainty modal terms was associated with increased combined discounting (Studies 2 and 3, Chapter 4), and use of the future tense was associated with decreased temporal discounting (Chapter 3). (See also Appendix G for correlations between other linguistic structures and risky-intertemporal outcomes.) This suggests a kind of "dose effect", with the strong caveat that reciprocal causality is likely involved. It remains unclear whether language is the "dose", or speaker-level risk and time preferences are. In fact neither are really a dose in the medical sense. They are simply correlated. However, increased speaker-level use of low-certainty terms was associated with increased discounting. This suggests dose effects of a similar nature would emerge between languages which differ in the relative extent to which they oblige low-certainty modal FTR (see Section 6.6.2 for further discussion).

6.4.2 The issue of magnitude: Representative design and relativity effects

A compelling aspect of Chen's (2013) project is that his outcomes—and many outcomes in follow ups—involve critical behaviours which most people would agree are important. Effects include attitudes about climate change policy (Mavisakalyan et al., 2018), macroeconomic outcomes (Hübner & Vannoorenberghe, 2015b, 2015a), or propensity to commit suicide (Lien & Zhang, 2020) (also see Section 1.3.3.2). Additionally, reported effect sizes are large. Chen (2013) reports weak-FTR speakers being 30% more likely to have saved money in the past year! This research involves large effects on outcomes which have much to do with, as McWhorter puts it, how "people... manage existence" (2014, p. 28). As such, Chen's (and subsequent) work represents a unique challenge to critical perspectives which characterise relativity effects as trivial. It is perhaps telling that McWhorter—who wrote an entire book which might be summarised as a critique of the magnitude of relativity effects (2014)—does not criticise Chen's work from this perspective (2014, pp. 95–9). However, although this thesis addresses many shortcomings of Chen's (2013) approach, my work mainly involves tightly controlled experimentation involving discounting of relatively trivial monetary amounts (£/€10 in Chapter 3, and Study 1 of Chapter 4, and £/€35 in the other studies). Similarly, average differences between English and Dutch speakers in terms of the subjective evaluation of future rewards were not large is absolute terms, typically on the order £/€1–2 at middling delays and probabilities. Of course, future work using larger values could reveal larger effects; nonetheless, this work involved relatively trivial monetary amounts.

A pertinent question is therefore whether the present findings represent a robust refutation of magnitude-based criticisms of linguistic relativity. I humbly suggest the answer is yes. There are two main reasons that support this conclusion. Firstly, experimentally-elicited risk and time preferences consistently predict real-world outcomes, e.g. real spending (Bickel et al., 2010), and financial outcomes such as income levels and financial mismanagement (Hamilton & Potenza, 2012; Xiao & Porto, 2019) (see also Section 1.3.3.1). As such, the measures used in this thesis likely correlate with important real-world behaviour. The relative novelty of research using risky intertemporal measures means their predictive validity is less well understood than strictly time or strictly risk preferences. However, there is a plausible theoretical argument that such designs best represent real-world (risky) intertemporal decisions. Therefore, they may exhibit a high degree of predictive validity over real behaviour, although, to date, there is not much empirical evidence for this. The second reason my results form a counterpoint to magnitude-based criticisms of linguistic relativity is as follows. I have significantly revised the theoretical account for how FTR status is hypothesised to impact outcomes. My results call into question the temporal framework developed by Chen (2013). However, they are among the strongest evidence to date for the more general hypothesis that FTR grammar impacts (risky) intertemporal behaviour, and buttress real-world impacts of FTR status on intertemporal outcomes (see Section 1.3.3.2). Therefore, the present investigation suggests that language impacts important, real-world decisions. Such outcomes seem difficult to disregard as peculiarities of studiously artificial experimentation (cf. McWhorter, 2014).

6.4.3 The issue of scope: Risk and time preferences may represent fertile ground for relativity research

A relevant consideration to linguistic relativity research is the question of which—and to what extent—perceptual and cognitive processes may be impacted by language. Two concerns were identified as pertinent (Section 1.2.3): (a) to what extent perception may be (im)penetrable by language; and (b) concerns around the durability of relativity effects. With regard to (b), distinctions were made between whether relativity effects are limited to: (i) preparation for speech, i.e. "thinking for speaking" (Slobin, 1987); (ii) online speech production, i.e. thinking *with* language[5]; or (iii) whether they persist in offline cognition, i.e. thinking *after* language (see Wolff & Holmes, 2011). With regard to these considerations, it was suggested that Judgement and Decision-Making (JDM) processes have theoretically relevant properties. In particular, JDM processes involve higher-level cognition, which is theoretically expected to be permeable by language. Moreover, the effects of decisions are durable in a way that percepts are not. This means that the downstream consequences of online language effects may also be durable in the context of JDM processes. Below, I outline my results in light of these critical perspectives.

6.4.3.1 Perceptual (im)penetrability

In the introduction, I pointed out that a relevant distinction is between perception and cognition (Section 1.2.3.1). Here, perception is understood as involving lower-level processes which translate environmental features into stable representative percepts, and cognition is understood as involving the thoughts, beliefs, desires, and decisions about those percepts (Firestone & Scholl, 2015). I framed the relevant issue as being to what extent perceptual processes are encapsulated and therefore (im)penetrable by "top-down" effects of language. I pointed out that some authors have argued for pervasive influences of language on perception (Lupyan et al., 2020), while others have suggested that apparent evidence for top-down effects can be explained by other experimental or theoretical factors (Firestone & Scholl, 2015). With regards to temporal perception, a critical touch point was theoretical work by Montemayor (2019). He makes a distinction between early-stage temporal perception and late-stage temporal processing. By his account, early-stage perception involves short (sub 1 second) duration estimations and simultaneity judgements, and late-stage processing involves the "[integration of] temporal structures in terms of linguistically based representations and symbolic formats, paradigmatically, explicit judgements" (Montemayor, 2019, p. 138).

I suggest that JDM processes, as well as time-slider type tasks, are canonical examples of what Montemayor deems to involve late-stage processing. As such, my results do not bear on debate about penetrability of temporal perception by linguistic representations.[6] Rather, the approach of this dissertation has been to focus on higher-level JDM processes that *are* more likely to be influenced. My findings suggest that JDM processes—in particular, (risky) intertemporal decision making—are indeed affected by language. This buttresses an argument against a tacit assumption which permeates this debate. Particularly, accounts of perceptual

[5]The distinction between (i) and (ii) is described by Wolff and Holmes (2011) as the difference between "thinking before language" on one hand and "thinking with language" (p. 255) on the other. For example, non-linguistic cognitive impacts might be limited to the process of preparing to form a linguistic utterance (i), or might continue while language is used during a task (ii).

[6]In fact, my time-slider findings do not suggest late-stage processing is impacted by language. Specifically, while Dutch speakers rated temporal events as farther away (Study 2, Chapter 2; Study 1, Chapter 4), this was unrelated to any linguistic data.

impenetrability view perception as more fundamental (Montemayor, 2019) and may therefore trivialise linguistic impacts on other cognitive domains like attention, memory, judgement, and decision making. In this thesis, I have provided evidence that language impacts the psychological discounting processes which underpin decision making about such critical areas as investment in personal education (Figlio et al., 2016), climate change policy (Mavisakalyan et al., 2018), and suicide (Lien & Zhang, 2020). While none of these involve perception, they are not very narrow in scope (cf. Montemayor, 2019).

6.4.3.2 Issues around the durability of linguistic effects

Here, the relevant question is about when, and for how long, linguistic effects persist. In particular, an important question is: (a) whether effects are limited to the process of preparing for a linguistic utterance (Slobin, 1987); (b) whether they persist during online language use (Wolff & Holmes, 2011); or (c) whether effects endure when language is not being used (Wolff & Holmes, 2011).

Admittedly, it remains unclear which of these are involved in my results. In particular, the extent to which (risky) intertemporal decision making involves the engagement of linguistic representations appears to be understudied. It was not empirically addressed in this dissertation. One approach to developing this would be to use visual paradigms. For instance, probability and delay could be represented in terms of placement on x- and y-axes. Amount could be represented iconically, e.g. with stacks of paper currency. In such paradigms, fixation crosses could be used to identify whether cross-linguistic effects are stronger in the right visual field. Non-linguistic stimuli would allow for linguistic interference tasks to be added. These could be used to identify whether decision-making biases remain when the language-production system is otherwise engaged.

This discussion raises a possible limitation of this thesis. The linguistic nature of the testing paradigms could have inadvertently resulted in participants using linguistic representations to make judgements. Standard, linguistically-represented binary choice paradigms were implemented throughout. This means that the effects I report might not persist in the context of non-linguistic tasks. It should be remembered that the choices participants had to make in the (risky) intertemporal tasks did not use the linguistic structures (modal verbs, future tenses) under investigation, e.g. "Which would you prefer: _____ now or a(n) _____% chance of _____ in two-weeks?" Nonetheless, the tasks were represented linguistically. Future work might directly investigate whether my results are reproduced in visually-represented (risky) intertemporal-choice paradigms. A relevant caveat is whether real-world (risky) intertemporal choices are ever presented entirely non-linguistically. It is hard to think of any. Linguistically-represented tasks may therefore exemplify better representative design—and predictive validity—than a studiously constructed visual representation of a (risky) intertemporal decision.

6.4.4 Space-time metaphors and intertemporal decision making: Is cross-pollination plausible?

The growing body of work by economists which investigates the effects of FTR status on economic behaviour seems to have been largely disregarded by cognitive scientists and linguists who count linguistic relativity among their principal concerns. For instance, two recent reviews of linguistic relativity research make no mention of Chen's (2013) work, or any subsequent follow up (Bohnemeyer, 2020; Casasanto, 2016). In a contemporaneous review of Chen's early findings, Everett (2013) points to online criticism from linguistic circles (see Section 1.3.2.1 for a review of this criticism). He notes that Chen's, "claim is unlikely to be accepted without significant evidence beyond that which he presents" (p. 137). Much follow up does not provide new *kinds* of evidence, but rather uses Chen's methods with new outcomes. Everett's summary may therefore be a reasonable description of ongoing dismissive attitudes among linguists and cognitive scientists. A major goal of this thesis has been to provide the kind of evidence such critics justifiably expect. My hope is this thesis can begin to open avenues for fruitful cross-pollination between economists, cognitive scientists, and linguists working on relativity.

Most research to date into relativity effects on the perception and representation of time has involved work on space-time metaphors (see Section 1.2), but there may be grounds for crossover between this work and research on FTR grammar and psychological discounting. An obvious question is whether cross-linguistic differences in space-time metaphors might impact (risky) intertemporal decisions. An idea comes to mind. It involves the overlap between *evidential* and *epistemic* notions. In this context Nuyts (2000) provides a useful distinction between *evidentiality* on the one hand and *epistemic modality* on the other. By his account, *epistemic modality* involves the expression of what is believed to be true relative to a speaker's knowledge about the world (p. 28). *Evidentiality* involves, "relating a state of affairs to types of informational source-domains" (p. 26). While the two constructs may be separable, they are closely related. For instance, evidential marking tends to arise in conceptual domains where medium epistemic certainty is involved. In high- or low-certainty conceptual domains, evidential marking may simply add no new information (see Givón, 1982). Additionally, the grammatical encoding of evidential sources tends to track epistemic notional categories. For instance, hearsay tends to be marked as less reliable than visual perception (Nuyts, 2000, p. 27). In fact, this contrast evokes a cross-culturally robust metaphor: *Knowledge is vision* (Lakoff & Johnson, 1980; also see San Roque, Kendrick, Norcliffe, & Majid, 2018).

This becomes salient to cross-pollination between work on space-time metaphors and work on FTR status because there are cross-linguistic differences in the extent to which space-time metaphors encode epistemic and evidential notions about future time. English is prototypical of the cross-linguistically common metaphors, *the future is in front of ego* and *the past is* *behind ego*. This can be seen in a multitude of expressions, such as *the week ahead*, or *back in those days* (Núñez & Sweetser, 2006). The Amerindian language Aymara, which is spoken in the Andean upper country of Western Bolivia, is a relevant contrast (Núñez & Sweetser, 2006). In Aymara, the predominant space-time metaphor for time's arrow is reversed, i.e. *the future is behind (and the past in front) of ego*. In particular, the Aymaran past time marker also means "in front" and the future time marker also means "behind" Núñez and Sweetser (2006):

(6.2) a. *nayra* *mara*
eye/front/sight year
'last year'

b. *qhipa* *uru*
back/behind day
'a future day'

As evidence that this metaphor impacts speakers' construals of time's arrow, Núñez and Sweetser (2006) video-recorded Aymara speakers while they were talking. The researchers then coded the video recordings for whether speakers' co-speech gestures were directed in front or behind them when they were making temporal reference to the past and future. Speakers tended to gesture backwards when talking about the future, and frontwards when talking about the past. This was moderated by degree of exposure to Spanish, which uses future-forward, past-back space-time metaphors, like English (another *strength* effect).

The motivations for the Aymaran metaphor are a fairly transparent response to the fact that the past can be "seen" (in memory), while the future cannot. Such a metaphor may therefore implicitly encode the epistemic paucity people face when thinking about future events. The authors support this conclusion. They suggest that the reversal of the metaphor for time's arrow may be related to the Aymaran evidential marking system, which strictly requires speakers to mark evidential sources using inflectional or syntactic structures which indicate whether an event has been directly visually perceived or not (Núñez & Sweetser, 2006, p. 440). The parallel between this and a space-time metaphor derived from being unable to "see" the future should be obvious. Might this impact (risky) intertemporal outcomes? Initial evidence for the plausibility of this hypothesis has emerged. Romero, Craig, and Kumar (2019) found that manipulating the representation of time's arrow between the English-canonical *horizontal* axis, and the English non-canonical *vertical* axis impacted speakers' time preferences. The question becomes whether the Aymaran space-time metaphor spotlights future uncertainty and results in increased psychological discounting relative to groups which use more typical *future is in front of ego* space-time metaphors.

6.5 Back to the beginning: Epistemic modal FTR and the Hopi time controversy

It would not be appropriate, or relevant, to evaluate Whorf's claims about Hopi beliefs in light of the empirical findings of this thesis. However, a few things may be said about his account of the way European languages encode temporal notions. Whorf contrasts Hopi with "Standard Average European" (Whorf, 1956, p. 138) conceptions of time. The Hopi "future tense" suffix, *-ni* is suggested to mark the modally-flavoured "subjective" domain, best characterised by the *bouletic* modal notion of *hope*, but tinged also with notions of causality (Whorf, 1956, pp. 61–2). According to Whorf, the Hopi construe hope and prayer as causally involved in the ongoing emergence of the future (Whorf, 1956, pp. 61–2). This is contrasted with European notions which are basically Newtonian. Space is three-dimensional and static, time is one-dimensional and flowing, and ternary tenses separate past, present, and future (Whorf, 1956, p. 59). How might the results of this thesis be brought to bear on this account?

Firstly, Whorf's easy agglomeration of all European languages into a uniform Standard Average European is undermined by the differences I identify between speakers of different European languages. However, this thesis is hardly the first relativity research which has involved contrasts between speakers of European languages (e.g. Majid, Gullberg, et al., 2007). A further relevant point can be made about Whorf's account of *-ni*. Whorf makes the case that *-ni* is not a future tense because it encodes modal notions. This account seems to suffer from essentialist reasoning about what constitutes a future tense. It fails to consider that the division of labour between modal, aspectual, and temporal meanings may constitute an essential feature of the FUTURE TENSE (Dahl, 2000a). Indeed, similar controversies have characterised discussion of the status of *will* and *zullen*. But such observations could be made from review of the typological literature on FTR and discussions of the specific future tenses involved (e.g. Dahl, 1985, 2000a; Bybee & Dahl, 1989; Bybee et al., 1994; Broekhuis & Verkuyl, 2014; Sarkar, 1998; Salkie, 2010; Fries, 1956; Huddleston, 1995; Enç, 1996).

However, the results in this dissertation do suggest English deals with (future) time using a variety of modal structures which are equal to the future tense in terms of frequency (i.e. the Reddit corpus analysis, Study 1, Chapter 5). Such multi-dimensional complexities are not well described by Whorf's Newtonian model. A relevant contrast may therefore be between which modal base is most important to FTR. Whorf argues that bouletic notions characterise the semantics of the Hopi suffix *-ni*. The results of the extensive FTR-elicitation tasks undertaken in this dissertation suggest "Standard Average European" FTR is overwhelmingly preoccupied with the encoding of epistemic notions (see also Giannakidou, 2014; Giannakidou & Mari, 2018). It should be noted that *will* grammaticised from *willan* 'to want', so it used to have bouletic semantics. At the time of Whorf's analysis, *-ni* could have been in the process of grammaticising from earlier bouletic to later temporal meanings. This is possible, though the fact that Whorf presents *-ni* as a bound morpheme indicates it was highly grammaticised already. Regardless, the present results suggest epistemic modality is a critical functional category to the expression of linguistic FTR in the Standard Average European languages analysed herein.

There is an argument to be made that the primacy of epistemic modal notions may be contingent on a Newtonian model of reality which sees the future as something to be predicted based on presently known facts about reality, and scientific approximations of physical processes. However, it should be considered that the FTR-elicitation questionnaire was not conceived to elicit bouletic language. This was because it seemed unclear why bouletic modal notions should impact (risky) intertemporal outcomes. This is pertinent to any conclusions about real-world language use made on the basis of the elicited data. Bouletic modal FTR was the second or third most frequent FTR type in the Reddit corpus (Study 1, Chapter 5). It was second only to the future tense, *Fig. 5.3*. This may further undermine the contrast between Hopi (putatively bouletic) FTR and English (putatively epistemic) FTR. In this perspective, Whorf's analysis suffers from simplification of FTR in general, and "European" FTR in particular. Both European and Hopi languages and cultures quite naturally seem to be concerned with encoding modal notions about future time, likely because the future cannot be known at the time of speech. The nature of Whorf's presentation of the Hopi culture may simply obscure underlying common ground (Pinker, 1994). Unfortunately, given the influence of colonial language and culture on Hopi, deciding anything conclusive about these ideas is longer possible.

6.6 Limitations and potential future directions

In this discussion, I pointed to various limitations as they have come up. For instance, the complex interactions which characterised the relationship between future tense use and intertemporal decisions in Dutch (Chapter 3) were never conclusively understood (I discussed this limitation above in Section 6.1.1.2). Additionally, (risky) intertemporal choices were presented linguistically. Do they persist in non-linguistic tasks? Are they impacted by linguistic interference? Such questions remain un-answered (discussed above in Section 6.4.3.2). I identified how future research could explore different speaking contexts. I suggested Jakobson's (1960) schema of *referential, emotive*, and *conative* contexts would be a good starting point (discussed above in Section 6.2.1.1). Limitations of the early drafts of the FTR-elicitation questionnaire were outlined in the introduction to this thesis (Section 1.4.1.3, and see Appendix I). Below I outline further limitations with a view to characterising how they could be addressed in future research.

6.6.1 Temporal and probability beliefs were not exhaustively tested

The time-slider tasks only tested the temporal-distance mechanism hypothesised by Chen (2013), not the temporal-precision mechanism. This leaves open the possibility that Chen's (2013) ideas about temporal precision are true to some extent. This would be inconsistent with the finding that English future tense was negatively associated with (risky) temporal discounting (Chapter 3). Nonetheless, future research could develop measures of temporal precision. For instance, participants could iteratively provide multiple ratings for a single objective distance. Spread statistics would capture the precision with which they represented temporal distance. Another potential limitation is that while an explicit measure of temporal representation was used—the time-slider tasks—there was no corresponding measure of subjective representations of probability. The modal hypothesis predicts that English speakers should construe objective future probabilities as less certain than Dutch speakers. This was not explicitly tested. Rather, I found that combined discounting outcomes were associated with low-certainty modal language use (Chapter 4). This suggests that beliefs about probability and/or risk were involved in mediating between language use and combined discounting, but this was not directly tested.

In fact, recent work has increasingly investigated how representations of time and probability factor into psychological discounting outcomes. This is compared with approaches which characterise non-constant discounting rates over objective time and probability (see L. Green & Myerson, 2004). Interestingly, curvilinear discounting rates over objective time are explained by log-scaled representations of temporal distance (Zauberman et al., 2009; B. K. Kim & Zauberman, 2009; Bradford et al., 2019), while frequency and/or probability is also characterised by log-scaled representations (Zhang et al., 2016).[7] Future research might use the techniques developed in this literature to investigate whether English and Dutch speakers represent probabilities differently, as is predicted by the modal hypothesis.

6.6.2 Language sample size and generalisability

An important limitation is simply the size of the language sample. The most general finding is that—insofar as the contrast between English and Dutch is concerned—Chen's (2013) temporal hypothesis was not supported. Rather, the results strongly suggest that higher obligatory encoding of low-certainty modality in English is responsible for English speakers' present-orientation—the modal hypothesis. Whether this is typical of strong-FTR languages is unclear. The generalisability of the findings is therefore unclear, though the results strongly suggest future work should not disregard modality. The expression of future time quite naturally tracks the epistemic deficit people face when linguistically predicating future events. Modal expressions arise. I would be surprised if this basic finding did not generalise. In

[7]See Appendix E for a discussion of how these findings informed my modelling decisions.

fact, it is well-known that the entanglement of future time with modality is cross-linguistically robust (Bybee et al., 1994; Dahl, 2000a; Giannakidou & Mari, 2018; Palmer, 2001).

Future research on FTR grammar and economic behaviour could use the methods developed in this thesis to expand the language sample while giving due treatment to modal complexities. Further cross-linguistic distinctions could be explored. In fact, Chen (2013) identifies four breakpoints in FTR grammaticisation: (a) languages with no future tense to speak of (e.g. Finnish, Mandarin); (b) languages with optional periphrastic future tenses (Dutch, German); (c) languages with obligatory periphrastic future tenses (English); and (d) languages with obligatory bound morphological future tenses (French, Spanish). This thesis investigated the contrast between two of these levels (b versus c). Future work might use similar techniques to explore how future tenses and modal grammar correlate with (risky) intertemporal outcomes in languages which typify other categories.

A further point is that the temporal mechanisms hypothesised by Chen (2013) might affect intertemporal decision making in languages which have graded future tense systems. In such FTR systems, different future tense markers are used for temporally proximal vs. temporally distal FTR. For instance, the Northeastern Bantu language Gĩkũyũ in Kenya makes use of such a system (Cable, 2013; also see Levinson & Majid, 2013). Chen's temporal hypothesis seems more appropriate to graded future tenses, which encode notions of temporal distance and/or precision.

In general, future exploration of cross-linguistic contrasts will progress best in small steps. Additional investigations should give due attention to the FTR systems under investigation (see Mavisakalyan & Weber, 2018). Large-sample statistical approaches to testing relativity hypotheses can produce useful insights and motivate future investigation (Roberts & Winters, 2013). However, such methods lack the granularity necessary to develop causal evidence or mechanistic understanding of effects of cross-linguistic differences on cognition and behaviour (Roberts et al., 2015).

6.6.3 Cultural confounding and bilinguals

Relativity research must untangle cultural, environmental, and linguistic factors which contribute to cognitive differences (see Section 1.2.1). Linguistic training paradigms are a common strategy used to demonstrate that language *causes* observed impacts. A good example comes from relativity work on psychophysical perceptions of musical pitch (Dolscheid, Shayan, Majid, & Casasanto, 2013). Like English, Dutch speakers describe musical pitch using height-based metaphors, e.g. *hoog* 'high' or *laag* 'low'. On the other hand, Farsi speakers use width-based metaphors, describing high pitches as *nāzok* 'thin' and low pitches as *koloft* 'thick'. To identify the impacts of this cross-linguistic difference, Dolscheid et al. (2013) used a psychophysical paradigm based on work with space-time metaphors (see Section 1.2). Participants heard a tone and were presented with a vertical line which varied either in height (the height condition) or thickness (the thickness condition). Tone pitch and line height/thickness were permuted to systematically vary. Participants reproduced the tone by singing. Dutch participants' pitch reproductions were impacted in the height condition, and Farsi participants' responses were impacted in the thickness condition—the effects reflected the language-canonical musical pitch metaphors (Dolscheid et al., 2013). While implicating language, the authors point out that this result was nonetheless correlational. As such, they conducted a language-training experiment. When Dutch speakers were trained to use Farsi-like pitch metaphors, a significant effect appeared in the thickness condition (Dolscheid et al., 2013). These methods provide causal evidence that musical pitch metaphors impacted participants' responses. In general, such training paradigms are a good way of demonstrating that language plays a causal role in producing an outcome of interest, because they allow researchers to exogenously manipulate language.

Such training methods were not felicitous to my research question. In particular, the relevant differences between English and Dutch FTR were less a matter of quality as extent. The languages share a system of modal auxiliary future time markers (which include future tenses and low-certainty modal verbs). English grammar simply places an increased obligation on speakers to use low-certainty modals for main-clause predictions about the future. While I did employ a future vs. present production task (Study 3, Chapter 3), training Dutch speakers to use *kunnen* 'may' more frequently would have risked transparently revealing the purposes of the experiment. Moreover, it would have rendered the elicited language data difficult to interpret. I made use of matched probabilities in the FTR-elicitation and risky intertemporal tasks to merge data and estimate mediated effects. Directing participants to use low-certainty language would have undermined such methods.

Another way to clarify the causal role of language is using linguistic interference tasks (see Section 1.2.3.2). However, such methods were also not suitable for my research. Principles of representative design led to the use of linguistically-represented (risky) intertemporal choices to measure psychological discounting. Adding linguistic interference to a linguistic measure would have placed incoherent task demands on participants. As I have pointed out above (Section 6.4.3.2), future research could implement non-linguistic discounting measures which could be used under linguistic interference.

In light of these issues, one strategy employed in this dissertation to isolate the role of language specifically was to work with bilingual participants (Study 3, Chapter 4). Dutch-English bilinguals completed the FTR-elicitation task and a risky-intertemporal-choice task, first in one language, then in the other. Since such individuals arguably share a culture, these methods go some way to isolating the effects of language from culture.

However, a limitation of this approach is that bilingual individuals are to some extent also bicultural (see S. X. Chen, 2015). The idea is that language learning is usually associated with a specific cultural context. When bilinguals switch between languages, this may activate the values, beliefs, attitudes, and construals which are associated with that cultural context (Grosjean, 1982; Ervin, 1964; for a critique of Chen's original findings from this perspective, see Sedivy, 2012). In this perspective, manipulations of language are to some extent also manipulations of culture. There is evidence to support this proposition. For instance, it is sometimes the case that outcomes pursuant to language manipulations in bilinguals are moderated by time spent accommodating to the second-language cultural environment, and not linguistic facility (Akkermans, Harzing, & van Witteloostuijn, 2010). This suggests that cultural—and not linguistic—exposure is sometimes responsible for driving effects when bilinguals switch between languages. On the other hand, work which demonstrates language-switching effects in "compound" bilinguals—who learn two languages in the same cultural environment—indicates this is not always the case (e.g. S. X. Chen, Benet-Martínez, Wu, Lam, & Bond, 2013; S. X. Chen & Bond, 2010).

How might my bilingual experiment be characterised in the context of these issues? Firstly, it involved the elicitation of language. Elicited FTR differences were shown to impact risky intertemporal outcomes. This indicates that language is involved in the transmission of (risky) time preferences, if not that cross-linguistic grammatical differences are solely driving outcomes. Secondly, English proficiency reflected cross-linguistic differences. Participants with higher English proficiency used more low-certainty modals, and this moderated the indirect effect of language condition on risky intertemporal decisions. Although I did not measure accommodation to Anglophone culture, this suggests that linguistic exposure was involved in producing outcomes.

This discussion brings up a relevant theoretical consideration: It would be incoherent to give a language-independent account of the idea that bilinguals are bicultural. It would collapse into what some have called the core idea of linguistic relativity: the notion that culture influences cognition *via* language (Gumperz & Levinson, 1996, p. 1). How could language switching activate cultural values and beliefs, if language were not implicated in the maintenance and transmission of such values? Some accounts go so far as to suggest that language and culture are inseparable (S. X. Chen, 2015). As such, the idea that culture confounds effects of bilingual language switching relies on the premise that language is a unique transmitter of cultural and individual beliefs, values, and attitudes—a central idea of linguistic relativity (Everett, 2013; Gumperz & Levinson, 1996). In fact, my language elicitation methods may have identified the grammatical feature most likely to transmit present-orientation among English speaking people: the obligation to use a modal verb when making predictions about the future. Regardless, future research could develop paradigms to further develop causal evidence that cross-linguistic differences in FTR grammar affect psychological discounting. One option would be to prime people using a comprehension task: Participants could read texts with more of one sort of construction than another. If they are not producing text, demand characteristics might not be the same as with the production-based priming paradigms employed in this dissertation.

6.6.4 Causal order: Could discounting have caused elicited FTR differences?

An important consideration in mediation modelling is whether there are plausible alternative accounts for how variables relate (see Hayes, 2013, pp. 172–83). If there are, researchers should specify and compare models which test alternative hypotheses (Hayes, 2013). A critical point is that, in many cases, answers are not in the data: Theoretical and practical considerations of the causal process involved must be used to prune the large number of model specifications which are usually statistically possible (Hayes, 2013; also see Pearl & Mackenzie, 2018). In this section, I discuss my findings in light of such considerations.

Issues of causal order are simplified when X is an exogenous experimental factor, as in the bilingual study. I can be confident that discounting and FTR differences did not affect experimental assignment to language condition. However, the $M \rightarrow Y$ is correlational. It could have been the case that $X \rightarrow Y \rightarrow M$, rather than $X \rightarrow M \rightarrow Y$. In other words, high discounters could have been using more low-certainty language, rather than the other way around. Several things may be said about this. The fact that English proficiency moderated the $X \rightarrow M \rightarrow Y$ path supports the conclusion that language is impacting discounting not the other way around. It is difficult to conceive of how discounting and English proficiency might interact to produce differences in FTR use, were exposure to English not involved in producing discounting differences. Additionally, my specification of $language \rightarrow FTR \rightarrow discounting$ is based on strong theoretical priors that language-level grammatical differences should impact speaker-level FTR. Finally, the FTR-elicitation task temporally preceded the discounting task, which suggests language more likely impacted discounting results than the other way around. For these reasons, I do not deem it necessary to specify alternate models for this finding.

These issues are even more complicated when X is not an exogenous experimental factor, as in the cross-cultural experiments involving English and Dutch speakers. In such cases, alternative specifications such as $M \rightarrow Y \rightarrow X$ or $Y \rightarrow X \rightarrow M$ should be considered (Hayes, 2013). Luckily, in the case of my research, there are good theoretical reasons to treat the English-Dutch contrast as exogenous. Even it if may be confounded by non-linguistic cultural differences, how could an individual's risk preferences, time preferences, or FTR habits impact the language they first learn? Further, it is difficult to give a plausible account of how individual-differences could affect language-level grammatical features. A relevant consideration is the speed of grammatical change. Languages evolve at different speeds depending on a variety of factors, such as the size of the speech community (Nettle, 1999). However, English FTR grammar is slow-changing. For instance, the English *de andative* future time marker *be going to* has been used as a future marker since at least 1490, but by the late 20$^{\text{th}}$ century, it was still infrequent in some dialects, accounting for only 10–15% of FTR contexts (Tagliamonte, Durham, & Smith, 2014). In fact, FTR status is within the top 6% of the most stable linguistic features in the World Atlas of Language Structures[8] (Roberts et al., 2015). Even trait-like characteristics such as psychological discounting (Ohmura et al., 2006) develop over shorter durations, i.e. the lifespan of an individual. It is difficult to see how such relatively ephemeral phenomena could impact durable language-level aspects of grammar. A second relevant consideration is that grammatical change appears to arise from frequency effects which are emergent in nature (Bybee, 2006). Populations, not individuals, give rise to grammar. While individual differences may collectively bring about cultural ones, the influence of single individuals is negligible. For these reasons it is difficult to provide an account of how speaker-level discounting could impact language-level grammar.

As such, I can be fairly confident about my specification of language as X. However, it could still have been the case that my language variable (English - Dutch) indexed cultural differences. Such cultural differences could have driven increased discounting, which might have caused higher use of low-certainty terms in English speakers. In other words, the $M \rightarrow Y$ relationship could have been reversed. I therefore tested the alternate account that *English* → *discounting* → *low-certainty language*. The relevant finding is Study 2 of Chapter 4, in which English speakers used more low-certainty language and consequently discounted more. This alternative specification resulted in a null indirect effect (Appendix E). This supports the conclusion that English grammar affects psychological discounting rather than Anglophone culture affecting FTR usage.

6.7 Conclusions

In conclusion, this thesis broadly attests to the complicated entanglement of temporal and modal notions which arises as a function of our inability to perceive the future. The grammaticisation of FTR may involve the encoding not of *when* but *if* a future outcome is going to occur. The consequences to psychological discounting appear to be concordant with this. Future time reference grammar may impact psychological discounting processes through the obligation to use low-certainty modals, rather than future tenses. Such grammatical differences may impact risky intertemporal outcomes through causing increased <u>combined</u>, rather than <u>temporal</u>, discounting in strong-FTR speakers. The temporal account given in Chen (2013), and widely reproduced, should therefore be reconsidered. It remains to be seen

[8]This is a large database of phonological, grammatical, and lexical properties of languages (Dryer & Haspelmath, 2013).

whether the patterns I have identified as characterising English and Dutch generalise. It is not clear that they *will*, but they *might*.

References

Addessi, E., Paglieri, F., & Focaroli, V. (2011). The ecological rationality of delay tolerance: Insignts from capuchin monkeys. *Cognition*, *119*(1), 142–147. doi: 10.1016/j.cognition.2010.10.021

Aguinis, H., Gottfredson, R., & Culpepper, S. (2013). Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. *Journal of Management*, *39*(6), 1490–1528. doi: 10.1177/0149206313478188

Ainslie, G., & Herrnstein, R. J. (1981). Preference reversals and delayed reinforcement. *Animal Learning & Behaviour*, *9*, 476–482. doi: 10.3758/BF03209777

Akkermans, D., Harzing, A. W., & van Witteloostuijn, A. (2010). Cultural accommodation and language priming: Competitive versus cooperative behavior in a prisoner's dilemma game. *Management International Review*, *50*(5), 559–583. doi: 10.1007/s11575-010-0053-0

Al-Fadhel, S. Z., Al-Hakeim, H. K., Al-Dujaili, A. H., & Maes, M. (2019). IL-10 is associated with increased mu-opioid receptor levels in major depressive disorder. *European Psychiatry*, *57*, 46–51. doi: 10.1016/j.eurpsy.2018.10.001

Al-Mosaiwi, M., & Johnstone, T. (2018). In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science*, *6*(4), 529–542. doi: 10.1177/2167702617747074

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). Washington: American Psychiatric Publishing.

American Psychiatric Association. (2015). *Anxiety disorders: DSM-5 selections*. Washington: American Psychiatric Publishing.

Anderson, H., & Goolishian, H. A. (1988). Human systems as linguistic systems: Preliminary and evolving ideas about the implications for clinical theory. *Family Process*, *27*(4), 371–393. doi: 10.1111/j.1545-5300.1988.00371.x

Andreoni, J., & Sprenger, C. (2012). Risk preferences are not time preferences. *American Economic Review*, *102*(7), 3357–3376. doi: 10.1257/aer.102.7.3357

Araújo, D., Davids, K., & Passos, P. (2007). Ecological validity, representative design, and correspondence between experimental task constraints and behavioral setting. *Ecological Psychology*, *19*(1), 69–78. doi: 10.1080/10407410709336951

Athanasopoulos, P., & Casaponsa, A. (2020). The Whorfian brain: Neuroscientific approaches to linguistic relativity. *Cognitive Neuropsychology*, *37*(5–6), 393–412. doi: 10.1080/02643294.2020.1769050

Austin, J. L. (1962). *How to do things with words.* Oxford: Oxford University Press.

Banerjee, A., & Urminsky, O. (2017). What you are getting and what you will be getting: Testing whether verb tense affects intertemporal choices. *Advances in Consumer Research*, *45*, 512–513.

Bard, E. G., Robertson, D., & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, *72*(1), 32–68. doi: 10.2307/416793

Barlow, M. (2013). Individual differences and usage-based grammar. *International Journal of Corpus Linguistics*, *18*(4), 443–478. doi: 10.1075/ijcl.18.4.01bar

Barrault, S., & Varescon, I. (2013). Cognitive distortions, anxiety, and depression among regular and pathological gambling online poker players. *Cyberpsychology, Behavior, and Social Networking*, *16*(3), 183–188. doi: 10.1089/cyber.2012.0150

Bartoń, K. (2020). *MuMIn: Multi-model inference.* Retrieved from `https://cran.r-project.org/package=MuMIn`

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). doi: 10.18637/jss.v067.i01

Bergs, A., & Heine, L. (2010). Mood in English. In B. Rothstein & R. Thieroff (Eds.), *Mood in the languages of Europe* (pp. 103–116). Amsterdam: John Benjamins. doi: 10.1075/slcs.120.06ber

Bernhofer, J., Costantini, F., & Kovacic, M. (2019). Risk attitudes, investment behavior and linguistic variation. *University Ca' Foscari of Venice, Department of Economics Research Paper Series*, *34*(15), 1–50.

Białaszek, W., Ostaszewski, P., Green, L., & Myerson, J. (2019). On four types of devaluation of outcomes due to their costs: Delay, probability, effort, and social discounting. *The Psychological Record*, *69*, 415–424. doi: 10.1007/s40732-019-00340-x

Bickel, W. K., Jones, B. A., Landes, R. D., Christensen, D. R., Jackson, L., & Mancino, M. (2010). Hypothetical intertemporal choice and real economic behavior: Delay discounting predicts voucher redemptions during contingency-management procedures. *Experimental and Clinical Psychopharmacology*, *18*(6), 546–552. doi: 10.1037/a0021739

Bickel, W. K., Odum, A. L., & Madden, G. J. (1999). Impulsivity and cigarette smoking: Delay discounting in current, never, and ex-smokers. *Psychopharmacology*, *146*(4), 447–454. doi: 10.1007/PL00005490

Blaszczynski, A., & McConaghy, N. (1989). Anxiety and/or depression in the pathogenesis of addictive gambling. *Substance Use and Misuse*, *24*(4), 337–350. doi: 10.3109/10826088909047292

Bloom, P., & Keil, F. C. (2001). Thinking through language. *Mind and Language*, *16*(4), 351–367. doi: 10.1111/1468-0017.00175

Bochnak, M. R. (2019). Future reference with and without future marking. *Language and Linguistics Compass*, *13*(1), 1–23. doi: 10.1111/lnc3.12307

Boe, B., PythonCoderAS, Damgaard, Jarhill0, LilSpazJoekp, Voussoir, … Tmels (2020). *Python Reddit API wrapper.* Retrieved from `https://github.com/praw-dev/praw`

Bohnemeyer, J. (2000). *Time reference across languages.* Tilburg. Retrieved from `http://www.linguistics.berkeley.edu/$\sim$avelino/FieldLinguistics/Bohnemayer_lot.pdf`

Bohnemeyer, J. (2020). Linguistic relativity: From Whorf to now. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, & T. E. Zimmerman (Eds.), *The Wiley Blackwell companion to semantics.* Hoboken: John Wiley & Sons. doi: 10.1126/science.202.4366.409

Boroditsky, L. (2006). Linguistic relativity. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 917–921). Hoboken: John Wiley & Sons, Ltd.

Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently? *Cognition*, *118*(1), 123–129. doi: 10.1016/j.cognition.2010.09.010

Boroditsky, L., & Gaby, A. (2010). Remembrances of times east: Absolute spatial representations of time in an Australian aboriginal community. *Psychological Science*, *21*(11), 1635–1639. doi: 10.1177/0956797610386621

Boroditsky, L., & Schmidt, L. A. (2000). Sex, syntax, semantics. In *Proceedings of the annual meeting of the Cognitive Science Society* (Vol. 22, pp. 92–105). doi: 10.1515/9783112316009-010

Bouma, L. (1973). *The semantics of the modal auxiliaries in contemporary German.* The Hague: Moulten.

Bouma, L. (1975). On contrasting the semantics of the modal auxiliaries of German and English. *Lingua*, *37*, 313–339.

Bowlby, J. (1969). *Attachment and loss* (2nd ed.). New York: Basic.

Bradford, W. D., Dolan, P., & Galizzi, M. M. (2019). Looking ahead: Subjective time perception and individual discounting. *Journal of Risk and Uncertainty*, *58*, 43–69. doi: 10.1007/s11166-019-09298-1

Branigan, H. P., & Pickering, M. J. (2017). An experimental approach to linguistic representation. *Behavioral and Brain Sciences*, *40, E282*. doi: 10.1017/S0140525X16002028

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, *1*(3), 185–216.

Broekhuis, H., & Verkuyl, H. J. (2014). Binary tense and modality. *Natural Language and Linguistic Theory*, *32*(3), 973–1009. doi: 10.1007/s11049-013-9213-9

Bross, F. (2012). Modal particles and the common ground. *Helikon. A Multidisciplinary Online Journal*, *2*, 182–209. doi: 10.1075/pbns.114.19kar

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

Brunswik, E. (1952). *Conceptual framework of psychology.* Chicago: University of Chicago Press.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley: University of California Press.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *10*(1), 1–28. doi: 10.18637/jss.v080.i01

Bybee, J. (2006). From usage to grammar: The mind's response to repitition. *Language*, *82*(4), 711–733. doi: 10.1353/lan.2006.0186

Bybee, J., & Dahl, Ö. (1989). The creation of tense and aspect systems in the languages of the world. *Studies in Language*, *13*(1), 51–103.

Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar.* Chicago: University of Chicago Press.

Byloo, P., Kastein, R., & Nuyts, J. (2007). On *certainly* and *zeker*. In M. Hannay & G. Steen (Eds.), *Structural-functional studies in English grammar: In honour of Lachlan Mackenzie* (pp. 35–58). Amsterdam: John Benjamins. doi: 10.1075/hcp.5

Bylund, E., & Athanasopoulos, P. (2017). The Whorfian time warp: Representing duration through the language hourglass. *Journal of Experimental Psychology: General*, *146*(7), 911–916. doi: 10.1037/xge0000314

Cable, S. (2013). Beyond the past, present, and future: Towards the semantics of 'graded tense' in Gĩkũyũ. *Natural Language Semantics*, *21*(3), 219–276. doi: 10.1007/s11050-012-9092-3

Cariani, F., & Santorio, P. (2018). *Will* done better: Selection semantics, future credence, and indeterminacy. *Mind*, *127*(505), 129–165. doi: 10.1093/mind/fzw004

Carleton, R. N., Mulvogue, M. K., Thibodeau, M. A., McCabe, R. E., Antony, M. M., & Asmundson, G. J. G. (2012). Increasingly certain about uncertainty: Intolerance of uncertainty across anxiety and depression. *Journal of Anxiety Disorders*, *26*(3), 468–479. doi: 10.1016/j.janxdis.2012.01.011

Carroll, J. B., & Casagrande, J. B. (1958). The function of language classification in behaviour. In E. Maccoby, T. Newcomb, & E. Hartlet (Eds.), *Readings in social psychology* (pp. 18–31). New York: Holt.

Casasanto, D. (2005). Crying "Whorf". *Science*, *307*(5716), 1721–1722. doi: 10.1126/science.307.5716.1720b

Casasanto, D. (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Language Learning*, *58*(SUPPL. 1), 63–79. doi: 10.1111/j.1467-9922.2008.00462.x

Casasanto, D. (2016). Linguistic relativity. In N. Riemer (Ed.), *Routledge handbook of semantics* (pp. 159–174). New York: Routledge. doi: 10.4324/9781315670522-217

Casasanto, D., & Boroditsky, L. (2003). Do we think about time in terms of space? In *Proceedings of the 25th annual meeting of the Cognitive Science Society* (pp. 216–221).

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, *106*(2), 579–593. doi: 10.1016/j.cognition.2007.03.004

Casasanto, D., Boroditsky, L., Phillips, W., Green, J., Goswami, S., Bocanegra-Thiel, S., … Gil, D. (2004). How deep are the effects of language on thought: Time estimation in speakers of English, Indonesian, Greek, and Spanish. In *Proceedings of the 26th annual meeting of the Cognitive Science Society* (pp. 575–580).

Chang, A. X., & Manning, C. D. (2012). SUTIME: A library for recognizing and normalizing time expressions. In *Proceedings of the 8th international conference on language resources and evaluation, LREC 2012* (pp. 3735–3740).

Chen. (2012, February). Whorfian economics. *Language Log.* Retrieved from `https://languagelog.ldc.upenn.edu/nll/?p=3792`

Chen. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review*, *103*(2), 690–731. doi: 10.1257/aer.103.2.690

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*, 1–12. doi: 10.1016/j.euroecorev.2019.103307

Chen, S., Cronqvist, H., Ni, S., & Zhang, F. (2017). Languages and corporate savings behavior. *Journal of Corporate Finance*, *46*, 320–341. doi: 10.1016/j.jcorpfin.2017.07.009

Chen, S. X. (2015). Toward a social psychology of bilingualism and biculturalism. *Asian Journal of Social Psychology*, *18*(1), 1–11. doi: 10.1111/ajsp.12088

Chen, S. X., Benet-Martínez, V., Wu, W. C., Lam, B. C., & Bond, M. H. (2013). The role of dialectical self and bicultural identity integration in psychological adjustment. *Journal of Personality*, *81*(1), 61–75. doi: 10.1111/j.1467-6494.2012.00791.x

Chen, S. X., & Bond, M. H. (2010). Two languages, two personalities? Examining language effects on the expression of personality in a bilingual context. *Personality and Social Psychology Bulletin*, *36*(11), 1514–1528. doi: 10.1177/0146167210385360

Chi, J. D., Su, X., Tang, Y., & Xu, B. (2018). Is language an economic institution? Evidence from R&D investment. *SSRN Electronic Journal*, 1–48. doi: 10.2139/ssrn.3262822

Chomsky, N. (1956). *Aspects of the theory of syntax.* Massachesetts: MIT Press.

Clark, H. (1973). Space, time, semantics, and the child. In *Cogntitive devlopment and the aquisition of language* (pp. 27–63). New York: Academic Press.

Collins, M. (2002). Discriminative training methods for hidden Markov models: Theory and experiments with perceptron algorithms. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)* (pp. 1–8). Philadelphia: Association for Computational Linguistics.

Comrie, B. (1984). Review: Hopi time by Ekkehart Malotki. *Australian Journal of Linguistics*, *4*(1), 128–141. doi: 10.1080/07268608408599323

Comrie, B., Haspelmath, M., & Bickel, B. (2015). *The Leipzig glossing rules.* Retrieved 2021-03-08, from `https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf`

Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K., & Mitchell, M. (2015). CLPsych 2015 shared task: Depression and PTSD on Twitter. In *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality* (pp. 31–39). Denver: Association for Computational Linguistics. doi: 10.3115/v1/w15-1204

Cox, D. J., & Dallery, J. (2016). Effects of delay and probability combinations on discounting in humans. *Behavioural Processes*, *131*, 15–23. doi: 10.1016/j.beproc.2016.08.002

Croijmans, I., Speed, L. J., Arshamian, A., & Majid, A. (2020). Expertise shapes multimodal imagery for wine. *Cognitive Science*, *44*(5), 1–16. doi: 10.1111/cogs.12842

Cyffer, N., Ebermann, E., & Ziegelmeyer, G. (2009). *Negation patterns in West African languages and beyond.* Amsterdam: John Benjamins.

Dabrowska, E. (2010). Naive v. expert intuitions: An empirical study of acceptability judgments. *Linguistic Review*, *27*(1), 1–23. doi: 10.1515/tlir.2010.001

Dabrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism*, *2*(3), 219–253. doi: 10.1075/lab.2.3.01dab

Dahl, Ö. (1985). *Tense and aspect systems.* Oxford: Blackwell.

Dahl, Ö. (2000a). The grammar of future time reference in European languages. In Ö. Dahl (Ed.), *Tense and aspect in the languages of Europe* (pp. 309–328). Berlin/New York: Mouton de Gruyter. doi: 10.1515/9783110197099.1.3

Dahl, Ö. (Ed.). (2000b). *Tense and aspect in the languages of Europe.* Berlin/New York: Mouton de Gruyter. doi: 10.1515/9783110197099.1.3

Dahl, Ö. (2000c). The tense-aspect systems of European languages in a typological perspective. In Ö. Dahl (Ed.), *Tense and aspect in the languages of Europe* (pp. 3–25). Berlin/New York: Mouton de Gruyter. doi: 10.1515/9783110197099.1.3

Dahl, Ö. (2013, March). Stuck in the futureless zone. *Diversity Linguistics Comment.* Retrieved from `https://dlc.hypotheses.org/360`

Dahl, Ö., & Kós Dienes, D. (1984). *Selected working papers from the Tense-Mood-Aspect Project.* Stockholm: Institute of Linguistics, University of Stockholm.

Davis, F. C., Neta, M., Kim, M. J., Moran, J. M., & Whalen, P. J. (2016). Interpreting ambiguous social cues in unpredictable contexts. *Social Cognitive and Affective Neuroscience*, *11*(5), 775–782. doi: 10.1093/scan/nsw003

De Choudhury, M., Counts, S., Horvitz, E. J., & Hoff, A. (2014). Characterizing and predicting postpartum depression from shared Facebook data. *Proceedings of the ACM conference on computer supported cooperative work, CSCW*, 625–637. doi: 10.1145/2531602.2531675

de Haan, F. (1999). Evidentiality and epistemic modality: Setting boundaries. *Southwest Journal of Linguistics*, *18*(1), 1–34.

Dewa, C. S., & McDaid, D. (2011). Investing in the mental health of the labor force: Epidemiological and economic impact of mental health disabilities in the workplace. In I. Z. Schultz & E. Sally Rogers (Eds.), *Work accommodation and retention in mental health* (pp. 33–51). Berlin: Springer Science+Business Media. doi: 10.1007/978-1-4419-0428-7

Dilling, C. A., & Rabin, A. I. (1967). Temporal experiences in depressive states and schizophrenia. *Journal of Counsulting Psychology*, *31*(6), 604–608.

Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science*, *24*(5), 613–621. doi: 10.1177/0956797612457374

Donaldson, B. (2017). *Dutch: A comprehensive grammar.* Abingdon: Routledge.

Dryer, M., & Haspelmath, M. (Eds.). (2013). *The world atlas of language structures.* Leipzig. Retrieved from `https://wals.info/`

Du, W., Green, L., & Myerson, J. (2002). Cross-cultural comparisons of discounting delayed and probabilistic rewards. *The Psychological Record*, *52*, 479–492. doi: 10.1007/BF03395199

Durkin, C., Hartnett, E., Shohamy, D., & Kandel, E. R. (2020). An objective evaluation of the beholder's response to abstract and figurative art based on construal level theory. *Proceedings of the National Academy of Sciences*, *117*(33), 19809–19815. doi: 10.1073/pnas.2001772117

Efstathiadi, L. (2009). *The use of epistemic modality markers as a means of hedging and boosting by L1 and L2 speakers of Modern Greek: A corpus-based study in informal letter-writing* (Unpublished doctoral dissertation). Aristotle University of Thessaloniki.

Enç, M. (1996). Tense and modality. In S. Lappin (Ed.), *The handbook of contemporary semantic theory* (pp. 345–58). Oxford: Blackwell.

Ervin, S. M. (1964). Language and TAT content in bilinguals. *Journal of Abnormal and Social Psychology*, *68*, 500–507.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. London: St. Martins Press.

Evans, N., Bergqvist, H., & San Roque, L. (2018). The grammar of engagement I: Framework and initial exemplification. *Language and Cognition*, *10*(1), 110–140. doi: 10.1017/langcog.2017.21

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–48; discussion 448–494. doi: 10.1017/S0140525X0999094X

Everett, C. (2013). *Linguistic relativity: Evidence across languages and cognitive domains* (G. Kristiansen, I. de Mendoza, & F. J. Ruiz, Eds.). Berlin: de Gruyter.

Explosion AI. (2020). *spaCy*. Berlin: Explosion AI. Retrieved from `https://spacy.io/`

Eyal, T., & Liberman, N. (2012). Morality and psychological distance: A construal level theory perspective. In M. Mikulincer (Ed.), *The social psychology of morality: A construal level theory perspective* (pp. 185–202). Washington D.C.: APA.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, *133*(4), 1645–1692. doi: 10.1093/qje/qjy013

Fasan, M., Gotti, G., Kang, T., & Liu, Y. (2016). Language FTR and earnings management: International evidence. *SSRN Electronic Journal*, 1–49. doi: 10.2139/ssrn.2763922

Fehringer, C. (2018). Internal constraints on the use of *gaan* versus *zullen* as future markers in spoken Dutch. *Nederlandse Taalkunde*, *22*(3), 359–387. doi: 10.5117/nedtaa2017.3.fehr

Felton, J. W., Strutz, K. L., McCauley, H. L., Poland, C. A., Barnhart, K. J., & Lejuez, C. W. (2020). Delay discounting interacts with distress tolerance to predict depression and alcohol use disorders among individuals receiving inpatient substance use services. *International Journal of Mental Health and Addiction*, *18*(5), 1416–1421. doi: 10.1007/s11469-019-00163-5

Figlio, D., Giuliano, P., Özek, U., & Sapienza, P. (2016). Long-term orientation and educational performance. *National Bureau of Economic Research Working Paper Series*, *10147*. doi: 10.3386/w22541

Firestone, C., & Scholl, B. J. (2015). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences*, *39, E229*, 1–77. doi: 10.1017/S0140525X15000965

Forder, L., He, X., & Franklin, A. (2017). Colour categories are reflected in sensory stages of colour perception when stimulus issues are resolved. *PLoS ONE*, *12*(5), 1–16. doi: 10.1371/journal.pone.0178097

Frahling, G. (2017). *Linguee German English dictionary*. Köln: DeepL GmbH.

Freeman, K. B., Green, L., Myerson, J., & Woolverton, W. L. (2009). Delay discounting of saccarin in rhesus monkeys. *Behavioural Processes*, *82*, 214–218. doi: 10.1016/j.beproc.2009.06.002

Freeman, K. B., Nonnemacher, J., Green, L., Myerson, J., & Woolverton, W. L. (2012). Delay discounting in rhesus monkeys: Equivalent discounting of more and less preferred sucrose concentrations. *Learning & Behavior*, *40*, 54–60. doi: 10.3758/s13420-011-0045-3

Fries, C. C. (1956). The expression of the future. *Language Learning*, *7*(3-4), 125–133. doi: 10.1111/j.1467-1770.1956.tb01207.x

Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit non-linguistic task. *Cognitive Science*, *34*, 1430–1451. doi: 10.1111/j.1551-6709.2010.01105.x

Fuhrman, O., Mccormick, K., Chen, E., Jiang, H., Shu, D., Mao, S., & Boroditsky, L. (2011). How linguistic and cultural forces shape conceptions of time: English and Mandarin time in 3D. *Cognitive Science*, *35*, 1305–1328. doi: 10.1111/j.1551-6709.2011.01193.x

Galor, O., Özak, Ö., & Sarid, A. (2016). Geographical origins and economic consequences of language structures. *IZA Discussion Paper Series*, *10379*.

Garami, J., & Moustafa, A. A. (2019). Probability discounting of monetary gains and losses in opioid-dependent adults. *Behavioural Brain Research*, *364*, 334–339. doi: 10.1016/j.bbr.2019.02.017

Gay, V., Hicks, D. L., Santacreu-Vasut, E., & Shoham, A. (2018). Decomposing culture: An analysis of gender, language, and labor supply in the household. *Review of Economics of the Household*, *16*(4), 879–909. doi: 10.1007/s11150-017-9369-x

Gelman, A., Goodrich, B., Gabry, J., & Vehtari, A. (2018). *R*-squared for Bayesian regression models. *The American Statistician*, *73*(3), 307–309. doi: 10.1080/00031305.2018.1549100

Gentner, D., & Boroditsky, L. (2001). Individuation, reality, and early word learning. In M. Bowerman & S. C. Levinson (Eds.), *Language aquisition and conceptual development* (pp. 215–256). Cambridge: Cambridge University Press.

Gerra, G., Zaimovic, A., Zambelli, U., Timpano, M., Reali, N., Bernasconi, S., & Brambilla, F. (2000). Neuroendocrine responses to psychological stress in adolescents with anxiety disorder. *Neuropsychobiology*, *42*(2), 82–92. doi: 10.1159/000026677

Giannakidou, A. (2014). The futurity of the present and the modality of the future: A commentary on Broekhuis and Verkuyl. *Natural Language and Linguistic Theory*, *32*, 1011–1032. doi: 10.1007/s11049-014-9234-z

Giannakidou, A. (2017). Epistemic future and epistemic MUST: Nonveridicality, evidence, and partial knowledge. In J. Blaszczak, A. Giannakidou, D. Klimek-Jankowska, & K. Migdalski (Eds.), *Mood, aspect, modality revisited: New answers to old questions.* Chicago: Chicago Scholarship Online.

Giannakidou, A., & Mari, A. (2018). A unified analysis of the future as epistemic modality: The view from Greek and Italian. *Natural Language and Linguistic Theory*, *36*(1), 85–129. doi: 10.1007/s11049-017-9366-z

Gibson, E., & Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, *28*(1-2), 88–124. doi: 10.1080/01690965.2010.515080

Gipper, H. (1977). Is there a linguistic relativity principle? In R. Pinxten (Ed.), *Special university lecture in linguistics at the University College of the University of London.* London. Retrieved from `https://www.iai.spk-berlin.de/fileadmin/dokumentenbibliothek/Indiana/Indiana_5/IND_05_Gipper.pdf`

Givón, T. (1982). Evidentiality and epistemic space. *Studies in Language*, *6*, 265–337.

Gjelsvik, B., Lovric, D., & Williams, J. M. G. (2018). Embodied cognition and emotional disorders: Embodiment and abstraction in understanding depression. *Journal of Experimental Psychopathology*, *9*(3). doi: 10.5127/pr.035714

Gkotsis, G., Oellrich, A., Velupillai, S., Liakata, M., Hubbard, T. J., Dobson, R. J., & Dutta, R. (2017). Characterisation of mental health conditions in social media using informed deep learning. *Scientific Reports*, *7*(45141). doi: 10.1038/srep45141

Gleitman, L., & Papafragou, A. (2012). New perspectives on language and thought. In K. Holyoak & R. Morrison (Eds.), *The Oxford handbook of thinking and reasoning.* New York: Oxford University Press. doi: 10.1093/oxfordhb/9780199734689.013.0028

Goldstone, R. L., & Lupyan, G. (2016). Discovering psychological principles by mining naturally occurring data sets. *Topics in Cognitive Science*, *8*(3), 548–568. doi: 10.1111/tops.12212

Goldwin, M., & Behar, E. (2012). Concreteness of idiographic periods of worry and depressive rumination. *Cognitive Therapy and Research*, *36*(6), 840–846. doi: 10.1007/s10608-011-9428-1

Gotti, G., Roberts, S. G., Fasan, M., & Robertson, C. (2020). Language in economics and accounting research: The role of linguistic history. *SSRN Electronic Journal.* doi: 10.2139/ssrn.3628501

Green, L., & Myerson, J. (1996). Exponential versus hyperbolic discounting of delayed outcomes: Risk and waiting time. *American Zoologist*, *36*(4), 496–505. doi: 10.1093/icb/36.4.496

Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin*, *130*(5), 769–792. doi: 10.1037/0033-2909.130.5.769

Green, L., Myerson, J., & Ostaszewski, P. (1999). Amount of reward has opposite effects on the discounting of delayed and probabalistic outcomes. *Joural of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 418–427. doi: 10.1037//0278-7393.25.2.418

Green, L., Myerson, J., & Vanderveldt, A. (2014). Delay and probability discounting. In *The Wiley Blackwell handbook of operant and classical conditioning* (pp. 374–409). doi: 10.1002/9781118468135.ch13

Green, P., & MacLeod, C. J. (2016). simr: An *R* package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. doi: 10.1111/2041-210X.12504

Grillon, C., Baas, J. P., Lissek, S., Smith, K., & Milstein, J. (2004). Anxious responses to predictable and unpredictable aversive events. *Behavioral Neuroscience*, *118*(5), 916–924. doi: 10.1037/0735-7044.118.5.916

Groen, R. N., Ryan, O., Wigman, J. T., Riese, H., Penninx, B. W., Giltay, E. J., ... Hartman, C. A. (2020). Comorbidity between depression and anxiety: Assessing the role of bridge mental states in dynamic psychological networks. *BMC Medicine*, *18*(1), 1–17. doi: 10.1186/s12916-020-01738-z

Grosjean, F. (1982). *Life with two languages: An introduction to bilingualism.* Cambridge, MA: Harvard University Press.

Grupe, D., & Nitschke, J. (2013). Uncertainty and anticipation in anxiety: An integrated neurobiological and psychological perspective. *Nature Neuroscience*, *14*(7), 488–501. doi: 10.1038/nrn3524

Gu, Y., Mol, L., Hoetjes, M., & Swerts, M. (2017). Conceptual and lexical effects on gestures: The case of vertical spatial metaphors for time in Chinese. *Language, Cognition and Neuroscience*, *32*(8), 1048–1063. doi: 10.1080/23273798.2017.1283425

Guin, B. (2017). Culture and household saving. *European Central Bank Working Paper Series*, *2069*, 1–53.

Gumperz, J. J., & Levinson, S. C. (Eds.). (1996). *Rethinking linguistic relativity.* Cambridge: Cambridge University Press.

Günther, F., Müller, H., Schmid, H.-j., & Geyer, T. (2016). *Constructions in cognitive contexts: Why individuals matter in linguistic relativity research.* Berlin: de Gruyter.

Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: An integrative review. *Current Opinion in Behavioral Sciences*, *18*, 43–49. doi: 10.1016/j.cobeha.2017.07.005

Gutzmann, D. (2009). Hybrid semantics for modal particles. *Sprache und Datenverarbeitung*, *33*(1-2), 45–59.

Hamilton, K. R., & Potenza, M. N. (2012). Relations among delay discounting, addictions, and money mismanagement: Implications and future directions. *American Journal of Drug and Alcohol Abuse*, *38*(1), 30–42. doi: 10.3109/00952990.2011.643978.Relations

Hart, K. L., Brown, H. E., Roffman, J. L., & Perlis, R. H. (2019). Risk tolerance measured by probability discounting among individuals with primary mood and psychotic disorders. *Neuropsychology*, *33*(3), 417–424. doi: 10.1037/neu0000506

Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. *Language*, *86*(3), 663–687. doi: 10.1353/lan.2010.0021

Hayes, A. (2013). *Introduction to mediation, moderation, and conditional process analysis.* New York: The Guilford Press.

Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, *93*(1), 10–20. doi: 10.1037/h0032606

Heider, E. R., & Olivier, D. C. (1972). The structure of the color space in naming and memory for two languages. *Cognitive Psychology*, *3*(2), 337–354. doi: 10.1016/0010-0285(72)90011-4

Heine, B. (1995). On the German *werden* future. In W. Abraham, T. Givón, & S. A. Thompson (Eds.), *Discourse, grammar and typology: Papers in honor of John W.M. Verhaar* (pp. 119–138). Amsterdam: John Benjamins. doi: 10.1075/slcs.27.11hei

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3), 61–83; discussion 83–135. doi: 10.1017/S0140525X0999152X

Hinton, L. (1988). Review: Hopi time by Ekkehart Malotki. *American Indian Quarterly*, *12*(4), 361–364.

Hinvest, N. S., & Anderson, I. M. (2010). The effects of real versus hypothetical reward on delay and probability discounting. *Quarterly Journal of Experimental Psychology*, *63*(6), 1072–1084. doi: 10.1080/17470210903276350

Hirsh, J. B., Mar, R. A., & Peterson, J. B. (2012). Psychological entropy: A framework for understanding uncertainty-related anxiety. *Psychological Review*, *119*(2), 304–320. doi: 10.1037/a0026767

Hockett, C. F. (1963). The problem of universals in language. In J. H. Greenberg (Ed.), *Universals of language* (pp. 1–22). Cambridge, MA: MIT Press.

Hodgins, D. C., & Engel, A. (2002). Future time perspective in pathological gamblers. *Journal of Nervous and Mental Disease*, *190*(11), 775–780. doi: 10.1097/00005053-200211000-00008

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*(2008), 1593–1623. doi: 10.5555/2627435.2638586

Hogeweg, L., Ramachers, S., & Wottrich, V. (2011). *Doch*, *toch*, and *wel* on the table. *Linguistics in the Netherlands*, *28*(2007), 50–60. doi: 10.1075/avt.28.05hog

Honnibal, M., & Johnson, M. (2015). An improved non-monotonic transition system for dependency parsing. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1373–1378). Lisbon: Association for Computational Linguistics. doi: 10.18653/v1/D15-1162

Hopper, P. J. (1996). Some recent trends in grammaticalization. *Annual Review of Anthropology*, *25*(1), 217–236. doi: 10.1146/annurev.anthro.25.1.217

Hübner, M., & Vannoorenberghe, G. (2015a). Patience and inflation. *Munich Personal RePEc Archive*, *65811*. doi: 10.1227/01.NEU.0000349921.14519.2A

Hübner, M., & Vannoorenberghe, G. (2015b). Patience and long-run growth. *Economics Letters*, *137*, 163–167. doi: 10.1016/j.econlet.2015.10.011

Huddleston, R. (1995). The case against a future tense in English. *Studies in Language*, *19*(2), 399–446. doi: 10.5555/2627435.2638586

Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language.* Cambridge: Cambridge University Press.

Huitink, J. (2008). *Modals, conditionals and compositionality* (Unpublished doctoral dissertation). Radboud.

Huitink, J. (2012). Modal concord: A case study of Dutch. *Journal of Semantics*, *29*(3), 403–437. doi: 10.1093/jos/ffr012

Ibarretxe-Antuñano, I. (2008). Path salience in motion events. In J. Guo, E. Lieven, N. Budwig, S. Ervin-tripp, & K. Nakamura (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 403–414). New York: Psychology Press.

Ireland, M. E., Schwartz, H. A., Chen, Q., Ungar, L. H., & Albarracín, D. (2015). Future-oriented tweets predict lower county-level HIV prevalence in the United States. *Health Psychology*, *34*(Supplement), 1252–1260. doi: 10.1037/hea0000279

Jakiela, P., & Ozier, O. (2018). Gendered language. *World Bank Policy Research Working Papers*, *8464*. doi: 10.1596/1813-9450-8464

Jakobson, R. (1960). Concluding statement: Linguistics and poetics. In T. A. Sebeok (Ed.), *Style in language* (pp. 350–377). Cambridge: MIT Press.

Janssen, T. (1989). Die Hilfsverben *werden* (Deutsch) und *zullen* (Niederländisch): Modal oder temporal? In W. Abraham & T. Janssen (Eds.), *Tempus-Aspekt-Modus: Die lexikalischen und grammatischen Formen in den germanischen Sprachen* (pp. 65–84). Tübingen: Neimeyer.

Jenks, C. W., & Lawyer, S. R. (2015). Using delay discounting to understand impulsive choice in socially anxious individuals: Failure to replicate. *Journal of Behavior Therapy and Experimental Psychiatry*, *46*, 198–201. doi: 10.1016/j.jbtep.2014.10.010

Johnson, A. (2021). *Interpretation of products of paths (and Est.) when multiple link functions are combined in a single model.* Retrieved 2021-05-06, from `https://discourse.mc-stan.org/`

Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it. *Molecular Psychiatry*, *17*(12), 1174–1179. doi: 10.1038/mp.2012.105

Karagjosova, E. (2004). *The meaning and function of German modal particles* (Unpublished doctoral dissertation). Saarland University.

Kay, P. (1996). Intra-speaker relativity. In J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 97–114). Cambridge: Cambridge University Press.

Kemmer, S. (1993). *The middle voice*. Amsterdam/Philadelphia: Benjamins.

Kim, B. K., & Zauberman, G. (2009). Perception of anticipatory time in temporal discounting. *Journal of Neuroscience, Psychology, and Economics*, *2*(2), 91–101. doi: 10.1037/a0017686

Kim, J., Kim, Y., & Zhou, J. (2017). Languages and earnings management. *Journal of Accounting and Economics*, *63*(2-3), 288–306. doi: 10.1016/j.jacceco.2017.04.001

Kim, S., & Filimonau, V. (2017). On linguistic relativity and pro-environmental attitudes in tourism. *Tourism Management*, *63*, 158–169. doi: 10.1016/j.tourman.2017.06.012

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology: General*, *128*(1), 78–87. doi: 10.1037//0096-3445.128.1.78

Kirsner, R. S. (1969). The role of *zullen* in the grammar of modern standard Dutch. *Lingua*, *24*, 101–154. doi: 10.1016/0024-3841(70)90070-7

Kish, L. (1965). *Survey sampling*. New York: John Wiley & Sons Ltd.

Kissine, M. (2008). Why *will* is not a modal. *Natural Language Semantics*, *16*(2), 129–155. doi: 10.1007/s11050-008-9028-0

Klecha, P. (2014). Diagnosing modality in predictive expressions. *Journal of Semantics*, *31*(3), 443–455. doi: 10.1093/jos/fft011

Klein, W. (1995). A time-relational analysis of Russian aspect. *Language*, *71*(4), 669–695. doi: 10.2307/415740

Koo, T., Carreras, X., & Collins, M. (2008). Simple semi-supervised dependency parsing. In *Proceedings of ACL-08: HLT* (pp. 595–603). Columbus, Ohio: Association for Computational Linguistics.

Kovacic, M., & Orso, C. E. (2016). Why do some countries fear immigration more than others? Evidence from Europe. *University Ca' Foscari of Venice, Department of Economics Research Paper Series*, *05*. doi: 10.2139/ssrn.2742037

Kratzer, A. (2012). *Modality and conditionals: New and revised perspectives*. Oxford: Oxford University Press.

Lai, M. H., & Kwok, O. M. (2015). Examining the rule of thumb of not using multilevel modeling: The "design effect smaller than two" rule. *Journal of Experimental Education*, *83*(3), 423–438. doi: 10.1080/00220973.2014.907229

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lassiter, D. (2015). Epistemic comparison, models of uncertainty, and the disjunction puzzle. *Journal of Semantics*, *32*(4), 649–684. doi: 10.1093/jos/ffu008

Le Poidevin, R. (2019). The experience and perception of time. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy.* doi: 10.5860/choice.41sup-0181

Leavitt, J. (2011). *Linguistic relativities.* Cambridge: Cambridge University Press.

Ledgerwood, A., & Callahan, S. P. (2012). The social side of abstraction: Psychological distance enhances conformity to group norms. *Psychological Science*, *23*(8), 907–913. doi: 10.1177/0956797611435920

Lee, A. V. D. (2016). *Cultural differences in the use of epistemic modality in the European Parliament* (Unpublished doctoral dissertation). Radboud.

Legler, J., & Roback, P. (2019). *Generalized linear models and multilevel models.* Boston: Bookdown.

Levinson, S. C. (2012). Foreword. In J. B. Carroll, S. C. Levinson, & P. Lee (Eds.), *Language, thought, and reality: Selected writings of Benjamin Lee Whorf, second edition* (pp. vii–xxiii). Cambridge: MIT Press.

Levinson, S. C., & Majid, A. (2013). The island of time: Yélî Dnye, the language of Rossel Island. *Frontiers in Psychology*, *4*, 1–11. doi: 10.3389/fpsyg.2013.00061

Liang, H., Marquis, C., Renneboog, L., & Sun, S. L. (2018). Future-time framing: The effect of language on corporate future orientation. *Organization Science*, *29*(6), 1093–1111. doi: 10.1287/orsc.2018.1217

Liberman, N., & Förster, J. (2009). Distancing from experienced self: How global-versus-local perception affects estimation of psychological distance. *Journal of Personality and Social Psychologysychology*, *97*(2), 203–216. doi: 10.1037/a0015671

Lien, D., & Zhang, S. (2020). Words matter life: The effect of language on suicide behavior. *Journal of Behavioral and Experimental Economics*, *86*, 1–12. doi: 10.1016/j.socec.2020.101536

Liu, W., & Aaker, J. (2007). Do you look to the future or focus on today? The impact of life experience on intertemporal decisions. *Organizational Behavior and Human Decision Processes*, *102*(2), 212–225. doi: 10.1016/j.obhdp.2006.02.004

Locey, M. L., Jones, B. A., & Rachlin, H. (2011). Real and hypothetical rewards. *Judgement and Decision Making*, *6*(6), 552–564. doi: 10.1016/S0003-4975(10)62476-9

Lousdal, M. L. (2018). An introduction to instrumental variable assumptions, validation and estimation. *Emerging Themes in Epidemiology*, *15*(1), 1–7. doi: 10.1186/s12982-018-0069-7

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the Depression Anxiety and Stress Scales* (2nd ed.). Sydney: Psychology Foundation.

Luckman, A., Donkin, C., & Newell, B. R. (2018). Can a single model account for both risky choices and inter-temporal choices? Testing the assumptions underlying models of risky inter-temporal choice. *Psychonomic Bulletin and Review*, *25*(2), 785–792. doi: 10.3758/s13423-017-1330-8

Lucy, J. A. (1992). *Language diversity and thought.* Cambridge: Cambridge University Press.

Lucy, J. A. (1997). Linguistic relativity. *Annual Review of Anthropology*, *26*, 291–312.

Lucy, J. A. (2016). Recent advances in the study of linguistic relativity in historical context: A critical assessment. *Language Learning*, *66*(3), 487–515. doi: 10.1111/lang.12195

Lüdecke, D. (2019). *ggeffects: Marginal effects of regression models.* Retrieved from `https://strengejacke.github.io/ggeffects/articles/ggeffects.html`

Luguri, J. B., Napier, J. L., & Dovidio, J. F. (2012). Reconstruing intolerance: Abstract thinking reduces conservatives' prejudice against nonnormative groups. *Psychological Science*, *23*(7), 756–763. doi: 10.1177/0956797611433877

Lupyan, G., Rahman, R. A., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception. *PsyArxiv.* doi: 10.31234/osf.io/pztmc

Lyons, J. (1968). Grammatical categories. In *Introduction to theoretical linguistics* (pp. 270–333). Cambridge: Cambridge University Press.

Macchia, L., Plagnol, A. C., & Reimers, S. (2018). Does experience with high inflation affect intertemporal decision making? Sensitivity to inflation rates in Argentine and British delay discounting choices. *Journal of Behavioral and Experimental Economics*, *75*, 76–83. doi: 10.1016/j.socec.2018.05.006

Madden, G. J., Begotka, A. M., Raiff, B. R., & Kastern, L. L. (2003). Delay discounting of real and hypothetical rewards. *Experimental and Clinical Psychopharmacology*, *11*(2), 139–145. doi: 10.1037/1064-1297.11.2.139

Majid, A. (2018). Language and cognition. In H. Callan (Ed.), *The international encyclopedia of anthropology.* Hoboken: John Wiley & Sons, Ltd. doi: 10.1002/9781118924396.wbiea2364

Majid, A., Bowerman, M., Van Staden, M., & Boster, J. S. (2007). The semantic categories of cutting and breaking events: A crosslinguistic perspective. *Cognitive Linguistics*, *18*(2), 133–152. doi: 10.1515/COG.2007.005

Majid, A., Gullberg, M., Van Staden, M., & Bowerman, M. (2007). How similar are semantic categories in closely related languages? A comparison of cutting and breaking in four Germanic languages. *Cognitive Linguistics*, *18*(2), 179–194. doi: 10.1515/COG.2007.007

Majstro. (2017). *Multilingual translation dictionary.* Retrieved from `http://www.majstro.com/dictionaries/`

Malotki, E. (1983). *Hopi time: A linguistic analysis of the temporal concepts in the Hopi language* (W. Winter, Ed.). Berlin: Moulten.

Mavisakalyan, A., Tarverdi, Y., & Weber, C. (2018). Talking in the present, caring for the future: Language and environment. *Journal of Comparative Economics*, *46*(4), 1370–1387. doi: 10.1016/j.jce.2018.01.003

Mavisakalyan, A., & Weber, C. (2018). Linguistic structures and economic outcomes. *Journal of Economic Surveys*, *32*(3), 916–939. doi: 10.1111/joes.12247

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *The effect of delay and of intervening events on reinforcement value: Quantitative analyses of behavior, Vol. 5* (pp. 55–73). Hillsdale, NJ: Erlbaum.

McKerchar, T. L., & Renda, C. R. (2012). Delay and probability discounting in humans: An overview. *Psychological Record*, *62*(4), 817–834. doi: 10.1007/BF03395837

McWhorter, J. H. (2014). *The language hoax: Why the world looks the same in any language.* Oxford: Oxford University Press.

Mejía-Cruz, D., Green, L., Myerson, J., Morales-Chainé, S., & Nieto, J. (2016). Delay and probability discounting by drug-dependent cocaine and marijuana users. *Psychopharmacology*, *233*(14), 2705–2714. doi: 10.1007/s00213-016-4316-8

Mezhevich, I. (2008). A Time-relational approach to tense and mood. In *Proceedings of the West Coast conference on formal linguistics* (Vol. 27, pp. 326–334). Somerville, MA: Cascadilla Proceedings Project.

Mindt, D. (1995). *An emprical grammar of the English verb: Modal verbs.* Berlin: Cornelson Verlag.

Montani, I., & Honnibal, M. (2020). *Prodigy: A new annotation tool for radically efficient machine learning.* Berlin: Explosion AI. Retrieved from `https://prod.gy`

Montemayor, C. (2019). Early and late time perception: On the narrow scope of the Whorfian hypothesis. *Review of Philosophy and Psychology*, *10*(1), 133–154. doi: 10.1007/s13164-018-0405-4

Mortelmans, T. (2000). Konjunktiv II and epistemic modals in German. In A. Foolen & F. V. D. Leek (Eds.), *Constuctions in cognitive linguistics: Selected papers from the fifth international Cognitive Linguistics conference* (pp. 191–215). Amsterdam: John Benjamins. doi: 10.1075/cilt.178.12mor

Mortelmans, T. (2009). Modal verbs, modal adverbs, and modal particals in English, German and Dutch: What markers of epistemic necessity and inferential evidentiality might express. In *Modality at work* (pp. 1–2). Lisbon. Retrieved from `http://www.paolapietrandrea.altervista.org/workshop/abstracts/mortelmans.pdf`

Mortelmans, T., Boye, K., & Auwera, J. V. D. (2009). Modals in the Germanic languages. In B. Hansen & F. de Haan (Eds.), *Modals in the languages of Europe: A reference work* (pp. 12–70). Berlin: De Gruyter Mouton.

Moss, S. (2015). On the semantics and pragmatics of epistemic vocabulary. *Semantics and Pragmatics*, *8*(5), 1–81. doi: 10.3765/sp.8.5

Murphy, J. G., Vuchinich, R. E., & Simpson, C. A. (2001). Delayed reward and cost discounting. *Psychological Record*, *51*(4), 571–588.

Muthén, B. O. (2011). Applications of causally defined direct and indirected effects in mediation analysis using SEM in Mplus. *Psychological Bulletin*, *107*(2), 238–246.

Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, *25*, 267–316. doi: 10.2307/271070

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. doi: 10.1901/jeab.2001.76-235

Nakagawa, S., Johnson, P. C., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134). doi: 10.1098/rsif.2017.0213

National Institute of Mental Health. (2019). *Mental illness.* Retrieved from `https://www.nimh.nih.gov/health/statistics/mental-illness`

Nettle, D. (1999). Is the rate of linguistic change constant? *Lingua*, *108*(2-3), 119–136. doi: 10.1016/s0024-3841(98)00047-3

Núñez, R. E., & Sweetser, E. (2006). With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science*, *30*(3), 401–450. doi: 10.1207/s15516709cog0000_62

Nurse, D. (2008). *Tense and aspect in Bantu.* New York: Oxford University Press.

Nuyts, J. (2000). *Epistemic modality, language, and conceptualization.* Amsterdam: John Benjamins. doi: 10.1075/hcp.5

Nuyts, J., & Vonk, W. (1999). Epistemic modality and focus in Dutch. *Linguistics*, *37*(4), 699. doi: 10.1515/ling.37.4.699

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. doi: 10.1126/science.aax2342

Odum, A. L., & Rainaud, C. P. (2003). Discounting of delayed hypothetical money, alcohol, and food. *Behavioural Processes*, *64*, 305–313. doi: 10.1016/S0376-6357(03)00145-1

Ohmura, Y., Takahashi, T., Kitamura, N., & Wehr, P. (2006). Three-month stability of delay and probability discounting measures. *Experimental and Clinical Psychopharmacology*, *14*(3), 318–328. doi: 10.1037/1064-1297.14.3.318

Olmen, D. V., Mortelmans, T., & Auwera, J. V. D. (2009). Grammaticalization and subjectification of the future: The case of English, Dutch and German. In W. Olesky &

P. Stalmaszcyk (Eds.), *Cognitive approaches to language and linguistic data: Studies in honor of Barbara Lewandowska-Tomaszczyk* (pp. 285–306). Frankfurt: Peter Lang AG.

Ostaszewski, P., Green, L., & Myerson, J. (1998). Effects of inflation on the subjective value. *Psychological Bulletin & Review*, *5*(2), 324–333.

Ouimet, A. J., Gawronski, B., & Dozois, D. J. A. (2009). Cognitive vulnerability to anxiety: A review and an integrative model. *Clinical Psychology Review*, *29*(6), 459–470. doi: 10.1016/j.cpr.2009.05.004

Oxford University Press. (2017). *Oxford English dictionary* (M. Proffitt, Ed.). Oxford: Oxford University Press.

Palmer, R. F. (2001). *Mood and modality* (2nd ed.). Cambridge: Cambridge University Press.

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence for eye movements. *Cognition*, *108*, 155–184. doi: 10.1016/j.cognition.2008.02.007

Papastamatelou, J., Unger, A., Giotakos, O., & Athanasiadou, F. (2015). Is time perspective a predictor of anxiety and perceived stress? Some preliminary results from Greece. *Psychological Studies*, *60*(4), 468–477. doi: 10.1007/s12646-015-0342-6

Pearl, J., & Mackenzie, D. (2018). *The book of why.* London: Alen Lane.

Pereltsvaig, A. (2011, October). You save what you speak? *Languages of the World.* Retrieved from `https://www.languagesoftheworld.info/language-and-mind/you-save-what-you-speak.html`

Pérez, E. O., & Tavits, M. (2017). Language shapes people's time perspective and support for future-oriented policies. *American Journal of Political Science*, *61*(3), 715–727. doi: 10.1111/ajps.12290

Petré, P. (2013). Passive auxiliaries in English and German: Decline versus grammaticalisation of bounded language use. In *Comparative studies in early Germanic languages: With a focus on verbal categories* (pp. 71–100). Amsterdam: Benjamins. doi: 10.1075/SLCS.138.04PET

Petry, N. M., Bickel, W. K., & Arnett, M. (1998). Shortened time horizons and insensitivity to future consequences in heroin addicts. *Addiction*, *93*(5), 729–738. doi: 10.1046/j.1360-0443.1998.9357298.x

Pinker, S. (1994). *The language instinct.* London: Penguin.

Poldi, F., Zacharias, C., Levin, M., Hpiedcoq, Urbanski, A., Nestor75, … Aldou (2019). *Twint.* Retrieved from `https://github.com/twintproject/twint`

Pulcu, E., Trotter, P. D., Thomas, E. J., McFarquhar, M., Juhasz, G., Sahakian, B. J., … Elliott, R. (2014). Temporal discounting in major depressive disorder. *Psychological Medicine*, *44*(9), 1825–1834. doi: 10.1017/S0033291713002584

Pullum, G. K. (2012, February). Keith Chen, Whorfian economist. *Language Log.* Retrieved from `http://languagelog.ldc.upenn.edu/nll/?p=3756`

Python Software Foundation. (2017). *Python language reference.* Retrieved from http://www.python.org

R Core Team. (2020). *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rachlin, H., Raineri, A., & Cross, D. (1991). Subjective probability and delay. *Journal of the Experimental Analysis of Behavior*, *55*(2), 233–44. doi: 10.1901/jeab.1991.55-233

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137–141. doi: 10.1007/s11747-019-00710-5

Raineri, A., & Rachlin, H. (1993). The effect of temporal constraints on the value of of money and other commodities. *Journal of Behavioural Decision Making*, *94*, 77–94. doi: 10.1002/bdm.3960060202

Ratcliffe, M. (2015). *Experiences of depression: A study in phenomenology.* Oxford: Oxford University Press.

Rau, J. (2011). *Semantic and syntactic differences between finite and infinitival complements in German* (Unpublished doctoral dissertation). Tübingen.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis and methods.* London: Sage.

Reddit Inc. (2020). *Discover what makes Reddit ads unique.* Retrieved 2020-12-22, from https://www.redditinc.com/advertising/audience

Reddit Inc. (2020). *Reddit.* Retrieved 2020-12-22, from https://www.reddit.com

Roberts, S., & Winters, J. (2013). Linguistic diversity and traffic accidents: Lessons from statistical studies of cultural traits. *PLoS ONE*, *8*(8), 1–13. doi: 10.1371/journal.pone.0070902

Roberts, S., Winters, J., & Chen. (2015). Future tense and economic decisions: Controlling for cultural evolution. *PLoS ONE*, *10*(7), 1–46. doi: 10.1371/journal.pone.0132145

Robertson, C. (2019a). *The FTR-elicitation questionnaire.* Retrieved from https://github.com/cbjrobertson/ftr_questionnaire_master

Robertson, C. (2019b). *The FTR-type classifier.* Retrieved from https://pypi.org/project/ftr-classifier/

Romero, M., Craig, A., & Kumar, A. (2019). Mapping time: How the spatial representation of time influences intertemporal choices. *Journal of Marketing Research*, *56*(4), 620–636. doi: 10.1177/0022243719827967

Rounds, J. S., Beck, J. G., & Grant, D. M. M. (2007). Is the delay discounting paradigm useful in understanding social anxiety? *Behaviour Research and Therapy*, *45*(4), 729–735. doi: 10.1016/j.brat.2006.06.007

Rozenman, M., Sturm, A., McCracken, J. T., & Piacentini, J. (2017). Autonomic arousal in anxious and typically developing youth during a stressor involving error feedback.

European Child and Adolescent Psychiatry, *26*(12), 1423–1432. doi: 10.1007/s00787-017-1001-3

Salkie, R. (2010). *Will*: Tense or modal or both? *English Language and Linguistics*, *14*(2), 187–215. doi: 10.1017/S1360674310000055

Samuel, S., Cole, G., & Eacott, M. J. (2019). Grammatical gender and linguistic relativity: A systematic review. *Psychonomic Bulletin and Review*, *26*(6), 1767–1786. doi: 10.3758/s13423-019-01652-3

Samuelson, P. A. (1937). A note on measurement of utility. *Review of Economic Studies*, *4*(2), 155–161. doi: 10.2307/2967612

San Roque, L., Kendrick, K. H., Norcliffe, E., & Majid, A. (2018). Universal meaning extensions of perception verbs are grounded in interaction. *Cognitive Linguistics*, *29*(3), 371–406. doi: 10.1515/cog-2017-0034

Santorio, P., & Romoli, J. (2017). Probability and implicatures: A unified account of the scalar effects of disjunction under modals. *Semantics and Pragmatics*, *10*(13), 1–61. doi: 10.3765/sp.10.13

Sarkar, A. (1998). The conflict between future tense and modality: The case of *will* in English. *University of Pennsylvania Working Papers in Liguistics*, *5*(2), 91–117.

Saulnier, K. G., Allan, N. P., Raines, A. M., & Schmidt, N. B. (2019). Depression and intolerance of uncertainty: Relations between uncertainty subfactors and depression dimensions. *Psychiatry*, *82*(1), 72–79. doi: 10.1080/00332747.2018.1560583

Schellen, M. (2019). *Laguage and wealth management* (Unpublished doctoral dissertation). University of St.Gallen.

Sedivy, J. (2012, February). Thought experiments on language and thought. *Language Log*. Retrieved from `https://languagelog.ldc.upenn.edu/nll/?p=3797`

Shepherd, J., Rossner, R., & Taylor, J. (1984). *Ways to grammar*. London: Macmillan.

Shoham, A., & Lee, S. M. (2018). The causal impact of grammatical gender marking on gender wage inequality and country income inequality. *Business and Society*, *57*(6), 1216–1251. doi: 10.1177/0007650317696231

Silvas, E. (2018). *Football matches odds: Changes in football winning odds - 32k+ matches, 500+leagues, 4.5k+ teams.* Retrieved from `https://www.kaggle.com/eladsil/football-games-odds`

Simon-Vandenbergen, A.-M., & Aijmer, K. (2007). *The semantic field of modal certainty: A corpus-based study of modal adverbs*. Berlin: de Gruyter.

Slobin, D. (1987). Thinking for speaking. In *Proceedings of the thirteenth annual meeting of the Berkeley Linguistics Society* (pp. 435–445). doi: 10.3765/bls.v13i0.1826

Slobin, D. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In M. Hickman & S. Robert (Eds.), *Space in languages: Linguistic systems and cognitive categories* (pp. 59–81). Amsterdam: John Benjamins.

Slobin, D., Ibarretxe-Antuñano, I., Kopecka, A., & Majid, A. (2014). Manners of human gait: A crosslinguistic event-naming study. *Cognitive Linguistics*, *25*(4), 701–741. doi: 10.1515/cog-2014-0061

Sluijs, R. V. (2011). *The modal system of 20th century Negerhollands* (Unpublished doctoral dissertation). Radboud University.

Smart, R. G. (1968). Future time perspectives in alcoholics and social drinkers. *Journal of Abnormal Psychology*, *73*(1), 81–83. doi: 10.1037/h0025449

Sorace, A., & Keller, F. (2005). Gradience in linguistic data. *Lingua*, *115*, 1497–1524. doi: 10.1016/j.lingua.2004.07.002

Sovey, A. (2011). Instrumental variables estimation in political science: A readers' guide. *American Journal of Political Science*, *55*(1), 188–200. doi: 10.1111/j.1540-5907.2010.00477.x

Stan Development Team. (2020). *Stan modeling language users guide and reference manual.* Retrieved from `https://mc-stan.org`

Steinglass, J. E., Lempert, K. M., Choo, T. H., Kimeldorf, M. B., Wall, M., Walsh, B. T., … Simpson, H. B. (2017). Temporal discounting across three psychiatric disorders: Anorexia nervosa, obsessive compulsive disorder, and social anxiety disorder. *Depression and Anxiety*, *34*(5), 463–470. doi: 10.1002/da.22586

Stevens, C. M. (1995). On the grammaticalization of German *können, dürfen, sollen, mögen, müssen,* and *wollen.* *American Journal of Germanic Linguistics & Literatures*, *7*(2), 179–206. doi: 10.1017/S1040820700001591

Stevens, J. R., Halliman, E. V., & Hauser, M. D. (2005). The ecology and evolution of patience in two New World monkeys. *Biology Letters*, *1*(2), 223–226. doi: 10.1098/rsbl.2004.0285

Stevens, J. R., & Mühlhoff, N. (2012). Intertemporal choice in lemers. *Behavioural Processes*, *89*(2), 121–127. doi: 10.1016/j.beproc.2011.10.002.Epub2011

Sutter, M., Angerer, S., Glätzle-rützler, D., & Lergetporer, P. (2015). The effect of language on economic behavior: Experimental evidence from children's intertemporal choices. *Institute for the Study of Labor Discussion Paper Series*, *9383*, 1–47.

Suvak, M. K., & Barrett, L. F. (2011). Considering PTSD from the perspective of brain processes: A psychological construction approach. *Journal of Traumatic Stress*, *24*(1), 3–24. doi: 10.1002/jts

Tagliamonte, S. A., Durham, M., & Smith, J. (2014). Grammaticalization at an early stage: Future *be going to* in conservative British dialects. *English Language and Linguistics*, *18*(1), 75–108. doi: 10.1017/S1360674313000282

Takano, K., & Tanno, Y. (2010). Concreteness of thinking and self-focus. *Consciousness and Cognition*, *19*(1), 419–425. doi: 10.1016/j.concog.2009.11.010

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and semantic description* (pp. 36–149). Cambridge: Cambridge University Press.

Tate, L. M., Tsai, P. F., Landes, R. D., Rettiganti, M., & Lefler, L. L. (2015). Temporal discounting rates and their relation to exercise behavior in older adults. *Physiology and Behavior*, *152*(Pt. A), 295–299. doi: 10.1016/j.physbeh.2015.10.003

Te Winkel, L. A. (1866). Over de wijzen en tijden der werkwoorden. *De Taalgids*, *8*(66-75).

Teich, E. (2003). *Cross-linguistic variation in system and text.* Berlin/New York: Mouton de Gruyter.

Thieroff, R. (2010a). Mood in German. In B. Rothstein & R. Thieroff (Eds.), *Mood in the languages of Europe* (pp. 133–154). Amsterdam/Philadelphia: John Benjamins Publishing Company. doi: 10.1075/slcs.120.08thi

Thieroff, R. (2010b). Moods, moods, moods. In B. Rothstein & R. Thieroff (Eds.), *Mood in the languages of Europe* (pp. 1–32). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Thiese, M. S., Ronna, B., & Ott, U. (2016). *P* value interpretations and considerations. *Journal of Thoracic Disease*, *8*(9), E928–E931. doi: 10.21037/jtd.2016.08.16

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. doi: 10.1111/cogs.12525

Thorstad, R., Nie, A., & Wolff, P. (2015). Representations of time affect willingness to wait for future rewards. In *Proceedings of the annual meeting of the Cognitive Science Society* (pp. 2392–2397).

Thorstad, R., & Wolff, P. (2018). A big data analysis of the relationship between future thinking and decision-making. *Proceedings of the National Academy of Sciences*, *E1741*. doi: 10.1073/pnas.1706589115/

Thorstad, R., & Wolff, P. (2019a). Predicting future mental illness from social media: A big-data approach. *Behavior Research Methods*, *51*(4), 1586–1600. doi: 10.3758/s13428-019-01235-z

Thorstad, R., & Wolff, P. (2019b). Using big data to understand memory and future thinking. In *The annual meeting of the Cognitive Science Society.* Retrieved from `https://cogsci.mindmodeling.org/2019/papers/0499/0499.pdf`

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software*, *59*(5), 1–38. doi: 10.18637/jss.v059.i05

Traugott, E. C. (1978). On the expression of spatio-temporal relations in language: Volume 3, word structure. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language* (pp. 369–400). Stanford: Stanford University Press.

Trautmann, S. T. (2019). Distance from a distance: The robustness of psychological distance effects. *Theory and Decision*, *87*(1), 1–15. doi: 10.1007/s11238-019-09696-6

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440–463. doi: 10.1037/a0018963

Twisk, J. (2006). *Applied multilevel analysis: A practical guide.* Cambridge: Cambridge University Press.

Van Dongen, S. (2006). Prior specification in Bayesian statistics: Three cautionary tales. *Journal of Theoretical Biology*, *242*(1), 90–100. doi: 10.1016/j.jtbi.2006.02.002

van der Auwera, J., & Plungian, V. A. (1998). Modality's semantic map. *Linguistic Typology*, *2*(1), 79–124. doi: 10.1515/lity.1998.2.1.79

Vanderveldt, A., Green, L., & Myerson, J. (2015). Discounting of monetary rewards that are both delayed and probabilistic: Delay and probability combine multiplicatively, not additively. *Journal of Experimental Psychology: Learning Memory and Cognition*, *41*(1), 148–162. doi: 10.1037/xlm0000029

Vanderveldt, A., Green, L., & Rachlin, H. (2017). Discounting by probabilistic waiting. *Journal of Behavioral Decision Making*, *30*(1), 39–53. doi: 10.1002/bdm.1917

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P. C. (2019). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC. *arXiv*, 1–27. doi: 10.1214/20-ba1221

Verkerk, A. (2013). Scramble, scurry and dash: The correlation between motion event encoding and manner verb lexicon size in Indo-European. *Language Dynamics and Change*, *3*(2), 169–217. doi: 10.1163/22105832-13030202

Verkuyl, H. J. (2008). Binary tense. In *Vol. 187 of CSLI lecture notes.* Stanford: CSLI Publications.

Volkers, A. C., Tulen, J. H., Van Den Broek, W. W., Bruijn, J. A., Passchier, J., & Pepplinkhuizen, L. (2003). Motor activity and autonomic cardiac functioning in major depressive disorder. *Journal of Affective Disorders*, *76*(1-3), 23–30. doi: 10.1016/S0165-0327(02)00066-6

Vuchinich, R. E., & Simpson, C. A. (1998). Hyperbolic temporal discounting in social drinkers and problem drinkers. *Experimental and Clinical Psychopharmacology*, *6*(3), 292–305. doi: 10.1037/1064-1297.6.3.292

Vygotsky, L. S. (1987). *Cognition and language: The collected works of L. S. Vygotsky, Vol. 1, problems of general psychology* (R. W. Rieber & A. S. Carton, Eds.). New York & London: Plenum Press.

Wallace, M. (1956). Future time perspective in schizophrenia. *Journal of Abnormal and Social Psychology*, *52*(2), 240–245. doi: 10.1037/h0039899

Watkins, E., Taylor, R. S., Byng, R., Baeyens, C., Read, R., Pearson, K., & Watson, L. (2012). Guided self-help concreteness training as an intervention for major depression in primary care: A phase II randomized controlled trial. *Psychological Medicine*, *42*(7), 1359–1371. doi: 10.1017/S0033291711002480

Werner-Seidler, A., & Moulds, M. L. (2012). Mood repair and processing mode in depression. *Emotion*, *12*(3), 470–478. doi: 10.1037/a0025984

Wetherly, J. N., Plumm, K., & Derenne, A. (2011). Delay discounting and social policy issues. *The Psychological Record*, *61*, 527–546. doi: 10.1007/BF03395775

White, R., & Wild, J. (2016). "Why" or "how": The effect of concrete versus abstract processing on intrusive memories following analogue trauma. *Behavior Therapy*, *47*(3), 404–415. doi: 10.1016/j.beth.2016.02.004

Whorf, B. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf* (J. B. Carroll, Ed.). Cambridge, MA: MIT Press.

Wilhelm, F. H., Trabert, W., & Roth, W. T. (2001). Physiologic instability in panic disorder and generalized anxiety disorder. *Biological Psychiatry*, *49*(7), 596–605. doi: 10.1016/s0006-3223(00)01000-3

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785. doi: 10.1073/pnas.0701644104

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 253–265. doi: 10.1002/wcs.104

Woltin, K. A., Corneille, O., Yzerbyt, V. Y., Förster, J., Luguri, J. B., Napier, J. L., & Dovidio, J. F. (2011). Narrowing down to open up for other people's concerns: Empathic concern can be enhanced by inducing detailed processing. *Journal of Experimental Social Psychology*, *23*(7), 418–424. doi: 10.1016/j.jesp.2010.11.006

World Values Survey Association. (2014). *World values survey: Round three v. 20180912.* Madrid: JD Systems Institute. Retrieved from `www.worldvaluessurvey.org`

World Values Survey Association. (2021). *What we do.* Retrieved from `https://www.worldvaluessurvey.org/WVSContents.jsp`

Xiao, J. J., & Porto, N. (2019). Present bias and financial behavior. *Financial Planning Review*, *2*(2), 1–14. doi: 10.1002/cfp2.1048

Yi, R., de la Piedad, X., & Bickel, W. K. (2006). The combined effects of delay and probability in discounting. *Behavioural Processes*, *73*(2), 149–155. doi: 10.1016/j.beproc.2006.05.001

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, *14*(4), 301–322. doi: 10.1037/a0016972

Zauberman, G., Kim, B. K., Malkoc, S. A., & Bettman, J. R. (2009). Discounting time and time discounting: Subjective time perception and intertemporal preferences. *Journal of Marketing Research*, *46*(4), 543–556. doi: 10.1509/jmkr.46.4.543

Zhang, Z., Zheng, C., Kim, C., Poucke, S. V., Lin, S., & Lan, P. (2016). Causal mediation analysis in the context of clinical research. *Annals of Translational Medicine*, *4*(21). doi: 10.21037/atm.2016.11.11

Zhu, J., Hu, S., Wang, J., & Zheng, X. (2020). Future orientation promotes climate concern and mitigation. *Journal of Cleaner Production*, *262*, 1–10. doi: 10.1016/j.jclepro.2020.121212

Zimmermann, M. (2019). Discourse particles. *Semantics*, *2*, 2011–2038. doi: 10.1515/9783110589863-015

English Summary

People tend to devalue future outcomes when they are risky or far away in time. For instance, most people would prefer \$20 immediately over \$20 in five years, because the delayed \$20 is worth less subjectively. Similarly, many would prefer a guaranteed \$100 over a 50% chance of \$200, because they would not consider the risk to be worth the reward. However, the extent to which risk and delay affect subjective value differs between people, so some might opt for the guaranteed \$100, while others would choose the risky \$200. The subjective devaluation of such outcomes due to risk and delay is called "psychological discounting." Differences in psychological discounting correlate with numerous behaviours which involve weighing present versus risky and/or delayed outcomes, e.g. exercising for better future health, investing in savings, or managing one's finances.

This thesis explores the idea that such behaviours can be affected by language. In particular, it uses language elicitation and experimental measurement of psychological discounting with cohorts of English, Dutch, and German speakers to shed light on how differences between these languages' Future Time Reference (FTR) grammar affect psychological discounting.

Previous work has focussed on how future tenses—e.g. *will*, *be going to*—affect discounting. However, this tense-based account is not supported by the evidence presented in this thesis. Rather, this thesis develops the hypothesis that low-certainty modal verbs—e.g. *may*—affect psychological discounting. Specifically, English obliges speakers to use a modal verb when they make predictions about the future—e.g. *may*, *might*, or *could*—whereas Dutch allows speakers to use the present tense. This difference caused English participants to devalue risky, delayed rewards to a greater degree than Dutch participants. The same was not true of the future tense—e.g. *will*. Given correlations between psychological discounting and real-world behaviours, this thesis argues that the grammatical obligation to use low-certainty modal verbs—and not the future tense—is driving reported correlations between languages' FTR systems and numerous important behaviours. These include national efforts to mitigate climate change; personal differences in investment in savings; and several indexes of corporate behaviour including investment in research and development, accounting practices, and cash holdings.

Nederlandse samenvatting

Mensen hebben de neiging toekomstige resultaten te devalueren wanneer ze risicovol zijn of ver weg in de toekomst liggen. De meeste mensen zouden bijvoorbeeld verkiezen €20 onmiddellijk in handen te hebben over €20 in vijf jaar omdat de uitgestelde €20 subjectief minder waard is. Evenzo zouden velen een gegarandeerde €100 verkiezen over een 50% kans op €200 omdat ze niet zouden overwegen dat het risico de beloning waard is. De mate waarin echter het risico en uitstel de subjectieve waarden beïnvloeden, verschilt tussen de mensen waarbij sommigen zouden kiezen voor de gegarandeerde €100 terwijl anderen zouden kiezen voor de risicovolle €200. De subjectieve devaluatie van dergelijke resultaten als gevolg van risico en uitstel noemt men "psychologische verdiscontering". Verschillen in psychologische verdiscontering correleert met verscheidene gedragingen die inhouden dat huidige versus risicovolle en/of uitgestelde resultaten worden afgewogen, bijv. sporten voor een betere gezondheid in de toekomst, investeren in spaargelden, of het beheren van je eigen financiën.

Deze thesis verdiept zich in de idee dat zo'n gedragingen door taal kunnen worden beïnvloed. Het gebruikt met name taalontlokking en experimentele bepaling van psychologische verdiscontering met cohorten van Engelse, Nederlandse en Duitse sprekers om duidelijkheid te verschaffen over hoe de verschillen tussen de toekomende tijdsverwijzing (FTR[9]) in de grammatica van de talen de psychologische verdiscontering beïnvloedt.

Eerder werk heeft zich gericht op de manier waarop toekomende tijden—bijv. *zullen*, *gaan*—verdiscontering beïnvloeden. Dit op de werkwoordstijd gebaseerd verslag wordt niet ondersteund door bewijsmateriaal dat in deze thesis wordt voorgelegd. Deze thesis ontwikkelt eerder de hypothese dat modale werkwoorden met lage zekerheid —bijv. *kunnen*—psychologische verdiscontering beïnvloeden. Meer specifiek verplicht het Engels de sprekers een modaal werkwoord te gebruiken wanneer ze voorspellingen doen over de toekomst—bijv. *may* 'kunnen', *might* 'kunnen', of *could* 'kunnen'—terwijl Nederlandse sprekers de tegenwoordige tijd gebruiken. Dit verschil heeft ervoor gezorgd dat Engelse deelnemers risicovolle, uitgestelde beloningen in grotere mate devalueren dan Nederlandse deelnemers. Hetzelfde gold niet voor de toekomende tijd—bijv. *will* 'zullen'. Gezien de correlaties tussen psychologische verdiscontering en levensechte gedragingen, stelt deze thesis dat de grammaticale verplichting om modale

[9]Uit de Engelse *Future Time Reference*.

werkwoorden van lage zekerheid te gebruiken—en niet de toekomende tijd—aangegeven correlaties tussen de FTR-systemen van talen en verscheidene belangrijke gedragingen aandrijft. Deze omvatten nationale inspanningen om klimaatverandering tegen te gaan; persoonlijke verschillen in investering in spaargelden; en verscheidene indexen van ondernemingsgedrag met inbegrip van investering in onderzoek en ontwikkeling, boekhoudkundige praktijken en kasmiddelen.

Acknowledgments

Bronwyn, thank you for your wonderful, unflagging, and indispensable support. Your empathy as I wandered the sometimes dark forest has been so deep and inflected with your own experience as to be irreplaceable by any other means. Words don't capture how grateful I am for you, my smol Ham. Thank you.

My supervisors' support, feedback, and dedication has been invaluable.

Asifa, your assiduous eye for detail, your deep knowledge of the literature, and your skill as a writer have been essential in shaping my drafts into something a stranger would recognise as the presentation of academic work. I may now and forever carry with me a voice that chirps "Asifa won't like that" the instant my prose wanders into inadvertent laziness, and this mixed blessing indicates you've played your role superlatively. Thank you.

Seán, you waded into the weeds with me whenever I called upon you, and I am grateful. Your generosity with your time, attention, and skills was invaluable. Most of this thesis was completed in isolation, but it would have felt orders of magnitude more isolated without you. Thank you.

Robin, thank you for giving me the time, freedom, and support to follow my interests, and the data, where they slowly led me. Your sink or swim approach to supervision led me to find in myself whatever it took to transcend the depths of my own ignorance and bring my subject into whatever meagre illumination the previous pages may provide. Thank you.

This thesis was commenced with a view to test Keith Chen's idea that using the future tense makes people think of future events as far away in time, and thereby more temporally discounted. We sought to explain our findings through this lens for several years before I began to develop a modal account that better described my results. Without Chen's creative scholarship, this dissertation would not exist. Thanks Keith.

Nick and Verity, you were so kind and dedicated in your generous copy editing of long reams of dense academic text. Thank you. And, of course, for everything else!

Gen, thank you for your wonderful cover art and your patience with me as I fumbled to explain my aesthetic vision.

Nina, thank you for being so flexible with your time (and car!) when our climbing holiday had to work around my writing schedule. And thanks for the catches on Zombie Roof.

Peter and Jacquie, that week I spent reading and writing at Regenstein was when I realised I had something substantive to say about my nascent results. Thank you so much for your generosity, kindness and support through this long process.

James, thank you for your flexibility and friendship when we worked at Brunel, and for the job! Thank you also for your useful copy edits and feedback.

Arran Davis and Max van Duijn, thank you both for your feedback on drafts and the FTR-elicitation task, respectively, and for graciously agreeing to be my paranymphs, though as a foreigner to Dutch traditions, the nature of your roles may remain shrouded from me in perpetuity.

Stu, thank you for kindly designing the jacket.

Finally, I thank Elise Hamerslag, Peter Hamerslag, Saskia van Putten, Josje de Valk, Birgit Rauchbaur, Gabriela Jank, Mary Kempnich, Hilário de Sousa, and Leonard Michaels for their response coding, translation, and copy editing.

Curriculum vitae

Cole Robertson (1984, Vancouver) completed the Bachelor of Arts in Western Society and Culture (Concordia University, Montreal), followed by the Master of Science in Cognitive and Evolutionary Anthropology (University of Oxford). In 2016, he began his PhD research at the Centre for Language Studies at Radboud University. He has been supported by a research fellowship from the Individual Behaviour to the Socio-Technical Man project and by a doctoral scholarship from the National Science and Engineering Council of Canada. He is currently living in Oxford and employed as the Chief Technology Officer for Texture AI.

List of publications

Robertson, C. B. J., & Roberts, S. G. (under review). Modality and future time reference in English, Dutch, and German. *Cognitive Science*.

Robertson, C. B. J., Roberts, S. G., Majid, A., & Dunbar, R. I. M. (in prep.). The effect of language on economic behaviour: Future tense use causes less not more temporal discounting.

Robertson, C. B. J., Roberts, S. G., Majid, A., & Dunbar, R. I. M. (in prep.). Not when but if: Low-certainty modals not future tenses cause increased psychological discounting.

Robertson, C. B. J., Carney, J., & Trudell, S. (in prep.). A big-data and experimental analysis of the relationship between time perspective, anxiety, and depression.

Appendices

Appendix A

Author contributions

In this appendix, I clarify my unique contribution to Chapters 2–5, which are based on multi-author publications, see *Table A.1*. Overall, I wrote the FTR-elicitation questionnaire items, did the reported data analysis, and wrote the preponderance of Chapters 2–5. All supervisors, Asifa Majid (AM), Seán Roberts (SR), and Robin Dunbar (RD) provided feedback on the FTR-elicitation questions, especially the early drafts; on drafts of Chapters 1–6; and on experimental design for Chapters 2–4. I wrote the FTR-type classifier. This includes the principled decisions to have it prioritise semantic categories and the modal dominance structure. It also includes the background research to compile keyword lists, and the practical aspects of packaging it as a Python library. SR provided feedback on the Github repository readme and install behaviour, and conducted data analyses which helped me resolve early strange behaviour in the present-tense category. The rest of this appendix provides chapter-specific breakdowns of author contributions.

A.1 Chapter 2

I analysed the data, produced the figures and tables, and wrote Chapter 2. I designed the experiments, principally in collaboration with SR. SR also provided feedback on the analytic approach taken and contributed data analysis of his own, especially in the early stages of the work. Anonymous reviewers provided feedback on drafts.

Table A.1

Publications on which empirical chapters are based

chapter	basis publication
Chapter 2	Robertson, C. B. J., & Roberts, S. G. (under review). Modality and future time reference in English, Dutch, and German. *Cognitive Science*.
Chapter 3	Robertson, C. B. J., Roberts, S. G., Majid, A., & Dunbar, R. I. M. (in prep.). The effect of language on economic behaviour: Future tense use causes less not more temporal discounting.
Chapter 4	Robertson, C. B. J., Roberts, S. G., Majid, A., & Dunbar, R. I. M. (in prep.). Not when but if: Low-certainty modals not future tenses cause increased psychological discounting.
Chapter 5	Robertson, C. B. J., Carney, J., & Trudell, S. (in prep.). A big-data and experimental analysis of the relationship between time perspective, anxiety, and depression.

A.2 Chapter 3

I analysed the data, produced the figures and tables, and wrote Chapter 3. I designed the experiments, principally in collaboration with SR. SR provided feedback on the analytic approach and contributed data analysis of his own, especially in the early stages of the work. Anonymous reviewers provided feedback on drafts.

A.3 Chapter 4

I analysed the data, produced the figures and tables, and wrote Chapter 4. I designed the experiments, with feedback principally from SR.

A.4 Chapter 5

I processed the data, analysed the data, produced the figures and tables, conceived of, and wrote the preponderance of Chapter 5. I designed the experiments with feedback from an academic collaborator, James Carney (JC). I conceived of the time-reference classifier, downloaded Twitter data, and annotated the FTR examples. JC annotated the majority of the PTR examples. JC downloaded the data from Reddit and wrote the section on construal level theory (Section 5.4.2). A second collaborator, Shane Trudell (ST), wrote the sections on functionalist perspectives (Section 5.4.1) and therapeutic applications (Section 5.4.3). I solicited these sections, and conceived of them collaboratively with JC and ST. I contributed some original writing, worked with ST and JC to refine their drafts, and integrated their work into mine.

Appendix B

Pre-registrations

This appendix contains pre-registrations for hypotheses and experimental designs for some studies presented in this thesis. I am committed to open science, which can include being open and transparent about what researchers expect to happen (predictions) and what actually does (results). These pre-registrations were made prior to the outset of data collection and were deposited online using AsPredicted. Once posted in the secure repository on AsPredicted, it is not possible to alter them. (I have used image-editing software to add study and chapter numbers at the top, but they are otherwise unchanged.) This ensures that hypotheses are not adapted to results in a post-hoc fashion, and guards against experimenters making analytic choices which unfairly support their hypotheses. Use the links provided to view these pre-registrations on AsPredicted, or view them on the following pages.

B.1 Pre-registration guide

Here I provide a guide to help match pre-registrations with experiments.

B.1.1 Study 3, Chapter 3 (#8489)

Future tense and patience, experiment 3, February, 2018 (2nd draft): In this study, Dutch participants completed the FTR-elicitation task and an intertemporal-choice task where they were directed to use either the future or present tense (presented in Section 3.4). The pre-registration can be viewed at `https://aspredicted.org/blind.php?x=3zu77e`.

B.1.2 Study 2, Chapter 4 (#59331)

FTR and discounting cross-linguistic experiment 8, 2021: In this study, Dutch and English participants completed the FTR-elicitation task and a risky intertemporal-choice task (Section 4.3). The pre-registration can be viewed on AsPredicted at `https://aspredicted.org/blind.php?x=q7jq9z`.

B.1.3 Study 3, Chapter 4 (#36973)

Bilingual FTR and probabilistic intertemporal choice, March, 2020: In this study, Dutch-English bilingual participants completed the FTR-elicitation task and a risky intertemporal-choice task, first in one language then in another (Section 4.4). The pre-registration can be viewed on AsPredicted at `https://aspredicted.org/blind.php?x=be43nf`.

CONFIDENTIAL - FOR PEER-REVIEW ONLY

Future tense and patience, experiment 3, February, 2018 (2nd draft) (#8489)

Created: 02/16/2018 05:44 AM (PT)
Shared: 04/19/2021 09:14 AM (PT)

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

Does use of the present tense to refer to events in the future cause speakers of Dutch to be more likely to express willingness to wait for future rewards?

3) Describe the key dependent variable(s) specifying how they will be measured.

The study will be conducted in two surveys. In an initial survey (t0) participants will complete 100 test sentences of the from described below (conditions section), but with no instructions as to which tense they should use. Participants will rather be told to give their answers "as though they were talking to a friend." They will then complete the delay discounting task described below. One week later, the same participants will complete a second survey (t1), this time randomly assigned to the below described manipulation before again completing the delay discounting task.

The dependent variable will be the differences scores of the dichotomous choice values in a standard delay discounting task between responses at t0 and t1. In the task, participants will choose between whether they would rather receive a "Smaller-Sooner Reward" now (scored as 0), or a "Larger-Later Reward" at various lengths of time in the future (scored as 1). The Larger-Later reward will invariably be €10. There will be 12 Smaller-Sooner Rewards (the sequence between €7 and €9.75 by steps of €0.25). There will be eight distances in the future (one through six months, nine months, and one year). Distances and amounts will be fully crossed for a total of 96 questions of the form: Difference scores will be calculated as follows: d = (c_t0-c_t1)*-1, where c_t0 and c_t1 are choice values at t0 and t1, meaning changes from smaller-sooner reward to larger-later reward will result in positive values (more patience), changes the other way will results in negative values (less patience), and no change will result in 0 values.

4) How many and which conditions will participants be assigned to?

Before completing the delay discounting task at (t1), participants will be randomly assigned to one of two conditions in which they complete the same 100 unconjugated future-time-reference sentences as at t0, but this time they will be directed to do so either using the present or future tense. Each sentence will be provided with a context, as well as a target sentence and the verb-to-be-conjugated, as in: CONTEXT: "Q: Do you think your dad will go to sleep early tonight?", TARGET: "A: Yes, he {BE} tired (by then)." Prior to completing all sentences, participants will be instructed to invariably use either the present tense ("He is tired") or the future tense ("He will be tired").

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We will use a multinomial hierarchical linear regression predicting change scores from condition interacted with ratio of future:present tense use in free completion sentences at t0. The model will be nested within participant, amount and temporal distance.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

We will exclude participants and sentence questions on a univariate basis by winzorizing participant's whose mean score across all observations (sentences and discount questions) is greater than the median +/- 3 times the median absolute deviation. We will also exclude any participants with Cooks values > 1 in the model described above.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

At t0, ww will attempt to recruit 360 L1 Dutch participants, which, assuming an attrition rate of 15-20%, will result in a final sample of ~288-306, matched between the t0 and t1 samples. If we have recruited more than 250 participants but not recruited 360 participants by March 9, 2018, we will cap the sample and continue with the number we will have thus far recruited. If we have recruited less than 250 participants we will continue recruiting until April 30, 2018 or until we reach 300 participants, whichever comes first.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We also collect a number of demographic variables as potential controls (sex, age, education level, employment status, income, relationship status, and whether participants speak a second language other than Dutch). These will be included in the final model only if they have any significant effects.

CONFIDENTIAL - FOR PEER-REVIEW ONLY

FTR and discounting cross-linguistic experiment 8, 2021. (#59331)

Created: 02/25/2021 08:19 AM (PT)
Shared: 04/19/2021 09:19 AM (PT)

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
1) Does the way someone talks about the future affect how much they discount uncertain future rewards?
a. We hypothesise that higher use of low-certainty modals in English speakers will result in more discounting compared with Dutch speakers.
b. We hypothesize that higher use of the future tense in English speakers will not be associated with more distal representations of future events, and therefore will not result in more discounting in English speakers.
2) Does probabilistic or temporal discounting drive population-level effects? We hypothesise that English speakers will probabilistically discount more, but that there will be no differences in temporal discounting between English and Dutch speakers.

3) Describe the key dependent variable(s) specifying how they will be measured.
1) Language usage in a future time reference elicitation task. Participants complete a number of incomplete sentences, which all refer to events in the future. Each sentence will be provided with a context, including temporal delay into the future, and some “certainty information”. Delays will be today, tomorrow, one week, three months, six months, and one year. Probabilities will be 100%, 90%, 80%, 70%, 60%, and 50%, and are fully crossed to produce a test battery of N = 36 items. Participants then conjugate a target sentence we provide, given the context. The dependent variables will be whether the participant uses a) the future tense, b) the present tense, or a modal term indicating either c) high-certainty, or d) low-certainty. Each are a separate dependent variable, scored as (1) to indicate use, otherwise (0).

2) Subjective value in a probabilistic temporal discounting task. Participants make binary choices between an uncertain future reward and a certain present reward, e.g. “Would you rather have £22 now or an 85% chance of receiving £35 in two months?” The delays and probabilities are fully crossed and match those in the language usage task. Participants complete 5 trials per delay/probability pair. The dependent variable is the subjective value of the uncertain delayed reward at a given probability and delay, i.e. the indifference point. One indifference point will be calculated for each delay/probability pair using the staircase method given in [Vanderveldt et al. 2015. Discounting of monetary rewards that are both delayed and probabilistic. J. Exp. Psych. Learn. Mem. Cogn., 41(1), 148-62]. This method has the advantage of “collapsing” into a simple intertemporal choice task when probability is 100% and into a simple risk-preference task when delays are today.

3) Perceived distance of future events. Participants indicate how far away a given event feels, i.e. “Manchester United has a 50% chance of winning today.” Delays and probabilities are fully crossed and match those in the other tasks. The DV is the participant ‘distance’ ratings using a slider ranging from “close to now (1)” to “far from now (1000).”

4) How many and which conditions will participants be assigned to?
The study will have a mixed-design. In addition to the within-subject factors of each task (delay, probability), there will be one between-subjects factor: language condition. In the English condition (British) English first language speaking participants will complete the survey in English; in the Dutch condition Dutch first language speaking participants will complete an otherwise identical survey in Dutch.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
1) We will use a Bayesian hierarchical path analysis to estimate the indirect effects of English on indifference points via language use and subjective future distance ratings. To test hypothesis 1a we will estimate a single indirect effect: English → low-certainty → indifference points. To test hypothesis 1b we will estimate a serial indirect effect: English → future tense → subjective temporal distance → indifference points. These indirect effects will be combined in a single model.

2) To test hypothesis 2, we will use regressions to estimate the impact of language on areas under the discounting curve (calculated following: Myerson, et al. (2001). Area under the curve. J. Exp. Anal. Behav., 76(2), 235–243.]). To test for differences in probabilistic discounting, we will use data where delay is today, and to test for differences in temporal discounting we will use data where the probability is 100%.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We will exclude responses in the FTR elicitation task which exhibit mixed modality or use negations. Participants who fail attention checks during the surveys will be ejected immediately and their data will be deleted. Any participants’ whose FTR elicitation task responses clearly indicate bad faith will be

excluded, e.g. repeatedly entering the same string of meaningless characters, providing colored commentary on the task rather than conjugating target sentences, only entering blank spaces, or repeatedly copy-pasting unaltered target sentences.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We will recruit approximately n = 200 participants to each language condition. If we have not reached that sample size by the end of March, 2021, we will stop recruitment.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
We will collect a number of demographic variables as potential controls (sex, age, education level, employment status, income, relationship status), and participants will be asked to rate their proficiency in (up to) three other languages apart from English/Dutch.

CONFIDENTIAL - FOR PEER-REVIEW ONLY

Bilingual FTR and probabilistic intertemporal choice, March, 2020. (#36973)

Created: 03/07/2020 06:46 AM (PT)
Shared: 04/19/2021 09:18 AM (PT)

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

Does switching the language between Dutch and English in a survey comprised of a language elicitation and probabilistic delay discounting task result in more absolute discounting in the English survey than Dutch, in a population of Dutch/English bilinguals.

3) Describe the key dependent variable(s) specifying how they will be measured.

1) Language usage in a future time reference elicitation task. In this task, participants complete a number of incomplete sentences, which all refer to events in the future. Each sentence will be provided with a context, as well as a target sentence and a verb-to-be-conjugated. Dependent variables will be whether the participant uses a) the future tense, b) the present tense, or a modal term indicating either c) certainty, or d) uncertainty.
2) Subjective value in a probabilistic temporal discounting task. In this task, participants choose between a Larger Later Reward (LLR) which they will be awarded at a given probability and delay in the future, or a Smaller Sooner Reward (SSR) given now with 100% probability. For example, "Would you rather have £22 now or an 85% chance of receiving £35 in two months?" In this task, the sets of delays and probabilities are fully crossed. Participants complete 5 trials per delay:probability pair. In the first trial, the SSR reward is ½ the LLR. If the participant chooses the LLR, the SRR is positively incremented by half the difference between them. If they choose the SRR, this increment is negative. This process is then iterated. The dependent variable is the value the SSR would take in the 6th trial (which never occurs).

4) How many and which conditions will participants be assigned to?

The study has a within-subjects design wherein the following conditions are implemented:

1) Survey language. Dutch/English bilinguals will complete the survey, initially in one language, and then after a period of 3-4 weeks, in the other. The only difference apart from the language of the surveys will be that the first survey will ask participants some demographic questions which will be eschewed in the second survey. Half the participants will be randomly assigned to complete the English survey first and Dutch second, while the order will be reversed for the other half.
2) Future time referencing elicitation task. In this task independent variables will be how far in the future the referenced event is, and how certain the participant is supposed to be about it. This information is included in each context. Delays will be later today, tomorrow, one week, three months, six months, and one year. Probabilities will be 100%, 90%, 80%, 70%, 60%, and 50%. Probabilities and delays will be fully crossed, and one question will be included per delay with no explicit probability in the context. All questions will be predictions. All participants will complete all questions.
3) Probabilistic temporal discounting task. Delays and probabilities are the IVs and will the same as in the future time referencing elicitation task. All participants will complete all questions.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We will use a hierarchical regression and structural equation models to predict probabilistic temporal discounting choices from both the probabilities and delays of the probabilistic discounting task, the results of the future time reference elicitation task, and from the language of the survey. We make the following hypotheses:

1) Usage differences. Compared with English, participants in the Dutch survey will use:
a. More present tense constructions
b. Fewer future tense constructions
c. More certain constructions
d. Fewer uncertain constructions
2) Discounting differences.
a. The area under the discounting curve will be more in the Dutch condition than the English one.
b. Dutch – English condition discounting change scores will be positive and significantly different from 0 (English will discount more).
c. There will be a negative three-way interaction between language condition (Dutch), probability, and delay.
d. There will be positive two-way interactions between language condition (Dutch) and, respectively, probability and delay.
3) Mediational hypotheses. While controlling for delay and probability:
a. Present tense will positively mediate a relationship between Dutch and discounting.

b. Future tense will negatively mediate a relationship between Dutch and discounting.
c. Uncertain language will positively mediate a relationship between Dutch and discounting.
d. Certain language will positively mediate a relationship between Dutch and discounting.
e. The total effect – the sum of the paths described in 3a-3d plus the direct effect – of Dutch on discounting will be positive and significant.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We will exclude participants on a model-by-model basis, excluding any participants whose data causes the model coefficients or significance tests to substantively change, i.e. change sign, or cross pcrit = .05 significance boundaries.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We will recruit approximately 300 participants for the first wave, and try to retain as many of these as is possible in the second. The only exclusion criteria will be that participants need to be over 18 and must be Dutch first language speakers who are better than beginner in English.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
We will collect a number of demographic variables as potential controls (sex, age, education level, employment status, income, relationship status), and participants will be asked to explicitly rate their English proficiency, as well as proficiency in (up to) three other languages apart from Dutch English, and where they currently reside. These additional variables will be included in the final models only if they have any significant effects on the DVs.

IMPORTANT
This pre-registration is part of a set of similar and related pre-registrations sharing at least one author. When one of these pre-registrations was shared by an author, the rest were shared automatically. Links to all of them, including this one, appear below:

https://aspredicted.org/blind.php?x=be43nf
https://aspredicted.org/blind.php?x=7v2g72

Appendix C

Supplementary materials to Chapter 2

In this appendix, I provide supplementary materials to Chapter 2, including statistical reportage and justification of various elements in the main text.

C.1 Participant and item exclusions by study

C.1.1 Study 1: Version A

C.1.1.1 Participant exclusions

There were two data collection waves in Study 1. The first wave was collected between July and October, 2016, from the MPI participant panel (Dutch and German), and Oxford University students (English). Final $n = 82$ ($n = 45$ in English, $n = 19$ in Dutch, and $n = 18$ in German). In a second wave of data collection in October 2019, and additional $n = 154$ German responses were collected in order to match the Dutch and English samples which had been collected in the interim. This wave was collected using a Qualtrics panel. In this wave, 8 participants were excluded because either they failed attention checks or their answers made it clear they were not providing good faith responses, e.g. repeatedly entering the same response, copy-pasting the target with no alteration, etc. Final $N = 228$ ($n = 45$ in English, $n = 19$ in Dutch, and $n = 164$ in German).

C.1.1.2 Item exclusions

Of the $N = 42$ items in the matched sample, 2 items needed to be cut because they were being persistently misunderstood as referring to present time, 4 items needed to be cut because they used modal or tense words in the prompt (e.g. think, expect, go) which prevented the automated classification program from accurately identifying whether participants had used a tense or modality word themselves, or whether they had simply copied the prompt, and 1 item needed to be cut because it was in both of these categories. This resulted in final sample of $n = 35$ items. Additionally, 3 items needed to be removed from the second wave of German data collection because they used a time prompt ("by 2018") which was, by that point, in the past.

C.1.2 Study 1: Version B

C.1.2.1 Participant exclusions

Data for Study 1 Version B was collected in two waves. The first wave was collected between January and February, 2017. An initial sample of $N = 240$ participants completed Version B. However, one participant was excluded because their answers to the FTR-elicitation task were given in bad-faith and four were excluded because of missing demographic data. This left a final sample of $n = 235$ ($n = 113$ in English, and $n = 122$ in Dutch). A second wave of data collection occurred in October and November 2018 (Dutch), and in May 2020 (English). An initial sample of $N = 625$ participants completed this second wave. However, $n = 7$ participants were excluded because inspection indicated their FTR-elicitation responses were given in bad faith and $n = 12$ participants were excluded because a survey error resulted in excessive missing data. After these exclusions, a final sample of $n = 606$ participants completed this wave ($n = 301$ in English, and $n = 305$ in Dutch). Final sample for Study 1 Version B was $N = 841$ ($n = 414$ in English, and $n = 427$ in Dutch). The final sample for Study 1 was therefore $N = 1069$ (Version A: $n = 228$; Version B: $n = 841$).

C.1.2.2 Item exclusions

Of the 29 items in the FTR-elicitation task, 2 needed to be removed from analysis because because it was later realised that the prompt included common modal or tense words (e.g. think, expect, go) which prevented the FTR-type classifier from accurately identifying whether participants had used a tense or modality word themselves, or whether they had simply copied the prompt, 2 items needed to be removed because inspection revealed that participants consistently interpreted the context as referring to present time, and 1 item was in both of these categories. A total of 3 items were therefore removed, and analysis was conducted on data from 26 of the 29 items participants were given. For the data collection that occurred in 2018 and later, 2 items which referred to present or past time ("by 2018") additionally needed to be removed, so only 24 items from these data were analysed.

Questions can be viewed at the FTR-elicitation task github repository, here: `https://github.com/cbjrobertson/ftr_questionnaire_master/tree/master/ftr_questionnaire/versions_1-2`. The README.md file there contains instructions on how to view the questions which were given to participants and analysed in the main text.

C.2 Balance of items by condition in the FTR-elicitation task (Study 1)

The conditions in the FTR-elicitation questionnaire in Study 1 were not balanced. *Tables C.1–C.3* present the count of the unique items for each language and each two-way combination of FTR-mode, probability condition, and temporal distance. German items are imbalanced because of the need to remove items which referred to the past at the time of data collection (e.g. “by 2018”).

C.3 Epistemic modality in English, Dutch, and German

C.3.1 FTR-type classifier keyword criteria

Tables C.4–C.6 present the FTR-type classifier’s taxonomy of FTR along with high-frequency keywords used as classification criteria, and for high-frequency terms, reference to the literature justifying each word’s encoding of the notion indexed by the class.

C.3.2 Review of relevant literature justifying the FTR-type classifier’s taxonomy

Modal expressions are an important way of referring to the future, and typically involve a speaker expressing degrees of possibility, probability, and certainty relative to a “modal base” which can range from the rules and norms governing social interaction (deontic modality) to a person or object’s inherent abilities (dynamic modality) (Palmer, 2001). Since it seems the most likely modal base to be implicated in psychological discounting processes, we are here most interested in epistemic modality, in which speakers express degrees of probability relative to what they know or believe (Palmer, 2001). We suggest this is the case because future predictions mostly involve expressing gradations of epistemic modality, which most closely reflects widely reported tendencies to discount as a function of probability. Indeed, it would be difficult to conceive of why differences in deontic modality (e.g. *You should always brush your teeth*) or dynamic modality (e.g. *It can be really hot here in August*) relate to either FTR or discounting.

English, Dutch, and German distribute the expression of epistemic modality across several shared word classes, these being: modal verbs (C.1a), modal adverbs (C.1b), predicative adjectives (C.1c), nouns (C.1d), and mental state predicates (C.1e) (Nuyts, 2000):

(C.1) a. *It <u>might/may/could/should</u> rain tomorrow.*

b. *It will <u>definitely/possibly</u> rain tomorrow.*

c. *It is <u>probable/certain</u> that it will rain tomorrow.*

d. *It is a certainty/possibility that it will rain tomorrow.*

e. *I think/believe/reckon it will rain tomorrow.*

We follow Nuyts (2000) and treat predicative adjectives, nouns, and modal adverbs as one meta-type, as they often share roots, and derived forms encode only subtle semantic differences. We terms these "modal modifiers". Below we briefly discuss each construction types.

Modal modifiers: These may be thought of as the "purest" expressions of epistemic modality, as they allow for the most precise expression of epistemic modal notions, and tend to have well-defined semantics that are relatively stable across differing pragmatic contexts (Nuyts, 2000). The three languages break the modal scale into three approximately equivalent points: 1) high certainty, 2) moderate certainty, and 3) low certainty (50/50). In English these are indexed with *certainly*, *probably*, and *possibly*, respectively (Dutch and German systems are similar, see *Tables C.5* and *C.6*).

Mental state predicates: English, Dutch, and German share similar systems of epistemic mental state predicates. However, there are between-language differences. For instance, the prototypical examples in English are *think*, *believe* and sometimes *guess* (American) (Nuyts, 2000). In Dutch *denken* 'think' is also prototypical, but the German cognate, *denken*, is rarely used epistemically, while *glauben* 'believe' is more prototypical (Nuyts, 2000), and *erwarten* 'expect' is the most frequent in our data, *Table C.6.* Similarly to epistemic modal modifiers, epistemic mental state predicates index gradations between certainty and uncertainty; however, their exact position on this scale is somewhat less well-defined, i.e. the relative differences between *believe, guess, suppose*, and *think* (and Dutch and German equivalents), is unclear (Nuyts, 2000). Differing from the modal modifiers, it seems that there are no clear high-certainty mental state predicates.

Modal verbs: In these languages, the modal verbs for which epistemic use is possible are a subset of the more general set of modal verbs (Nuyts, 2000). Similarly to mental state predicates, epistemic modal verbs tend to express gradations of uncertainty rather than high certainty. The main high-certainty modal verb, when used without a modal particle (see below), *must/mussen/moeten*, is mostly limited to deontic rather than epistemic modal bases. Apart from *will, zullen* and *werden*, which might be considered high certainty, this leaves a set of modal verbs which tend to encode either probability or possibility. These are described in Section 2.1.2.2.

Additionally to these shared ways of expressing epistemic modality, Dutch and German also have systems of modal particles which can affect the modal strength of FTR statements, often subtly changing the modal valence of an expression when they co-occur with other modal morphemes, or changing the modal base in which it is possible to use a modal verb (Fehringer, 2018; Nuyts, 2000). While a limited set of modal particles appear to independently epistemically affect statements (*see Tables C.5* and *C.6*), we will not spend much time discussing these as they appear to be a relatively minor way of expressing epistemic modality (Nuyts, 2000), and are generally more involved in marking the epistemic states of discourse participants, or expressing discourse attitudes (c.f. Bross, 2012; Gutzmann, 2009; Hogeweg, Ramachers, & Wottrich, 2011; Karagjosova, 2004; Zimmermann, 2019).

To get a sense of how actual participant responses map to the classification system of the FTR-type classifier, we present example responses from each classification category of the FTR-type classifier, in *Table C.7*.

C.4 FTR-type classifier performance metrics

Tables C.8 and *C.9* presents performance metrics for agreement between the predictions of the FTR-type classifier and trained human annotators in Studies 1 and 2, respectively. By any metric, agreement is high, indicating the FTR-type classifier agrees closely with human judgement.

C.5 Language sample demographic equivalence

The samples in Study 1 were not matched for demographic equivalence. Descriptive statistics for demographic sample composition are presented visually in *Figs. C.1* and *C.2*. In general, in Study 1 the Dutch and German samples were older, while the German sample tended to be more likely to be married, or retired. The English sample was more likely to have an undergraduate degree, while the German sample was more likely to have undertaken trade or vocational training. These differences probably reflect sampling from the Oxford student body on the one hand and panel providers (Prolific, Qualtrics) on the other. In Study 2, the English and Dutch samples were matched for age and sex, which were set to reflect UK population norms. In terms of the other demographics, the English sample tended to have slightly lower income and be more likely to have an undergraduate degree. It is difficult to conceive of why demographic differences would affect both first language spoken and language usage habits, suggesting it is unlikely that these demographic differences confounded usage (Pearl & Mackenzie, 2018). We nonetheless controlled for potential demographic confounding when inclusion of individual demographic variables improved model fit, $p < .05$.

C.6 Random components for mixed models from Studies 1 and 2

Table C.10 presents random components for the mixed models used to test hypotheses in Studies 1 and 2. Significant model fit improvement resulted in all cases from adding random intercepts by participant, random intercepts by item, and random slopes for language by item. By allowing intercepts and language slopes to vary by item, we were able to estimate parameters for fixed effects which were unbiased by idiosyncratic item level effects. Significant intercepts for participants indicated that participants tended to respond in similar ways to different items.

C.7 Levels for the multilingualism question (Study 2)

Before beginning the FTR-questionnaire and after other demographic questions, participants in Study 2 completed question about multilingualism. The levels for the multilingualism question were as follows:

1. can ask directions and answer simple questions
2. can have a basic conversation on familiar topics
3. functional in most contexts (e.g. could tell a story or fill out a form), but not fluent
4. fluent with occasional mistakes, clearly a foreigner
5. very fluent, can use the language as well as a native language

Participant used this scale to rate their proficiency for up to 3 languages other than their first language (English/Dutch). Results for this question were summed, see *Fig. C.3*, to create a scalar multilingualism score per participant. It is this score which was used as a control in analyses. The Dutch sample were far more multilingual, *Fig. C.3*.

C.8 Tables

Table C.1

Item balance across language, modality condition, and temporal distance (Study 1)

modality condition	temporal distance	Dutch	English	German
high-certainty	tomorrow	1	1	1
	one week	1	1	1
	six months	1	1	1
	one year	1	1	1
	two years	1	1	1
	twenty five plus	1	1	1
low-certainty	one week	1	1	1
	one month	1	1	1
	one year	3	3	3
	two years	2	2	2
	ten years	2	2	1
	twenty five plus	2	2	2
	ongoing prediction	1	1	1
	indeterminate	1	1	1
neutral	today	1	1	1
	tomorrow	1	1	
	one week	2	2	2
	one month	2	2	1
	six months	2	2	2
	one year	2	2	2
	two years	2	2	2
	ten years	1	1	1
	twenty five plus	3	3	1
	ongoing prediction	4	4	4
	indeterminate	1	1	1

Table C.2

Item balance across language, probability condition, and ftr mode (Study 1)

FTR-mode	modality condition	Dutch	English	German
prediction	high-certainty	4	4	4
	low-certainty	13	13	12
	neutral	17	17	13
intention	high-certainty	2	2	2
	neutral	4	4	4

Table C.3

Item balance across language, ftr mode, and temporal distance (Study 1)

FTR-mode	temporal distance	Dutch	English	German
prediction	today	1	1	1
	tomorrow	1	1	
	one week	1	1	1
	one month	3	3	2
	six months	2	2	2
	one year	5	5	5
	two years	5	5	5
	ten years	3	3	2
	twenty five plus	6	6	4
	ongoing prediction	5	5	5
	indeterminate	2	2	2
intention	tomorrow	1	1	1
	one week	3	3	3
	six months	1	1	1
	one year	1	1	1

Table C.4

English classification of FTR tense and modality, with word stem frequencies

meta category	category	word/expression	count[h]	percent[i]	justification
verbal-low-certainty	modal verbs	might	2141	7.48%	indicates %75 probability[a]
		may	1810	6.32%	indicates %50 probability/permission[a]
		could	1277	4.46%	indicates possibility/high probability[b]
		should	731	2.55%	indicates %75 probability[a]
		would	187	0.65%	indicates possibility/high probability[b]
		can	133	0.46%	indicates “possibility/high probability (with emphasis on possibility)”[b]
verbal-high-certainty	modal verbs	must	17	0.06%	expresses necessity/obligation[a]
other-low-certainty	modal modifiers	probable	681	2.38%	“in a way that seems likely to prove true; with likelihood (though not with certainty)”[c]
		likely	364	1.27%	“having a high chance of occurring; probable”[c]
		possible	236	0.82%	“qualifying a statement, and expressing contingency or uncertainty”[c]
		chance	111	0.39%	
		maybe	82	0.29%	“possibly, perhaps”[c]
		potential	74	0.26%	conveys uncertainty[d]
		perhaps	12	0.04%	
		unsure	11	0.04%	
		risk	8	0.03%	
		50/50	5	0.02%	
		it is unclear if	4	0.01%	
		supposed	4	0.01%	
	mental state predicates	think	842	2.94%	primary epistemic mental state predicate in English, encodes weakening[e]
		expect	91	0.32%	“to regard as probable or imminent; to envisage; to anticipate”[c]
		believe	45	0.16%	
		reckon	12	0.04%	“to expect, plan, or intend to do something”[c]
		doubt	11	0.04%	encodes weakening commitment to proposition[e]
		suppose	4	0.01%	
other-high-certainty	modal modifiers	definite	736	2.57%	determinate, fixed, certain[f]
		sure	294	1.03%	conveys certainty/sureness[d]
		certain	266	0.93%	in a certain manner; determined, fixed, settled[f]
		for sure	43	0.15%	as a certainty, for certain; without doubt; undoubtedly[f]
		absolute	20	0.07%	
		positive	18	0.06%	
		100%	10	0.03%	
		guarantee	8	0.03%	
future	grammar book “futures”	will	16508	57.66%	high frequency, central FTR auxiliary[b]
		be going to	1616	5.64%	indicates FTR[g]
		shall	63	0.22%	expresses “only future time orientation”[b]
		set to	14	0.05%	
present[l]	present tense response	-	-	-	-

[a]Shepherd et al. (1984); [b]Mindt (1995); [c]Oxford University Press (2017); [d]Lee (2016); [e]Nuyts (2000); [f]Huitink (2008); [g]Dahl (2000a).
[h] Counts are for all derived forms of the root given above, i.e. the word stem. For example, the count for “probable” is inclusive of “probable”, “probability”, and “probably”. Counts are not subjected to the dominance relationships described in the main text so will not align with percentages in other analyses. The purpose of *Tables C.4–C.6* is simply to give a sense of word frequency, so dominance subjected counts would not make sense.
[i]Percent, p, is equal to $p = count/n * 100$, where n is the number of responses in each language.
[k]Counts, percentages and word examples are not given for present tense, as the roots involved are simply the matrix verbs in the present tense, and would be meaningless unless subjected to the dominance hierarchy describe in the main text, as they would appear in almost every response. For this reason as well, percentages do not add up to 100%.
[l]For brevity, any word with 3 or less occurrences has been excluded.

Table C.5

Dutch classification of FTR tense and modality, with word stem frequencies

meta category	category	word/expression	count	percent	gloss	justification
verbal-low-certainty	modal verbs	kunnen	1038	3.69%	can/could	encodes epistemic possibility and/or dynamic ability[a b]
		zouden	351	1.25%	would/should	used to express the conditional tense in future time reference, and sometimes epistemic probability/possibility[b c]
		mogen	35	0.12%	may	primarily deontic/dynamic modal, but review of our data indicate peripheral epistemic uses are possible (rare)[b b]
verbal-high-certainty	modal verbs	moeten	88	0.31%	must	encodes obligation or certainty, generally used in deontic not epistemic contexts[b]
other-low-certainty	modal modifiers	waarschijnlijk	1328	4.72%	probably	encodes "intermediate value between necessity and possibility"[d b]
		misschien	820	2.91%	perhaps	conveys weak epistemic marking[e b]
		mogelijk	185	0.66%	possibly	indicates uncertainty about event obtaining[f b]
		kans	89	0.32%		
		wellicht	76	0.27%	perhaps	encodes uncertainty[d]
		vermoedelijk	31	0.11%	presumably	epistemic use weakening commitment possible[b]
		ongeveer	16	0.06%	about, roughly, approximately	encodes uncertainty[g]
		50%	6	0.02%		
		is er een kans	4	0.01%	there is a chance	indicates uncertainty
	mental state predicates	denken	623	2.21%	think	prototypical mental state predicate for expressing epistemic probability[b]
		verwacht	126	0.45%	perhaps	conveys epistemic possibility[e]
		volgens mij	83	0.29%		
		geloven	43	0.15%		
		vermoeden	21	0.07%		
		horen	15	0.05%	hear	primarily evidential, but with noted overlap with epistemic qualificational use[b]
		verwachten	14	0.05%		
other-high-certainty	modal modifiers	zeker	758	2.69%	certainly	conveys epistemic certainty[d]
		vast	92	0.33%	fixed, firm	encodes necessity, fixedness[h]
		sowieso	66	0.23%	in any event	conveys sureness/certainty[g]
		echt	55	0.2%	really	conveys sureness/certainty[g]
		absoluut	40	0.14%	absolutely	used to indicate sureness[g]
		hoogstwaarschijnlijk	35	0.12%	highly probable	conveys sureness/certainty[g]
		gegarandeerd	28	0.1%	guaranteed	conveys sureness/certainty[g]
		beslist	23	0.08%	definitely, surely	conveys epistemic certainty[d i]
		definitief	23	0.08%	definitely	conveys sureness/certainty[g b]
		100%	17	0.06%		
		nodig	16	0.06%		
		duidelijk	9	0.03%	clearly	low frequency predicational adjective to indicate certainty,lack of alternatives[b]
		ongetwijfeld	8	0.03%	doubtless	encodes high certainty[b]
		natuurlijk	6	0.02%	naturally, of course	encodes sureness[g]
future	grammar book "futures"	zullen	6967	24.75%	will	temporal auxiliary indicating future[j]
		gaan	3295	11.7%	be going to	indicates future time reference[c]
		staat op	8	0.03%	about to	indicates near future[k]
present	present tense response	-	-	-	-	-

[a]van der Auwera and Plungian (1998); [b]Nuyts (2000); [c]Donaldson (2017); [d]de Haan (1999); [e]Byloo et al. (2007); [f]Huitink (2012); [g]Lee (2016); [h]Mortelmans (2009); [i]Simon-Vandenbergen and Aijmer (2007); [j]Kirsner (1969); [k]Dahl (2000a).
See notes for *Table C.4*.

Table C.6

German classification of FTR tense and modality, with word stem frequencies

meta category	category	word/expression	count[o]	percent	gloss	justification
verbal-low-certainty	modal verbs	$\text{können}_{\text{SUBJ}}$	108	2.88%	may	conveys epistemic probability, weakened compared with $\text{können}_{\text{IND}}$[a b]
		$\text{können}_{\text{IND}}$	64	1.71%	may	indicates precarious chance of obtaining, i.e. 50/50 odds[c d e a]
		$\text{mögen}_{\text{IND}}$	19	0.51%	may	conveys epistemic possibility[a]
		$\text{sollen}_{\text{IND}}$	17	0.45%	should	some epistemic use indicating probability[a b]
		$\text{werden}_{\text{SUBJ}}$	12	0.32%	would	used to indicate uncertainty[f]
		$\text{sollen}_{\text{SUBJ}}$	4	0.11%	should	some epistemic use possible [f a]
other-low-certainty	modal modifiers	wahrscheinlich	316	8.43%	probably	conveys probability/likelihood[a h]
		möglicherweise	53	1.41%	possibly	conveys possibility[a h]
		unter umständen	48	1.28%	in some circumstances	conveys likelihood/possibility[i a i a]
		eventuell	29	0.77%	potentially/perhaps	conveys probability[j h]
		möglich	17	0.45%	possible/possibly	neutral epistemic qualification, i.e. 50/50[a]
		vermutlich	15	0.4%	presumably	epistemic meanings possible, weakening implied[a]
		womöglich	15	0.4%	possibly/maybe	encodes probability/possibility[h k]
	mental state predicates	erwarten	96	2.56%	expect	encodes weakened commitment[l]
		meinen	20	0.53%	mean	indicates probability[a]
		denken	13	0.35%	think	indicates probability[a]
		glauben	12	0.32%	believe	indicates probability[a a]
other-high-certainty	modal modifiers	sicher	144	3.84%	surely/must	conveys certainty[b a]
		definitiv	53	1.41%	definitely	conveys certainty[m]
		sicherlich	30	0.8%	certainly	conveys certainty[d h]
		bestimmen	25	0.67%	certainly	indicates high certainty[n a]
		auf jeden fall	8	0.21%	in any event	conveys certainty[h m]
future	grammar book "futures"	werden	1778	47.45%	will/shall	future time auxiliary[c]
present	present tense response	-	-	-	-	-

[a]Nuyts (2000); [b]Mortelmans (2009); [c]Bouma (1973); [d]C. M. Stevens (1995); [e]van der Auwera and Plungian (1998); [f]Advice from a native-speaking informant provided to the author; [g]Zimmermann (2019); [h]Majstro (2017); [i]Frahling (2017); [j]Teich (2003); [k]Efstathiadi (2009); [l]Rau (2011); [m]Simon-Vandenbergen and Aijmer (2007); [n]Karagjosova (2004).
[o]Counts for German modal verbs are of lemmas, not stems, i.e. subjunctive and indicative forms are counted separately.
See notes for *Table C.4*.

Table C.7

Participant responses from each category of the FTR-type classifier coding system for each language (Study 1)

coding category	English	Dutch	German
future tense	He will call me.	Hij zal me bellen.	Er wird mich anrufen.
present tense	No, I am living in San Francisco next year.	De woningmarkt valt om in de komende 2 jaar.	Der bruder kommt nächste woche.
other-low-certainty	The housing market will probably crash in the next two years.	Het regent waarschijnlijk.	Es wird wahrscheinlich regnen.
verbal-low-certainty	Next year, the housing market might crash.	Zij kunnen stijgen.	Der Immobilienmarkt könnte zusammenbrechen.
other-high-certainty	They are definitely going to rise.	Ik zeg nee deze keer absoluut.	Er steigt aufjedenfall.
verbal-high-certainty	I must say no this time.	Ze moeten zeker stijgen.	...ich muss jedes mal nein sagen.

Table C.8

Accuracy, precision, recall, and F1 for the FTR-type classifier (Study 1)

variable	accuracy	precision	recall	*F1*
future tense	0.98	0.98	0.99	0.98
present tense	0.96	0.97	0.98	0.97
low-certainty	0.98	0.99	0.98	0.98
high-certainty	0.99	0.99	0.99	0.99

Accuracy is defined as $a = (tp+tn)/(tp+fp+fn+tn)$ where tp is the number of true positives, tn is the number of true negatives, fp is the number of false positives and fn is the number of false negatives. Accuracy captures the classifier's performance without prioritising either positive or negative examples. Precision is $p = tp/(tp+fp)$; it captures the model's likelihood of being correct if it makes a positive prediction and is therefore sensitive to the model's type I error rate. Conversely, denominator in recall is the false negatives, $r = tp/(tp+fn)$; it therefore captures whether the model tends to miss true examples, and is sensitive to the model's type II error rate. *F1* is the harmonic mean of recall and precision, $F1 = (2rp)/(r+p)$, and attempts to balance these two priorities.

Table C.9

Accuracy, precision, recall, and F1 for the FTR-type classifier (Study 2)

usage variable	accuracy	precision	recall	*F1*
future tense	0.999	0.999	1	0.999
present tense	0.998	0.997	1	0.999
low-certainty	0.998	1	0.997	0.999
high-certainty	0.999	1	0.999	0.999

Table C.10

Random variance components for the mixed models in Studies 1 and 2

study	outcome	group	component	σ^2	SD	$slope : slope\ r$	$slope : slope\ r$	χ^2	DF	p
Study 1	low-certainty	participants	(Intercept)	1.52	1.23	-	-	359.72	1	$< .001$
		items	(Intercept)	0.81	0.9	-	-	14040.47	1	$< .001$
			Dutch slopes	0.37	0.61	0.07	-	310.44	5	$< .001$
			German slope	0.73	0.85	-0.12	0.32			
		exp. version	(Intercept)	0.11	0.33	-	-	6.48	1	.011
Study 2	low-certainty	participants	(Intercept)	1.87	1.37	-	-	2439.37	1	$< .001$
		items	(Intercept)	0.12	0.35	-	-	17627.77	1	$< .001$
			Dutch slopes	0.2	0.45	-0.04	-	354.33	20	$< .001$

For random variance components, significance testing was accomplished by using log-likelihood ratio testing to infer whether adding a random parameter improved model fit compared with a nested model without that parameter. This was performed using the *anova* method in R.

C.9 Figures

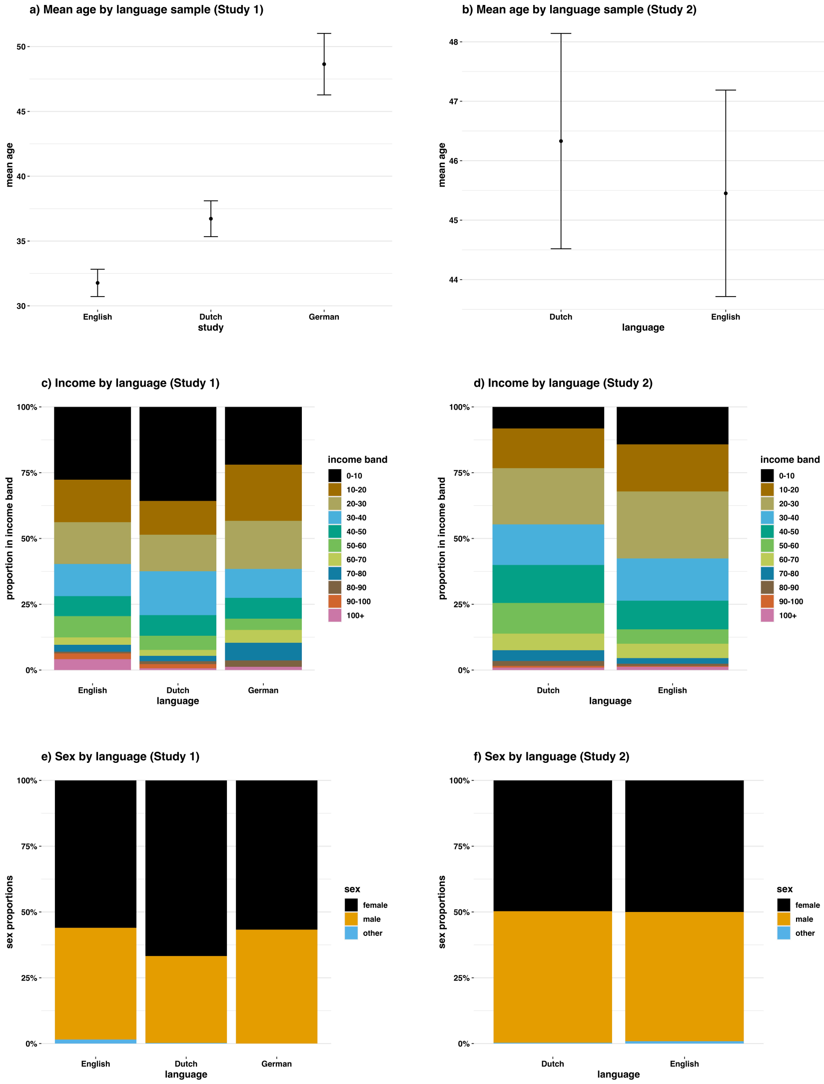


Figure C.1: Demographic sample composition by language for age, income, and sex (Studies 1 and 2). In the sex question, participants were given 4 options (male, female, do not identify as either, and prefer not to say). The latter two are collapsed into "other." Error bars in sex plot represent the mean and $\pm 1.96SE$. Other plots (c–f) depict proportions.

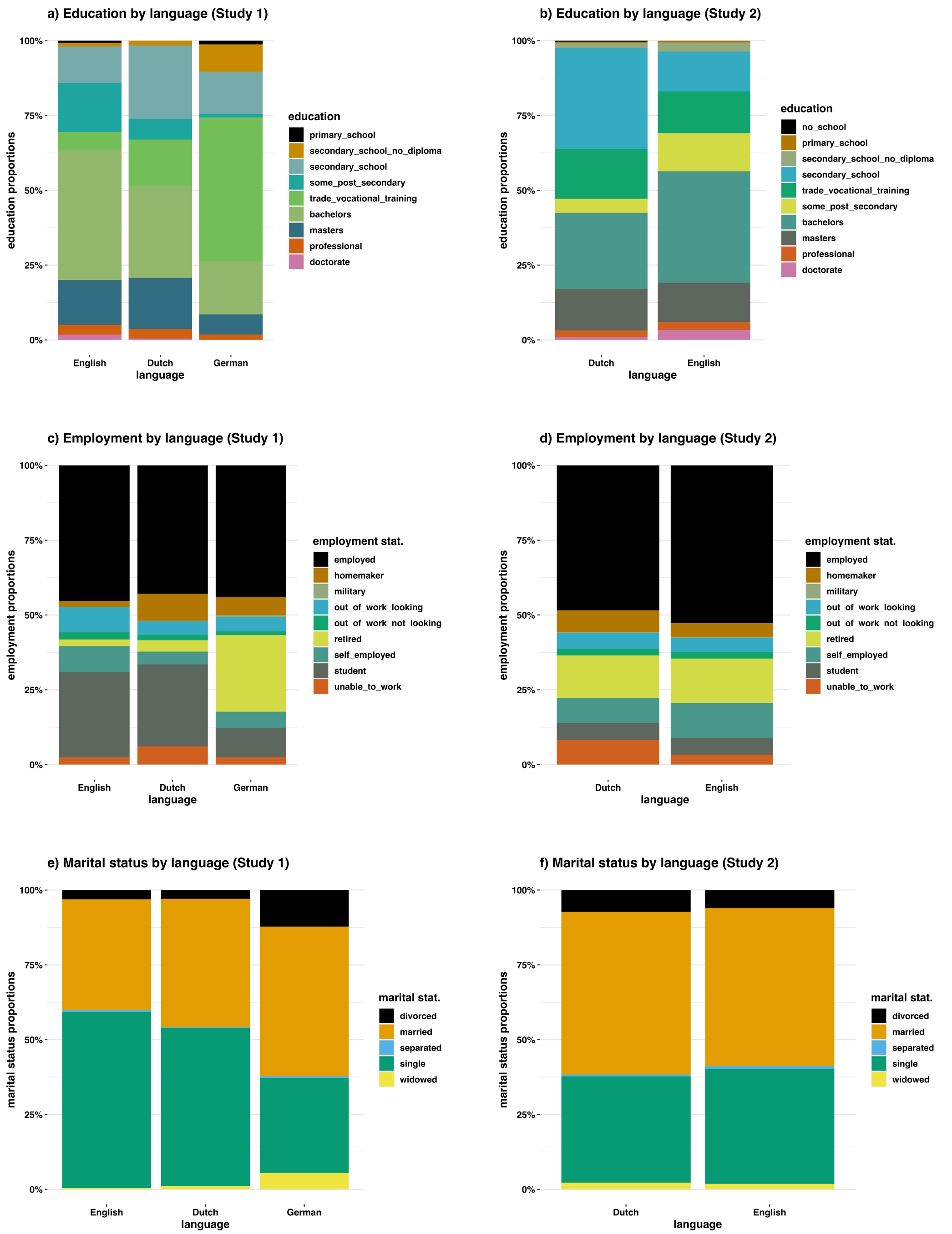


Figure C.2: Demographic sample composition by language for education, employment status, and marital status (Studies 1 and 2). See notes to *Fig. C.1*.

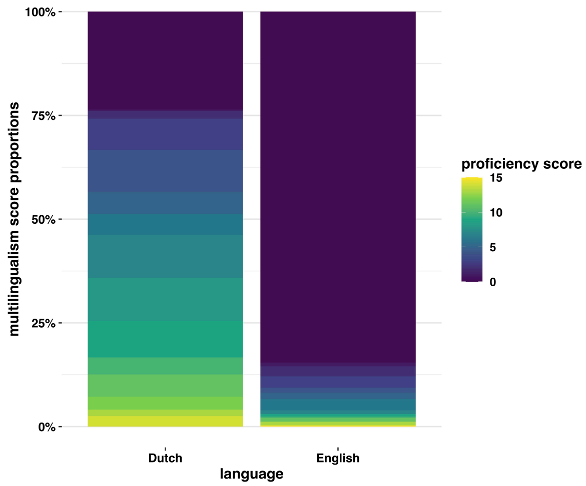


Figure C.3: Multilingualism score by language (Study 2). Multilingualism score is defined as $\sum_{n=1}^{k} s_i$, where k are the up to 3 language proficiency questions answered by participant i. The Dutch sample was much more multilingual than the English sample.

Appendix D

Supplementary materials to Chapter 3

In this appendix, I provide supplementary materials to Chapter 3.

D.1 Cross-study analyses

In this section, I present results which are pertinent to Studies 1–3. Following this, I present analyses which pertain to each study individually.

D.1.1 FTR-type classifier accuracy

See *Table D.1* for accuracy metrics for the predictions of the FTR-type classifier against human raters. Data are drawn from Studies 1–3.

D.1.2 FTR-elicitation task factor balance

Balance of FTR-mode, temporal distance, and probability condition in the FTR-elicitation task across experiments are given in *Tables D.2–D.4*. Tables present the counts of unique items per cross-tabulated factor of the FTR-elicitation task, for each of Studies 1–3.

D.1.3 Demographic differences and descriptions

Demographics are reported in *Fig. D.1*. In no case did adding any of these variables substantively change parameters of interest, or significantly improve model fit, $p > .05$. They were therefore disregarded.

D.2 Study 1

D.2.1 Item exclusions

Of the 29 items in the FTR-elicitation task, one needed to be removed from analysis because because it was later realised that the prompt included common modal or tense words (e.g. think, expect, go) which prevented the FTR classifier from accurately identifying whether participants had used a tense or modality word themselves, or whether they had simply copied the prompt. Additionally, one question needed to be removed because inspection revealed that participants consistently interpreted the context as referring to present time, and one question needed to be removed because it was in both these categories. A total of three questions were therefore removed, so analysis was conducted on data from 26 of the 29 items participants were given.

Questions can be viewed at the FTR-elicitation task github repository, here: `https://github.com/cbjrobertson/ftr_questionnaire_master/tree/master/ftr_questionnaire/versions_1-2`. There are instructions on how to view the questions which were given to participants and analysed in the main text.

D.2.2 Selecting items in the FTR-elicitation task

To chose the items which we would use in Study 1, we conducted a pilot where we collected English, Dutch, and German FTR usage data using the full (48 item) FTR-elicitation task. Results of this pilot are reported as Study 1 Version A, in Chapter 2. See Section 2.2 for experimental methods. In the present analyses, we discarded the German data because they were irrelevant. To choose the questions to be used in Study 1, we regressed future tense use over an intercept, which we allowed to vary by item, and participant. We also allowed slopes for language to vary by item so that random slopes would be estimated for both English and Dutch:

$$log_e(\frac{\pi_{ijk}}{1-\pi_{ijk}}) = \beta_0 + (\beta_1 + s_k)LANG_{ijk} + p_j + q_k \tag{D.1}$$

where π_{ijk} is the probability that $y_{FUT} = 1$ for observation i, participant p_j, question q_k, and language. Random intercepts and slopes, p_j, q_k, s_j, are assumed to be drawn from the normal distributions with means of 0 and observed standard deviations observed in the sample, i.e. $p, q, s \sim N(0, \sigma^2_{p,q,s})$. We then sorted in ascending order by the absolute value of the random slope term, s_k, for each language. The allowed us to select the top $n = 20$ items for each language for which the item slope was closest to 0, i.e. the items which had the least constrained contexts in terms of use of the future tense. To create a matched sample of items, we took the union of the results for English and Dutch. This resulted in $N = 29$ unique items. These can be see in Appendix I, *Table I.1.* The question numbers are: 1, 2, 3, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, 18, 19, 21, 25, 27, 28, 32, 33, 34, 35, 37, 39, 41, 45, 46, and 48. To get the question numbers in *Table I.1*, add 1000.

D.3 Study 2

D.3.1 Item exclusions

Final set of FTR-elicitation task questions was identical to Study 1, though an additional two items had to be excluded because they accidentally used an absolute time frame which at the time the experiment was conducted was in the present/past ("by 2018"). This means the final number of elicitation questions analysed is 24 items. Details can be viewed at the github link above.

D.3.2 Power analysis

In order to arrive at an optimal number of participants for Study 2, we conducted a power analysis using the R package *simr* (P. Green & MacLeod, 2016). The power of an hypothesis test can be defined as the "probability that a test will reject the null hypothesis, assuming that the null hypothesis is false" (P. Green & MacLeod, 2016, p. 493). In other words, it is the probability that a hypothesis test will detect a true effect. A standard acceptability rate is 80% (P. Green & MacLeod, 2016). The *simr* package (P. Green & MacLeod, 2016) calculates power for generalised mixed models estimated with the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015), so it was well-suited to our needs. Power calculations in *simr* are based on Monte Carlo simulations which allow researchers to simulate power under various hypothetical scenarios. In this way, results can inform trade-offs between sample size and power (P. Green & MacLeod, 2016). Basically, researchers can specify a mixed model in *lme4*, and then use *simr* to understand the power of the model given previous data, and across hypothetical values for fixed effects and sample sizes.

Our analytic approach at the time we were planning Study 2 was different from the analyses we present in the paper (we thank our reviewers for pointing us towards better analytic approaches). This means our power analysis was based on a different model structure than we report in the main text. However, since it informed our decisions about sample size, it is reported here for transparency. We were interested in the impact of future tense use on time preferences. At the time, we were using mixed logistic regressions to regress raw Inter Temporal Choice (ITC) values (SSR = 0; LLR = 1) over participant-level mean future tense use, and allowing intercepts to vary by participant, SSR amount, and LLR delay:

$$log_e(\frac{\pi_{ijkl}}{1-\pi_{ijkl}}) = \beta_0 + \beta_1 FUT_i + p_j + d_k + a_l \tag{D.2}$$

where π_{ijkl} is the probability that $y_{ITC} = 1$ for observation i, participant p_j, delay d_j, amount a_l. Random intercepts, $p_j,\ d_k,\ a_l$, are assumed to be drawn from the normal distributions with means of 0 and observed standard deviations observed in the sample, i.e. $p, d, a \sim N(0, \sigma^2_{p,d,a})$. When we estimated this model using the data from Study 1, the effect of future tense was $\beta_1 FUT = 0.26$. In our simulations, we therefore used 0.25 as our lower bound, and simulated potentially larger effects between $[0.25, 0.4, 0.75, 0.9]$. We then calculated power for a sample of $N = 0$–$N = 300$ participants (the largest sample our budget allowed) for each hypothetical $\beta_1 FUT$.

Results are reported in *Fig. D.3.* They are not particularly promising. They suggest that even with $N = 300$ participants, Study 2 may be underpowered for $\beta_1 FUT = 0.25$ and $\beta_1 FUT = 0.4$. As $\beta_1 FUT$ increases, so does power: for $\beta_1 FUT = 0.75$, approximately $N = 175$ participants are necessary to reach 80% power; for $\beta_1 FUT = 0.9$, this drops to $n = 120$. Since our observed $\beta_1 FUT$ was 0.25, we concluded a large sample was appropriate and decided to collect data from $n = 300$ participants in each language.

D.3.3 Delay and amount adjustment procedure

The amounts and delays in Study 1 appeared to be too distant from relevant decision boundary criteria. For instance, 82.64% of Dutch participants chose the larger-later reward 100% of the time for one-day delays. To choose the amounts and delays for Study 2, we used binary logistic regression to regress raw Inter Temporal Choices (ITCs) (SSR = 0; LLR = 1) over our language dummy, future tense use, LLR delay, and SSR amounts. We included all two- and three-way interactions, and allowed intercepts to randomly vary by participant:

$$
\begin{aligned}
log_e(\frac{\pi_{ij}}{1-\pi_{ij}}) = \beta_0 + \beta_1 LANG_i + \beta_2 FUT_i + \beta_4 DEL_i + \beta_4 AMT_i + \\
&\beta_5 LANG_i : FUT_i + \\
&\beta_6 LANG_i : DEL_i + \\
&\beta_7 LANG_i : AMT_i + \\
&\beta_8 FUT_i : DEL_i + \\
&\beta_9 FUT_i : AMT_i + \\
&\beta_{10} AMT_i : DEL_i + \\
&\beta_{11} LANG_i : FUT_i : DEL_i + \\
&\beta_{12} LANG_i : FUT_i : AMT_i + \\
&\beta_{13} LANG_i : DEL_i : AMT_i + \\
&\beta_{14} FUT_i : DEL_i : AMT_i + \\
&+ p_j
\end{aligned}
\tag{D.3}
$$

where π_{ij} is the probability that $y_{ITC} = 1$ for observation i and participant p_j. The random intercepts, p_j, are assumed to be drawn from the normal distributions with means of 0 and observed standard deviations drawn from the sample, i.e. $p \sim N(0, \sigma_p^2)$. Amount was operationalised as the raw SSR amount values; delay was operationalised as the natural log of the number of days. This allowed us to estimate the probability of choosing the LLR over future tense use across the LLR delays and SSR amounts under the study. Data were the data from Study 1. This analysis indicated amounts approaching the larger-later reward (i.e. > 7) and with delays between 2 and 6 months had the strongest correlations with future tense use, e.g. steeper slopes can be observed over future tense use in *Fig. D.2.* We therefore adjusted the amounts and delays in Study 2 to approximate these ranges.

D.4 Study 3

D.4.1 Certainty ratings correlations

To see whether some participants had rated the future as higher certainty and others rated the present as higher certainty, we plotted present tense certainty ratings over future tense certainty ratings, *Fig. D.4.* Data for this plot are from the "certainty slider task" in Study 2 of Chapter 2. See Section 2.3.1 for methods. Some participants rated the future tense as encoding relatively higher certainty, for others, it was the other way around, *Fig. D.4.*

D.4.2 Item exclusions

Of the 100 items in the updated FTR questionnaire, 17 had to be removed because it was later realised that the prompts included common modal or tense words (e.g. think, expect, go) which prevented the FTR-type classifier from accurately identifying whether participants had used a tense or modality word themselves, or whether they had simply copied the prompt. Two additional questions had to be removed because inspection revealed that participants consistently interpreted the context as referring to present time. Additionally, one question had to be removed because it accidentally used an absolute time frame which, at the time the experiment was conducted, was in the present ("by 2018"). This left a final sample of 80 items.

D.5 Tables

Table D.1

FTR-type classifier performance metrics by FTR-type variable

variable	accuracy	precision	recall	*F1*
future tense	0.98	0.98	0.99	0.98
present tense	0.96	0.97	0.98	0.97
low-certainty	0.98	0.99	0.98	0.98
high-certainty	0.99	1.00	0.99	1.00

Clearly by any metric, the performance of the FTR classifier is good. It should be noted that it is highly likely the FTR-type classifier is overfit to these data, though this is not a problem: The FTR-type classifier is designed to be used on data generated by the FTR questionnaire, so does not need to generalise beyond this constrained domain. Accuracy is defined as $a = (tp + tn)/(tp + fp + fn + tn)$ where tp is the number of true positives, tn is the number of true negatives, fp is the number of false positives and fn is the number of false negatives. Accuracy captures the classifier's performance without prioritising either positive or negative examples. Precision is $p = tp/(tp + fp)$; it captures the model's likelihood of being correct if it makes a positive prediction and is therefore sensitive to the model's type I error rate. Conversely, denominator in recall is the false negatives, $r = tp/(tp + fn)$; it therefore captures whether the model tends to miss true examples, and is sensitive to the model's type II error rate. *F1* is the harmonic mean of recall and precision, $F1 = (2rp)/(r + p)$, and attempts to balance the two.

Table D.2

FTR task cross-tabs: FTR-mode x certainty condition

study	FTR-mode	certainty condition	Dutch	English
Study 1	intention	high-certainty	2	2
		neutral	3	3
	prediction	high-certainty	3	3
		neutral	11	11
		low-certainty	6	6
	scheduling	neutral	1	1
Study 2	intention	high-certainty	2	2
		neutral	3	3
	prediction	high-certainty	2	2
		neutral	11	11
		low-certainty	5	5
	scheduling	neutral	1	1
Study 3	intention	high-certainty	1	-
		neutral	16	-
	prediction	high-certainty	11	-
		neutral	50	-
	scheduling	neutral	2	-

Table D.3

FTR task cross-tabs: FTR-mode x temporal distance

study	FTR-mode	temporal distance	Dutch	English
Study 1	intention	one week	3	3
		six months	1	1
		tomorrow	1	1
		one week	3	3
		six months	1	1
		tomorrow	1	1
	prediction	indeterminate	2	2
		one month	1	1
		one year	6	6
		ongoing prediction	2	2
		six months	2	2
		ten years	2	2
		today	1	1
		tomorrow	1	1
		twenty five plus	2	2
		two years	1	1
	scheduling	one year	1	1
Study 2	prediction	indeterminate	2	2
		one month	1	1
		one year	4	4
		ongoing prediction	2	2
		six months	2	2
		ten years	2	2
		today	1	1
		tomorrow	1	1
		twenty five plus	2	2
		two years	1	1
	scheduling	one year	1	1
Study 3	intention	one month	1	-
		one week	6	-
		one year	2	-
		six months	2	-
		today	3	-
		tomorrow	3	-
	prediction	indeterminate	4	-
		one month	3	-
		one week	2	-
		one year	8	-
		ongoing prediction	18	-
		six months	3	-

Continued on next page...

Table D.3 – *FTR task cross-tabs: FTR-mode x temporal distance continued*

study	FTR-mode	temporal distance	Dutch	English
Study 3	prediction	ten years	1	-
		today	6	-
		tomorrow	4	-
		twenty five plus	9	-
		two years	3	-
	scheduling	today	1	-
		tomorrow	1	-

Table D.4

FTR task cross-tabs: certainty condition x temporal distance

study	certainty condition	temporal distance	Dutch	English
Study 1	high-certainty	one week	1	1
		one year	2	2
		six months	1	1
		tomorrow	1	1
	neutral	indeterminate	1	1
		one month	1	1
		one week	2	2
		one year	2	2
		ongoing prediction	1	1
		six months	2	2
		ten years	1	1
		today	1	1
		tomorrow	1	1
		twenty five plus	2	2
		two years	1	1
	low-certainty	indeterminate	1	1
		one year	3	3
		ongoing prediction	1	1
		ten years	1	1
Study 2	high-certainty	one week	1	1
		one year	1	1
		six months	1	1
		tomorrow	1	1
	neutral	indeterminate	1	1
		one month	1	1
		one week	2	2
		one year	2	2
		ongoing prediction	1	1
		six months	2	2
		ten years	1	1
		today	1	1
		tomorrow	1	1
		twenty five plus	2	2
		two years	1	1
	low-certainty	indeterminate	1	1
		one year	2	2
		ongoing prediction	1	1
		ten years	1	1
Study 3	high-certainty	indeterminate	1	-
		one year	2	-

Continued on next page...

Table D.4 – *FTR task cross-tabs: certainty condition x temporal distance continued*

study	certainty condition	temporal distance	Dutch	English
		ongoing prediction	7	-
		six months	1	-
		tomorrow	1	-
	neutral	indeterminate	3	-
		one month	4	-
		one week	8	-
		one year	8	-
		ongoing prediction	11	-
		six months	4	-
		ten years	1	-
		today	10	-
		tomorrow	7	-
		twenty five plus	9	-
		two years	3	-

D.6 Figures

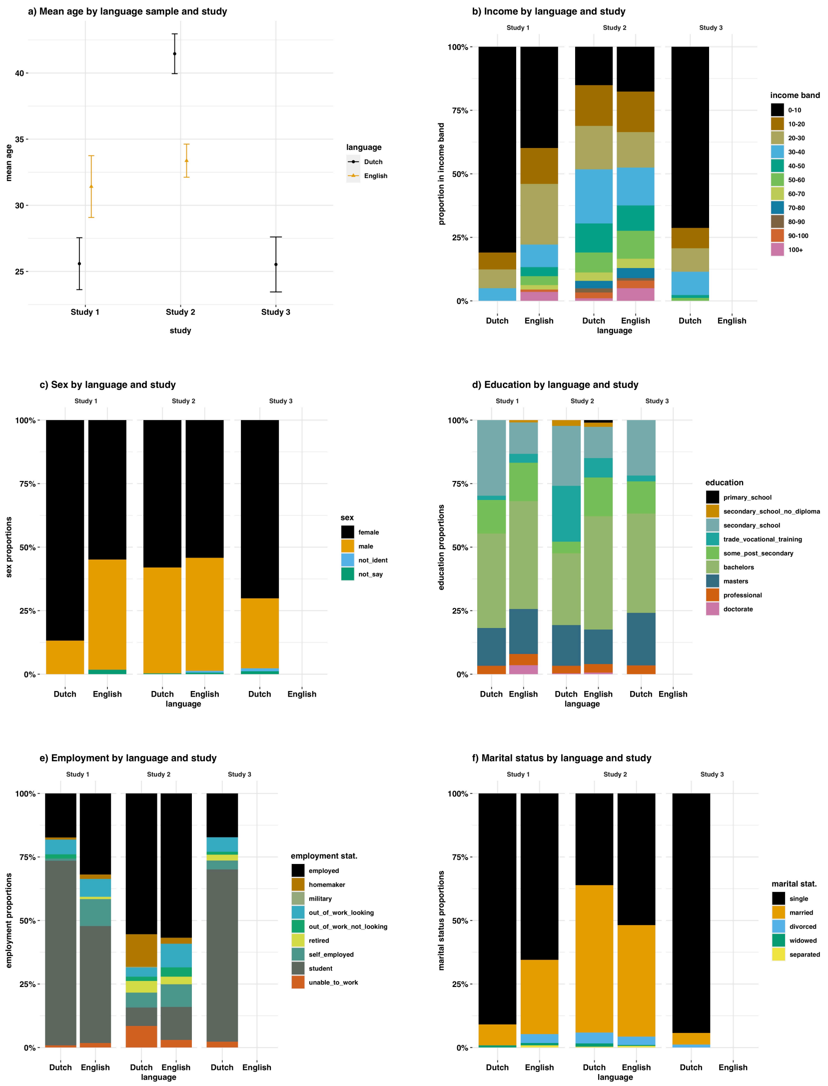


Figure D.1: Demographic sample composition by language and study. The mean ±1.96 *SE* is reported for sex (a). Other sub-figures (b–f) report proportions.

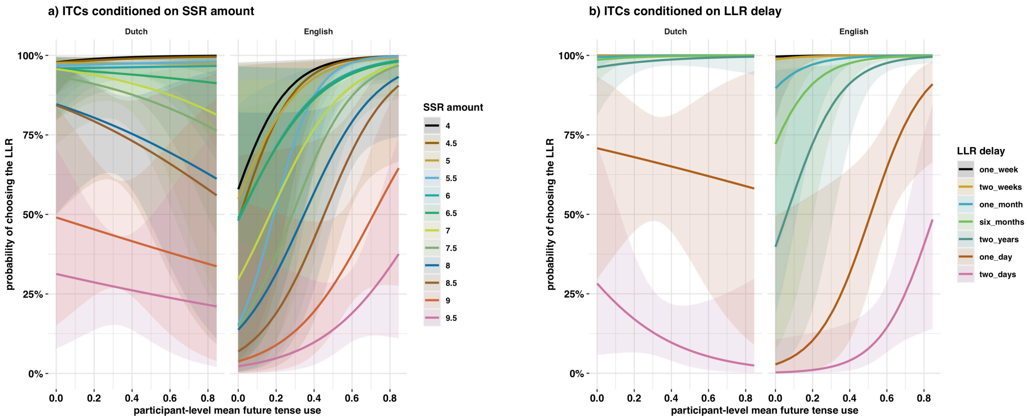
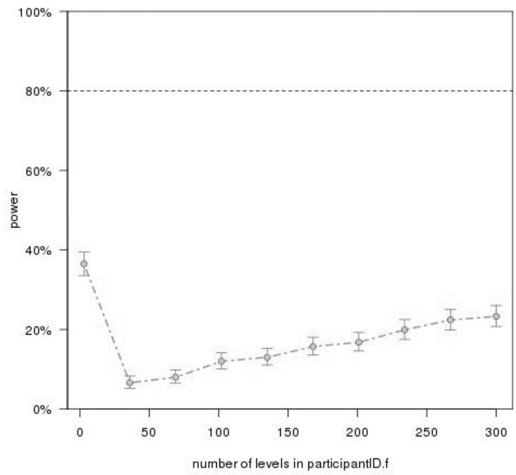
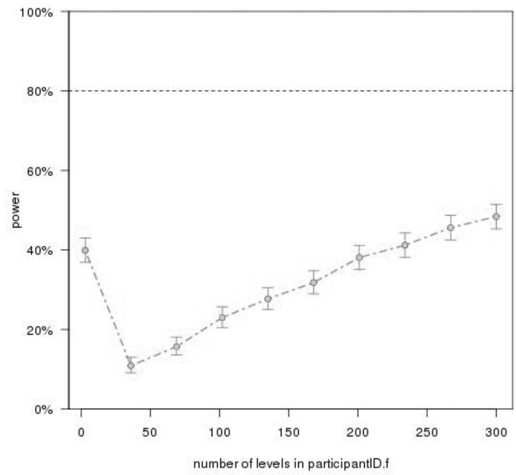


Figure D.2: Probability of choosing the LLR over participant-level mean future tense use in . Probabilities are conditioned on SSR amounts and delays. Plotted values are the marginal effects, ±1.96 SE (see Lüdecke, 2019). For ease of interpretation, adjectival delays are shown rather than $ln(days)$ in (b).

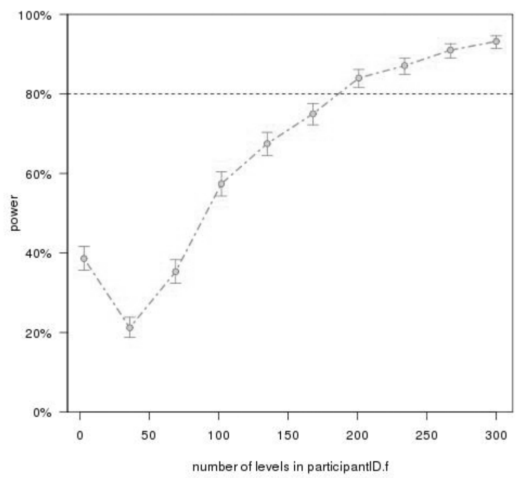
(a) $\beta FUT = 0.25$



(b) $\beta FUT = 0.4$



(c) $\beta FUT = 0.75$



(d) $\beta FUT = 0.9$

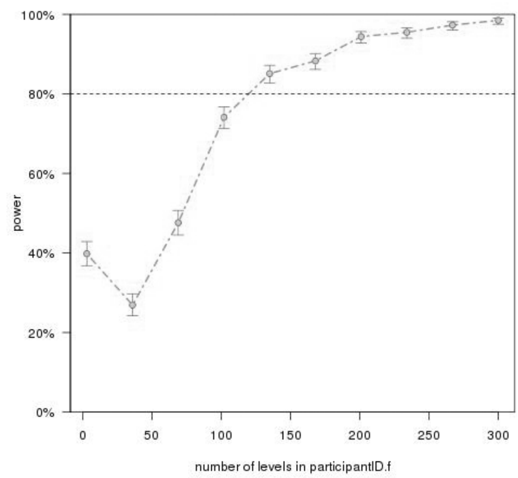


Figure D.3: Simulated power under different hypothetical sample and effect sizes, according to our *simr* analysis. The standard cut-off of 80% is depicted (P. Green & MacLeod, 2016). Simulated power is plotted on the y-axes. Simulated sample size (number of participants) is plotted on the x-axes.

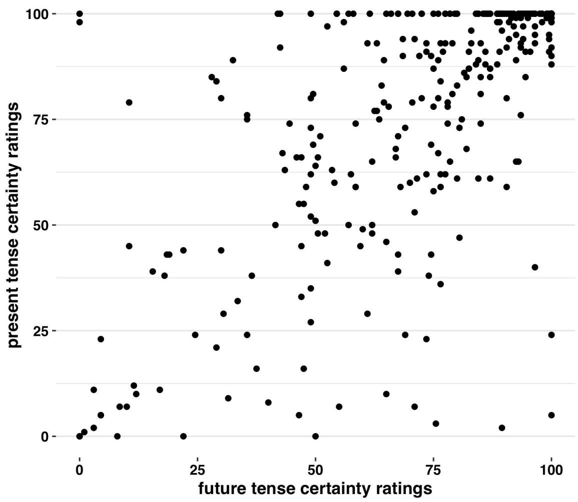


Figure D.4: Mean certainty ratings for the present tense over the future tense in Dutch. There were two items in the future tense category (i.e. *zullen* and *gaan*). This plot therefore depicts means per participant and FTR type, i.e. $CERT_{jk} = \sum_{n}^{i=1} \frac{CERT_{ij}}{n_{jk}}$, where $CERT$ is certainty rating (0-100), and n is the total number of responses by participant j within each FTR-type category k.

Appendix E

Supplementary materials to Chapter 4

E.1 Alternative methods for deriving indifference points in Study 1

To make the Study 1 outcome comparable with outcomes in Studies 2 and 3, we calculated indifference points for each participant and delay. This is unconventional. A more typical approach is to calculate k following Kirby et al. (1999) and then use $log_e(k)$ as a participant-level measure of discounting (because k is generally exponentially distributed). In fact, Kirby's (1999) method involved calculating indifference points and then using these to identify the k which best approximates participants' empirical choices. However, the resultant indifference points are directly proportionate to k given equation 4.5. They therefore assume the true discounting function is hyperbolic (Myerson, Green, & Warusawitharana, 2001), when temporal discounting is better described by hyperboloid functions (L. Green & Myerson, 2004). To better capture empirical discounting, we therefore adjusted the indifference points derived from Kirby's (1999) method. To do this in a principled way, we trialled several adjustment procedures and then evaluated these against two performance metrics.

E.1.1 Calculation of performance metrics

The first measure was simply the number of incorrectly predicted choices. However, this metric does not capture the extent of the error involved in an incorrect prediction. For instance, if a participant is predicted to choose the LLR (£10) and they choose an $SSR = 9.5$, this is "less wrong" than if they choose an $SSR = 4$. We therefore developed the following error metric, which captures such differences:

$$e = \sum_{i=1}^{n} \begin{cases} 0, & \text{if } c_p = c_a \\ \sqrt{(v_a - ip)^2}, & \text{otherwise} \end{cases} \tag{E.1}$$

where c_p is the predicted choice, c_a is the participant's actual choice, ip is the hypothetical indifference point, n is the total number of observations, and v_a is the monetary value of c_a:

$$v_a = \begin{cases} 10, & \text{if } c_a = LLR \\ V_{SSR}, & \text{otherwise} \end{cases} \tag{E.2}$$

where V_{SSR} is the monetary value of the SSR. In other words, when predictions were incorrect, we took the square root of the squared difference between the empirical choice value and the hypothetical indifference point. For correct predictions error is equal to zero. This captures degrees of incorrectness between incorrect predictions.

E.1.2 Alternative indifference point adjustment calculations

We then tested several methods for adjusting indifference points and picked the one which performed best according to these two measures of error. Our starting point was the indifferent points derived from inputting the Kirby et al. (1999) values of k into equation 4.5. We had one unique indifference point for each participant and delay. We then duplicated these across the SSR amounts. We then adjusted these according to three different methods. Method 1 was:

$$ip_1 = \begin{cases} ip, & \text{if } c_p = c_a \\ \left(\prod_{i=1}^{2} x_i\right)^{\frac{1}{2}}, & \text{otherwise} \end{cases} \tag{E.3}$$

where ip is the hypothetical indifference point from the Kirby et al. (1999) method, c_p and c_a are the predicted and actual choices and x_i are the $n = 2$ values comprising ip and the actual choice value (i.e. v_a above). This is simply saying that when predictions were correct, we retained the Kirby et al. (1999) indifference point and when predictions were incorrect, we took the geometric mean of the Kirby et al. (1999) indifference point and the actual choice value. Method 2 was identical, but used the arithmetic mean:

$$ip_2 = \begin{cases} ip, & \text{if } c_p = c_a \\ \sum_{i=1}^{2} \frac{x_i}{2}, & \text{otherwise} \end{cases} \tag{E.4}$$

This is identical to equation E.3, but uses the arithmetic mean instead of the geometric mean. Method 3 was even simpler:

$$ip_3 = \begin{cases} ip, & \text{if } c_p = c_a \\ v_a, & \text{otherwise} \end{cases} \tag{E.5}$$

Table E.1

Performance metrics for alternative indifference point calculation methods

metric	averaging calculation	Kirby indifference point*	method 1	method 2	method 3
error	geometric mean	2907.93	1647.52	1624.51	1549.7
	arithmetic mean	2907.93	1629.49	1577.84	1415.87
incorrect predictions	geometric mean	1298	1281	1283	1242
	arithmetic mean	1298	1283	1273	1226

Total number of observations was $N = 18,791$.
*Evaluation metrics for the original Kirby et al. (1999) indifference points.

where v_a is the actual choice value, as above. This simply says that where choices were predicted correctly, we retained the Kirby et al. (1999) indifference points, otherwise we substitutes these for the actual choice value.

We then permuted these with two other calculation techniques. The quantity of interest we needed was a single indifference point for each participant and delay. The calculations described in equations E.3–E.5 had resulted in a set of different adjusted indifference points within each delay and across each SSR amount. We therefore grouped data by participant and delay and either took the arithmetic mean or geometric mean of the set of indifference points. For the arithmetic mean this is equal to $\sum_{i=1}^{n} \frac{x_{ij}}{n}$, and for the geometric mean this is equal to $(\prod_{i=1}^{n} x_{ij})^{\frac{1}{n}}$, where x is the indifference points resulting from E.3–E.5 and n is the total number of indifferences points made by participant i within each delay j. This resulted in 6 different ways of calculating indifference points, i.e. $3_{ip.types} \times 2_{mean.types} = 6$. We used the error metrics given in equations E.1 and E.2 to evaluate which best captured participants responses.

E.1.3 Evaluation

Results are reported in *Table E.1.* Taking the arithmetic mean of the indifference points from Method 3 (equation E.5) performed the best. This is what we report and analyse in the main text.

E.2 Modelling decisions for objective factors of psychological discounting outcomes

In the reported analyses, we use mean-centred, z-scaled days as our measure of temporal distance, and mean-centred, z-scaled odds against as our measure of probability condition. This is unconventional but we do it for a principled reason, which we explain here. It is unconventional because discounting (whether it is over time or probability) is non-linear. Most approaches have understandably focussed on using non-linear regression techniques to establish the shape of discounting functions, e.g. establishing whether hyperbolic (Mazur, 1987) or hyperboloid (L. Green & Myerson, 2004) functions fit empirical indifference points the best.

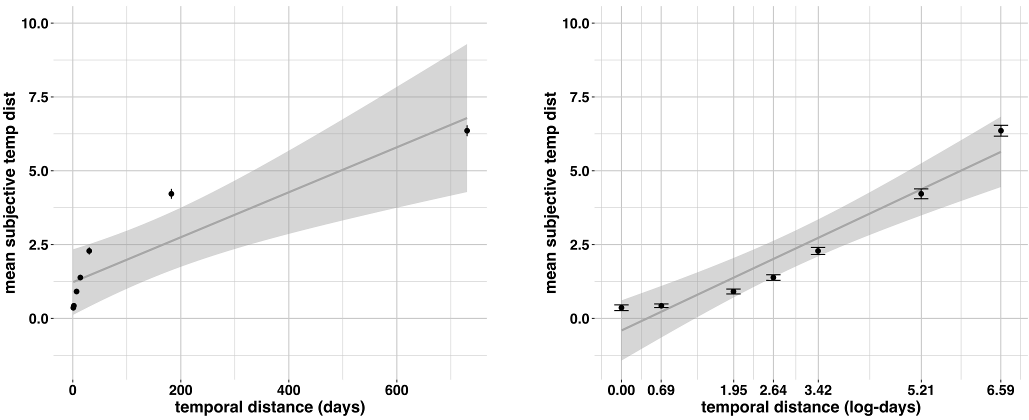


Figure E.1: Mean subjective-temporal-distance ratings ($\pm 1.96SE$) over objective temporal distance in days (left) and in log-days (right) in Study 1, with linear line of best fit.

Our approach is unconventional because we use linear regressions to regress indifference points over objective predictors (time, probability) even though we know that—to some extent—this breaks assumptions of linearity.

This is justified for two reasons. First, throughout analyses we include measures of subjective time (i.e. from the time-slider tasks) and/or subjective probability (accepting that elicited FTR data measure something to do with participants' representations of probability/certainty, which appears to be justified). A critical point is that temporal discounting over subjective representations of time appears to be mostly linear. Using time-slider tasks like the ones we implement Bradford et al. (2019) show the subjective representations of time appear to follow near logarithmic processes (similarly to psychophysical perceptions of heat, sound, and light). Apparently non-constant discounting over objective time may be driven by log-scaled relationships between objective and subjective time. Our data support this conclusion; over objective time, subjective time is curvilinear, whereas over log-time the relationship is basically linear, *Fig. E.1.* Data are from Study 1. Correspondingly, indifference points over objective time are non-linear, but are approximately linear over subjective time, *Fig. E.2.* This suggests that using non-linear regressions over objective time can serve to capture non-linearities in the psychophysical perception of time.

Our approach is different. We directly measure subjective representations of time and probability and seek to estimate discounting over these. Allowing discounting to be non-linear over objective inputs would make little sense. In fact, in unreported analyses, we trialled modelling non-linearities over objective time using both $log_e(t)$ and t^2, and over objective probability using both $log_e(p)$ and p^2. In general, using $log_e(t)$ and t^2 rendered effects of subjective temporal distance non-significant, and using $log_e(p)$ and p^2 renders effects of low-certainty language non-significant. This confirms our account that these measures are captur-

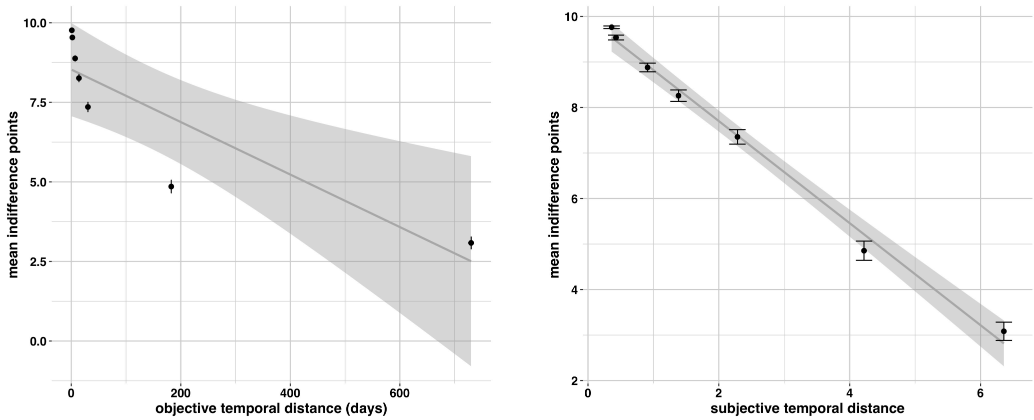


Figure E.2: Mean indifference points ($\pm 1.96SE$) over objective temporal distance in days (left) and subjective temporal distance (right) in Study 1, with linear line of best fit.

ing elements of participants' beliefs about time and probability which may drive non-constant discounting over objective measures. It was therefore appropriate to model discounting over objective time linearly and capture non-linearities via our subjective measures. Inspection of residuals over predicted values in analyses of data from Studies 1–3 suggested that this was generally reasonable. Some non-linearities were present. In particular, the models reported in the main text may underestimate discounting at near delays and low odds against, but overestimate discounting as delays grow longer and odds against grow higher. Nonetheless, violation of regression assumptions appeared to be within tolerable bounds, and the Bayesian methods we use make no assumptions about residual normality so significance inferences were not affected.

The second reason modelling untransformed objective time and probability is justified has to do with generalisability. One of the reasons that using linear models to capture non-linear relationships is problematic is that it makes it impossible to generalise beyond the observed data. However, since we avoid doing this in the main text, this is not really an issue. It would anyway be inappropriate to generalise our findings to other reward amounts and distances since both factors have an impact on discounting rates (L. Green & Myerson, 2004). In any case, the non-linearities are over objective measures of time and probability, which we do not interpret or analyse in the main text. For these reasons, our modelling decisions are sound.

E.3 Normalising model coefficients measured on different scales

In the main text, we report the untransformed model parameters. This means the substantive effect of Xs on Ys via Ms is difficult to interpret. This is because the $X \rightarrow M$ regressions in Studies 2 and 3 are logistic so the α parameters are measured in logits (the log of the odds ratio) and the $M \rightarrow Y$ regressions in Studies 2 and 3, and the $M_2 \rightarrow M_3$ regression in Study 2, are measures without a linking function in terms of unit change in outcomes as a function of unit change in predictors. Mediation models where all regression are ordinary afford conclusions such as, "the indirect effect meant that English speakers discounted ____% more than Dutch speakers as function of using more low-certainty modal verbs". This is not possible in models like ours which measure outcomes on different scales. This is why we avoid interpreting the extent of indirect effects in the main text.

There are various solutions to this complicated modelling issue, for instance using causally defined effects (Muthén, 2011), or latent variables (Johnson, 2021). A simpler solution is to standardise coefficients before computing path products. This results in changing the measurement scale, which means it would have been necessary to plot direct and total effects separately from indirect effects in the main text. This seemed cumbersome, so we here report our hypothesis testing but with standardised path estimated. Estimates were standardised in the following way (from Johnson, 2021). All α coefficients which capture $X \rightarrow M$ paths are standardised by:

$$\alpha \frac{SD_x}{SD_m} \tag{E.6}$$

where SD_x is the observed standard deviation of X and SD_m is the standard deviation of the binary outcome when analysed with a logit link:

$$SD_m = \sqrt{\alpha^2 + SD_x^2 + \frac{\pi^2}{3}} \tag{E.7}$$

and all β coefficients were standardised in a similar fashion:

$$\beta \frac{SD_m}{SD_y} \tag{E.8}$$

However, in this case SD_m is simply the observed standard deviation of the binary mediator M. Finally in Study 2, the $M_2 \rightarrow M_3$ pathway (i.e. *future tense* $\rightarrow$ *subjective temporal distance*) is normalised following equation E.8, by:

$$\delta \frac{SD_{m2}}{SD_{m3}} \tag{E.9}$$

Table E.2

Parameter estimates for Study 2 results with discounting as the mediator

outcome	predictor	estimate	SE	low-95%CI	high-95%CI
subjective value	Intercept	19.31	0.44	18.46	20.19
	English	-1.16	0.66	-2.50	0.16
	subjective temporal distance	-0.18	0.08	-0.33	-0.03
	odds against	-4.24	0.05	-4.34	-4.14
	temporal distance	-1.50	0.07	-1.64	-1.37
	item order	0.03	0.05	-0.06	0.13
	odds against * temporal distance	0.60	0.05	0.50	0.70
low-certainty language	Intercept	-1.64	0.18	-2.01	-1.29
	English	1.69	0.17	1.36	2.02
	subjective value	-0.01	0.00	-0.01	0.00
	odds against	1.17	0.12	0.94	1.41
	temporal distance	0.17	0.12	-0.06	0.41
	item order	0.11	0.02	0.07	0.15
	odds against * temporal distance	-0.05	0.12	-0.28	0.17

After these normalisation procedures, we checked all path products reported in the main text to ensure than significance was not affected. In no cases was it affected, which indicates the direction and significance of the path products can still be interpreted, which we do in the main text.

E.4 Testing the Study 2 results with discounting scores as the mediator

It is important when conducting mediation analyses to test theoretically-plausible alternative specifications (Hayes, 2013). Study 2 involved the analysis of naturalistic data; it is therefore plausible that English speakers might have discounted more and therefore used more low-certainty language, rather than the other way around. As such, we specified a simple mediation model to test the alternative account that *English* $\rightarrow$ *discounting* $\rightarrow$ *low-certainty language*. We therefore specified the following model:

$$\begin{aligned} subj.val._{ij} &= \lambda_1 + \alpha lang_i + \kappa_{1q} X_{1qi} + u_{1j} + e_i \\ log_e\left(\frac{\pi_{LC_{ijk}}}{1-\pi_{LC_{ijk}}}\right) &= \lambda_2 + \tau' lang_i + \beta subj.val._i + \kappa_{2q} X_{2qi} + u_{2j} + h_k \end{aligned} \tag{E.10}$$

where π_{ijk} is the probability that $y = 1$ for low-certainty language (LC); α is the parameter for the effect of language on discounting, τ' is the parameter for the effect of language on the probability of using low-certainty language, and β is the parameter for the effect of discounting on the probability of low-certainty language use. Intercepts, error, control vectors, and random terms are analogous to the Study 2 model in the main text. Parameter estimates are reported in *Table E.2*.

The critical question question was whether increased discounting in English resulted in increased low-certainty language use. To test this, we tested the one-tailed hypothesis that $\alpha\beta > 0$. Evidence for this was weak, $Est. = 0.01,\ CI_{90\%} = [0, 0.02],\ pp = .909$. This analysis indicates that English speakers discounted more as a function of using low-certainty language, not the other way around.

Appendix F

Supplementary materials to Chapter 5

F.1 Description of Twitter data on which the time-reference classifier was trained

A sample of $n = 3000$ future time reference (FTR) annotations for the time-reference classifier were initially drawn from a dataset of tweets. We then added $n = 1002$ examples from the Reddit data which forms the basis of Study 1. We describe the Twitter data here. Data were downloaded from Twitter using the open-source Python package *Twint* (Poldi et al., 2019) in June 2019, as part of research involving the prediction of football match outcomes from Twitter data (in prep.). Our starting point was the publicly-available "Football matches odds" dataset (Silvas, 2018), which can be downloaded from Kaggle. The dataset contains information on football match outcomes between December 2016 and May 2018 for approximately 32,000 matches, 500 leagues, and 4,500 teams. Data includes home and away team names and the date of each match. Boolean operators are permitted in *Twint*, so we created search terms by concatenating home and away team names with the Boolean operator AND, i.e. [NAME_{home}] AND [NAME_{away}]. Dates were constrained to the two-week period immediately preceding each match date, and we downloaded only English tweets. This allowed us to download FTR tweets made in advance of each match. This resulted in a dataset of $N = 141,428$ tweets made about upcoming football matches. The earliest tweet in this data was made on 10 December 2016 and the latest was made on 22 May 2016. These data were used as the source for the initial $n = 3000$ FTR annotations on which the time-reference classifier was trained.

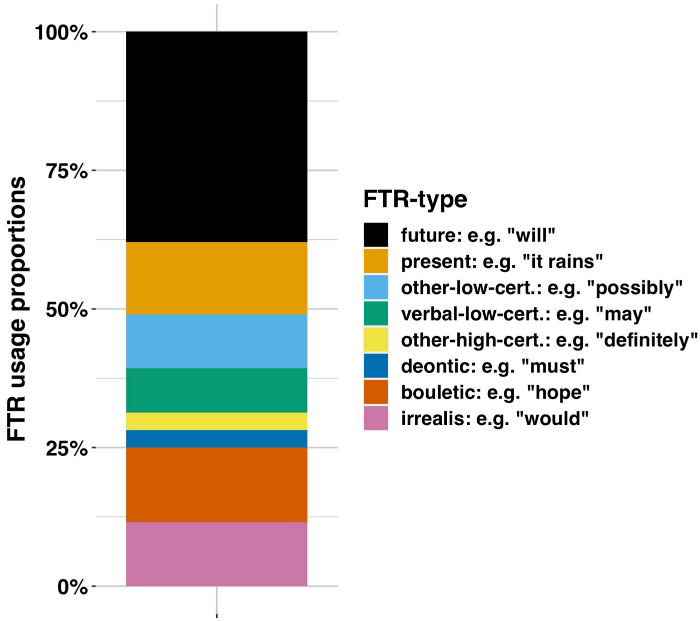


Figure F.1: Differences in probability of using different FTR types when referring to the future in Study 1. Above, any expression using *will, be going to*, *shall*, etc. is classed as future tense, regardless what other modal words are used. Data are the same as in *Fig. 5.3* in the main text.

F.2 FTR-type proportions without dominated future tense

The FTR-type classifier does not, by default, count expressions which use the future tense in combination with modal modifiers (e.g. *possibly*) or mental state predicated (e.g. *think*) as future tense. For example, *I think it will rain tomorrow* and *It will possibly rain tomorrow* would be classified as other-low-certainty and not also future tense, even though both use *will*. We comment in the main text that FTR-type usage proportions indicate English FTR is characterised by a mixture of FTR types, rather than a preponderance of future tense constructions. It was therefore appropriate to check whether this was an artefact of the modal dominance scheme. As such, we calculated usage proportions using non-dominated future tense, which counts any use of *will, shall,* or *be going to* as future tense, regardless of what other keywords are additionally present, *Fig. F.1.* Future tense use was higher than the dominated proportions reported in the main text, *Fig. 5.3*, but the overall pattern was similar. This suggests English FTR is characterised by a mixture of different FTR types, as we conclude in the main text.

Table F.1

DASS-21 items

sub-scale	question
depression	I couldn't seem to experience any positive feeling at all
	I found it difficult to work up the initiative to do things
	I felt that I had nothing to look forward to
	I felt down-hearted and blue
	I was unable to become enthusiastic about anything
	I felt I wasn't worth much as a person
	I felt that life was meaningless
anxiety	I was aware of dryness of my mouth
	I experienced breathing difficulty (e.g. excessively rapid breathing, breathlessness in the absence of physical exertion)
	I experienced trembling (e.g. in the hands)
	I was worried about situations in which I might panic and make a fool of myself
	I felt I was close to panic
	I was aware of the action of my heart in the absence of physical exertion (e.g. sense of heart rate increase, heart missing a beat)
	I felt scared without any good reason
stress	I found it hard to wind down
	I tended to over-react to situations
	I felt that I was using a lot of nervous energy
	I found myself getting agitated
	I found it difficult to relax
	I was intolerant of anything that kept me from getting on with what I was doing
	I felt that I was rather touchy

F.3 DASS-21 items

Items for the full DASS-21 are given in *Table F.1.* These are from Lovibond and Lovibond (1995). Participants rated these using a Likert scale between 'did not apply to me at all" (0), "applied to me to some degree, or some of the time" (1), "applied to me to a considerable degree, or a good part of time" (2), and "applied to me very much or most of the time" (3). Depression and anxiety scores were 2x the sum of scores within each sub-scale.

Appendix G

Study X: A flawed but useful replication

In this appendix, I report the results from a study which I conducted, but which is not reported in any of Chapters 2–5.[1] It is not reported because of flaws in the experimental design. I report it here for transparency, and because—despite flaws—it may have value. It originally was presented as Study 2 in Chapter 4. I will refer to it here as Study X. As in Study 2 of Chapter 4, the principal goal of Study X was to conduct a contrasting hypothesis test between the modal and temporal hypotheses. We measured low-certainty language use using the FTR-elicitation task (Version 3, see Appendix I); subjective temporal distance using a time-slider task; and risky intertemporal preferences by having participants make repeated choices between delayed, risky rewards and immediate non-risky rewards. We tested English and Dutch speakers, and used mediation analyses to test how speaking English impacted risky time preferences via these mediating variables. We hypothesised that English speakers would use more low-certainty language, and that this would in-turn result in less future orientation. Conversely, we hypothesised that English speakers would use more future tense constructions, but that this would NOT result in more distal temporal representations and therefore NOT impact future orientation.

G.1 Methods

G.1.1 Participants

Four participants were excluded because of missing demographic data. After this, the final study sample was $N = 648$ participants ($n = 330$ in English [$n = 165$ females, $n = 162$ males, $n = 3$ other], $n = 318$ in Dutch [$n = 158$ females, $n = 159$ males, $n = 1$ other]). Data were collected between September and November 2019. English participants were recruited from Prolific Academic (British English speakers), and Dutch participants were recruited from

[1] Rather, the risky intertemporal-choice data is not reported. Analysis of the FTR-elicitation and subjective-temporal-distance task data is given in Study 2 of Chapter 2.

Qualtrics. All participants completed the study online. Ethical approval for the study was granted by the Oxford internal review board, ref. no. R39324/RE001. All participants were remunerated.

G.1.2 Materials

Study X comprised three tasks: an FTR-elicitation task designed to establish free-text FTR language use; a risky-intertemporal-choice task designed to measure psychological discounting when rewards are both delayed *and* risky; and a subjective-temporal-distance task designed to establish how subjectively proximal or distal participants construed future outcomes.

The FTR-elicitation task: Participants were given a context and a target sentence and were tasked with typing in the conjugated target sentence. Before starting, participants were advised that there "were no correct answers," and that they should complete the questionnaire sentences, "as though they were speaking to a close friend." They were given two training items with example responses, and one trial item where they typed in a response. These were in the past tense in order to avoid biasing participants. There was one attention check: At a random point, participants were instructed to enter the word "dance" (Dutch "dans"). If they failed to do this, there were ejected from the survey immediately.

There were three within-subjects factors in the task: FTR-mode (predictions, intentions, scheduling); modality condition (high-certainty, low-certainty, neutral); and temporal distance (one month, two months, three months, six months, one year, five years). FTR-mode was operationalised by constructing contexts which matched these criteria. Temporal distance was operationalised using temporal adverbials in the contexts. Modality condition was operationalised by giving participants numerical "certainty information" above each target sentence, for example:

Context: Chris's brother {SEND} him some money next month. You never know with him... When he get it...

Certainty: 50% certain.

Target: ...he {SPEND} it at the bar.

Prior to starting, participants were told "there will be some 'certainty information' included in the context". They were informed that, "this indicates how certain you are about what you are saying". They were then directed to, "please imagine you are this certain and write down what you would say". For schedules and predictions, participants were told they were supposed to be "_____% certain", and for intentions they were told they were supposed to be "_____% decided" (this was because it was difficult to make "certain" agree with all intention contexts).

In the low-certainty condition, certainty information varied between 40%, 50%, and 60%. This was implemented to try to maintain participant engagement. In the high-certainty condition, certainty information was invariably 100%. In the neutral condition no certainty information was given. We changed our definition of FTR-mode slightly, and counted as intention any intention statement whether it was first or third person. This was to try to isolate language usage in more prototypical prediction contexts. See Chapter 2, *Table 2.2* for example items from a single frame distance—one month—with one example from each modality condition and FTR-mode.[2] See Appendix I for questionnaire items.

The FTR-elicitation task had balanced factors. In each temporal-distance by modality condition there were 5 items: 3 prediction items, 1 intention item, and 1 scheduling item. This means there were 15 items per temporal distance, $3_{prediction} \times 3_{prob.cond.} + 1_{intention} \times 3_{prob.cond.} + 1_{scheduling} \times 3_{prob.cond.} = 15$, and 90 items in total, $5_{temp.dist} \times 5_{FTR-mode} \times 3_{prob.}$. In order to address the possibility that idiosyncratic aspects of items were driving language usage, we systematically varied certainty information across highly similar items (unavoidable differences between FTR-modes prevented base items from being altered across FTR-mode). Semantic details (nouns, names, pronouns) were altered, but other details were only changed to ensure the certainty information did not clash with the certainty implied by the context of the item. Because of time constraints, each participant answered 60 randomly selected items. Item order was randomised and one item was displayed per page.

Data were classified using the FTR-type classifier. The classification categories were: (1) present tense, for responses in the present tense with no modal language, e.g. *Tomorrow it rains*; (2) future tense, for responses in the future tense with no modal language, e.g. *zullen* 'will'; (3) verbal-low-certainty, for responses with low-certainty semantics using a modal verb, e.g. *kunnen* 'may'; (4) other-low-certainty, for responses with low-certainty semantics using other construction types, e.g. *mogelijk* 'possibly', *denken* 'think'; (5) verbal-high-certainty, for responses with high-certainty semantics using a modal verb, e.g. *moeten* 'must'; and (6) other-high-certainty, for responses with high-certainty semantics using other constructions types, e.g. *zecker* 'certainly'. Responses in the present tense which express modal semantics (e.g. *Tomorrow it could rain*) are classed according to modal notions, and not also present tense. Similarly, responses which are in the future tense but express modal notions (e.g. *Tomorrow it will possibly rain*) are classed according to modal notions and not also the future tense. This is because *It will possibly rain* and *It will rain* express different semantics and this was reflected in the classification scheme. Because the FTR-type classifier uses keyword methods it cannot accurately classify negations, or when conflicting modal polarities are encoded, e.g. *I am certain there's a possibility it will rain.* Responses exhibiting these features were removed ($n = 471$ of $N = 38,638$ total responses; missing data resulted in N being less than the expected $N = 648_{part} \times 60_{items} = 38,880$). To test whether the FTR-type classifier was

[2]Study 2 in Chapter 2 presents (just) the FTR-elicitation task data from Study X.

accurate, we had linguistically-trained human annotators classify a subset of the data ($n = 501$ responses in Dutch and $n = 504$ responses in English). We then tested the accuracy of the FTR-type classifier against the human annotations. We found that all accuracy metrics were $> .98$.

The risky-intertemporal-choice task: In this task, participants made a series of binary choices between a smaller-sooner and a larger-later reward. For example, “Would you rather have £17.50 now or a 60% chance of £35 in two months?” (Dutch participants were given amounts in euros). The factors of this task were: (1) The amounts of the SSR (20.65, 24.15, 27.65, and 31.15); (2) the delays of the LLR (now, one month, two months, three months, six months, nine months, and one year); and (3) the probabilities of the LLR (40%-100% by increments of 10%). These factors were crossed. The $n = 4$ items where delay was now and probability was 100% were removed because subjective values for identical items would be the same. This resulted in a test battery of $7_{delays} \times 7_{probabilities} \times 4_{amounts} - 4 = 192_{items}$. The amount of the LLR was always £35. The choice to wait for the larger-later reward was scored (1), otherwise (0). Participants completed a random sample of $n = 128$ items. Prior to starting the task, participants were told to “try to answer quickly and honestly, without thinking about it too much” and completed a training example with an LLR of £10,000. Item order was randomised and one item was displayed per page.

The subjective-temporal-distance task: In this task, participants used a slider to rate whether they construed a given objective temporal distance as “very close to now” (0) or “very far from now” (10). To increase the sense of subjectivity, numbered slider intervals were not displayed to participants. The within-subjects factor of the subjective-temporal-distance task was the objective distances. These were the same as in the risky-intertemporal-choice task. For each item, participants were directed to “indicate with the slider how far away from NOW the given time feels to you.” Prior to starting the task, participants were given a training example involving a very far distant future (40 years), and were told to “try to answer quickly and honestly, without thinking about it too much.” Item order was randomised, and one item was displayed per page.

G.1.3 Procedure

The study was hosted on the Qualtrics survey platform and was conducted online. It had a mixed-design. The within-subjects factors for each task are described above. There was one between-subjects factor: survey language. There were two (otherwise identical) versions of the survey: one in English, the other in Dutch. At the beginning of the surveys, participants confirmed their first language and current residence. English speakers confirmed they were native English speakers currently residing in the United Kingdom and Dutch speakers confirmed they were native Dutch speakers currently residing in the Netherlands. If they did not, they were immediately ejected from the survey. Following this, participants completed the FTR-elicitation task, then the risky-intertemporal-choice task, then the subjective-temporal-distance task.

G.2 Results

While English speakers used more future tense and fewer present tense constructions across all factors of the FTR task, they also tended to use more low-certainty language (i.e. other-low-certainty + verbal-low-certainty), and this was mostly driven by grammatically obligatory low-certainty modal verbs, *Fig. G.1a*. Additionally, English speakers' use of low-certainty language as well as the future tense increased for intentions compared with schedules, and predictions compared with intentions, *Fig. G.1c*. A similar but not as pronounced pattern can be seen in Dutch. Together, *Figs. G.1a* and *G.1c* suggest that increased grammaticisation of FTR in English is characterised by increasingly obligatory linguistic encoding of modal notions under low-certainty conditions, *Fig. G.1a*, and sensitivity to the inherent differences in implied certainty across FTR-mode (predictions are less certain than intentions, which are less certain than schedules (see Dahl, 1985, 2000a)), *Fig. G.1c*. Finally, FTR-usage was not sensitive to temporal distance, which suggests that temporal notions are not the primary semantic domain encoded by these FTR structures, *Fig. G.1b*. In so far as it may generalise, this suggests that FTR grammaticisation has more to do with the encoding of future probability than time, and may involve increased obligatorisation of both low-certainty modals and the future tense.

We had predicted that use of low-certainty language would negatively mediate a relationship between English and probabilistic intertemporal choices (PITCS). To test our prediction, we created a new FTR usage variable ("low-certainty") which was (1) when either verbal-low-certainty or other-low-certainty terms were used, otherwise (0). We then calculated the mean of PITCs and low-certainty per participant, temporal distance and probability level, and merged the data on shared levels between tasks (distance: two months, three months, six months, one year; probability: 40%, 50%, 60%, and 100%). This resulted in $N = 10,273$ observations. Since these variables are dichotomous (0,1), this is simply the proportion of low-certainty language used, and the proportion of future PITCs, for each participant, temporal distance, and probability level.

As in Study 1, we used a Bayesian multilevel approach with random intercepts clustered by participant, and no priors. We estimated a simple mediation model, with mean low-certainty language as the mediator, and included objective probability (odds against, $OA = p/(1-p)$), temporal distance (days), and their interaction as predictors of the mediator and outcome variables, *"Model A", Fig. G.2b*.

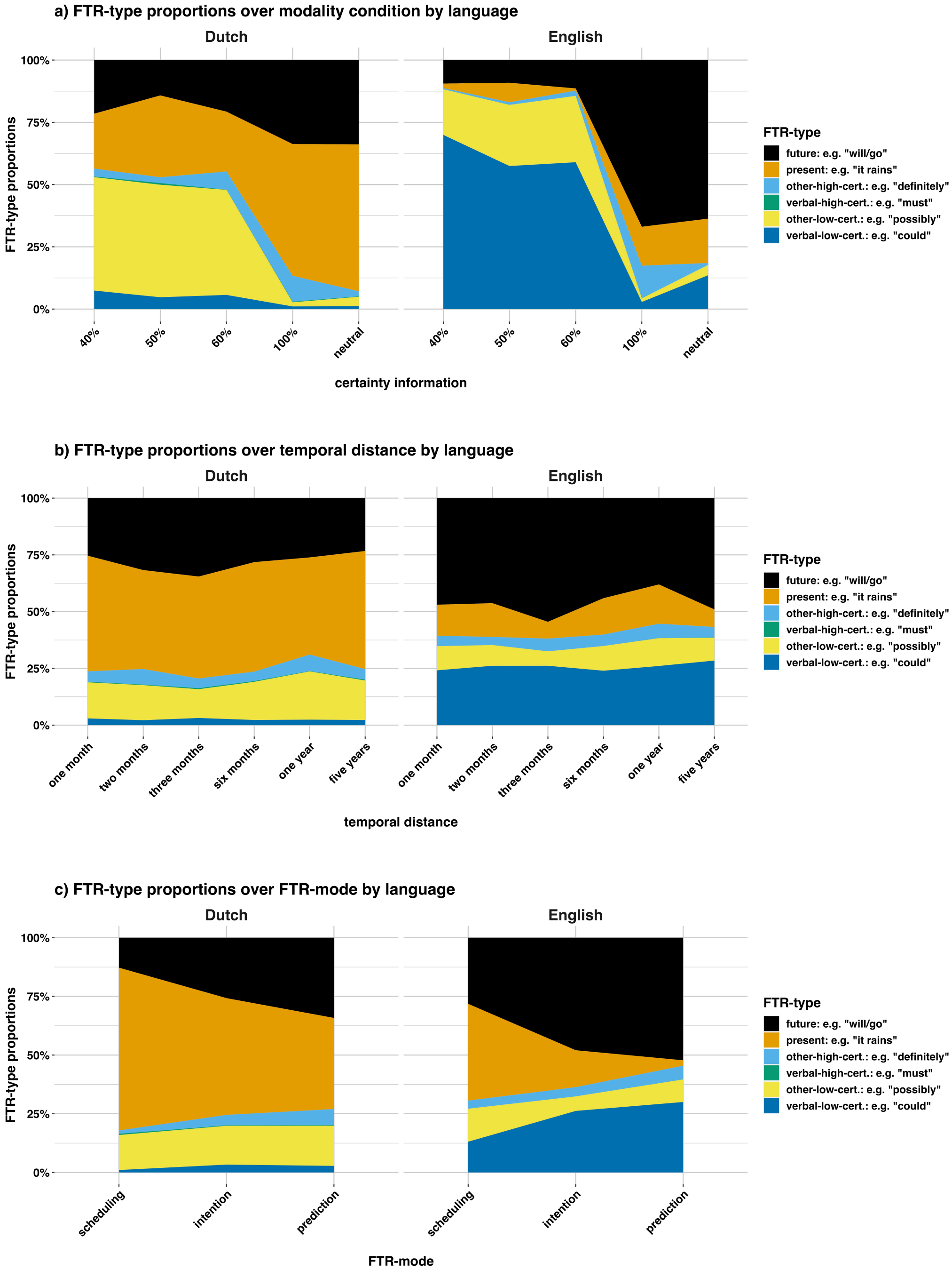


Figure G.1: Elicited FTR-type proportions over probability condition, temporal distance, FTR-mode in English and Dutch. For probability condition, 40% - 60% certainty represent the low-certainty condition, 100% is the "certain" condition, and "neutral" is the neutral condition, in which no certainty information was given. Reported values above are the proportions per language and FTR-task factor, i.e. means per level normalised to sum to 1 ($x = x_i / \sum_n x$).

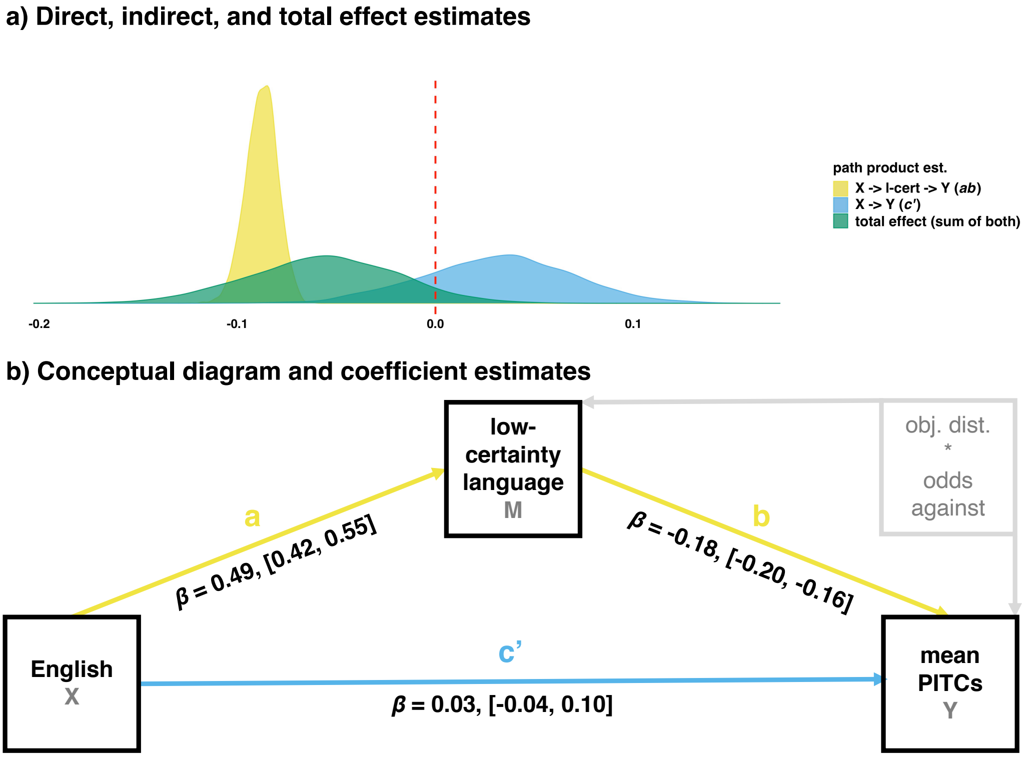


Figure G.2: a) Posterior estimates for paths and path products for the effect of speaking English on mean PITCs via use of low-certainty language (Model A). b) Conceptual diagram of model.

We found that English speakers used more low-certainty terms and that such usage was negatively associated with mean PITCs, *Fig. G.2a*. As predicted, the indirect effect was negative, $Est. = -0.09$, $CI_{90\%} = [-0.1, -0.08]$, $pp > .999$. This indicates that higher use of low-certainty language in English resulted in higher discounting and therefore lower mean PITCs. Did this result in more discounting overall? In exploratory analyses (i.e. not pre-registered), we found the posterior probability for the total effect being negative approached 95%, $Est. = -0.06$, $CI_{90\%} = [-0.12, 0]$, $pp = .938$. This indicates that it was about 15 times more likely to be negative than positive, but falls short of standard 95% cut-offs.

In further exploratory analyses, we carried on to test the predictions of the temporal hypothesis. Because we needed to use subjective-temporal-distance ratings as a mediating variable, and we only had one value per level of temporal distance, we calculated mean PITCs and future tense use per participant and temporal-distance level. This resulted in $N = 3010$ observations. We then estimated a path model which operationalised the predictions of the temporal hypothesis, with no priors and intercepts clustered by participant, and controlling for objective temporal distance (days), *"Model B", Fig. G.3b.*

However, we found no evidence that higher future tense use in English resulted in more distal construals of the future and therefore increased discounting, i.e. the probability that

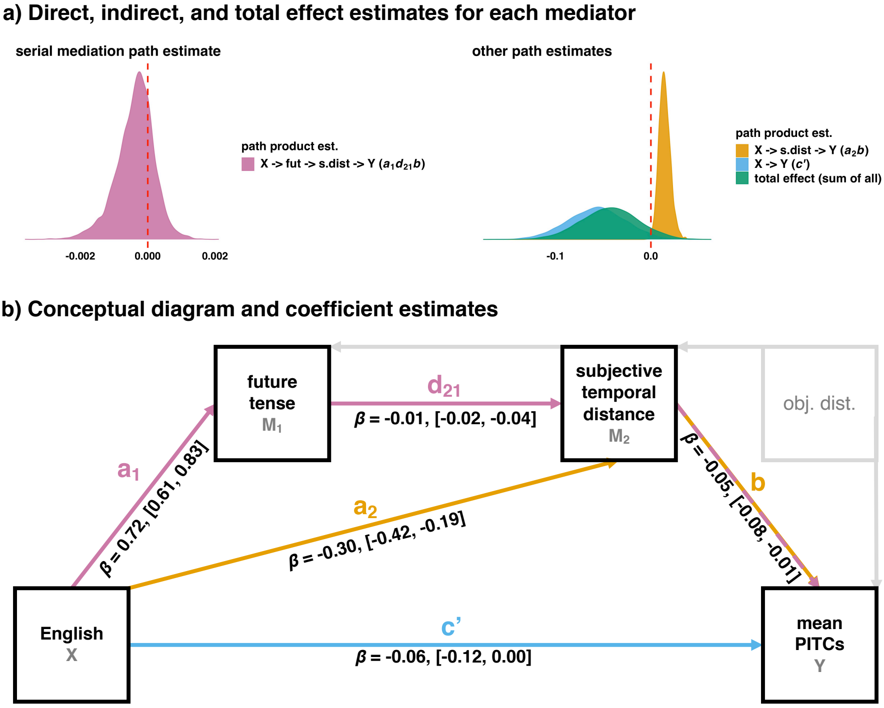


Figure G.3: a) Posterior estimates for paths and path products for the effect of speaking English on mean PITCs via use of the future tense and subjective ratings of temporal distance (Model B). The serial mediation path (left panel) is depicted separately for ease of visualisation (including it in one figure makes the other distributions difficult to discern). b) Conceptual diagram of model.

$a_1d_{21}b < 0$ was not different from zero, $Est. = 0$, $CI_{90\%} = [0, 0]$, $pp = .756$. English speakers rated the future as further away, and that therefore the $X_{lang} \rightarrow M_{subj.dist} \rightarrow Y_{PITCs}$ path (a_2b) was positive rather than negative, $Est. = 0.01$, $CI_{90\%} = [0.01, 0.02]$, $pp = .998$. These analyses provided no support for the predictions of the temporal hypothesis.

However, we had not yet explored how high-certainty terms, the present tense, or the future tense might be directly related to our discounting measures. The two hypotheses (modal, temporal) make different predictions about how different FTR structures should relate to discounting. These predictions depend on how the frameworks construe the semantics of common FTR constructions. The temporal hypothesis assumes the future and present tenses encode temporal distality and proximity. It therefore predicts that speaker-level use of the future tense should be associated with increased discounting, and speaker-level use of the present tense should be associated with decreased discounting. The modal hypothesis assumes that the present tense, the future tense, and high-certainty constructions encode high certainty, and low-certainty constructions encode low certainty. It therefore predicts that future, present, and high-certainty use should be associated with decreased discounting, and that low-certainty use should be associated with increased discounting.

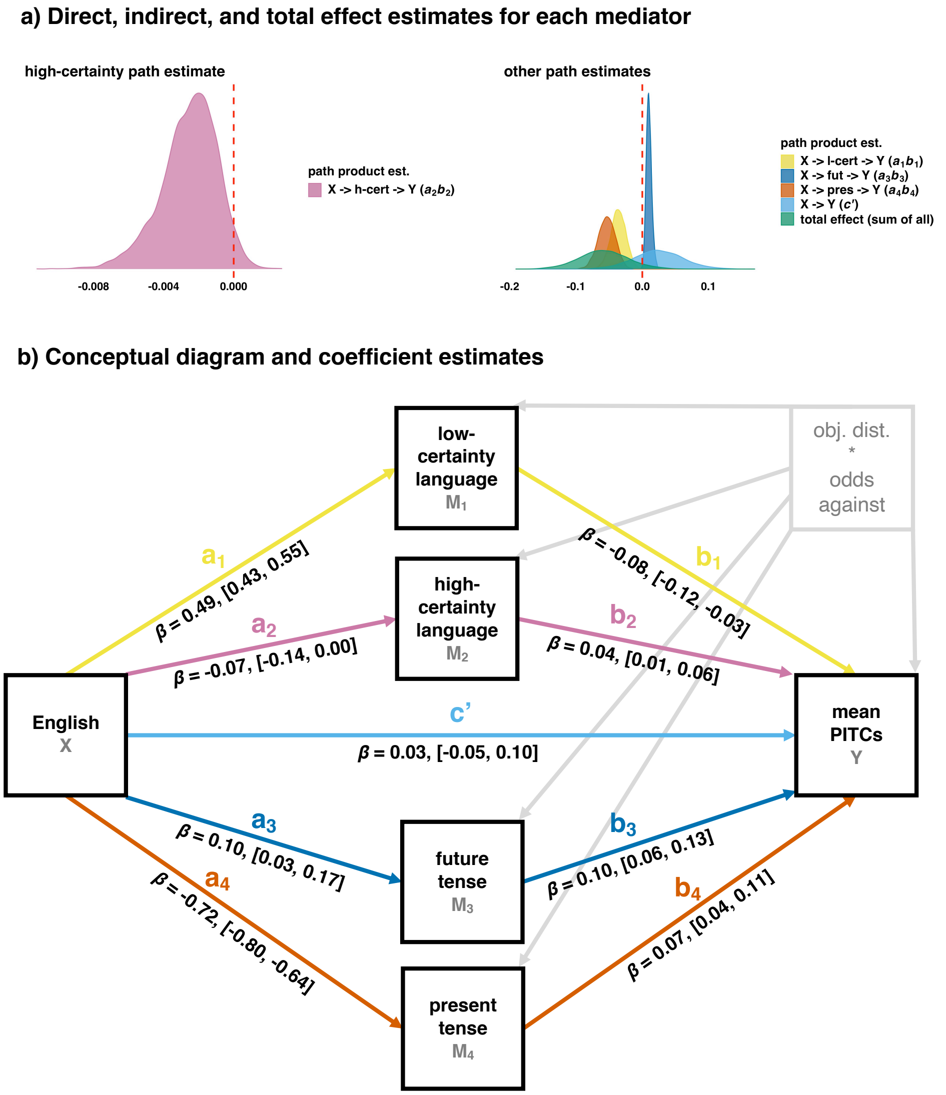


Figure G.4: a) Posterior estimates for paths and path products for the effect of speaking English on mean PITCs via use of low-certainty language, high-certainty language, the future tense, and the present tense (Model C). The high-certainty mediation path (left panel) is depicted separately for ease of visualisation (including it in one figure makes the other distributions difficult to discern). b) Conceptual diagram of model.

We therefore conducted exploratory analyses to test which of these alternate predictions was supported by our data. We used the probability by distance means (as in Model A) and specified a multiple mediator model with means for low-certainty, high-certainty, the future tense, and the present tense as mediators, *"Model C", Fig. G.4*.

Critically, we found that the $X_{lang} \rightarrow M_{fut} \rightarrow Y_{PITCs}$ path (a_3b_3) was positive as predicted by the modal (and not the temporal) hypothesis, $Est. = 0.01$, $CI_{90\%} = [0, 0.02]$,

$pp = .998$. This indicates speaker-level future tense use was positively associated with mean-PITCS, and therefore decreased discounting.

As before, the $X_{lang} \rightarrow M_{low-cert} \rightarrow Y_{PITCs}$ path (a_1b_1) was negative, $Est. = -0.04$, $CI_{90\%} = [-0.06, -0.02]$, $pp > .999$ (it is the same effect). The $X_{lang} \rightarrow M_{high-cert} \rightarrow Y_{PITCs}$ path (a_2b_2) was also negative, $Est. = 0$, $CI_{90\%} = [-0.01, 0]$, $pp = .963$ (see *Fig. G.4*, effect is small but does not contain zero). English speakers used fewer high-certainty terms and these were positively associated with PITCs. Finally, the $X_{lang} \rightarrow M_{pres} \rightarrow Y_{PITCs}$ path (a_4b_4) was negative, $Est. = -0.05$, $CI_{90\%} = [-0.08, -0.03]$, $pp > .999$. English speakers used fewer present tense constructions, and use was positively associated with mean PITCs. Taken together, the direction of these effects tends to support the conjecture that use of low-certainty terms is associated with increased discounting and use of high-certainty terms (including present and future tense FTR) is associated with decreased discounting.

G.3 Discussion

On the whole, we found support for the modal but not temporal hypothesis. However, there were several flaws which limit the conclusions which can be drawn from these results. Critically, we set fixed amounts for the SSR. This made calculating indifference points impossible, which meant we had to use the proportion of future-oriented choices as our dependent variable. The theoretical meaning of this measure is not well understood. Crucially, this also meant that the nominal value of the LLR, $LLR \times p$, was lower than the value of the SSR for approximately 1/2 of the data. For instance, if $SSR = 27.65$ and the probability given to receive the LLR was greater than $27.65/35 \times 100$ then $LLR \times p < SSR$. If participants were factoring the nominal value into their decisions, this meant that our task was confounded. In unreported analyses, we removed such data where $LLR \times p < SSR$. The direction of the $English \rightarrow low-cert. \rightarrow disc$ path was still negative but was no longer significant, $p = .1$.

Additionally, only some probabilities are represented in both discounting and FTR usage data (40%, 50%, 60%, and 100%). This meant that the path analyses were done on a subset of the data, which is not ideal. Finally, we only had 1 observation in the subjective-temporal-distance task per temporal-distance level which meant analyses involving these data (i.e. Model B) had to be conducted separately, on averaged data. This prevented us from testing both hypotheses using a single path analysis, which would have been preferable.

For these reasons, we decided to re-do this study with a design which addressed these confounds. This new study is now reported as Study 2 in Chapter 4.

Appendix H

FTR-type classifier word lists

In this appendix, I present the class criteria keyword word lists for the main categories of the FTR-type classifier (English: *Table H.1*; Dutch: *Table H.2*; and German: *Table H.3*). To keep tables concise, present tense keywords are not displayed. These are the present tense forms for all the verbs in the FTR-elicitation question so comprise a long list. Interested readers should consult the Github repository. Lists include common misspellings. There are additionally categories which are implemented in the FTR-type classifier but not—in the main—analysed in this dissertation and so not presented here. For instance, the past tense, the past perfect, bouletic modality, separate *gaan* 'be going to' and *zullen* 'will' futures, and the subjunctive. The present progressive is not treated differently to the present. While I tried to disambiguate these, keyword methods made this very difficult; *-ing* forms have too many other uses in English. Statistical natural language processing techniques might be needed to make progress. To view the source code, consult the FTR-type classifier github repository at `https://github.com/cbjrobertson/ftr_classifier`.

Table H.1

FTR-type classifier keyword lists (English)

future	verbal-low-cert	verbal-high-cert	other-low-cert	other-high-cert
is going	can	must	is a chance	for sure
are going	may		fairly certain	no chance
am going	could		rates are shakey	100%
going to	coud		likely to	guaranteed to
gonna	might		toss up	definitely
gona	should		predicted to	definetly
gonn	sould		it is unclear if	definately
theyll	ought		has to opportunity	definite
they‘ll	would		apparently	certainly
i’ll	wouls		dubiously	certain
you’ll			improbably	certainty
he’ll			likely	assuredly
she’ll			maybe	assured
we’ll			maby	positive
will			mayhap	clearly
wil			chance	doubtless
twill			chanche	indubitably
willbe			50%	indubitable
gonna			50/50	inevitably
wont			risk	infallibly
won’t			perchance	irrefutably
shall			perhaps	irrefutable
’ll			possibility	necessarily
ll			possible	necessary
youll			possibly	obviously
shell			potentially	obvious
ill			presumably	surely

Continued on next page...

Table H.1 – *FTR-type classifier keyword lists (English) continued*

future	verbal-low-cert	verbal-high-cert	other-low-cert	other-high-cert
about to			probable	sure
set to			probability	unavoidably
on the verge of			probably	unavoidable
			probaly	undeniably
			probarly	undeniable
			improbable	undoubtedly
			improbability	undoubted
			improbably	unquestionably
			questionably	unquestionable
			seemingly	absolutely
			somewhat	absolute
			supposedly	
			supposed	
			unsure	
			uncertainly	
			feel like	
			not too sure	
			i am not sure	
			not certain	
			not 100% certain	
			in my opinion	
			think	
			thinking	
			thinks	
			thought	
			believe	
			believes	
			believed	
			believing	

Continued on next page...

Table H.1 – *FTR-type classifier keyword lists (English) continued*

future	verbal-low-cert	verbal-high-cert	other-low-cert	other-high-cert
			reckon	
			reckons	
			reckoned	
			reckoning	
			expect	
			expects	
			expected	
			expecting	
			expectably	
			expectedly	
			aspect	
			doubt	
			doubts	
			doubted	
			doubting	
			suppose	
			supposes	
			supposed	
			supposing	
			guess	
			guesses	
			guessed	
			guessing	

Table H.2

FTR-type classifier keyword lists (Dutch)

future	verbal-low-cert	verbal-high-cert	other-low-cert	other-high-cert
ga	kunnen	moeten	in aanmerking komend	wel degelijk
gaat	gekund	moet	niet zeker	100%
gat	kan	moest	zich vast	twijfel er niet
gaar	kun	gemoeten	een maand of vijf	absoluut
gaan	kunt	moesten	50%	alleszins
zal	kon		de kans is	allicht
zullen	mag		niet duidelijk	bepaald
zult	mogen		is er een kans	beslist
zul	vermogen		lijkt erop	definitief
staat op	vermoogd		lijkt alsof	duidelijk
	zou		naar het schijnt	echt
	zouden		aannemelijk	eenvoudigweg
			allichet	essentieel
			bedenkelijk	evident
			blijkbaar	fietelijk
			denkelijk	gedwongen
			kans	gegarandeerd
			geschiktlijkend	glashelder
			hypothetisch	hoogstwaarschijnlijk
			misschien	helder
			msschien	inderdaad
			misscxhien	kennelijk
			mogelijk	klaarblijkelijk
			mogelijkerwijs	logisch
			ongeveer	natuurlijk
			onzeker	nodig

Continued on next page...

Table H.2 – *FTR-type classifier keyword lists (Dutch) continued*

future	verbal-low-cert	verbal-high-cert	other-low-cert	other-high-cert
			ogenschijnlijk	noodzakelijk
			klaarblijkelijk	normaal
			schijnbaar	ondenkbaar
			twijfelachtig	onmiskenbaar
			veelbelovend	onmogelijk
			vermoedelijk	onomstotelijk
			waarschijnlijk	ontwijfelbaar
			waarschijnljk	onvermijdelijk
			waarschynlijk	onwaarschijnlijk
			waarschijnlijkheid	onweerlegbaar
			wellicht	onwrikbaar
			volgens mij	overduidelijk
			volgens mijn	overtuigend
			houden voor	sowieso
			niet weten	stellig
			neit weet	uiteraard
			weet niet	vanzelfsprekend
			weten niet	vast
			denk	verplicht
			denken	voorgoed
			denkt	werkelijk
			geloven	wis
			gelooft	zeker
			geloof	ongetwijfeld
			meen	toch
			ment	
			menen	
			gokken	
			gok	

Continued on next page...

Table H.2 – *FTR-type classifier keyword lists (Dutch) continued*

future	verbal-low-cert	verbal-high-cert	other-low-cert	other-high-cert
			gokt	
			veronderstellen	
			veronderstel	
			veronderstelt	
			vermoed	
			vermoedt	
			vermoeden	
			gehoord	
			horen	
			hoor	
			hoort	
			betwijfel	
			betwijfelt	
			betwijfelen	
			annehm	
			annehmt	
			annehmen	
			verwacht	
			verwachten	
			verwachting	
			wel eens	
			wel	

Table H.3

FTR-type classifier keyword lists (German)

future	verbal-low-cert	verbal-high-cert	other-low-cert	other-high-cert
werde	dürfte	muss	unter umständen	auf jeden fall
werden	dürften	musst	annehmbar	klipp und klar
werdet	dürftest	muß	anscheinend	aufjedenfall
wird	dürftet	mußt	eventuell	augenscheinlich
wirst	kann	müssen	gegebenenfalls	bestimmt
wirdt	kannst	müsst	möglich	definitiv
	können	müßen	möglicherweise	deutlich
	könnt	müßt	offenbar	eindeutig
	könnte		scheinbar	gewiss
	könnten		vermutlich	jedenfalls
	könntest		vielleicht	klar
	könntet		wahrscheinlich	offensichtlich
	mag		womöglich	sicher
	magst		wahrscheinlichkeit	sicherlich
	mögen		nehme an	zweifellos
	mögt		nehmen an	zweifelsohne
	müsste		nehmt an	weiß
	müssten		nimmst an	weißt
	müsstest		nimmt an	wissen
	müsstet		denke	wisst
	müßte		denken	
	müßten		denkst	
	müßtest		denkt	
	müßtet		glaube	
	soll		glauben	
	sollen		glaubst	
	sollst		glaubt	
	sollt		meine	

Continued on next page...

Table H.3 – *FTR-type classifier keyword lists (German) continued*

future	verbal-low-cert	verbal-high-cert	other-low-cert	other-high-cert
	sollte		meinen	
	sollten		meinst	
	solltest		meint	
	solltet		rechne	
	würde		rechnen	
	würdest		rechnest	
	würden		rechnet	
	würdet		sage	
			sagen	
			sagst	
			sagt	
			vermute	
			vermuten	
			vermutest	
			vermutet	
			erwarte	
			erwartest	
			erwartet	
			erwarten	
			wohl	
			vielleicht	

Appendix I

FTR-elicitation tasks

This appendix presents various versions of the FTR-elicitation tasks which were used in this dissertation. The FTR-elicitation questionnaire was re-drafted three times. Some items were re-used, but many were rewritten, changed, discarded, and refined as various issue arose. I include the early drafts for transparency. The preferred version is the final draft (Version 3), *Table I.3*. Guides as to which items needed to be excluded from various analyses are provided on the FTR questionnaire github repository here: `https://github.com/cbjrobertson/ftr_questionnaire_master`, and in the supplementary materials for each chapter.

I.1 FTR-elicitation task guide

Below I provide a guide as to which versions of the task were used for which experiments. Following Dahl's (1985, 2000a) conventions, contexts are given in [brackets]. Version 1 is translated in English, Dutch and German. Versions 2 and 3 are translated into English and Dutch. Including all translations of all version would mean this appendix ran to more than 100 pages. As such, I only provide English translations. Readers should consult the Github repository for Dutch and German translations. Question numbers are provided to help map translated questions from one language to another, i.e. the thousands column indexes language. For instance, in *Table I.1*, question number 1001 is translated into Dutch as 2001, and German as 3001. This convention is used throughout (English: 1000; Dutch: 2000; German: 3000).

I.1.1 Version 1

The first version of the FTR-elicitation task is given in *Table I.1*. Items from this version are analysed in Study 1 of Chapter 2 and Studies 1 and 2 of Chapter 3.

Participants in Study 1 Version A of Chapter 2 completed the full 48 items, but because of survey uploading errors a matched sample had to be created, see Appendix C. In Study 1

Version B of Chapter 2 and Studies 1 and 2 of Chapter 3 participants completed a subset of items.[1] The question numbers which participants completed in these studies were: 1, 2, 3, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, 18, 19, 21, 25, 27, 28, 32, 33, 34, 35, 37, 39, 41, 45, 46, and 48.

There are many issues with this draft, which I have pointed out in the main text and appendices. Principally, factors were not balanced. Additionally, the lexical operationalisation of modality condition leaves something to be desired. For instance, in the low-certainty condition, information is included in the target in (parentheses) e.g. "(at least it says it's 50% likely)". A directive to not copy out text in parentheses was given above each item. I nonetheless thought it was appropriate to change this to a numerical operationalisation in later drafts, in order to avoid lexical priming. Another issue is that some questions made it impossible to tell whether participants actually were making future time reference, so these had to be excluded. Other items caused problems for the FTR-type classifier because class-criterion words were used in their non modal/temporal sense. It is perhaps surprising given these issues that many of the patterns I identified using this version of the questionnaire were later reproduced using Version 3. Nonetheless, this was the case.

If I could conduct this research again, I would have re-drafted earlier rather than using this flawed instrument to collect so much data. This dissertation was not commenced with the goal of exploring the impacts of modality on (risky) intertemporal decision making. Rather, our goal was to test Chen's (2013) hypothesis. It took much time and work to arrive at the modal framework I develop. The data led me there, not the other way around. My understanding of the issues of the FTR-elicitation task necessarily deepened as I read more widely and combed (by hand) through the massive elicited datasets.

I.1.2 Version 2

The second version of the FTR-elicitation task is given in *Table I.2*. This version was presented to Dutch participants in Study 3 of Chapter 3. All items in this version were presented to participants, though some had to be excluded (see Appendix D). Many of the same issues which affected Version 1 are still present in Version 2, which is what lead to the comprehensive re-draft in Version 3.

I.1.3 Version 3

The third version of the FTR-elicitation task is given in *Table I.3.* Items from this version are analysed in Study 2 of Chapter 2 and in Study X in Appendix G. The full $N = 174$ items were not presented to participants. Rather, participants completed all the items in the following temporal-distance levels: one month, two months, three months, six months,

[1] Studies 1 and 2 of Chapter 3 and Study 1 Version B of Chapter 2 present different aspects of the same experiments. The FTR-elicitation data are analysed in isolation in Chapter 2. In Chapter 3, discounting data are analysed in conjunction with FTR-elicitation data.

one year, five years. In Version 3, factors were balanced, and various issues which affected individual items were addressed as much as possible.

In order to address the possibility that idiosyncratic aspects of items were driving language usage, I systematically varied probability information across highly similar items (unavoidable differences between FTR-modes prevented base items from being altered across FTR-mode). Semantic details (nouns, names, pronouns) were altered, but other details were only changed to ensure the certainty information did not clash with the certainty implied by the context of the item. To address the possibility that linguistic certainty information was directly priming participants, I changed our operationalisation of modality condition. Rather than including verbal cues in parentheses, participants were given numerical "certainty information" above each target sentence. For schedules and predictions, participants were told they were supposed to be "_____% certain", and for intentions they were told they were supposed to be "_____% decided" (this was because it was difficult to make "certain" agree with all intention contexts). These certainty numbers are given in the "certainty number" column.

In the low-certainty condition, certainty information varied between 40%, 50%, and 60%; schedules and intentions were all 50%. Within each level of temporal distance, certainty information in the 3 prediction questions in the low-certainty condition was 40%, 50%, and 60%. This was implemented to try to maintain participant engagement. In the high-certainty condition, certainty information was invariably 100%. In the neutral condition no certainty information was given. I also changed our definition of FTR-modes slightly, and included as intention any intention statement whether it was first or third person. This was to try to isolate language usage in more prototypical prediction contexts. Finally, to allow us to model temporal distance as a ratio variable (i.e. number of days), I attempted as much as possible to regularise the temporal adverbials by which temporal distances were expressed.

The final version comprises $N = 174$ items translated into English and Dutch. There are three FTR-modes (prediction, intention, schedule). There are three modality conditions (low-certainty, high-certainty, neutral), and there are (today, tomorrow, one week, one month, two months, three months, six months, one year, two years, five years, ten years, and ongoing predictions). These are balanced with FTR-mode and modality condition.

In Studies 2 and 3 of Chapter, 4, I added modality conditions by imputing a wider range of certainty numbers into the certain and neutral condition, i.e. 70%, 80%, 90%, and 100%. This does not constitute a new version, as the items themselves are an unchanged subset of Version 3. The only difference is the certainty information. Items are presented as Version 3.1 in *Table I.4*. All items are predictions, and modality condition is eschewed in favour of certainty numbers.

I.2 Tables

Table I.1

FTR-elicitation questionnaire Version 1 (English)

question	question number	FTR-mode	temporal distance	modality condition
[Q: Do you think your dad will go to sleep?] A: Yes, he {BE} tired.	1001	prediction	today	neutral
[Q: What your brother {DO} if you don't go to see him today, do you think?] A: He {CALL} me (on the phone).	1002	prediction	today	neutral
[It's no use trying to swim in the lake tomorrow...] ...The water {BE} cold (then).	1003	prediction	tomorrow	neutral
The sun {RISE} at six O'clock tomorrow.	1004	scheduling	tomorrow	neutral
[The boy's father sent him a sum of money some days ago and it will arrive next week.] (The boy {GET} the money next week. When {GET} it...) ...he {BUY} a present for the girl.	1005	intention	one week	neutral
[My brother {SAY} yesterday...] ...that he {COME} here next week.	1006	intention	one week	neutral
[Don't invest in derivatives. The market is fraudulent...] ...It {CRASH} within a month.	1007	prediction	one month	neutral
[Father to son: If you {PUT} your allowance in a savings account...] ...next month it {BE} worth more.	1008	prediction	one month	neutral

Continued on next page...

Table I.1 – *FTR-elicitation questionnaire Version 1 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition
[Talking about a third person's plan for the summer (it is presently December):] (In the summer) he {TRAVEL} to Morocco.	1009	intention	six months	neutral
[Talking about the summer's upcoming weather (it is presently December):] (There {BE} a warm current coming from the tropics this year...) ...It {MAKE} this summer hot and rainy.	1010	prediction	six months	neutral
[If I keep eating pizza every day...] ...I {BE} fat by the end of the year.	1011	prediction	one year	neutral
[Q: You {STAY} here, next year?] A: No, I {LIVE} in San Francisco (next year).	1012	intention	one year	neutral
[Don't bother investing in real estate...] ...the housing market {CRASH} in the next 2 years.	1013	prediction	two years	neutral
[Don't buy those shoes...] ...they {WEAR OUT} in a couple of years.	1014	prediction	two years	neutral
[Q: Are you {SELL} both your houses? A: We {WAIT}. Maybe in another ten years. Real estate always goes up in value...] ...By then, they {BE} much more valuable.	1015	prediction	ten years	neutral

Continued on next page...

Table I.1 – *FTR-elicitation questionnaire Version 1 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition
[Q: What {BE} your ten year projection for Coca-Cola stock? A: I {THINK} they have good fundamentals. I {THINK} in ten years...] ...they {GAIN} 10% or even 20%.	1016	prediction	ten years	neutral
[Q: Do you {WANT} a cigarette? A: No, I don't smoke...] ...If you smoke you {GET} cancer (when you are old).	1017	prediction	twenty five plus	neutral
[Q: You {BE} saving for retirement? A: Yes, I {PUT} away £100 every month. By the time I {RETIRE}...] ...it {BE} worth £50,000, with interest.	1018	prediction	twenty five plus	neutral
[You should {PUT} money into a retirement savings plan; when you {BE} older...] ...you {THANK} yourself.	1019	prediction	twenty five plus	neutral
[If you {PUT} a stone into this bag...] ...the bag {BREAK}.	1020	prediction	ongoing prediction	neutral
[A: I have a headache. B: Take this medicine...] ...It {MAKE} you feel better.	1021	prediction	ongoing prediction	neutral
[Q: What {HAPPEN} if you eat this mushroom?] A: You {DIE}.	1022	prediction	ongoing prediction	neutral
[A sexual health instructor speaking to class: If you {HAVE} sex without protection...] ...you {HAVE} a baby.	1023	prediction	ongoing prediction	neutral

Continued on next page...

Table I.1 – *FTR-elicitation questionnaire Version 1 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition
[Q: What you DO this evening? A: I'm not sure...] ...I {GO} out to dinner with a friend (but I'm waiting for her to call).	1024	intention	today	low certainty
[Talking about a third person's plans for the evening: I {BE} not sure...] ...He {WORK} in the garden. (He usually {DO} on Sunday).	1025	prediction	today	low certainty
[Q: What you {DO} tomorrow evening? A: It depends how I feel. There {BE} a film playing that I {WANT} see at 7...] ...I {GO} with Mary.	1026	intention	tomorrow	low certainty
[Q: Are you sure that you want to go? A: I've decided. I {LEAVE} tomorrow...] ...In two days, I {BE} in Australia (whether you {BE} with me or not).	1027	intention	tomorrow	neutral
[Q: When does your mother arrive? A: Ug. In a week. She {INSIST} on {BUY} the kids too much sugar, like always...] ...I {SAY} no this time (definitely).	1028	intention	one week	neutral
[Q: What is the weather forecast for next week?] A: It {RAIN} (at least it says it's 50% likely).	1029	prediction	one week	low certainty
[Q: How do expect the markets {BEHAVE} next month? A: I {THINK}...] ...they {RISE} (but it is uncertain).	1030	prediction	one month	low certainty

Continued on next page...

Table I.1 – *FTR-elicitation questionnaire Version 1 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition
[Q: How do expect the markets {BEHAVE} next month? A: Next month, I {EXPECT}...] ...they {RISE} (but I am not sure).	1031	prediction	one month	low certainty
[Q: How do expect the markets {BEHAVE} in the next 6 months?] A: They {RISE} (I am certain of it!).	1032	prediction	six months	neutral
[Don't bother investing in real estate...] ...next year, the housing market {CRASH} (it is a possibility).	1033	prediction	one year	low certainty
[Don't bother investing in real estate...] ...the housing market {CRASH} (next year, potentially).	1034	prediction	one year	low certainty
[Don't bother investing in real estate...] ...(I'm very sure) the housing market {CRASH} (next year).	1035	prediction	one year	neutral
[Q: How do expect the markets {BEHAVE} next year?] A: They {RISE} (it is likely, in my opinion).	1036	prediction	one year	low certainty
[Don't bother investing in the tech industry...] ...Silicon Valley {CRASH} by 2018 (probably but not certainly).	1037	prediction	two years	low certainty
[Don't bother investing in the tech industry...] ...Silicon Valley {CRASH} by 2018 (it's possible).	1038	prediction	two years	low certainty
[Don't bother investing in the tech industry...] ...Silicon Valley {CRASH} by 2018 (definitely).	1039	prediction	two years	neutral

Continued on next page...

Table I.1 – *FTR-elicitation questionnaire Version 1 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition
[Don't bother investing in real estate...] ...the housing market {COLLAPSE} (in the next ten years, probably).	1040	prediction	ten years	low certainty
[Don't bother investing in real estate...] ...the housing market {COLLAPSE} (in the next ten years, it's a potential but not certain).	1041	prediction	ten years	low certainty
[Q: Do you {WANT} a cigarette? A: No, I Don't smoke...] ...If you smoke you {GET} cancer (when you are old, I'm fairly sure).	1042	prediction	twenty five plus	low certainty
[Q: Do you {WANT} a cigarette? A: No, I Don't smoke...] ...If you smoke you {GET} cancer (when you are old, definitely. I'm certain).	1043	prediction	twenty five plus	neutral
[Q: Do you {WANT} a cigarette? A: No, I Don't smoke...] ...If you smoke you {GET} cancer (when you are old, probably).	1044	prediction	twenty five plus	low certainty
[Don't bother investing in real estate...] ...the housing market {CRASH} (it's possible).	1045	prediction	indet.	low certainty
[Q: Do you {WANT} a cigarette? A: No, I Don't smoke. I {BE} afraid...] ...I {GET} cancer.	1046	prediction	indet.	neutral

Continued on next page...

Table I.1 – *FTR-elicitation questionnaire Version 1 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition
[The boy thinks he will perhaps get a sum of money: If the boy {GET} the money...] ...he {BUY} a present for the girl.	1047	intention	ongoing prediction	neutral
[Q: Do you {WANT} a cigarette? A: No, I Don't smoke...] ...If you smoke you {GET} cancer (it's possible).	1048	prediction	ongoing prediction	low certainty

Table I.2

FTR-elicitation questionnaire Version 2 (English)

question	question number	FTR-mode	temporal distance	modality condition
[Q: Do you think your dad will go to sleep?] A: Yes, by tonight, he {BE} tired.	1001	prediction	today	neutral
[Q: Do you think your mom will go to bed early tonight?] A: Yes, by tonight, she {BE} tired.	1002	prediction	today	neutral
[Q: What your brother {DO} if you don't go to see him today, do you think?] A: He {CALL} me (on the phone).	1003	prediction	today	neutral
[Q: What your sister {DO} if you don't go to see her later, do you think?] A: She {CALL} me (on the phone).	1004	prediction	today	neutral
[It's no use trying to swim in the pool tomorrow...] ...the water {BE} cold (then).	1005	prediction	tomorrow	neutral
[It's no use trying to swim in the pool tomorrow...] ...(then) the water {BE} cold.	1006	prediction	tomorrow	neutral
The sun {RISE} at six O'clock tomorrow.	1007	scheduling	tomorrow	neutral
The moon {RISE} at ten o'clock tonight.	1008	scheduling	today	neutral

Continued on next page...

Table I.2 – *FTR-elicitation questionnaire Version 2 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition
[The boy's father {SEND} him a sum of money some days ago and it {ARRIVE} next week.] (The boy {GET} the money next week. When he {GET} it...) ...he {BUY} a present for the girl.	1009	intention	one week	neutral
[The girl's uncle {SEND} her a sum of money and it {ARRIVE} in a couple of weeks.] (The girl {GET} the money soon. When she {GET} it...) ...she {BUY} a new pair of skis.	1010	intention	one month	neutral
[My brother {SAY} yesterday...] ...that he {COME} here next week.	1011	intention	one week	neutral
[My sister {SAY} this morning...] ...that she {DRIVE} down here at the end of next week.	1012	intention	one week	neutral
[Don't invest in derivatives. The market is fraudulent...] ...it {CRASH} within a month.	1013	prediction	one month	neutral
[Don't invest in commodities. The market is shakey...] ...it {CRASH} within a month.	1014	prediction	one month	neutral
[Father to son: If you {PUT} your allowance in a savings account...] ...next month it {BE} worth more.	1015	prediction	one month	neutral
[Mother to daughter: If you {PUT} the extra earnings from your weekend job in an RRSP...] ...it {BE} worth a lot more (when you retire).	1016	prediction	twenty five plus	neutral

Continued on next page...

Table I.2 – *FTR-elicitation questionnaire Version 2 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition
[Talking about a third person's plan for the summer (it is presently December):] (In the summer) he {TRAVEL} to Morocco.	1017	intention	six months	neutral
[Talking about a friend's plan for the autumn (it is presently the spring):] She {TRAVEL} to South America.	1018	intention	six months	neutral
[Talking about the summer's upcoming weather (it is presently December):] (There {BE} a warm current coming from the tropics this year.) It {MAKE} this summer hot and rainy.	1019	prediction	six months	neutral
[Talking about the winter's upcoming weather (it is presently June):] (There {BE} a cold current coming from the arctic this year.) There {BE} lots of snow in the mountains.	1020	prediction	six months	neutral
[If I don't start getting some exercise...] ...I {GAIN} a lot of weight (by the end of the year).	1021	prediction	one year	neutral
[If you keep eating take out all the time...] ...You {BE} fat by the end of the year.	1022	prediction	one year	neutral
[Q: You {LIVE} here, next year?] A: No, I {MOVE} to San Francisco (next year).	1023	intention	one year	neutral
[Q: You {LIVE} here, next year?] A: No, I {MOVE} back in with my parents (in the summer).	1024	intention	one year	neutral

Continued on next page...

Table I.2 – *FTR-elicitation questionnaire Version 2 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition
[Q: Do you {THINK} Johanson {PUT} chapter seven on the exam?] A: It {BE} on there, definitely.	1025	prediction	ongoing prediction	certain
[Q: Do you {THINK} Neidermeyer {PUT} econometrics on the final paper?] A: Chapter nine {BE} on it, (definitely, though maybe not ten).	1026	prediction	ongoing prediction	certain
[Don't {BUY} those shoes...] ...they {WEAR OUT} in a couple of years.	1027	prediction	two years	neutral
[Don't {BUY} that car...] ...the engine {WEAR OUT} in a couple of years.	1028	prediction	two years	neutral
[Q: Are you {SELL} both your houses? A: We {WAIT}. Maybe in another ten years. Real estate always goes up in value...] ...by then, they {BE} much more valuable.	1029	prediction	ten years	neutral
[Q: Are you {SELL} both your cars? A: Yes. Cars always {GO} down in value...] ...soon, they {BE} much less valuable.	1030	prediction	ongoing prediction	neutral
[Q: What {BE} your ten year projection for Coca-Cola stock? A: I {THINK} they have good fundamentals. In ten years...] ...they {GAIN} 10% or even 20%.	1031	prediction	ten years	neutral

Continued on next page...

Table I.2 – *FTR-elicitation questionnaire Version 2 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition
[Q: What {BE} your ten year projection for Apple stock? A: I {THINK} they are a great company. In ten years...] ...they {GAIN} 30% or even 40%.	1032	prediction	ten years	neutral
[Q: Do you {WANT} a cigarette? A: No, I don't smoke...] ...if you smoke you {GET} cancer (when you are old).	1033	prediction	twenty five plus	neutral
[Q: Do you {WANT} a drink? A: No, I don't {DRINK} during the week...] ...if you drink too much, you {BE} very unhealthy (when you're older).	1034	prediction	twenty five plus	neutral
[Q: You {BE} saving for retirement? A: Yes, I {PUT} away £200 every month. By the time I {RETIRE}...] ...it {BE} worth £50,000, with interest.	1035	prediction	twenty five plus	neutral
[Q: You {BE} saving every month? A: Yes, I {PUT} away £100 every paycheque. In 10 years] ...it {BE} worth £30,000, with interest.	1036	prediction	twenty five plus	neutral
[You should {PUT} money into a retirement savings plan; when you {BE} older...] ...you {THANK} yourself.	1037	prediction	twenty five plus	neutral
[You should {SAVE} your money now; when you {BE} older...] ...you {THANK} yourself.	1038	prediction	twenty five plus	neutral
[If you {PUT} that giant stone into this bag...] ...the bag {BREAK}.	1039	prediction	ongoing prediction	neutral

Continued on next page...

Table I.2 – *FTR-elicitation questionnaire Version 2 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition
[If you {PUT} too much stuff into your backpack...] ...it {BREAK}.	1040	prediction	ongoing prediction	neutral
[A: I have a headache. B: Take this medicine...] ...It {MAKE} you feel better.	1041	prediction	ongoing prediction	neutral
[A: I have a stomach ache. B: Drink this ginger tea...] ...It {MAKE} you feel better.	1042	prediction	ongoing prediction	neutral
[Q: What {HAPPEN} if you eat this mushroom?] A: You {DIE}.	1043	prediction	ongoing prediction	neutral
[Q: What {HAPPEN} if this spider bites you?] A: You {DIE}.	1044	prediction	ongoing prediction	neutral
[A sexual health instructor speaking to class: If you {HAVE} sex without protection...] ...you {HAVE} a baby eventually.	1045	prediction	ongoing prediction	neutral
[A mother to her daughter: If you don't use condoms...] ...eventually, you {HAVE} a baby.	1046	prediction	ongoing prediction	neutral
[Q: What you {DO} tonight?] A: I {DINE OUT} with a friend, (this evening).	1047	intention	today	neutral
[Q: What you {DO} later?] A: Tonight, I {GO OUT} with John. (He finally asked me!)	1048	intention	today	neutral
[Q: What is Sam doing later tonight?] A: He {WORK} in the garden.	1049	intention	today	neutral

Continued on next page...

Table I.2 – *FTR-elicitation questionnaire Version 2 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition
[Q: What is Sam doing later tonight?] A: He {WORK} late again (at his new job).	1050	intention	today	neutral
[Q: What you {DO} tomorrow evening? A: Avengers {PLAY} at 7pm...] ...I {SEE} it with Mary.	1051	intention	tomorrow	neutral
[Q: What you {DO} tomorrow? A: Imagine Dragons {PLAY} tomorrow...] ...I {SEE} them with Tom.	1052	intention	tomorrow	neutral
[Q: Are you sure that you want to go? A: I've decided. I {LEAVE} tomorrow...] ...in two days, I {BE} in Australia (whether you {BE} with me or not).	1053	intention	tomorrow	certain
[Q: Are you sure that you want to go? A: I've decided. I {LEAVE} next week...] ...in ten days, I {BE} in South America! (I'm so excited!)	1054	intention	one week	neutral
[Q: When does your mother arrive? A: Ug. In a week. She {INSIST} on {BUY} the kids too much sugar, like always...] ...I {SAY} no this time.	1055	intention	one week	neutral
[Q: When does your mother-in-law arrive? A: Tonight. She {BUY} the kids too many toys, like always...] ...I {SAY} no this time.	1056	intention	one week	neutral
[Q: What is the weather forecast for next week?] A: They say it {RAIN}.	1057	prediction	one week	neutral

Continued on next page...

Table I.2 – *FTR-elicitation questionnaire Version 2 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition
[Q: What is the weather forecast for tomorrow?] A: They say it {BE} very hot.	1058	prediction	tomorrow	neutral
[Q: How do you expect the markets {BEHAVE} next month? A: I {THINK}...] ...they {RISE}.	1059	prediction	one month	uncertain
[Q: How do you expect energy stocks {BEHAVE} next year? A: I {THINK}...] ...they {FALL}.	1060	prediction	one year	uncertain
[Q: What do you expect the sea levels {DO} in the future?] A: ...they {RISE}.	1061	prediction	indet.	neutral
[Q: What do you expect the sea levels {DO} in the next 50 years?] A: They {STAY} the same (I think).	1062	prediction	twenty five plus	uncertain
[Q: Do you think John and Celine still {BE} together in 6 months?] A: They {BREAK UP}. (I am certain of it!)	1063	prediction	six months	certain
[Q: Do you think Hamish and Natalie still {BE} together in a year?] A: They still {BE} together! (They are really happy).	1064	prediction	one year	neutral
[Don't bother investing in real estate...] ...next year, the housing market {CRASH}.	1065	prediction	one year	neutral
[Don't invest in commodities...] ...next year, domestic demand {FALL} (and prices {GO} down).	1066	prediction	one year	neutral

Continued on next page...

Table I.2 – *FTR-elicitation questionnaire Version 2 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition
[I wouldn't buy bitcoins...] ...the market {FALL} very soon.	1067	prediction	ongoing prediction	neutral
[I would buy some cryptocurrencies as soon as possible...] ...the market {EXPLODE} very soon.	1068	prediction	ongoing prediction	neutral
[You shouldn't put money into real estate...] ...(I'm very sure) the housing market {CRASH} (next year).	1069	prediction	one year	certain
[You should {PUT} money into gold...] ...(I'm very sure) the stock market {FALL} (soon).	1070	prediction	ongoing prediction	certain
[Q: How do expect the Dutch economy {BEHAVE} next year?] A: It {GROW}.	1071	prediction	one year	neutral
[Q: How do expect the European economy {BEHAVE} next year? A: I {BE} afraid...] ...it {FALL}. (I {BE} worried about my kids).	1072	prediction	one year	neutral
[If I were you, I {BUY} a house as soon as possible...] ...next year, the market {GO} up even more.	1073	prediction	one year	neutral
[If I were you, I {SELL} your house as soon as possible...] ...next year, the market {GO} down even more.	1074	prediction	one year	neutral
[I {ADVISE} that you don't {PURCHASE} oil stocks...] ...renewable energy {BE} cheaper by 2020.	1075	prediction	two years	neutral

Continued on next page...

Table I.2 – *FTR-elicitation questionnaire Version 2 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition
[I {ADVISE} that you {PURCHASE} oil futures...] ...renewable energy never {REPLACE} oil.	1076	prediction	ongoing prediction	certain
[Don't bother investing in the tech industry...] ...Silicon Valley {CRASH} by 2018 (definitely).	1077	prediction	two years	certain
[You should invest in the tech industry...] ...Apple {DOUBLE} again in the next 25 years.	1078	prediction	twenty five plus	neutral
[I would definitely {BRING} some sun cream on the weekend, the weather forcast {SAY}...] ... temperature {HIT} forty degrees!	1079	prediction	one week	neutral
[I would definitely {BRING} a hat tomorrow, I {HEARD} on the news that...] ...UV levels {BE} very high this summer.	1080	prediction	tomorrow	neutral
[Did you bring your umbrella? I {READ} that...] ...a storm {BLOW IN} this afternoon.	1081	prediction	today	neutral
[Did you bring your warm jacket? I {HEARD} that...] ...it {SNOW} this afternoon.	1082	prediction	today	neutral
[Q: Do you {WANT} to {ORDER} pizza? A: No, I don't {EAT} takeaway...] ...you {BE} fat (before you turn 30 if you keep it up.)	1083	prediction	indet.	neutral

Continued on next page...

Table I.2 – *FTR-elicitation questionnaire Version 2 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition
[Q: Do you {WANT} to {ORDER} pizza? A: Yes, {GET} one with pepperoni. I {MAKE} a salad. By the time the pizza {ARRIVE}...] ...the salad {BE} ready.	1084	prediction	today	neutral
[You better {PUT} some ice on that sprained ankle...] ...it {SWELL UP} if you don't.	1085	prediction	ongoing prediction	neutral
[You better {PUT} some anti inflamatories on that sprained wrist...] ...it {SWELL UP} if you don't.	1086	prediction	ongoing prediction	neutral
[We should definitely {GO} to John's party this weekend. Everyone {BE} going...] ...it {BE} really fun.	1087	prediction	one week	neutral
[We should definitely not {GO} to Christine's party next week. Peter and James {BE} going...] ...it {NOT BE} fun at all.	1088	prediction	one week	neutral
[Don't vote for the liberals...] ...they {LOSE} in August (definitely).	1089	prediction	six months	certain
[You should vote for the liberals...] ...the conservatives {LOSE} in August anyway.	1090	prediction	six months	neutral
[Q: Do you {WANT} a cigarette? A: No, I Don't smoke. I {BE} afraid...] ...I {GET} cancer.	1091	prediction	indet.	neutral

Continued on next page...

Table I.2 – *FTR-elicitation questionnaire Version 2 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition
[Q: Do you {WANT} to go for a run? A: Yes, I {TRY} to get more exercise. I {BE} afraid...] ...I {GET} fat (when I {BE} old).	1092	prediction	twenty five plus	neutral
[Hendrich {GET} a whole bunch of money for his inheritance from his grandfather...] ...he {WASTE} it in no time, like always.	1093	prediction	ongoing prediction	neutral
[Jess {INHERIT} a lot of money from her mother...] ...she {SAVE} it (I am sure; she is a smart girl).	1094	prediction	ongoing prediction	certain
[Talking about the world cup: Q: You {THINK} The Netherlands {ADVANCE} to the finals this year?] A: They {MAKE} it (I'm certain).	1095	prediction	one year	certain
[Talking about the world cup: Q: You {THINK} England {ADVANCE} past Brasil in the next round?] A: They {LOSE}. (I'm certain).	1096	prediction	indet.	certain
[We definitely {NEED} to {STUDY} chapter three...] ...there {BE} a question on it (for sure).	1097	prediction	ongoing prediction	certain
[We definitely should not {STUDY} chapter three...] ...he {NOT ASK} a question on it. (He hasn't for the last three years).	1098	prediction	ongoing prediction	certain
[We should {GO} to the party tonight...] ... John {BE} there (and I {BE} certain he {LIKE} you).	1099	prediction	today	neutral

Table I.3

FTR-elicitation questionnaire Version 3 (English)

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Q: What you {DO} tonight?] A: I {DINE OUT} with John.	1001	intention	today	neutral	-999
[Q: What do you think about tonight's game?] A: Manchester {LOSE}.	1002	prediction	today	neutral	-999
[Q: Should I bring my umbrella? A: Yes...] ...a storm {BLOW IN} this afternoon.	1003	prediction	today	neutral	-999
[Speaking to a friend who fancies Claire: You should come to the party tonight...] ...Claire {BE} there.	1004	prediction	today	neutral	-999
[Q: When {BE} the exam tonight?] A: It {BE} at 7pm.	1005	scheduling	today	neutral	-999
[Q: What you {DO} tomorrow evening? A: The new Avengers film {BE} playing...] ...I {SEE} it with Mary.	1006	intention	tomorrow	neutral	-999
[It's no use trying to swim in the lake...] ...the weather {BE} cold tomorrow.	1007	prediction	tomorrow	neutral	-999
[Q: What your brother {DO} if you don't go to see him tomorrow, do you think?] A: He {CALL} me.	1008	prediction	tomorrow	neutral	-999

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Q: Do you think the exam tomorrow {BE} hard?] A: We {PASS}.	1009	prediction	tomorrow	neutral	-999
[Q: Do you know when dawn {BE}, tomorrow?] A: The sun {RISE} at 6.	1010	scheduling	tomorrow	neutral	-999
[Q: When does your mother arrive? A: Ug. She always buys the kids too much sugar. When she gets here next week...] ...I {SAY} no.	1011	intention	one week	neutral	-999
[Q: What is the weather forecast for next week?] A: It {RAIN}.	1012	prediction	one week	neutral	-999
[A to B: Bring some sun cream this weekend...] ...it {BE} sunny!	1013	prediction	one week	neutral	-999
[A to B: We should study chapter three for the exam next week...] ...there {BE} a question on it.	1014	prediction	one week	neutral	-999
[My brother's...] ...flight {LEAVE} in one week.	1015	scheduling	one week	neutral	-999
[Chris's father {SEND} him some money next month. When he {GET} it...] ...he {BUY} a present for Amelie.	1016	intention	one month	neutral	-999

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[A to B: Don't invest in derivatives. The market is fraudulent...] ...It {CRASH} within a month.	1017	prediction	one month	neutral	-999
[Father to son: Start investing now. In even a month...] ...your money {BE} worth more.	1018	prediction	one month	neutral	-999
[Q: How do you think the commodities markets {ACT} next month?] A: They {RISE}.	1019	prediction	one month	neutral	-999
[In November: Q: When do you fly to Mexico?] A: My flight {LEAVE} on 15 December!	1020	scheduling	one month	neutral	-999
[Q: Are you going on holiday this summer? A: Yes. My annual review is in two months. After that...] ...I {TRAVEL} to Thailand.	1021	intention	two months	neutral	-999
[Salesman to customer: I'd come back in a couple of months. Next year's stock is arriving soon...] ...the price on these {DROP}.	1022	prediction	two months	neutral	-999
[Doctor to panel: At current infection rates...] ...ebola {KILL} another 10,000 in the next two months.	1023	prediction	two months	neutral	-999

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Teacher to parent: It can take some time for kids to adjust to a new school. In a couple of months...] ...Johnny {IMPROVE}.	1024	prediction	two months	neutral	-999
[Q: When is the Radiohead show?] A: They {PLAY} in a couple months.	1025	scheduling	two months	neutral	-999
[Girl: My birthday is in three months... are you coming to my party?] Boy: I {BE} there.	1026	intention	three months	neutral	-999
[Doctor to patient: try to sleep and get some gentle exercise. In a few months...] ...you {FEEL} better.	1027	prediction	three months	neutral	-999
[A to B: The price of gold always goes up in the autumn. Let's invest now...] ... we {MAKE} a profit in a few months.	1028	prediction	three months	neutral	-999
[In the autumn, a weather presenter: Good news for all you skiers out there...] ...we {SEE} a lot of snow in the next few months.	1029	prediction	three months	neutral	-999
[Q: When do you defend your thesis?] A: My viva {BE} in three months!	1030	scheduling	three months	neutral	-999

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[In December, talking about a friend's plans: In the summer...] ...he {TRAVEL} in Morocco.	1031	intention	six months	neutral	-999
[Talking about the summer's upcoming weather (it is presently December): There {BE} a warm current coming from the tropics this year...] ...it {MAKE} this summer hot and rainy.	1032	prediction	six months	neutral	-999
[Talking about upcoming elections (it is presently July): Don't waste a vote on the liberal party...] ...they {LOSE} in January.	1033	prediction	six months	neutral	-999
[A to B: Don't bother investing in the tech industry...] ...Silicon Valley {CRASH} within six months.	1034	prediction	six months	neutral	-999
[Q: When do you move house?] A: Our lease {EXPIRE} in six months.	1035	scheduling	six months	neutral	-999
[Q: Will you stay here next year?] A: I {MOVE} to San Francisco.	1036	intention	one year	neutral	-999
[Q: Do you think Hamish and Natalie still {BE} together in a year?] A: Yes, they {STAY} together.	1037	prediction	one year	neutral	-999

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[(Talking about the world cup) Q: Do you think The Netherlands {ADVANCE} to the finals this year?] A: They {MAKE} it.	1038	prediction	one year	neutral	-999
[Q: What do you think {HAPPEN} in the American elections next year?] A: The Republicans {LOSE}.	1039	prediction	one year	neutral	-999
[Q: How long do you have on your contract?] A: It {EXPIRE} in a year.	1040	scheduling	one year	neutral	-999
[Q: Are you going to stay in the UK? A: Yes, when my documents come through in two years...] ...I {APPLY} for residency.	1041	intention	two years	neutral	-999
[Don't buy those shoes...] ...they {WEAR OUT} in a couple of years.	1042	prediction	two years	neutral	-999
[Don't purchase oil stocks...] ...renewable energy {BE} cheaper in a couple of years.	1043	prediction	two years	neutral	-999
[You shouldn't put money into real estate. In the next couple of years...] ...the market {CRASH}.	1044	prediction	two years	neutral	-999
[Q: How much longer do you have on your degree?] A: I {GRADUATE} in two years.	1045	scheduling	two years	neutral	-999

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Q: Are you selling you house? A: Maybe in another five years...] ...by then, it {BE} more valuable.	1046	prediction	five years	neutral	-999
[Q: What do you expect the sea levels {DO} in the next five years?] A: They {RISE}.	1047	prediction	five years	neutral	-999
[You should put money into a retirement savings plan. In just five years...] ...it {BE} worth a lot more.	1048	prediction	five years	neutral	-999
[Q: Do you have have a five-year-plan? A: In five years...] ...I {MAKE} partner at my firm.	1049	intention	five years	neutral	-999
[Father to son:] Your bond {MATURE} in five years.	1050	scheduling	five years	neutral	-999
[Q: Are you saving for retirement? A: Yes, I save £100 every month. In ten years...] ...it {BE} worth £40,000.	1051	prediction	ten years	neutral	-999
[Board director: How long will it take to build the new bridge?] Project manager: It {TAKE} ten years.	1052	prediction	ten years	neutral	-999
[Q: Will you stay at your company? A: Yes, in ten years...] ...my stock options {BE} worth enough to retire.	1053	prediction	ten years	neutral	-999

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Father: What do after you graduate from college?] Young teen: I {TRAIN} as a firefighter.	1054	intention	ten years	neutral	-999
[Q: How long until you pay off your mortgage?] A: The payment schedule {FINISH} in ten years.	1055	scheduling	ten years	neutral	-999
[A to B: You better put some ice on that sprained ankle...] ...it {SWELL UP}.	1056	prediction	ongoing prediction	neutral	-999
[A: I have a headache. B: Take this medicine...] ...It {MAKE} you feel better.	1057	prediction	ongoing prediction	neutral	-999
[Q: What {HAPPEN} if you eat this mushroom?] A: You {DIE}.	1058	prediction	ongoing prediction	neutral	-999
[Q: You want to see a film tonight? A: Sorry, I can't...] ...I {DINE OUT} with Ellie.	1059	intention	today	certain	100
[Q: What do you think about the football this evening?] A: Madrid {LOSE}.	1060	prediction	today	certain	100
[Q: Should I being my sun hat to the beach? A: Yes...] ...it {BE} very hot this afternoon.	1061	prediction	today	certain	100
[Speaking to a friend who fancies Annie: You should come to the movie tonight...] ...Annie {BE} there.	1062	prediction	today	certain	100

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Q: When {BE} the exam tonight? A: They have been the same all week...] ...it {BE} at 7pm.	1063	scheduling	today	certain	100
[Q: What you {DO} tomorrow evening? A: The new James Bond film {BE} playing...] ...I {SEE} it with Mary.	1064	intention	tomorrow	certain	100
[It's no use skiing tomorrow...] ...the snow conditions {BE} bad.	1065	prediction	tomorrow	certain	100
[Q: What your dad {DO} if you don't go to see him tomorrow, do you think?] A: He {CALL} me.	1066	prediction	tomorrow	certain	100
[Q: Do you think the exam tomorrow {BE} difficult? A: Professor Green is easy...]...we {PASS}.	1067	prediction	tomorrow	certain	100
[Q: Do you konw when the sun {RISE} tomorrow?] A: It {RISE} at 6:17.	1068	scheduling	tomorrow	certain	100
[Q: When does your mother-in-law arrive? A: Ug. She buys the kids too many toys, like always. I can't stand it...] ...I {SAY} no this time.	1069	intention	one week	certain	100

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Q: Do you know what the weather {BE} like next week? A: A big storm {BE} coming...] ...It {BE} bad.	1070	prediction	one week	certain	100
[A to B: Bring an umbrella to Germany next week...] ...it {RAIN} in Berlin.	1071	prediction	one week	certain	100
[A to B: We should definitely study the digestive system for the exam in a week...] ...there {BE} a question on it.	1072	prediction	one week	certain	100
[Christine {SAY} yesterday that my brother {LEAVE) tomorrow, but...] ...his flight {LEAVE} next week.	1073	scheduling	one week	certain	100
[Jen's uncle {SEND} her some money next month. She just loves skiing. When she gets it...] ...she {BUY} a new pair of skis.	1074	intention	one month	certain	100
[A to B: Don't invest in commodities. The market is very shaky...] ...it {CRASH} within a month.	1075	prediction	one month	certain	100
[Mother to daughter: Put your allowance in a savings account. Next month...] ...it {BE} worth much more.	1076	prediction	one month	certain	100

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Q: What is your opinion for oil futures next month?] A: They {RISE}.	1077	prediction	one month	certain	100
[In June: Q: When do you and Jen fly to France? A: She {SAID} it {BE} the 17th, but I just {CHECK} the schedule...] ...we {LEAVE} on the 15th.	1078	scheduling	one month	certain	100
[Q: Are you going on holiday this summer? A: Yes, in a couple of months. My boyfriend doesn't want to, but...] ...we {TRAVEL} to Georgia.	1079	intention	two months	certain	100
[Salesman to customer: I wouldn't buy for a couple of months. Next year's stock is arriving...] ...the price on these {DROP}.	1080	prediction	two months	certain	100
[Doctor to panel: You can disregard eveything my ill-informed colleague just said. At current infection rates...] ...swine flu {KILL} 5,000 people in the next two months.	1081	prediction	two months	certain	100
[Teacher to parent: Peter just needs time to settle in. In a couple of months...] ...he {IMPROVE}.	1082	prediction	two months	certain	100

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Q: Are you excited for the Billie Eilish gig next week? A: What? I just checked online...] ...she {PLAY} in two months.	1083	scheduling	two months	certain	100
[Boy: My birthday is in three months... are you coming to my party?] Girl: I {BE} there.	1084	intention	three months	certain	100
[Doctor to patient: This is just stress. I see it all the time. Try to sleep well and get exercise. In a few months...] ...you {FEEL} better.	1085	prediction	three months	certain	100
[A to B: The price of oil always goes up in the summer. Let's invest now...] ... we {MAKE} a profit in a few months.	1086	prediction	three months	certain	100
[In the spring, a weather presenter: There hasn't been such a sustained series of high pressure systems in years...] ...we {SEE} a lot of nice weather this summer.	1087	prediction	three months	certain	100
[Supervisor to student: So you defend next month?] Student: No, my viva {BE} in three months.	1088	scheduling	three months	certain	100

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[In the spring, talking about a friend's plans: I don't know why you think otherwise! She told me yesterday she has {BOUGHT} her ticket...] ...she {TRAVEL} in South East Asia this autumn.	1089	intention	six months	certain	100
[Talking about the winter's upcoming weather (it is presently June): There {BE} a cold current coming from the arctic this year...] ...it {SNOW} a lot this winter.	1090	prediction	six months	certain	100
[Talking about elections in the spring (it is presently October): A: Don't waste a vote on Labour...] ...they {LOSE} in April.	1091	prediction	six months	certain	100
[A to B: Don't bother investing in the China. There is a bubble...] ...it {CRASH}, within six months.	1092	prediction	six months	certain	100
[Q: Are you sure about the lease?] A: It {EXPIRE} in six months.	1093	scheduling	six months	certain	100
[Q: Will you live in London next year?] A: Yes. I {MOVE} to Camden.	1094	intention	one year	certain	100

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Q: Do you think Amy and John still {BE} together in a year? A: I heard he proposed...] ...they {STAY} together.	1095	prediction	one year	certain	100
[Talking about the world cup: Q: Do you think England {ADVANCE} past Brasil in the next round?] A: They {LOSE}.	1096	prediction	one year	certain	100
[Q: What do you think {HAPPEN} in the American elections next year? A: Have you seen their candidate play the crowds?] ...the Republicans {WIN}.	1097	prediction	one year	certain	100
[Q: How long do you have on your contract? A: I checked yesterday...] ...it {EXPIRE} in a year.	1098	scheduling	one year	certain	100
[Q: Are you going to stay in the UK? A: I love it here...] ...I {APPLY} for residency in two years.	1099	intention	two years	certain	100
[A to B: Don't buy Adidas. I had some...] ...they {WEAR OUT} in a couple of years.	1100	prediction	two years	certain	100
[A to B: I advise that you purchase renewables...] ...renewable energy {REPLACE} oil in the next two years.	1101	prediction	two years	certain	100

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[A to B: You should invest in gold. In the next couple of years...] ...the stock market {FALL}.	1102	prediction	two years	certain	100
[Q: How much longer you have on your degree?] A: I {GRADUATE} in two years.	1103	scheduling	two years	certain	100
[Q: Are you selling you car? A: I cannot wait 5 years for the least to expire...] ...it {BE} worth much less by then.	1104	prediction	five years	certain	100
[Q: What do you expect the sea levels to do in the next five years? A: I mean...] ...they {RISE}.	1105	prediction	five years	certain	100
[There is no excuse not to invest in a retirement savings plan. In five years...] ...your money {BE} worth a lot more.	1106	prediction	five years	certain	100
[Q: Do you have have a five-year-plan? A: Yes, in five years...] ...I {MAKE} partner.	1107	intention	five years	certain	100
[Father to son: I have just checked...] ... your bond {MATURE} in five years.	1108	scheduling	five years	certain	100
[Q: Are you saving any money? A: Yes, I save £100 every month. In ten years...] ...it {BE} worth it.	1109	prediction	ten years	certain	100

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Board director: How long will it take to build the new tower?] Project manager: It {TAKE} ten years at least.	1110	prediction	ten years	certain	100
[Q: Will you stay with Deutsche Bank for another ten years? A: Yes. By then...] ...my stock options {BE} enough to retire.	1111	prediction	ten years	certain	100
[Grandmother: Have you thought about working for your grandfather at the factory after you graduate from college? Young teen: No...] ...I {ENROL} in the police academy.	1112	intention	ten years	certain	100
[Q: How long until you pay your mortgage? A: Jane said yesterday it {BE} eight years, but...] ...it {FINISH} in ten.	1113	scheduling	ten years	certain	100
[You better put some ice on your back. It looks bad...] ...it {SWELL UP}.	1114	prediction	ongoing prediction	certain	100
[A: I have a headache. B: These pills are amazing. Take some...] ...they {HELP}.	1115	prediction	ongoing prediction	certain	100
[Q: What {HAPPEN} if you eat this mushroom? A: Wow! That is an Amanita Pantherina...] ...you {GET} sick.	1116	prediction	ongoing prediction	certain	100

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Q: What you {DO} tonight? A: Hmmm...] ...I {DINE OUT} with Christine.	1117	intention	today	uncertain	50
[Q: What do you think about the match tonight? A: It {BE} hard to say...] ...Arsenal {LOSE}.	1118	prediction	today	uncertain	40
[Q: Should I bring and extra warm jacket? A: It be your decision...] ...it {SNOW} this afternoon.	1119	prediction	today	uncertain	60
[Speaking to a friend who fancies John: You should come to the dinner tonight...] ...John {BE} there.	1120	prediction	today	uncertain	50
[Q: When {BE} the exam tonight? A: I have to check the class portal, but...] ...it {BE} at 7pm.	1121	scheduling	today	uncertain	50
[Q: What you {DO} tomorrow? A: Imagine Dragons {PLAY} tomorrow. They are not bad, but the tickets {BE} expensive...] ...I {SEE} them with Tom.	1122	intention	tomorrow	uncertain	50
[I have some doubts about {TRYING} to sail tomorrow...] ...the wind {BE} too high.	1123	prediction	tomorrow	uncertain	50

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Q: What your sister {DO} if you don't go to see her tomorrow, do you think?] A: She {CALL} me.	1124	prediction	tomorrow	uncertain	40
[Q: Do you think the exam tomorrow will be hard? A: We've been studying, but I never know with Professor Johnson...]...we {PASS}.	1125	prediction	tomorrow	uncertain	60
[Q: Do you know when the moon {RISE} tomorrow night?] A: It {RISE} at 17:00.	1126	scheduling	tomorrow	uncertain	50
[Q: When does your father arrive? A: Ug. He always buys the kids too much junk food. But they {LOVE} him so much...] ...I {SAY} no this time.	1127	intention	one week	uncertain	50
[Q: Do you know what the weather {BE} like next week?] A: It {BE} sunny.	1128	prediction	one week	uncertain	40
[A to B: I would bring your bathing suit this weekend...] ...the tempurature {HIT} 45 degrees.	1129	prediction	one week	uncertain	50
[A to B: Don't you think we should study the extra material for next week's test...] ...there {BE} a question on it.	1130	prediction	one week	uncertain	60

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[A to B : John {SAY} yesterday that Mary {LEAVE} in two weeks, but...] ...her flight {LEAVE} next week.	1131	scheduling	one week	uncertain	50
[Chris's brother {SEND} him some money next month. You never know with him. When he gets it...] ...he {SPEND} it at the bar.	1132	intention	one month	uncertain	50
[A to B: Don't invest in Latin America. Conditions are unstable...] ...Brasil {CRASH} within a month.	1133	prediction	one month	uncertain	50
[Father to son: Don' t throw out that action figure. We {GO} to the comic convention next month...]...it {BE} worth something there.	1134	prediction	one month	uncertain	60
[Q: How do you expect the stock markets {BEHAVE} next month? A: It is unclear...] ...they {RISE}.	1135	prediction	one month	uncertain	40
[In February: Q: When do you fly to Spain? A: I have to check the ticket...] ...my flight {LEAVE} 15 March.	1136	scheduling	one month	uncertain	50

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Q: Are you going on holiday this summer? A: I don't know. My annual review is in two months. After that...] ...I {TRAVEL} to Italy.	1137	intention	two months	uncertain	50
[Salesman to customer: I'd try coming back in a couple of months. People aren't buying too many...] ...the price {DROP}.	1138	prediction	two months	uncertain	40
[Doctor to panel: We need more research on the transmission vectors...] ...the H1N1 virus {KILL} 10,000 people in the next two months.	1139	prediction	two months	uncertain	60
[Teacher to parent: Ellie worries me but I think she just needs some time to adjust. In a couple of months...] ...she {IMPROVE}.	1140	prediction	two months	uncertain	50
[Q: When is the Lady Gaga concert?] A: She {PLAY} in a couple of months.	1141	scheduling	two months	uncertain	50
[Boy: My birthday is in three months... are you coming to my party? Girl: I have to check with my parents...] ...I {COME}.	1142	intention	three months	uncertain	50

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Doctor to patient: These results concerning, but they can be caused by stress. Try to sleep and get some exercise. In a few months...] ...you {FEEL} better.	1143	prediction	three months	uncertain	40
[A to B: Gold usually goes up when markets are skittish. Don't worry...] ... we {START} to make money in a few months.	1144	prediction	three months	uncertain	50
[In the spring, a weather presenter: The broad outlook over the next few months is fairly unstable, but don't give up hope...] ...we {SEE} sunny weather this summer.	1145	prediction	three months	uncertain	60
[Student to supervisor: So when do I defend? Supervisor: Let me check...] ...your viva {BE} in three months.	1146	scheduling	three months	uncertain	50
[In October, talking about a friend's plans: I do not think he is happy living here...] ...he {TRAVEL} in Mexico in the spring.	1147	intention	six months	uncertain	50
[Talking about the winter's upcoming weather (it is presently June): I hear there {BE} a warm current coming up from the tropics...] ...it {MAKE} this winter warm and wet.	1148	prediction	six months	uncertain	50

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Talking about upcoming elections (it is presently January) Q: What do you think will happen to the Tories?] A: They {LOSE} this summer.	1149	prediction	six months	uncertain	40
[A to B: Are you sure about investing in Africa? The next six months {LOOK} unstable...] ...it {CRASH}.	1150	prediction	six months	uncertain	60
[Q: When do you move house? A: I need to check the documents...] ...our lease {EXPIRE} in six months.	1151	scheduling	six months	uncertain	50
[Q: Will you live here, next year? A: I don't know...] ...I {MOVE} to L.A..	1152	intention	one year	uncertain	50
[Q: Do you think Delia and Chris still {BE} together in a year? A: I don't know. They are fighting a lot, but they {LOVE} each other...] ...they {STAY} together.	1153	prediction	one year	uncertain	50
[Q: Do you think Man U {ADVANCE} to the finals this year? A: They had a weak start, but they are playing well...] ...they {MAKE} it.	1154	prediction	one year	uncertain	60

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Q: What do you think {HAPPEN} in the American elections next year? A: The Republicans have a strong field, but you never know... with the Democrat base...] ...the Demograts {WIN}.	1155	prediction	one year	uncertain	40
[Q: How long do you have on your contract? A: I am not sure...] ...it {EXPIRE} in a year.	1156	scheduling	one year	uncertain	50
[Q: Are you going to stay in the UK? A: My visa {EXPIRE} in two years. After that...] ...I {APPLY} for residency.	1157	intention	two years	uncertain	50
[A to B: Be careful buying that old Ford...] ...the engine {WEAR OUT} in a couple of years.	1158	prediction	two years	uncertain	60
[A to B: Be careful purchasing oil stocks now...] ...renewable energy {BE} cheaper in a couple of years.	1159	prediction	two years	uncertain	40
[A to B: Be cautious investing in crypto currencies. In the next couple of years...] ...the market {CRASH}.	1160	prediction	two years	uncertain	50

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Q: How much longer you have on your degree? A: I need to check how many credits I have...] ...I {GRADUATE} in two years.	1161	scheduling	two years	uncertain	50
[Q: Are you selling you house? A: We are not sure. The market is shakey right now. In five years...] ...it {BE} more valuable.	1162	prediction	five years	uncertain	40
[Q: What do you expect the sea levels to do in the next five years? A: The science is unclear, right?] ...they {RISE}.	1163	prediction	five years	uncertain	60
[You should put money into a retirement savings plan. Interest rates are shaky now, but in five years...] ...it {BE} worth quite a bit more.	1164	prediction	five years	uncertain	50
[Q: Do you have have a five-year-plan? A: I really don't know what I want to do. I love my job, but...] ...I {MOVE} to the country.	1165	intention	five years	uncertain	50
[Father to son: I need to check when exactly, but...] ...your bond {MATURE} in five years.	1166	scheduling	five years	uncertain	50
[Q: Are you saving every month? A: Yes, I put away £100 every paycheque. In 10 years...] ...it {BE} worth as much as £30,000.	1167	prediction	ten years	uncertain	50

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[CEO: How long will it take to build the resort? Project manager: It depends on the supply chain...] ...it {TAKE} as much as ten years.	1168	prediction	ten years	uncertain	60
[Q: Will you {STAY} at your firm for good? A: I {THINK} so. In ten years...] ...my stock options {BE} worth enough to retire.	1169	prediction	ten years	uncertain	40
[Mother: What do you want to do after college? Young teen: Why do you always ask, mom! I don't know...] ...I {JOIN} the army.	1170	intention	ten years	uncertain	50
[Q: How long until you pay off your mortgage? A: I don't know. My wife manages the payment schedule...] ...it {FINISH} in ten years.	1171	scheduling	ten years	uncertain	50
[A to B: Do you need anti-inflammatories for you wrist?] ...they {HELP} it heal faster.	1172	prediction	ongoing prediction	uncertain	60
[A: I have a stomach ache. B: Have you tried ginger tea? I have been drinking it lately...] ...it {HELP}.	1173	prediction	ongoing prediction	uncertain	50

Continued on next page...

Table I.3 – *FTR-elicitation questionnaire Version 3 (English) continued*

question	question number	FTR-mode	temporal distance	modality condition	certainty number
[Q: What {HAPPEN} if you eat this berry? A: It looks sort of poisonous...] ...it {MAKE} you sick.	1174	prediction	ongoing prediction	uncertain	40

Table I.4

FTR-elicitation questionnaire Version 3.1 (English)

question	question number	temporal distance	certainty number
[Q: Should I bring my umbrella? A: Yes...] ...a storm {BLOW IN} this afternoon.	1003	today	70
[Speaking to a friend who fancies Claire: You should come to the party tonight...] ...Claire {BE} there.	1004	today	80
[Q: What do you think about the football this evening?] A: Madrid {LOSE}.	1060	today	90
[Q: What do you think about the match tonight? A: It {BE} hard to say...] ...Arsenal {LOSE}.	1118	today	60
[Q: Should I bring and extra warm jacket? A: It be your decision...] ...it {SNOW} this afternoon.	1119	today	50
[A to B: Bring some sun cream this weekend...] ...it {BE} sunny!	1013	one week	70
[A to B: We should study chapter three for the exam next week...] ...there {BE} a question on it.	1014	one week	80
[A to B: Bring an umbrella to Germany next week...] ...it {RAIN} in Berlin.	1071	one week	90
[A to B: We should definitely study the digestive system for the exam in a week...] ...there {BE} a question on it.	1072	one week	100
[A to B: I would bring your bathing suit this weekend...] ...the tempurature {HIT} 45 degrees.	1129	one week	60

Continued on next page...

Table I.4 – *FTR-elicitation questionnaire Version 3.1 (English) continued*

question	question number	temporal distance	certainty number
[A to B: Don't you think we should study the extra material for next week's test...] ...there {BE} a question on it.	1130	one week	50
[Father to son: Start investing now. In even a month...] ...your money {BE} worth more.	1018	one month	70
[Q: How do you think the commodities markets {ACT} next month?] A: They {RISE}.	1019	one month	80
[Mother to daughter: Put your allowance in a savings account. Next month...] ...it {BE} worth much more.	1076	one month	90
[Q: What is your opinion for oil futures next month?] A: They {RISE}.	1077	one month	100
[Father to son: Don' t throw out that action figure. We {GO} to the comic convention next month...]...it {BE} worth something there.	1134	one month	60
[Q: How do you expect the stock markets {BEHAVE} next month? A: It is unclear...] ...they {RISE}.	1135	one month	50
[A to B: The price of gold always goes up in the autumn. Let's invest now...] ... we {MAKE} a profit in a few months.	1028	three months	70
[In the autumn, a weather presenter: Good news for all you skiers out there...] ...we {SEE} a lot of snow in the next few months.	1029	three months	80
[Doctor to patient: This is just stress. I see it all the time. Try to sleep well and get exercise. In a few months...] ...you {FEEL} better.	1085	three months	90

Continued on next page...

Table I.4 – *FTR-elicitation questionnaire Version 3.1 (English) continued*

question	question number	temporal distance	certainty number
[In the spring, a weather presenter: There hasn't been such a sustained series of high pressure systems in years...] ...we {SEE} a lot of nice weather this summer.	1087	three months	100
[Doctor to patient: These results are concerning, but they can be caused by stress. Try to sleep and get some exercise. In a few months...] ...you {FEEL} better.	1143	three months	60
[In the spring, a weather presenter: The broad outlook over the next few months is fairly unstable, but don't give up hope...] ...we {SEE} sunny weather this summer.	1145	three months	50
[Talking about upcoming elections (it is presently July): Don't waste a vote on the liberal party...] ...they {LOSE} in January.	1033	six months	70
[A to B: Don't bother investing in the tech industry...] ...Silicon Valley {CRASH} within six months.	1034	six months	80
[Talking about the winter's upcoming weather (it is presently June): There {BE} a cold current coming from the arctic this year...] ...it {SNOW} a lot this winter.	1090	six months	90
[A to B: Don't bother investing in the China. There is a bubble...] ...it {CRASH}, within six months.	1092	six months	100
[Talking about the winter's upcoming weather (it is presently June): I hear there {BE} a warm current coming up from the tropics...] ...it {MAKE} this winter warm and wet.	1148	six months	60

Continued on next page...

Table I.4 – *FTR-elicitation questionnaire Version 3.1 (English) continued*

question	question number	temporal distance	certainty number
[A to B: Are you sure about investing in Africa? The next six months {LOOK} unstable...] ...it {CRASH}.	1150	six months	50
[(Talking about the world cup) Q: Do you think The Netherlands {ADVANCE} to the finals this year?] A: They {MAKE} it.	1038	one year	70
[Q: What do you think {HAPPEN} in the American elections next year?] A: The Republicans {LOSE}.	1039	one year	80
[Talking about the world cup: Q: Do you think England {ADVANCE} past Brasil in the next round?] A: They {LOSE}.	1096	one year	90
[Q: What do you think {HAPPEN} in the American elections next year? A: Have you seen their candidate play the crowds?] ...the Republicans {WIN}.	1097	one year	100
[Q: Do you think Man U {ADVANCE} to the finals this year? A: They had a weak start, but they are playing well...] ...they {MAKE} it.	1154	one year	60
[Q: What do you think {HAPPEN} in the American elections next year? A: The Republicans have a strong field, but you never know... with the Democrat base...] ...the Democrats {WIN}.	1155	one year	50

People tend to devalue future outcomes when they are risky or far away in time. For instance, most people would prefer $20 immediately over $20 in five years, because the delayed $20 is worth less subjectively. Similarly, many would prefer a guaranteed $100 over a 50% chance of $200, because they would not consider the risk to be worth the reward.

However, the extent to which risk and delay affect subjective value differs between people, so some might opt for the guaranteed $100, while others would choose the risky $200. Such subjective devaluation is called "psychological discounting." Differences in psychological discounting correlate with numerous behaviours which involve weighing present versus risky and/or delayed outcomes, e.g. exercising for better future health, investing in savings, and managing one's finances.

This thesis uses language elicitation with cohorts of English and Dutch speakers to reveal that, relative to Dutch, English obliges speakers to use more low-certainty future-shifting modal verbs – e.g. *may, might, could, should*. Causal modelling and experimental methods revealed that this caused increased psychological discounting in English participants. Given correlations between psychological discounting and real-world behaviours, it is argued that the grammatical obligation to use low-certainty modal verbs is driving reported correlations between languages' FTR systems and numerous important behaviours. These include national efforts to mitigate climate change; personal differences in investment in savings; and several indexes of corporate behaviour including investment in research and development, accounting practices, and cash holdings.