PNAS Nexus, 2024, 3, pgae218
https://doi.org/10.1093/pnasnexus/pgae218
Advance access publication 3 June 2024
Research Report

Foreign-language effects in cross-cultural behavioral research: Evidence from the Tanzanian Hadza

Duncan N. E. Stibbard-Hawkes[a,*], Linda Abarbanell[b], Ibrahim A. Mabulla[c], Endeko S. Endeko[d], Cristine H. Legare[e] and Coren L. Apicella[f]

[a]Department of Anthropology, Durham University, Durham, County Durham, DH1 3LE, UK
[b]Department of Psychology, College of Sciences, San Diego State University, Imperial Valley Campus, Calexico, CA 92231-09021, USA
[c]Independent Researcher, Dar Es Salaam, Tanzania
[d]Independent Researcher, Yaida, Tanzania
[e]Department of Psychology, Center for Applied Cognitive Science, University of Texas at Austin, Austin, TX 78712, USA
[f]Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA
*To whom correspondence should be addressed: Email: Duncan.Stibbard-Hawkes@durham.ac.uk
Edited By: Michele Gelfand

Abstract

Behavioral research in traditional subsistence populations is often conducted in a non-native language. Recent studies show that non-native language-use systematically influences behavior, including in widely used methodologies. However, such studies are largely conducted in rich, industrialized societies, using at least one European language. This study expands sample diversity. We presented four standard tasks—a "dictator" game, two sacrificial dilemmas, a wager task, and five Likert-risk tolerance measures—to 129 Hadza participants. We randomly varied study languages—Hadzane and Kiswahili—between participants. We report a moderate impact of study language on wager decisions, alongside a substantial effect on dilemma decisions and responses to Likert-assessments of risk. As expected, non-native languages fostered utilitarian choices in sacrificial dilemmas. Unlike previous studies, non-native-language-use decreased risk preference in wager and Likert-tasks. We consider alternative explanatory mechanisms to account for this reversal, including linguistic relativity and cultural context. Given the strength of the effects reported here, we recommend, where possible, that future cross-cultural research should be conducted in participants' first language.

Keywords: foreign language effects, economic games, risk tolerance, moral dilemmas, cross-cultural research

Significance Statement

Cross-cultural researchers working among small-scale societies have often found it expedient to conduct research using accessible national majority languages, not participants' first languages. However, recent research shows that presenting tasks in a non-native language systematically changes participant behavior: the foreign language effect (FLE). However, in FLE research, sample diversity has been lacking. The present study addressed both issues. We explore foreign language effects among the Hadza, a multilingual population who have historically lived through foraging, where school participation is low. We found clear evidence of foreign language effects in risk-sensitivity and sacrificial dilemma responses. Risk effects, particularly, were large, and could substantively confound cross-societal comparative studies and meta-analyses. Results highlight the importance of attending to language in cross-cultural study design.

Introduction

Although there are more people alive today than ever before, linguistic diversity is at a historic low. Mass communication and other avenues for enculturation—schooling, market participation, and organized religion—continue to foster linguistic homogeneity. The current ubiquity of national monolingualism, especially in industrialized anglophone nations, obscures the fact that multilingualism is the norm, both throughout the globe, and perhaps most times and places throughout human history. For instance, among the Tanzanian Hadza—an ethnolinguistic group in northern Tanzania of approximately 1,000 speakers, who have traditionally subsisted as hunter-gatherers—most individuals interact regularly with other linguistic groups: Almost everyone (1) reports some competency in a non-native language, and some individuals speak up to six. Though not studied systematically (2), this pattern is not unusual globally. Multilingualism is common in many small-scale and traditional-subsistence societies (e.g. see Refs. (2–9)), and interactions with other groups can

Competing Interest: The authors declare no competing interest.
Received: February 16, 2024. **Accepted:** May 23, 2024

profoundly shape linguistic evolution (3, 7). Multilingualism, and its influence on behavior and decision-making, is thus highly consequential for behavioral scientists conducting cross-cultural research among small-scale societies.

Much research into the influence of language on behavior has attended to linguistic relativism (i.e. the Sapir-Whorf hypothesis)—the notion that human perception, reason, and phenomenology (including non/extra-linguistic phenomenology) are structured by language (10–12). A wealth of behavioral research has explored relativism vs. universalism in color psychology (11, 13, 14), psychology of emotion (15, 16), conceptions of time (17, 18), object categorization (19), and spatial cognition (20–23).

More recent research has instead examined the influence specifically of *non-native* languages on human behavior and decision-making in experimental contexts, independent of linguistic relativism. Presenting information in a non-native language influences perception and mental imagery (24), risk-taking (25), and moral judgements (26–28); and is further associated with reduced risk and loss aversion in wager experiments (25, 29), more utilitarian solutions to moral dilemmas (27, 30–32), decreased generosity in public goods games (33), and reduced susceptibility to gambling fallacies (34). These findings are typically robust and persist across contexts (27, 32).

They are typically termed "foreign language effects" (FLEs). The term "foreign" was used instead of "second" as it implies less fluency (see discussion in (35)). As many countries have numerous languages, however, and as identities may transcend language, "non-native language effects" may be more apposite in certain contexts. When discussing the Hadza specifically, we use the term "non-native" languages. However, for the sake of consilience with the existing the literature, we use "foreign language effects" throughout and speak of "foreign languages" when discussing existing FLE literature.

There are multiple mechanistic explanations for the impact of foreign and non-native language-use on behavior and decision-making. Some (e.g. (36, 37)) propose foreign languages might elicit less pronounced emotional responses. For instance, lying, praying, swearing, and expressive phrases such as "I love you" elicit less emotional resonance in foreign languages ((38), for review). In the case of moral reasoning, foreign language-use might render normative decision rules less accessible (32, 39). Last, FLEs have been considered in light of dual-systems theory (itself reviewed (40, 41)); Due to the increased baseline demands of linguistic processing (e.g. (42)), information presented in a foreign language may elicit greater deliberation compared to the more intuitive responses elicited by one's native language (33, 43).

Foreign language effects are salient for anthropologists and other cross-cultural researchers. Many methods customarily used in cross-cultural human behavioral research—giving tasks (e.g. (44–47)), risk tasks (e.g. (48–50)), and moral dilemmas (e.g. (51–53))—are also those which yield the clearest evidence for foreign language effects (25–27, 29, 30, 32, 39, 54). Moreover, widely spoken *lingua francae* or other national majority languages are often used in cross-cultural research, and are expedient. They are typically more accessible to learners, especially nonlocal researchers. Moreover, conducting research in a national majority language broadens the available talent pool when hiring research assistants. In consequence, much influential human behavior research has variously been conducted in participants' first language (52, 53, 55, 56), a non-native language (46, 57, 58), multiple languages (e.g. (45, 59)), a choice of languages (e.g. (60)) or, in some comparative work, participants' first language at some sites and a non-native language at others (e.g. (48, 61)).

Table 1. Mean posterior probabilities of choosing the consequentialist response by dilemma type and game language, with 90 percent highest density intervals [HDI].

Dilemma type	Game language	Probability	90% HDI
Footbridge	Hadza	0.81	0.74–0.87
Footbridge	Kiswahili	0.87	0.82–0.93
Switch	Hadza	0.82	0.75–0.88
Switch	Kiswahili	0.87	0.82–0.93

Sometimes, study language is not reported or under-reported (e.g. (1, 51)), even where first languages may have been used. This is potentially consequential as, were foreign language effects to persist in nonindustrialized contexts, it would represent a notable confound in these literatures.

Where existing cross cultural-research has often not accounted for foreign language effects, research on foreign language effects has often lacked cross-cultural sample diversity. Almost all existing FLE studies have been conducted in rich, industrialized nations (27). Moreover, in a plurality of cases, one study language is English (e.g. (27, 33)), and in almost all cases at least one language is European in origin (e.g. (32, 62), and see Table 2 in Ref. (63)). Experimental findings from industrialized contexts do not unfailingly generalize and often result from hidden contextual peculiarities (64). It is possible, for instance, that speaking a minority language might importantly change the context of decision making in non-native languages. Moreover, participants in most FLE studies typically attained foreign language proficiency through formal education (27, 33), and such studies cannot account for conditioning effects, norms or heuristics introduced in learning contexts.

There is a clear need to expand sample diversity to a broader range of language groups and to investigate whether foreign language effects are also found in societies where multilingualism is normal, where non-native-language acquisition is not primarily constrained to educational contexts, where neither of the study languages is European and where the first language is a minority language, as is the case in many traditional subsistence groups (e.g. see Ref. (65)). The present study does just this.

The Hadza are an ethnolinguistic group in the north of Tanzania. While most research has attended primarily to their engagement with foraging (8, 66–72), the Eyasi region is also ideally suited to investigating multilingualism. Nearly, all Hadza individuals speak at least two languages: Hadzane, a linguistic isolate with no undisputed (73) phylogenetic relationship to any other extant language, and Kiswahili, the Bantu-Arabic *lingua franca* for much of East Africa. Almost all Hadza speak Kiswahili. Moreover, many individuals learn Kiswahili not exclusively through formal education but through interacting with members of neighboring groups (e.g. Sukuma, Iraqw, Maasai, Datoga). While most individuals speak Hadzane at home, Kiswahili is the language of trade, school, and interactions with government officials and medical professionals. Human behavioral research with Hadza participants has frequently employed those exact methodologies—moral reasoning tasks (46), giving games (1, 57, 74), and risk tasks (49)—where foreign language effects are found (54).

In the present study, among a sample of 129 Hadza participants (66 female), aged 15–75, we employed four separate research tasks selected due to (a) their centrality in cross-cultural research (e.g. (44, 49, 51, 53)) and (b) their proneness to foreign language effects (25, 29, 33, 39, 54, 75, 76). These were: (i) a single-player "dictator" game where individuals were given the opportunity keep 0–10

120 g cups of a valued food resource (maize), or donate them to another Hadza camp; (ii) two binary-outcome moral (sacrificial) dilemma vignettes, adapted from the switch and trolley problem to be contextually relevant; (iii) a one-player triple-or-nothing coin-flip wager-game, where participants could wager 0–3 120 g cups of maize; (iv) a set of three-point Likert-style questions assessing the perceived danger of five contextually salient risks to a friend (tree-climbing, traveling alone between camps, habitual smoking, habitual alcohol-use, and covid-transmission). We randomly varied study language by coin-flip between participants—such that each participant conducted all tasks in either Hadzane or Kiswahili. We also conducted a short interview in Hadzane about language-learning context (62, 77), assessing languages spoken, self-reported language fluency, frequency of "thinking in" a non-native language, and attitudes towards Kiswahili speakers.

Previous research into foreign language effects provides clear directional expectations for each set of results. Due to the proposed decreased emotional salience of non-native-language prompts (27, 36–38), inaccessibility of decision heuristics (32), and increased deliberation ((26), but see (76)), we expect that responses to the sacrificial dilemmas should be more utilitarian when presented in Kiswahili (27, 32, 63). For these reasons, we also expect that loss aversion should be diminished in Kiswahili in the wager game, where riskier decisions have a probabilistically more favorable outcome, and in Likert-measures where harms might be less salient (39). Further, like previous studies (33), in our dictator games we expect lower levels of giving in the non-native language condition. Finally, we expect that lower fluency will accentuate the strength of any foreign language effects (see (63)).

Results

To conduct these analyses, we used Bayesian binomial and cumulative regression models, alongside formal model comparison. A full account is of the analyses and modeling strategy is provided in the Methods section. Here, we report each set of results in turn, first Moral Dilemmas, then Giving Games, then Wager Games, then Likert Risk measures.

Moral (sacrificial) dilemma vignettes

As in previous studies employing sacrificial dilemmas (53), Hadza participants preferred consequentialist (utilitarian) to deontologist responses: ~86% of participants across both vignettes opted to sacrifice one person to save five. Notably, in our null (i.e. mean-only) model, there were no differences between the two dilemmas, and the estimated mean probability of choosing the consequentialist decision in the adapted switch dilemma—0.84 [90% HDI = 0.79–0.89]—was effectively identical to the adapted footbridge dilemma—0.84 [90% HDI = 0.79–0.89].

Language of instruction had a clear influence on game decisions. In line with predictions, participants were, in absolute terms, 6.2 percentage points [90% HDI = −0.45–12.76] more likely to choose the consequential response across both vignettes when presented in Kiswahili. This effect was statistically robust, and 94.02% of the contrast distribution was above zero (Fig. 1). The difference in probabilities of choosing the consequentialist response between Hadzane and Kiswahili was similar for both dilemmas (see Table 1). The model including game language outperformed the mean only model in a leave-one-out (LOO)

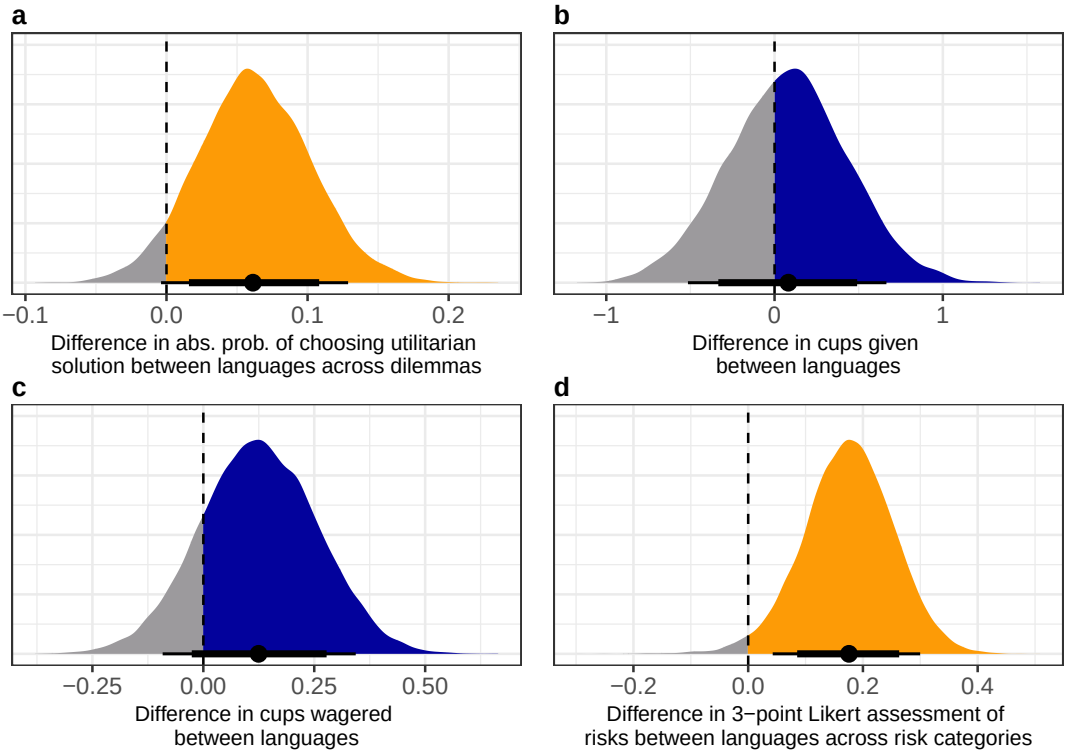

Fig. 1. Best-fitting outcome-scale study-language contrast distributions with 90% HDIs showing, left to right, top to bottom, a) A substantive increase in *p* of making utilitarian decisions in moral dilemmas moving from Hadzane to Kiswahili; b) No strong increase in cups given in a dictator game moving from Kiswahili to Hadzane; c) A moderate increase in number of cups wagered on a triple or nothing coin-flip moving from Kiswahili to Hadzane; d) A substantive increase in Likert-assessment of five perceived risks to a friend, moving from Hadzane to Kiswahili. Right-side fill represents higher MAP fitted estimate, Kiswahili in yellow, Hadzane in blue.

Table 2. Leave-one-out model selection results including expected log-predictive density differences, standard errors and Akaike weights for all four tasks (numbered by their appearance in text).

Method	Outcome	Definition	ELPD difference	SE Difference	Weights
2.1 Moral Dilemmas	Consequentialist	1 + Language + (" | Dilemma)	0.00	0.00	0.30
		1 + Language + Language*Thoughts + (" | Dilemma)	−0.42	0.89	0.20
		1 + (" | Dilemma)	−0.44	1.40	0.19
		1 + Language + Language*Fluency + (" | Dilemma)	−0.49	1.52	0.18
		1 + Language + Gender + (" | Dilemma)	−0.86	0.18	0.13
2.2 Dictator Game	Cups Given	1 + Gender + (1 | Camp)	0.00	0.00	0.46
		1 + Language + Gender + (1 | Camp)	−0.87	0.22	0.19
		1 + Language + Language*Fluency + Gender + (1 | Camp)	−1.24	0.40	0.13
		1 + Language + Language*Thoughts + Gender + (1 | Camp)	−1.26	1.02	0.13
		1 + Language + (1 | Camp)	−2.45	1.89	0.04
		1	−2.97	2.87	0.02
		1 + Language	−3.70	2.88	0.01
2.3 Wager Game	Cups Wagered	1 + (1 | Camp)	0.00	0.00	0.34
		1 + Language + (1 | Camp)	−0.05	0.81	0.32
		1 + Language + Gender + (1 | Camp)	−0.68	0.97	0.17
		1 + Language + Language*Thoughts + (1 | Camp)	−1.02	1.26	0.12
		1 + Language + Language*Fluency + (1 | Camp)	−1.86	1.02	0.05
2.4 Perceived Risks	Likert Score	1 + Language + Language*Fluency + (" | RiskType) + (1 | Camp)	0.00	0.00	1.00
		1 + Language + Language*Thoughts + (" | RiskType) + (1 | Camp)	−11.65	4.63	0.00
		1 + Language + (1 + Language | Risk_Type) + (1 | Camp)	−14.80	5.62	0.00
		1 + Language + Gender + (" | Risk_Type) + (1 | Camp)	−16.05	5.54	0.00
		1 + (" | Risk_Type) + (" | Camp)	−17.50	6.34	0.00

Left side model definitions provided in LLMs/BRMs syntax. Ditto marks (") represent repetition of nonparenthesized code.

model selection (Table 2 (2.1)). Including gender in the model did not substantively alter findings and did not improve fit (Table 2 (2.1)).

We also explored the influence of self-reported "non-native language fluency" and self-reported "frequency of thinking in non-native language" as interactions with study language. Our "fluency" model did not yield strong evidence of an interaction, and did not improve upon either the null model, or the noninteraction non-native language model in a LOO model selection (Table 2 (2.1)). Frequency of thinking in a non-native language showed some evidence of an interaction with study language in the expected direction; our interaction model predicted that those who reported never thinking in Kiswahili were 6.38 [90% HDI = −3.21–16.14] percentage points more likely to prefer the utilitarian option when vignettes were presented in Kiswahili, while those who reported often thinking in Kiswahili had substantially overlapping estimated probabilities of choosing the utilitarian option in either language (i.e. −2.93 [90% HDI = −16.11–10.8] percentage points less likely). However, this effect was not strong, and this model did not outperform the model with language only in a LOO model selection (Table 2 (2.1)).

One-player dictator game

Like previous Hadza giving games, most people gave a small amount to the community, but kept the majority of resources for themselves. In our mean only model, participants opted to give an estimated average of 3.2/10 cups [90% HDI = 2.84–3.56] of grain away and keep the remainder. There were clear gender differences in giving decisions. In our model including gender only as a fixed predictor, men gave an estimated 0.77 [90% HDI = 0.17–1.38] more cups to the community than did women. In the model including game only language as a predictor, in line with expectations, there was a small increase in the MAP estimated number of cups of grain given when the study was presented in Hadzane relative to Kiswahili (0.2). However, the estimate substantially crossed zero (90% HDI = −0.41–0.8), this model did not outperform the mean only model in a LOO model selection (Table 2 (2.2)), and the effect all but disappeared when residential camp and gender were included in the model (0.08; 90% HDI = −0.53–0.65).

We included both self-reported frequency of thinking in a non-native language and self-reported non-native language fluency, in two separate models, as interactions with game language. Both models weakly predicted that those less fluent in Kiswahili or who less often thought in Kiswahili gave more when the study was presented in Kiswahili, relative to those who were more fluent. The fluency interaction model also weakly predicted a reversal of the main effect for those who were highly fluent in Kiswahili—giving more in the Hadzane condition. However, estimates were wide, and only constituted very weak evidence of an interaction. Neither interaction model outperformed the model without an interaction term in a LOO model selection (Table 2 (2.2)).

Wager game

In our triple-or-nothing wager game participants, on average, opted to wager 1.85 cups of maize (90% HDI = 1.67–2.02), or above half of their resources (99.7% of the distribution was above that expected by chance). Unusually, we observed only minor gender differences in risk preference. In our full model including gender, women wagered approximately the same amount (i.e. only 0.07 [90% HDI = −0.15–0.29] more cups) as men. There was substantial overlap between distributions, and models including gender did not outperform those without (Table 2 (2.3)).

Against expectations, in the model including game language, participants wagered 0.13 [90% HDI = −0.09–0.35] more cups of grain when the game was presented in Hadzane, than when it was presented in Kiswahili, with 69.5% of the contrast distribution above zero (Figure 1). In a LOO model selection, the model including game language was deemed effectively equivalent to the model without. The language model had an ELPD difference of only −0.05 (Table 2 (2.3)) and commanded almost equal (i.e. 49% of) model weight when both were compared.

As before, we ran two models including self-reported "frequency of thinking in a non-native language", and self-reported fluency as interactions with game language. The interaction model including

"language of thought" followed expectations—the main finding that individuals wagered more when the game was presented in Hadzane was accentuated for those who reported never thinking in Kiswahili (estimated MAP difference = 0.25 [90% HDI = −0.04–0.55]), and diminished, perhaps weakly reversed (estimated MAP difference = −0.11 [90% HDI = −0.53–0.32]), for those who reported often doing so. Unexpectedly, the fluency model showed the opposite trend, weakly accentuating the main effect among those with greater Kiswahili fluency. However, there was great uncertainty in estimates and neither interaction model outperformed any non-interaction model in a LOO model selection (Table 2 (2.3)).

Perceived-risk Likert measures

Each of the five potential risks in our survey were perceived as being somewhat dangerous. The activities perceived as least risky were regular smoking, with an estimated MAP Likert-score of 2.16 [90% HDI = 1.96–2.35], followed by regular drinking (2.27; 90% HDI = 2.1–2.45). Walking alone out of camp (2.4; 90% HDI = 2.24–2.57), contracting coronavirus (2.41; 90% HDI = 2.26–2.59), and climbing trees, e.g. to get honey (2.43; 90% HDI = 2.27–2.59), were each perceived as similarly highly risky.

While most previous studies predict that non-native languages should reduce the perceived salience of risk, we again found the opposite trend. When questions were presented in Kiswahili, perceived risk increased by a MAP estimated 0.17 [90% HDI = 0.05–0.3] Likert points across all questions, relative to Hadzane, with 98.1% of the contrast distribution above zero. These findings persisted across risk categories, and were statistically robust—those models including game language substantially outperformed those without in a LOO model selection (Table 2 (2.4)). Including gender in the model did not substantively impact estimates, and models including gender did not outperform those without in a LOO model selection (Table 2 (2.4)).

We also ran models including frequency of thinking in a non-native language and self-reported non-native language fluency as interactions with study language. The "language of thought" interaction model aligned with expectations: the main effect was greatly accentuated among those who reported never thinking in Kiswahili (i.e. a 0.32 [90% HDI = 0.18–0.46] increase in perceived risk across categories), and less extreme with a minor MAP reversal of the effect among those who reported often thinking in Kiswahili (a −0.11 [90% HDI = −0.3–0.08] decrease in perceived risk across categories) with greater overlap between distributions (16.9% of the contrast distribution above zero). The model including fluency (Fig. 2) showed an even greater accentuation of the effect of game language among those with the lowest Kiswahili fluency (i.e. a 0.45 [90% HDI = 0.25–0.66] estimated increase in perceived risk across categories). However, contrary to expectations, this model predicted a moderate reversal of the trend in those reporting the greatest Kiswahili fluency (a −0.25 [90% HDI = −0.41–0.09] decrease in perceived risk across categories). Both interaction models outperformed the non-interaction models in a LOO model-selection, the fluency model very substantially (Table 2 (2.4)).

Discussion

Here, we report three key findings. First, this study yielded clear statistical evidence of an effect of task language on decision-making in sacrificial dilemmas and Likert-assessments of perceived risk to a second-party. Second, we found moderate evidence of an effect of study language on willingness to risk resources in a favorable wager. Third, we find little compelling statistical evidence of any influence of game language on generosity in a public goods game—and any effect disappeared entirely when conditioning on gender.

That 3/4 measures showed statistically real differences between languages, even in a sample of 129 individuals, provides compelling evidence for a general effect of study language on study outcomes in at least three important and widely used methods. This finding, alongside the existing literature on foreign language effects, highlights the need for anthropologists and cross-cultural researchers to systematize study language, using first languages where possible. Findings are especially consequential because the effects here were in several cases large, and could substantively confound cross-societal comparative studies and meta-analyses.

Across all measures, we also tested for an interaction effect of fluency (i.e. self-reported fluency and frequency of "thinking in a non-native language") on these trends. Interaction effects were mostly in the expected direction—accentuating the primary finding for less fluent speakers (see (63, 78) but see (79, 80)). However, effects were weak in both the sacrificial dilemmas and wager games, and only improve upon the noninteraction model in the Likert-style risk assessments. Results thus provide only tentative support for the notion that lower fluency accentuates foreign language effects. This may result from the homogeneity of fluency in the study sample relative to other settings, or noise introduced via self-report measures. We predict that the effect would have a stronger statistical signature in larger or more diverse samples and results certainly do not constitute evidence against a general interaction of FLEs with language fluency (63, 78).

Although we find strong evidence for an influence of non-native language on study outcomes, despite careful translation, for both risk measures effects were not in the direction predicted by the existing "foreign language effect" literature. This merits further consideration. We discuss each set of results in more detail—first moral dilemmas, then social preference tasks and finally our two risk preferences measures. We finish by considering two potential alternative mechanisms— linguistic relativism (i.e. small yet consequential language-specific differences in meaning and semantic valance) and linguistic context specific norms—which might account for the unexpected direction of effects in the two risk preference measures.

Moral dilemmas

While foreign language effects have been investigated in multiple contexts, moral dilemmas are the most widely researched (27, 32, 63, 76, 79). In most studies, presenting sacrificial dilemmas in a non-native language increased the probability of making utilitarian choices (i.e. electing to sacrifice one person to save several), perhaps as consequence of reduced emotional saliency of the vignettes or of greater deliberation (see (76)). The literature appears largely free from publication bias (see (79, 80)), the effect has broad support (63, 79, 80) and has been replicated numerous times. In a review of 111 studies, bilingual individuals were 1.6 times more likely to choose the utilitarian option in moral dilemma presented in a non-native language (63). Unlike the majority of previous studies, the present study was conducted between two non-European languages; two, moreover, that share no close phylogenetic relationship. Also, unlike previous studies, we adapted the wording of our sacrificial dilemmas to make them coherent for individuals who might be unfamiliar with the concept of a trolley car—substituting the trolley with a moving boulder and the switch with an unattached log. Despite the novel context, our study aligns with previous findings.

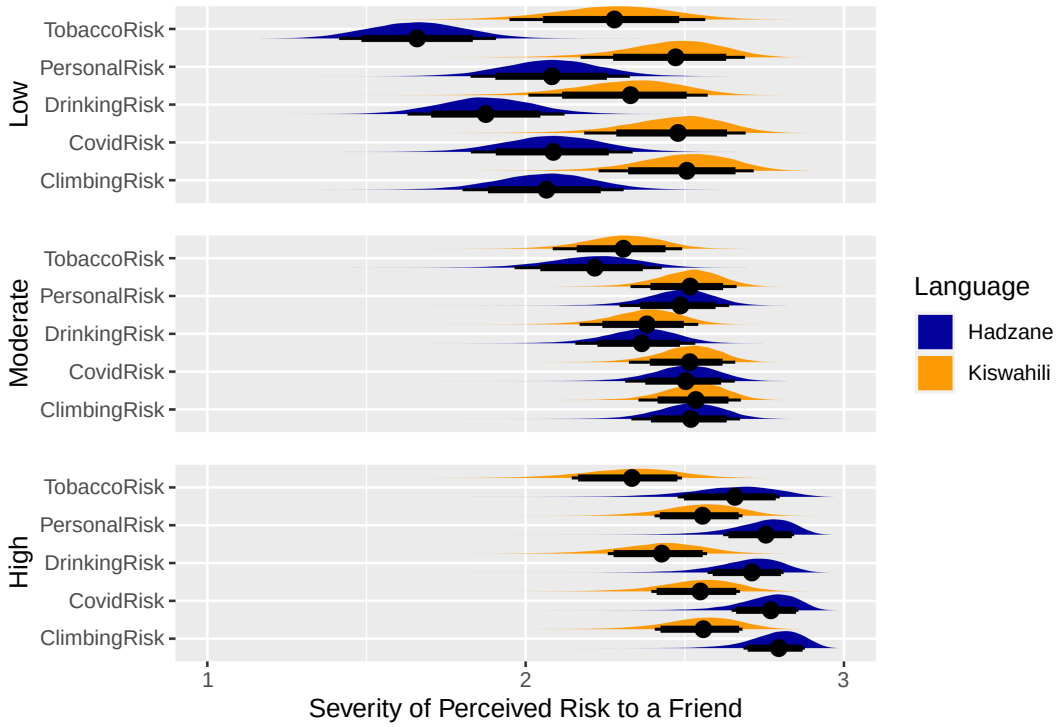

Fig. 2. Fitted predictions from our best-fitting interaction model, showing study-language differences in risk estimation (not worried; a little worried; very worried) across five risk categories among low-fluency, conversant (i.e. moderate fluency) and high-fluency Kiswahili speakers.

In a previous Hadza moral dilemma study, people did not prefer deontological over consequentialist (i.e. utilitarian) solutions, and so were more utilitarian than Western respondents, who favored deontological solutions (53). Here, we replicate the finding that deontological solutions are not preferred, though, in fact, present results were far stronger. Unlike Smith and Apicella (53), who found no statistically real population-wide preference for either deontological and consequentialist responses, we instead report a strong preference for consequentialist solutions across dilemmas and languages. This dissimilarity is difficult to account for but is perhaps a consequence of methodological difference: The 2022 study invoked third-person decisions and posed more complex moral scenarios involving theft and predation risk.

In line with previous foreign language effect research (27, 32, 63), presenting dilemmas in a non-native language increased the probability of choosing utilitarian solutions. This effect was fairly substantial—a MAP increase in probability of 6.5 percentage points. Although a proportion of the fitted contrast distribution crossed zero (see Fig. 1), some 17% of the contrast distribution density was above 10 percentage points. To give some context for this effect, the MAP divergence between language conditions reported here *within* one population would match or exceed the differences *between* most pairs of European countries (81). Moral decision-making, of various kinds, has been a key topic in cross-cultural research (51, 53, 56, 82). Gladly, many previous studies already employed first languages (e.g. (53, 56)). The strength of the effect here demonstrates why it is important to continue doing so, especially in cross-cultural comparisons.

Giving games

Compared to moral dilemmas, there is less research into the influence of non-native languages on behavior in giving games (33, 83). However, the one study that explicitly draws on the FLE framework reports that presenting economic games in a foreign language decreases generosity (33)—a fact the authors attributed to greater deliberation. This is potentially consequential, as economic giving games are routinely used in cross-cultural research (45, 46, 57, 84, 85), including in some of the fields' most high-profile comparative studies (55, 59).

Like Li et al. (83), this study finds scant statistical evidence for an effect of game language on incentivized social preference tasks. Here, in our best-fitting model—accounting for the effect of gender and study language on game decisions—MAP estimates were close to zero (0.08; 90% HDI = −0.53–0.64). This is possibly a result of the fact that, while certain tasks like the moral dilemmas, involved much verbal reasoning, giving games largely involved token allocation, with a shorter rubric. However, while more complex, the wager game also involved primarily numerical reasoning, and did show evidence of FLEs, both here and elsewhere (25, 29, 75). It may simply be that decisions of this type are less prone to foreign language effects, although more data are needed to make any such determination.

Present results imply that studies which have previously employed economic games using either non-native languages or multiple languages (e.g. (45, 46, 57, 59)), may not be compromised. However, it is difficult to prove a negative, and our model did not rule out changes in giving of ±0.5/10 cups between language conditions—which would constitute a small, but not insubstantial effect. To more decisively rule out foreign language effects in giving game behavior, it would be worthwhile to repeat the present methodology in a wider variety of contexts.

Risk preferences

The last two measures both explored sensitivity to different forms of risk. As with moral dilemmas, there is extensive research in most cases demonstrating that non-native-languages diminishes

risk preference in the context of experimental games (25, 29, 75), and in Likert-style responses to hazards (39). This trend has broad literature-wide support (79) albeit with some notable replication failures (75). Further, as with moral dilemmas, there are multiple proposed mechanisms which might explain this phenomenon: It is proposed that non-native languages may reduce risk aversion by reducing the emotional saliency (a.k.a., "affect") of loss (25, 29, 39) and also by rendering normative decision rules or heuristics less accessible (25).

In both experiments, there was evidence that game language altered sensitivity to risk. The impact on risk preference in the wager game was modest, a MAP increase of 0.13/3 cups wagered in the non-native language condition relative to the Hadzane condition. This model was deemed equivalent to the null model in a model selection. For Likert-style risk sensitivity questions, the effect was substantial. Averaged across five locally salient risks (tobacco, alcohol, contracting covid, tree-climbing, and walking along at night) MAP-estimated Likert assessment of risk to a friend increased by 0.17/3 points—and almost half a point among those with the lowest Kiswahili fluency in the best-fitting language-fluency interaction model.

These results provide clear evidence for a directional effect of study language on risk preference. This is notable. Risk preference has been an important paradigm in both the evolutionary anthropology of hunting decisions (86–89), and broadly in cross-cultural experimental work (e.g. (48, 49)). The realities of cross-cultural comparative experimental work sometimes engender pragmatic trade-offs between minimization of confounders and successful data-collection. Yet these findings again highlight the importance of conducting such experiments in a first language where possible.

In wager-games, men and women were similarly risk-seeking; an unexpected finding as men are usually more risk seeking both in previous Hadza studies (49, 60) and across cultures (50, 90). The incentive structure here made wagering additional resources favorable—much more so than prior maize risk tasks (49)—and it's possible that these high potential gains overwhelmed any gender differences in loss-aversion. Likert results showed the expected pattern: women across language conditions provided higher assessments of risk than men.

Perhaps the most unexpected finding was not the absence of gender differences in wager games, but the direction of the effects. The FLE literature expects, and in most cases finds (39, 79, 80) that non-native languages systematically *decrease* sensitivity to risk and loss aversion. Here, we report the opposite. Presenting the study in a non-native language—Kiswahili—resulted in participants wagering less in a favorable bet, and greatly increased their assessment of five different types of risk. This contradicts the narrative that non-native languages diminish the "affect" of different stimuli. Given that fluency had an only minor impact on the strength of the main study effect in vignette measures—the most linguistically complex tasks—this is unlikely to have resulted from differences in comprehension between languages. It is also worth highlighting that, for ease of participant comprehension, the risks considered in the Likert questionnaires (e.g. smoking, drinking alcohol, climbing) were exclusively proximate, while previous Likert questionnaire studies primarily explored more abstract risks like nuclear power or chemical fertilizers (39). However, Hadjichristidis et al. (39) also found foreign language effects for proximate risks including alcohol consumption. Moreover, this does not explain why wager game results also ran contrary to the predicted direction. Here, we consider two other possible alternative mechanisms which might explain this finding.

The first is linguistic relativism (11, 19): Although care was taken to ensure the meaning of the translations matched as closely as possible, we cannot discount the possibility that Kiswahili relates semantic concepts in a way that engenders higher risk-sensitivity. The Hadza language, by dint of its small number of speakers, is poorly described. This precludes a thorough exploration of the semantic mappings between Hadzane and Kiswahili. In support of linguistic relativism, there is some evidence that language-similarity diminishes foreign language effects (78, 79, 91). However, were linguistic relativism a key concern, we might expect more evidence for changes in polarity between language pairs than is observed in the existing literature (63, 79, 80). Most existing studies employ at least one European language, usually two, and research in different linguistic contexts and between a greater variety of language pairs—both closely and more distantly related—would address this question.

The second possibility is that the context in which a language is spoken might, itself, imbue a language with salient emotional content, engender different sets of decision heuristics, or bring to mind different norms. There is evidence from multiple sources that bilinguals take on board the cultural norms associated with the language they are speaking (92–94) to the extent it may measurably influence personality perception ratings (92). Normative differences in risk perception between Swahili and Hadza people have not been directly investigated; and such investigations are complicated by the fact that Kiswahili is a trade language. Yet there is evidence of high risk-tolerance in Hadza foraging decisions (87, 95), and wagering is commonplace (49, 96). The minor fluency effects in the Likert-tasks also support the possible influence of language-bound normativity.

Another related explanation is not cultural normative content *sensu lato* (i.e. injunctive norms (97)), but heuristics formed in the specific contexts in which the language is spoken (i.e. descriptive norms (97)). Most FLE studies have been conducted in industrialized societies where linguistic diversity is low, and languages are acquired in the classroom (79, 80). The present study is unique in that, while Hadzane is spoken with peers and neighbors, Kiswahili is the language used to interact with government officials, missionaries, educators, and members of other tribes, some of whom have historically competed for land and resources.

There are several reasons to suppose, then, that interactions specifically in Kiswahili might systematically engender greater risks than do interactions in Hadzane, associating the language with descriptive norms (*sensu* (97)) of risk aversion. In formal educational contexts (see (65)), for example, Hadza children are often exposed to greater risk of disciplinary punishment. In contrast, Hadza pedagogical practices involve less formal instruction, with greater emphasis on autonomy, play, and autodidacty (61, 98), and are typically more permissive of errors. Market interactions involve relationships of trust with unknown third parties (46, 57)—and in consequence greater risk of being cheated without recourse (though, for counterpoint, see (48)). Similarly, interactions with proselytizers often depend on carefully choosing responses in order to receive gifts of food or clothing (57, 99). Kiswahili is also the language of healthcare, where failure to communicate might result in less effective treatment. Last, interactions with state officials have sometimes historically been negative, e.g. engendering risk of forced relocation (100, 101).

It is possible, therefore, that language-learning contexts which are relevant to task decisions might moderate foreign language effects. In the present study, we tried to capture this possibility indirectly by asking people's opinions of Kiswahili speakers. Few participants reported having a negative view of Kiswahili speakers

(7.3%). These responses were too invariant for meaningful statistical analysis. To further address this, it would be worthwhile to investigate the relationship between language-learning context, language and risk preference among other societies for whom schooling, healthcare and interactions with markets, government officials or missionaries is primarily conducted in a non-native language.

Conclusions

Here, we find clear evidence of foreign language effect in risk sensitivity and responses to ethical dilemmas. In two of three cases, the effect was substantial and, e.g. difference between language conditions for sacrificial dilemmas were comparable or greater than between most pairs of countries (81). Risk preference and ethical dilemmas are both important topics in the human behavioral sciences, and while many cross-cultural studies have employed first languages (52, 53, 55, 56), not all have.

It is important to highlight that, while foreign language effects have only been tested using a limited number of instruments (80), the mechanisms thought to underly foreign language effects—diminishing the salience of normative decision heuristics (28, 32, 102), and dampening the emotive content of study measures (25, 103)—are context general. Here, we find no strong evidence of a foreign language effect on behavior in giving games, and so this important literature (45, 46, 55, 57, 59, 84, 85) at least might not be systematically confounded by failure to standardize study language. Still, cross-cultural researchers should assume, unless demonstrated otherwise, that study language may substantively influence results. Gladly, the solution is straightforward: unless researchers are specifically interested in behavior in non-native-language contexts, where possible, methods should be presented in a first language, especially in comparative work. Where unfeasible, the potential influence of study language should be clearly highlighted as a study limitation.

There has been some previous debate concerning the extent to which fluency diminishes foreign language effects (63, 79, 80). Here we find mixed evidence for an interaction. In each of the three experiments where there was a foreign language effect, we found at least some trending evidence that decreasing language fluency accentuates the main study effect. However, in only one case (Likert risk assessment questions) did our interaction models improve model fit. We suspect that the interaction effect of fluency may be real, but weak, which would account for the variability of evidence for this phenomenon both here and elsewhere (79).

Our last key finding is that in both risk tasks, effects were not in the expected direction. Unlike nearly all other FLE studies (63, 79), presenting methods in a non-native language decreased risk tolerance. This might be an outcome of linguistic relativism—the capacity of different languages to map semantic content (or even numerical reasoning) differently (11). However, we consider an alternative scenario: It has been proposed that first languages might be the primary carriers "of emotions and socio-moral norms which in turn govern judgments and choices" ((102), p. 253). The present evidence suggests instead that non-native languages, rather than universally reducing the salience of decision rules, might sometimes carry *different* norms or heuristic content. This may be because people adopt the heuristics or cognitive styles of the culture associated with the language they are speaking (see (92, 93)). Independent of normative content, it might also reflect the context in which that language is spoken. For many Hadza, Kiswahili is spoken in contexts—market interactions, student–teacher interactions, medical contexts and interactions with proselytizers—which engender greater risks than those where Hadzane typically is spoken. This could plausibly account for the reversal in the direction of the effect.

Although this explanation is compelling, data from other contexts are needed to test this hypothesis further. It would be enlightening to investigate whether the same effect is observed in other populations where, e.g. school, healthcare, and government interactions are typically conducted in a non-native language. Our study highlights a clear need for further research exploring foreign language effects in a greater variety of cross-cultural comparative contexts, beyond the wealthy, industrialized, university samples that represent most existing literature (80).

In the meantime, it is sufficient to highlight that foreign language effects persist outside of Western, industrialized societies, and may influence methodologies that have been widely used. Thus, there is clear cause for cross-cultural researchers to better account for the demonstrable influence of language on behavior when designing, conducting and reporting cross-cultural behavioral research.

Methods

Study population, demographics, and sampling strategy

The Hadza are an ethnolinguistic group in the Northern Tanzania, numbering perhaps 1,000 people. They have traditionally subsisted as hunter-gatherers, and many continue to, though today most people augment foraging incomes with horticulture, apiculture, trade, and/or engagement in Tanzania's large tourist economy. Identity is normally defined by fluency in Hadzane, the Hadza language (8). Hadzane is a click language (104, 105), a phonetic property it shares with Khoe and Sandawe (106), although any phylogenetic links are disputed and Hadzane is probably a linguistic isolate (104). In the 1950s, most Hadza spoke Isanzu, the Bantu language of their neighbors (107) while, today almost all Hadza speak Kiswahili, the Arabic-influenced Bantu *lingua franca* of East Africa, to varying degrees of fluency. For some Hadza, Kiswahili fluency is attained in school, but usually it is learned informally through interactions with members of other groups. Both formal education and interactions with Kiswahili speakers have increased in recent decades, as the region has become more accessible by road.

Data were collected between September and December of 2023. We recruited 129 participants, 66 women, 63 men, aged 15–75, from five different camps/settlements. These were chosen to represent a roughly balanced sample of individuals from ephemeral bush-camps, settled villages, semisedentary village-adjacent foraging camps and tourist-economy camps. To maximize geographical diversity in the sample, we selected camps/settlements in five different locations, in an arc from the southern Yaeda valley to the eastern region of Mangola. To locate ephemeral camps, we relied on knowledge from local informants and participants.

Analyses and modeling strategy

All statistical analyses were conducted in the R programming language (108), using the Bayesian regression models package (109) and STAN (110). We employ Bayesian methods because, although computationally expensive and still less widely used and understood than their frequentist analogs, they allow for more straightforward quantification, description, and visualization of statistical uncertainty. To ease comprehension for readers unfamiliar with Bayesian methods, we have reported (fitted) results

in their natural units, such that they may be interpreted straightforwardly.

Binary response data were analyzed using binomial (Bernoulli) regression. Both Likert data and giving/wager game outcomes were analyzed using cumulative models which capture bounded yet ordered data where not every step is similarly meaningful (111). We employed wide but weakly regularizing priors, penalizing extreme values, as these typically result in more accurate out-of-sample estimates (112).

We added measures in order of their theoretic importance, first a mean-only (null) model (including outcome grouping variables where relevant), then adding language of instruction as a predictor, then gender. We next explored both self-reported language of thought, and self-reported language fluency as interaction effects on language of instruction. We centered both interaction variables on zero to prevent directional biases in our estimates. For model comparison we used leave-one-out model selection, which trades-off improvement in fit against the overfitting risk inherent in adding additional parameters (for a useful primer on model selection, please see Chapter 7 of Statistical Rethinking (112).

Where we conducted repeat measures using the same task (i.e. sacrificial dilemmas; perceived-risk Likert measures), rather than running individual models, we included all measures in the same model with partial pooling, setting varying effects on both intercepts and slopes. To account for camp-level effects, where it improved model fit, we included residential camp as a grouping factor variable with varying effects on the intercept.

Model results are reported as outcome-scale fitted estimates in text to allow easier and more intuitive interpretation of findings. Analyses scripts and precomputed model fits are provided in the accompanying open data archive (https://osf.io/8ec3s/?view_only=543b67ecd87c43279fcfbcb99368624a).

Translation procedure and study design

Study rubric for each method was translated from English to Kiswahili and Hadzane by study authors I.M. (fluent in English and Kiswahili) and E.E. (fluent in English, Kiswahili, and Hadzane) respectively. Both have extensive experience in translation and both have completed accredited post-graduate degrees at master's level. E.E. has conducted paid work for multiple internationally prestigious linguistics departments and is among the most qualified Hadza-English translators living. As no Hadza-English back-translators were available, after initial translations were complete, E.E. compared the Hadzane and the Kiswahili wording, and minorly amended the Kiswahili rubric to ensure parity with Hadzane. We then back-translated the revised Kiswahili rubric into English. The resulting English back-translation was deemed sufficiently close to require no further amendment.

Research was administered by E.E. with supervision from D.S.H. To minimize the risk of errors, we presented each method in the same order, first moral dilemma vignettes, then giving games, then wager games, then Likert measures. The ordering of the two moral dilemmas was randomized by coin-flip. To prevent confusion, we opted for a between-participant design and all tasks were presented to each participant in either Hadzane or Kiswahili, determined by coin-flip.

Moral dilemma vignettes

We presented each participant with two sacrificial dilemma vignettes, randomly varying order via coin-flip. These were both adapted "trolley problems" (113) where, after being presented with a short vignette, participants were given the binary choice to actively sacrifice one person to save five by diverting the trajectory of a moving object. Trolley-carts are unfamiliar to many Hadza, and we instead chose a recognizable naturally occurring scenario—a falling boulder—which shared the same intuitive physical properties (momentum, high mass) as a moving trolley-car. The first was a footbridge dilemma (push a person) the second a switch dilemma (flip a switch). As railroad switches are unfamiliar, we substituted the switch in these stories with a log, which similarly diverted the boulder's trajectory. We varied game language between participants by coin-flip; wording of both vignettes in Kiswahili and Hadzane are provided in Section 1.1 alongside an English translation. To aid comprehension, simple pictographic visualizations were provided alongside each story (Section 2) in laminated A5.

One-player giving game

Each participant played a one-player giving game (i.e. a "dictator game"). Each was presented with ten 120 g cups of maize, and received ten plastic tokens representing each cup. Participants were given the choice to either donate each cup to a community of Hadza living in another camp, or keep it for themselves. Allocations were made by placing tokens one-by-one into one of two cups representing either the self or the community. We systematically varied left and right cups between games. We also randomly varied the language of study rubric between participants by coin-flip. Wording in Kiswahili and Hadzane, alongside English translation, are provided in Section 1.2.

Wager game

To assess risk tolerance, we conducted a straightforward triple-or-nothing coin-flip wager game, where the probabilistically "rational" strategy was to wager everything. The rules were similar to lukuchuko, a traditional Hadza betting game where individuals flip two-sided bark discs and wager arrows or nails on the outcome (96). We presented each participant with three 120 g cups of maize, which we placed in front of them. We instructed each participant that they could wager any number of cups, and we would flip a coin. If heads, they would keep everything and win two additional cups for each cup wagered; if tails they would win nothing and lose any cups wagered. We randomly varied language of study rubric between participants by coin-flip. Study rubric in both languages, and an English translation, is provided in Section 1.3.

Perceived-risk Likert measures

To further assess risk tolerance, we asked participants a series of three-choice questions, assessing risk of harm to a second party—a friend—in five different scenarios. Scenarios were each chosen to be familiar and salient to participants, and to represent genuine risks that community members might encounter. These were perceived risks to a friend of (i) "drinking every day", (ii) "smoking tobacco every day", (iii) "catching covid", (iv) "climbing to the top of a tall tree every day to get honey", (v) "often traveling out of camp at night alone in the dark". Participants were asked to choose between three responses: "very worried"; "a little worried"; "not worried". The first three scenarios are known to be broadly salient across communities and contexts. The latter two are particular to the Hadza, although are known to be salient both from discussions with community members and from previous ethnographic accounts (see, e.g. (8, 114)). Study language was randomly varied

between participants. English, Kiswahili, and Hadzane rubric is provided in Section 1.4.

General questionnaire

Last, we conducted a short survey in the participants' first language, gathering basic demographic information (age, gender), and a list of languages spoken, self-reported Kiswahili fluency ("I speak it like a native speaker"; "I can hold a conversation"; "I get by, but do not always understand"; "I often do not understand", "My Kiswahili is limited"), how often participants "thought" in Kiswahili ("Often", "Occasionally", "Never"), and attitudes towards Kiswahili speakers ("I like them", "I have no strong feelings about them", "I do not like them").

Ethics, consent, and participant remuneration

Fieldwork was approved by the Tanzanian Commission for Science and Technology (#2022-879-NA-2022-229), and the Ethics Committee of the Durham University Anthropology department (#ANTH-2021-01-26T16_11_21-vpft62).

Letters of permission were attained from the regional government offices of the Karatu and Manyara regions. Local research permits were also attained from the Dorobo Fund and the Mangola Village District Council. Informed consent to conduct research was attained both collectively at the camp level, and individually from each study participant. Parental consent was attained for those under the age of 18. Participants were remunerated with gifts of clothing (shawls, Maasai blankets, shirts, & shorts) and grain (maize and sembe), alongside those smaller awards provided in the two incentivized games.

Acknowledgments

We would like to extend our sincere thanks to those who have participated in this research. Thanks to Kris Smith for his advice in selecting appropriate priors and modeling strategies for our analysis. Thanks to Boaz Keysar for early conversations and the suggestion to expand sample diversity in this research area. Thanks also to Catherine Caldwell-Harris and one other anonymous reviewer for their knowledgable, thoughtful, constructive commentaries.

Funding

This research was part of the Culture of Schooling Project funded by Templeton Trust and the Issachar Fund (Grant title: "The Consequences of Formal Education for Science and Religion"). Additional support in the preparation of the final manuscript was provided by The Leverhulme Trust (Award Number: RF170291)

Author Contributions

C.A. and D.S.H. conceived the study; D.S.H. and C.A. designed the experiments, with input and revision from L.A.; D.S.H., I.M., and E.E. conducted the experiments; D.S.H. conducted statistical analyses; D.S.H. wrote and revised the manuscript with input, review and commentary from C.A., L.A., and C.L.

Data Availability

Supplementary materials, including all anonymized study data and annotated R scripts, are available at https://osf.io/8ec3s/?view_only=492337108e8c4ccc938a12b937fee499.

References

1 Marlowe FW. 2005. Dictators and ultimatums in an egalitarian society of hunter-gatherers: the Hadza of Tanzania. In: Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, editors. Foundations of human sociality: economic experiments and ethnographic evidence from fifteen small-scale societies. Oxford: Oxford University Press. p. 167–192.
2 Pawley A, Ross M. 1993. Austronesian historical linguistics and culture history. *Annu Rev Anthropol*. 22(1):425–459.
3 Bahuchet S. 2012. Changing language, remaining Pygmy. *Hum Biol*. 84(1):11–43.
4 Cristia A, Farabolini G, Scaff C, Havron N, Stieglitz J. 2020. Infant-directed input and literacy effects on phonological processing: non-word repetition scores among the Tsimane'. *PLoS One*. 15(9):e0237702.
5 Güldemann T. 1998. *San languages for education: a linguistic short survey and proposal*. Okahandja, Namibia: NIED, Ministry of Basic Education and Culture.
6 Güldemann T, McConvell P, Rhodes RA. 2020. Hunter-gatherer anthropology and language. In: Güldemann T, McConvell P, Rhodes RA, editors. *The language of hunter-gatherers*. Cambridge: Cambridge University Press. p. 3–48.
7 Hauck JD. 2018. The origin of language among the Aché. *Lang Commun*. 63:76–88.
8 Marlowe FW. 2010. *The Hadza: hunter-gatherers of Tanzania*. Los Angeles: University of California Press.
9 Reid LA. 1994. Unravelling the linguistic histories of Philippine Negritos. In: Tryon DT, editor. Language contact and change in the austronesian world. Berlin: Mouton de Gruyter. p. 443–475.
10 Gleitman L, Papafragou A. 2005. Language and thought. In: Holyoak KJ, Morrison RG, editors. The Cambridge handbook of thinking and reasoning. Cambridge: Cambridge University Press. p. 633–661.
11 Kay P, Kempton W. 1984. What is the Sapir-Whorf hypothesis? *Am Anthropol*. 86(1):65–79.
12 Whorf BL. 2012. The relation of habitual thought and behavior to language. In: Caroll JB, editor. Language, thought, and reality: selected writings of Benjamin Lee Whorf. Cambridge, MA: MIT Press.
13 Lindsey DT, Brown AM, Brainard DH, Apicella CL. 2016. Hadza color terms are sparse, diverse, and distributed, and presage the universal color categories found in other world languages. *Iperception*. 7(6):1–6.
14 Regier T, Kay P. 2009. Language, thought, and color: Whorf was half right. *Trends Cogn Sci*. 13(10):439–446.
15 Guttfreund DG. 1990. Effects of language usage on the emotional experience of Spanish-English and English-Spanish bilinguals. *J Consult Clin Psychol*. 58(5):604–607.
16 Perlovsky L. 2009. Language and emotions: emotional Sapir–Whorf hypothesis. *Neural Netw*. 22(5-6):518–526.
17 Boroditsky L. 2011. How languages construct time. In: Dehaene S, Brannon EM, editors. *Space, time and number in the brain*. London: Academic Press. p. 333–341.
18 Casasanto D, *et al.* 2004. How deep are effects of language on thought? Time estimation in speakers of English, Indonesian, Greek, and Spanish. *Proc Annu Meeting Cogn Sci Soc*. 26(6): 186–191.
19 Ji L-J, Zhang Z, Nisbett RE. 2004. Is it culture or is it language? Examination of language effects in cross-cultural research on categorization. *J Pers Soc Psychol*. 87(1):57–65.

20 Haun DBM, Rapold CJ, Janzen G, Levinson SC. 2011. Plasticity of human spatial cognition: spatial language and cognition covary across cultures. *Cognition*. 119(1):70–80.
21 Levinson SC, Kita S, Haun DBM, Rasch BH. 2002. Returning the tables: language affects spatial reasoning. *Cognition*. 84(2): 155–188.
22 Li P, Abarbanell L. 2019. Alternative spin on phylogenetically inherited spatial reference frames. *Cognition*. 191:103983.
23 Li P, Abarbanell L, Gleitman L, Papafragou A. 2011. Spatial reasoning in tenejapan mayans. *Cognition*. 120(1):33–53.
24 Hayakawa S, Keysar B. 2018. Using a foreign language reduces mental imagery. *Cognition*. 173:8–15.
25 Keysar B, Hayakawa SL, An SG. 2012. The foreign-language effect: thinking in a foreign tongue reduces decision biases. *Psychol Sci*. 23(6):661–668.
26 Cipolletti H, McFarlane S, Weissglass C. 2016. The moral foreign-language effect. *Philos Psychol*. 29(1):23–40.
27 Costa A, *et al.* 2014. Your morals depend on language. *PLoS One*. 9(4):e94842.
28 Hennig M, Hütter M. 2021. Consequences, norms, or willingness to interfere: a proCNI model analysis of the foreign language effect in moral dilemma judgment. *J Exp Soc Psychol*. 95:104148.
29 Costa A, Foucart A, Arnon I, Aparici M, Apesteguia J. 2014. "Piensa" twice: on the foreign language effect in decision making. *Cognition*. 130(2):236–254.
30 Białek M, Paruzel-Czachura M, Gawronski B. 2019. Foreign language effects on moral dilemma judgments: an analysis using the CNI model. *J Exp Soc Psychol*. 85:103855.
31 Brouwer S. 2019. The auditory foreign-language effect of moral decision making in highly proficient bilinguals. *J Multiling Multicult Dev*. 40(10):865–878.
32 Geipel J, Hadjichristidis C, Surian L. 2015. The foreign language effect on moral judgment: the role of emotions and norms. *PLoS One*. 10(7):1–17.
33 Urbig D, Terjesen S, Procher V, Muehlfeld K, Van Witteloostuijn A. 2016. Come on and take a free ride: contributing to public goods in native and foreign language settings. *Acad Manag Learn Edu*. 15(2):268–286.
34 Gao S, Zika O, Rogers RD, Thierry G. 2015. Second language feedback abolishes the "hot hand" effect during even-probability gambling. *J Neurosci*. 35(15):5983–5989.
35 Harris CL, Gleason JB, Ayçiçeği A. 2006. When is a first language more emotional? Psychophysiological evidence from bilingual speakers. In: Pavlenko A, editor. Bilingual minds. Bristol, UK: Multilingual Matters. p. 257–283.
36 Harris CL. 2004. Bilingual speakers in the lab: psychophysiological measures of emotional reactivity. *J Multiling Multicult Dev*. 25(2-3):223–247.
37 Harris CL, Ayçiçeği A, Gleason JB. 2003. Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Appl Psycholinguist*. 24(4):561–579.
38 Caldwell-Harris CL. 2014. Emotionality differences between a native and foreign language: theoretical implications. *Front Psychol*. 5:1–4.
39 Hadjichristidis C, Geipel J, Savadori L. 2015. The effect of foreign language in judgments of risk and benefit: the role of affect. *J Exp Psychol*. 21(2):117–129.
40 Evans JSBT, Stanovich KE. 2013. Dual-process theories of higher cognition: advancing the debate. *Perspect Psychol Sci*. 8(3): 223–241.
41 Frankish K. 2010. Dual-process and dual-system theories of reasoning: dual-process and dual-system theories of reasoning. *Philos Compass*. 5(10):914–926.
42 Thierry G, Wu YJ. 2007. Brain potentials reveal unconscious translation during foreign-language comprehension. *Proc Natl Acad Sci USA*. 104(30):12530–12535.
43 Costa A, Vives M-L, Corey JD. 2017. On language processing shaping decision making. *Curr Dir Psychol Sci*. 26(2):146–151.
44 Henrich J, *et al.* 2005. "Economic man" in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *Behav Brain Sci*. 28(6):795–815. discussion 815–855.
45 Smith D, *et al.* 2019. A friend in need is a friend indeed: need-based sharing, rather than cooperative assortment, predicts experimental resource transfers among Agta hunter-gatherers. *Evol Hum Behav*. 40(1):82–89.
46 Smith KM, Mabulla IA, Apicella CL. 2022. Hadza hunter-gatherers with greater exposure to other cultures share more with generous campmates. *Biol Lett*. 18(7):20220157.
47 Wiessner P. 2009. Experimental games and games of life among the Ju/'hoan Bushmen. *Curr Anthropol*. 50(1):133–138.
48 Amir D, *et al.* 2020. The developmental origins of risk and time preferences across diverse societies. *J Exp Psychol Gen*. 149(4): 650–661.
49 Apicella CL, Crittenden AN, Tobolsky VA. 2017. Evolution and human behavior hunter-gatherer males are more risk-seeking than females, even in late childhood. *Evol Hum Behav*. 38(5): 592–603.
50 Gray PB. 2004. Evolutionary and cross-cultural perspectives on gambling. *J Gambl Stud*. 20(4):347–371.
51 Barrett HC, *et al.* 2016. Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proc Natl Acad Sci USA*. 113(17):4688–4693.
52 Fessler DMT, *et al.* 2015. Moral parochialism and contextual contingency across seven societies. *Proc R Soc B Biol Sci*. 282(1813):20150907.
53 Smith KM, Apicella CL. 2022. Hadza hunter-gatherers are not deontologists and do not prefer deontologists as social partners. *J Exp Soc Psychol*. 101:104314.
54 Hayakawa S, Costa A, Foucart A, Keysar B. 2016. Using a foreign language changes our choices. *Trends Cogn Sci (Regul Ed)*. 20(11): 791–793.
55 Henrich J, *et al.* 2006. Costly punishment across human societies. *Science*. 312(5781):1767–1770.
56 Purzycki BG, *et al.* 2018. The cognitive and cultural foundations of moral behavior. *Evol Hum Behav*. 39(5):490–501.
57 Stagnaro MN, Stibbard-Hawkes DNE, Apicella CL. 2022. Do religious and market-based institutions promote cooperation in Hadza hunter-gatherers? *Religion Brain Behav*. 12(1-2):171–189.
58 Stibbard-Hawkes DNE, Amir D, Apicella CL. 2023. A cost for signaling: do Hadza hunter-gatherers forgo calories to show-off in an experimental context? *Evol Hum Behav*. 44(5):398–410.
59 Henrich J. 2000. Does culture matter in economic behavior? Ultimatum game bargaining among the Machiguenga of the Peruvian Amazon. *Am Econ Rev*. 90(4):973–979.
60 Stibbard-Hawkes DNE, Smith KM, Apicella CL. 2022. Why hunt? Why gather? Why share? Hadza assessments of foraging and food-sharing motive. *Evol Hum Behav*. 43(3):257–272.
61 Lew-Levy S, *et al.* 2019. Inter- and intra-cultural variation in learning-through-participation among Hadza and BaYaka forager children and adolescents from Tanzania and the Republic of Congo. *J Psychol Africa*. 29(4):309–318.
62 Costa A, Duñabeitia JA, Keysar B. 2019. Language context and decision-making: challenges and advances. *Q J Exp Psychol*. 72(1):1–2.

63 Stankovic M, Biedermann B, Hamamura T. 2022. Not all bilinguals are the same: a meta-analysis of the moral foreign language effect. *Brain Lang*. 227:105082.
64 Henrich J, Heine SJ, Norenzayan A. 2010. The weirdest people in the world. *Behav Brain Sci*. 33(2-3):61–83.
65 Hays J, Ninkova V, Dounias E. 2019. Hunter-gatherers and education: towards a recognition of extreme local diversity and common global challenges. *Hunter Gatherer Res*. 5(1-2):13–38.
66 Apicella CL. 2013. Upper-body strength predicts hunting reputation and reproductive success in Hadza hunter-gatherers. *Evol Hum Behav*. 35(6):508–518.
67 Berbesque JC, Marlowe FW. 2009. Sex differences in food preferences of Hadza hunter-gatherers. *Evol Psychol*. 7(4):601–616.
68 Jones NGB. 1984. A selfish origin for human food sharing: tolerated theft. *Ethol Sociobiol*. 5(1):1–3.
69 Crittenden AN, Conklin-Brittain NL, Zes DA, Schoeninger MJ, Marlowe FW. 2013. Juvenile foraging among the Hadza: implications for human life history. *Evol Hum Behav*. 34(4):299–304.
70 Stibbard-Hawkes DNE, Attenborough RD, Mabulla IA, Marlowe FW. 2020. To the hunter go the spoils? No evidence of nutritional benefit to being or marrying a well-reputed Hadza hunter *Am J Phys Anthropol*. 173(1):61–79.
71 Stibbard-Hawkes DNE, Attenborough RD, Marlowe FW. 2018. A noisy signal: to what extent are Hadza hunting reputations predictive of actual hunting skills? *Evol Hum Behav*. 39(6):639–651.
72 Wood BM. 2006. Prestige or provisioning? A test of foraging goals among the Hadza. *Curr Anthropol*. 47(2):383–387.
73 Sands BE. 1998. *Eastern and Southern African Khoisan: evaluating claims of distant linguistic relationships*. Vol. 14. Cologne: Rüdiger Köppe.
74 Smith KM, Larroucau T, Mabulla IA, Apicella CL. 2018. Hunter-gatherers maintain assortativity in cooperation despite high levels of residential change and mixing. *Curr Biol*. 28(19): 3152–3157.e4.
75 Hayakawa S, Lau BKY, Holtzmann S, Costa A, Keysar B. 2019. On the reliability of the foreign language effect on risk-taking. *Q J Exp Psychol*. 72(1):29–40.
76 Hayakawa S, Tannenbaum D, Costa A, Corey JD, Keysar B. 2017. Thinking more or feeling less? explaining the foreign-language effect on moral judgment. *Psychol Sci*. 28(10):1387–1397.
77 de Bruin A. 2019. Not all Bilinguals are the same: a call for more detailed assessments and descriptions of bilingual experiences. *Behav Sci*. 9(3):33.
78 Miozzo M, *et al*. 2020. Foreign language effect in decision-making: how foreign is it? *Cognition*. 199:104245.
79 Circi R, Gatti D, Russo V, Vecchi T. 2021. The foreign language effect on decision-making: a meta-analysis. *Psychon Bull Rev*. 28(4):1131–1141.
80 Del Maschio N, Crespi F, Peressotti F, Abutalebi J, Sulpizio S. 2022. Decision-making depends on language: a meta-analysis of the foreign language effect. *Bilingualism Lang Cogn*. 25(4): 617–630.
81 Awad E, Dsouza S, Shariff A, Rahwan I, Bonnefon J-F. 2020. Universals and variations in moral decisions made in 42 countries by 70,000 participants. *Proc Natl Acad Sci USA*. 117(5): 2332–2337.
82 Eriksson K, *et al*. 2021. Perceptions of the appropriate response to norm violation in 57 societies. *Nat Commun*. 12(1):1481.
83 Li KK. 2017. How does language affect decision-making in social interactions and decision biases? *J Econ Psychol*. 61:15–28.
84 Gurven M. 2004. Economic games among the Amazonian Tsimane: exploring the roles of market access, costs of giving, and cooperation on pro-social game behavior. *Exp Econ*. 7(1): 5–24.
85 Hill K, Gurven M. 2004. Economic experiments to examine fairness and cooperation among the ache Indians of paraguay. In: Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, editors. Foundations of human sociality. Oxford: Oxford University Press. p. 382–410.
86 Bird DW, Bird RB. 2002. Risk and reciprocity in Meriam food sharing. *Evol Hum Behav*. 23(4):297–321.
87 Codding BF, Bird RB, Bird DW. 2011. Provisioning offspring and others: risk-energy trade-offs and gender differences in hunter-gatherer foraging strategies. *Proc R Soc B Biol Sci*. 278(1717): 2502–2509.
88 Hawkes K, O'Connell J, Blurton Jones N. 2018. Hunter-gatherer studies and human evolution: a very selective review. *Am J Phys Anthropol*. 165(4):777–800.
89 Stibbard-Hawkes DNE. 2019. Costly signaling and the handicap principle in hunter-gatherer research: a critical review. *Evol Anthropol*. 28(3):144–157.
90 Charness G, Gneezy U. 2012. Strong evidence for gender differences in risk taking. *J Econ Behav Organ*. 83(1):50–58.
91 Dylman AS, Champoux-Larsson M-F. 2020. It's (not) all Greek to me: boundaries of the foreign language effect. *Cognition*. 196: 104148.
92 Chen SX, Bond MH. 2010. Two languages, two personalities? examining language effects on the expression of personality in a bilingual context. *Pers Soc Psychol Bull*. 36(11):1514–1528.
93 Kashima Y, Kashima E, Kidd E. 2014. Language and culture. In: The Oxford handbook of language and social psychology, Oxford Library of Psychology. New York (NY), USA: Oxford University Press. p. 46–61.
94 McCrae RR, Yik MSM, Trapnell PD, Bond MH, Paulhus DL. 1998. Interpreting personality profiles across cultures: bilingual, acculturation, and peer rating studies of Chinese undergraduates. *J Pers Soc Psychol*. 74(4):1041–1055.
95 Hawkes K, O'Connell JF, Blurton Jones NG. 1991. Hunting income patterns among the Hadza: big game, common goods, foraging goals and the evolution of the human diet. *Philos Trans R Soc B Biol Sci*. 334(1270):243–251. discussion 250–251.
96 Woodburn J. 1970. *Hunters and gatherers: the material culture of the nomadic Hadza*. London: British Museum.
97 Reno RR, Cialdini RB, Kallgren CA. 1993. The transsituational influence of social norms. *J Pers Soc Psychol*. 64(1):104–112.
98 Lew-Levy S, Reckin R, Lavi N, Cristóbal-Azkarate J, Ellis-Davies K. 2017. How do hunter-gatherer children learn subsistence skills?: A meta-ethnographic review. *Hum Nat*. 28(4):367–394.
99 Apicella CL. 2018. High levels of rule-bending in a minimally religious and largely egalitarian forager population. *Religion Brain Behav*. 8(2):133–148.
100 McDowell W. 1984. A brief history of the Mangola Hadza. Technical Report, Mbulu District Development Directorate, Mbulu District, Arusha Region.
101 Ndagala DK. 1985. Attempts to develop the Hadzabe of Tanzania. *Nomad People*. 18(18):17–26.
102 Hadjichristidis C, Geipel J, Keysar B. 2019. Chapter 11 - The influence of native language in shaping judgment and choice. In: Srinivasan N, editor. *Progress in brain research. Emotion and cognition*. Vol. 247. Cambridge, MA: Elsevier. p. 253–272.
103 Caldwell-Harris CL. 2015. Emotionality differences between a native and foreign language: implications for everyday life. *Curr Dir Psychol Sci*. 24(3):214–219.
104 Sands B, Güldemann T. 2009. What click languages can and can't tell us about language origins. In: Botha R, Knight C,

editors. The cradle of language. Vol 12. Oxford: Oxford University Press. p. 204–218.
105 Sands B, Maddieson I, Ladefoged P. 1996. The phonetic structures of Hadza. *Stud African Linguist*. 25(2):171–204.
106 Güldemann T, Stoneking M. 2008. A historical appraisal of clicks: a linguistic and genetic population perspective. *Annu Rev Anthropol*. 37(1):93–109.
107 Woodburn J. 1964. The social organisation of the Hadza of North Tanganyika. [PhD thesis]. University of Cambridge.
108 R Core Team. 2023. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/
109 Burkner P-C. 2017. BRMs: an R package for Bayesian multilevel models using stan. *J Stat Softw*. 80(1):1–28.
110 Stan Development Team. 2023. Stan Modeling Language Users Guide and Reference Manual, 2.3.1.
111 Bürkner P-C, Vuorre M. 2019. Ordinal regression models in psychology: a tutorial. *Adv Meth Pract Psychol Sci*. 2(1): 77–101.
112 McElreath R. 2016. *Statistical rethinking: a Bayesian course with examples in R and Stan*. London: Chapman and Hall/CRC.
113 Thomson JJ. 1985. The trolley problem. *Yale Law J*. 94(6):1395.
114 Crittenden AN. 2011. The importance of honey consumption in human evolution. *Food Foodways*. 19(4):257–273.