

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

Do international students learn foreign preferences? The interplay of language, identity and assimilation

Paul Clist^{a,*}, Ying-yi Hong^b^a University of East Anglia, UK^b The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

Dataset link: <https://doi.org/10.7910/DVN/Z9CANV>

JEL classification:

Z13

F22

J15

Keywords:

Preferences

Assimilation

Language

Migration

ABSTRACT

Every year millions of students study at foreign universities, swapping one set of cultural surroundings for another. This may reveal whether measured preferences are fixed or flexible, whether they can be altered in the short run by moving country, or learning a new language. We disentangle these influences by measuring international students' preferences. For Chinese students in the UK (who arrived up to five years previously) we randomise a survey's language. We add reference groups in each country, doing the survey in the relevant language. Simple comparisons provide a causal estimate of language's effect and observational estimates of differences by country, location and nationality. We find language has a large causal effect on a range of survey responses. The effect size is similar to differences by country or nationality (at 0.4 standard deviations), and larger than differences by location (at 0.1 standard deviations). Assimilation theories predict any movement in measured preferences for Chinese students in the UK would be towards those of UK students, even if they may be small. We do not find this. In Mandarin, Chinese students hardly differ from those in Beijing. Yet in English, they are not close to either Chinese students in Beijing or British students in the UK. This can be explained by a model of identity priming with monocultural subjects. For Chinese students in the UK, speaking English reduces the pull of a Chinese frame without increasing the pull of a British one. International students do not so much learn foreign preferences as learn to ignore old ones. Our reliance on mostly stated preferences enables a rich dataset covering many domains; future work is needed to see if such large effects are also found for a wide range of revealed preferences.

1. Introduction

From trust to time, economic preferences differ between countries. A rich literature seeks to explain such differences, examining various elements of biology, geography and culture (Falk et al., 2018; Gächter & Schulz, 2016; Galor & Özak, 2016; Henrich et al., 2010; Rieger, Wang, & Hens, 2014). These are natural comparisons between people who differ in their location, language and nationality. Yet whilst these comparisons are static, people are not. They move countries, learn languages and even change nationality. International students are an interesting case. In 2017, there were 5.3 million international students, of whom a little under a million came from mainland China (UNESCO Institute for Statistics, 2020).¹ International students are relatively short-term

* Corresponding author.

E-mail address: paul.clist@uea.ac.uk (P. Clist).URL: <http://paulclist.github.io> (P. Clist).¹ For context, UNESCO Institute for Statistics (2020) report that in 2017 there were 25.5 million students in China studying a bachelors or higher.<https://doi.org/10.1016/j.joep.2023.102658>

Received 8 September 2021; Received in revised form 13 June 2023; Accepted 28 July 2023

Available online 2 August 2023

0167-4870/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

migrants, but they are also receptive (Cameron, Erkal, Gangadharan, & Zhang, 2015). They travel for thousands of miles with the intention of learning, suggesting they are a useful group to understand the short-run dynamics of cultural assimilation.

In this paper, we ask how international students' measured preferences differ by location, language and nationality. Does moving country close the country-level gap; is speaking another language sufficient; or is there an irreducible difference by nationality? Comparing the magnitude of differences gives each one greater context. For example whilst several experimental papers have found statistically significant differences in preferences by language (e.g. Clist & Verschoor, 2017; Lambarraa & Riener, 2015; Li, 2017), it is not clear if they are large compared to cross-country effects.

The direction and correlation of differences by country, location, language and nationality help us to test relevant theories. *Slow Assimilation* implies that recent migrants, like those we study, should have relatively small differences in preferences from their sending country, with any changes in the direction of the host country (Abramitzky, Boustan, & Eriksson, 2020; Algan, Bisin, & Verdier, 2012; Bisin & Verdier, 2011; Cameron et al., 2015). As such, the international students' preferences should become more like those of the UK host group, whereas the language used should not matter much. By contrast, with bicultural subjects *Identity Priming* implies that moving to the UK or speaking English could make a British identity more salient relative to a Chinese one, and so predict larger differences by location and/or language (Akerlof & Kranton, 2000; Benjamin, Choi, & Fisher, 2016; Benjamin, Choi, & Strickland, 2010; Chakravarty & Fonseca, 2014; Hong, Morris, Chiu, & Benet-Martinez, 2000). However, in our data the average length of stay in the UK is a little over a year, so subjects are unlikely to be bicultural; they would still be monocultural subjects, without an internalised British cultural frame or British identity. For them, moving to the UK or speaking English are primes which could reduce the pull of a Chinese identity without increasing the pull of a British one. Compared to *Slow Assimilation*, *Identity Priming* in our setting expects smaller differences by country and greater differences by location and/or language, but makes no predictions about the direction of these differences.

At the core of our between-subject experiment are Chinese students studying in the UK, for whom we randomise the language of a survey. A simple comparison between those two groups gives a causal estimate of the (English/Mandarin) language effect: does language change elicited preferences? We add reference groups in Beijing and the UK: students in each place doing the survey in the relevant language. Together these four groups allow observational estimates of how survey responses differ by country (comparing Beijing and UK groups), location (comparing the Beijing group to Chinese students in the UK speaking Mandarin) and nationality (comparing the UK group to Chinese students in the UK speaking English). These effects are not clean causal estimates as we do not randomly assign these elements. Rather they describe the size and direction of differences, in exactly the same way as previous cross-country research. Together these four effects are sufficient to test relevant theories, whilst also giving greater context to each effect.

On preferences, we cover several domains. The survey comprises 19 outcome questions covering eight topics. We use an incentivised die-roll to measure lying, and elicit stated preferences regarding cheating, patience, negative reciprocity, altruism, positive reciprocity, trust and ethics. This means each effect (country, location, language and nationality) can be estimated 19 times, using over 6000 observations in total. Previous work on language effects has tended to focus on one revealed preference (e.g. Angerer, Glätzle-Rützler, Lergetporer, & Sutter, 2021; Clist & Verschoor, 2017; Lambarraa & Riener, 2015; Sutter, Angerer, Glätzle-Rützler, & Lergetporer, 2018), whereas we use a wide range of mainly stated preferences.

Looking at the average magnitude of each effect we find randomising the language of the survey causes differences of 0.40 standard deviations. This is similar in size to the standard cross-country comparison (0.39) and the nationality effect (0.42), whilst the difference by location is much smaller (0.12). These effects are often significant, with half of the questions finding a significant language effect at the 5% level, after controlling the family-wise error rate (Jones, Molitor, & Reif, 2019; Westfall & Young, 1993). Effects of interest are on average much larger than controls, with the language effect around 3 times larger than gender differences. In short, language has a significant and meaningful effect on measured preferences. The relatively large language effect is perhaps surprising as it is causal, whilst other comparisons are observational estimates between people that differ in a variety of ways. The difference by location is small, despite being positively biased by any selection into migration caused by preference similarity. In the one or two years that most Chinese students spend in the UK, their elicited preferences are very similar to those elicited in China, at least in Mandarin.

Our results are consistent with *identity priming* with monocultural subjects, and inconsistent with *slow assimilation* in the few years we observe for our sample. (The average time in the UK in our sample is a little over a year, and ranges from 3 weeks to 5 years.) Whilst UK-based international students' responses are very close to those of Beijing-based students in Mandarin, they are not in English. When answering in English, Chinese students are not close to others in either Beijing or the UK. This is in contrast to assimilation theories, which predict new influences gradually replace old ones. Rather than small effects in the direction of UK-style preferences (as in Cameron et al., 2015), we find large differences in inconsistent directions. *Identity priming* models with monocultural subjects provide the best explanation of the evidence (Akerlof & Kranton, 2000; Benjamin et al., 2016, 2010; Hong et al., 2000). Whilst language activates an associated cultural frame for bicultural subjects, Chinese students in the UK do not have a British identity which speaking English can make more salient. The effect of speaking English is not to activate a new frame, but to reduce the pull of an old one. In a sense, English appears to free international students from answering according to their Chinese identity.

Our paper contributes to several literatures. There is persistent interest in documenting cross-country differences in preferences (Falk et al., 2018; Gächter & Schulz, 2016; Henrich et al., 2001, 2010, 2006; Hugh-Jones, 2016; Rieger et al., 2014; Roth, Prasnikar, Okuno-Fujiwara, & Zamir, 1991; Vieider et al., 2015). We put the size of such effects into context. While these are large, they are slightly smaller than simply randomising language for bilingual subjects. Explanations of cross-country differences should not be incompatible with large language differences. For example, explanations of cross-country differences that imply elicited

preferences are very hard to move are somewhat at odds with evidence that they differ so much simply by changing the language of the question.

Why does changing the language of a survey have such large effects? [Lambarraa and Riener \(2015\)](#), [Li \(2017\)](#) and [Clist and Verschoor \(2017\)](#) all find large effects from randomising the language of experiments for bilingual subjects, and argue they can be explained by different associations, norms or expectations. We contribute to this literature by measuring language effects for migrants as well as reference groups, i.e. groups of students in China and the UK. This allows us to compare effect sizes; we show the language effect is comparable in size to cross-country differences. It also allows us to see whether Chinese students in the UK move from Chinese-style preferences towards UK-style preferences as they move country and speak a different language. We do not find subjects merely copy their new reference group, as perhaps implied in other work. As these students were not in the UK for long enough to internalise a UK frame, speaking English seems to deactivate a Chinese frame without activating a UK one. Language can activate a related frame or identity ([Akerlof & Kranton, 2000](#); [Clist & Verschoor, 2017](#); [Fehr & Hoff, 2011](#); [Hong et al., 2000](#)), but only where it is already internalised.

Some of our findings echo slow assimilation. This research shows that moving country does not eliminate cross-country differences, which can persist for generations ([Abramitzky et al., 2020](#); [Bisin & Verdier, 2011](#); [Cameron et al., 2015](#); [Fernández, 2011](#)). This fits with the large nationality and small location effects in our sample. Indeed slow assimilation could still be an accurate description of what happens in the long run. However, the impression of gradual and unavoidable convergence in preferences is misleading for what occurs in the short run. If we only observed the responses of Chinese students in English, we would conclude that they diverge from Chinese-style preferences quickly. By measuring this group's preferences in Mandarin, we see their elicited preferences can look very similar to their counterparts in Beijing. This opens up the intriguing possibility that assimilation research would benefit from examining preferences in more than one language.

A popular competing reason why language could influence preferences relates to the idea that language features may make certain distinctions more salient and therefore influence preferences ([Angerer et al., 2021](#); [Chen, 2013](#); [Chen, He, & Riyanto, 2019](#); [Falk et al., 2018](#); [Galer, Ozak, & Sarid, 2020](#); [Jakiela & Ozier, 2018](#); [Roberts, Winters, & Chen, 2015](#); [Sutter et al., 2018](#)). We can test three of these popular neo-Whorfian ideas: that 'pronoun-drop' correlates with trust ([Tabellini, 2008](#)), that 'pronoun-drop' correlates with patience ([Chen & He, 2021](#)), and 'future time reference' correlates with patience ([Chen, 2013](#)). On trust the four relevant tests are all significant at 5%, but in the opposite direction to the original hypothesis ([Davis & Abdurazokzoda, 2016](#); [Kashima & Kashima, 1998](#)). On time preferences, patience is one of two domains in which there are no significant language or country effects. Of course, given our limited sample and focus on only one pair of languages we do not wish to overstate this evidence. However, we find evidence that is either insignificant or contradicts the relevant neo-Whorfian predictions.

Another potential confound, the Foreign Language Effect, finds that people tend to be less biased and more rational in their second language ([Circi, Gatti, Russo, & Vecchi, 2021](#); [Del Maschio, Crespi, Peressotti, Abutalebi, & Sulpizio, 2022](#); [Stankovic, Biedermann, & Hamamura, 2022](#)). Plausible mechanisms include being more emotional in one's native language, with effects established in two domains: utilitarian moral judgement and risk-taking. As we do not include either topic amongst our questions, we cannot easily test for this effect. However, it does not appear to be driving our results, as questions that are more plausibly related to a Foreign Language Effect (ethics, cheating and patience) do not differ from the other domains. We also discuss whether the language effects should be understood as evidence of elicitation bias (which would have implications for survey research) or evidence of context-dependent preferences (which would have more fundamental implications for our understanding of preferences).

We proceed as follows. Section 2 presents the experimental design, theory and implementation details. Section 3 presents the results, focusing on the magnitude, direction and robustness of effects. Section 4 considers alternative explanations of the language effects, and Section 5 concludes.

2. Experimental design, empirical approach & theoretical predictions

The experiment comprises a short survey, with 19 outcome questions covering nine themes (see [Clist, 2023a](#), for the full text). Our interest is not primarily in the preferences themselves, so we use well-established questions wherever possible. The first question is incentivised, using [Fischbacher and Föllmi-Heusi's \(2013\)](#) die-roll. Subjects are paid according to their report, having rolled a six-sided die in private. Reports 1–5 receive increasing rewards (£1-5 in the UK samples), with 6 receiving no reward (and so is coded as zero).

We use [Falk et al.'s \(2018\)](#) experimentally-validated questions to cover patience, negative reciprocity, altruism, positive reciprocity and trust. We augment the trusting theme with the standard wording from the World Values Survey so that each stated preference theme has at least two questions. For cheating, we use three standard questions from World Values Survey to assess how justifiable subjects feel in cheating on government benefit claims, transport fare, and tax evasion. For Ethics, we use four items from the Ethical Position Questionnaire ([Forsyth, 1980](#)) to assess endorsement of ethical relativism.² We discuss the costs and benefits of eliciting mostly stated preferences in Section 2.3.

At the heart of the experiment are the groups *Mandarin* and *English*, as shown in [Fig. 1](#). The survey was either in British English or Simplified Mandarin Chinese, and for these groups we randomly assign the survey language to Chinese students studying in the

² To be more precise, questions 16–22 and 24–26 are from the Global Preference Survey ([Falk et al., 2018](#)). We did not include risk, as the Global Preference Survey only asks one self-reported question on the topic, with the other difficult to implement in a paper and pen survey. For patience to be covered by two questions we employ our question 24, which comes from [Falk et al. \(2018\)](#) but is not normally used. Questions 13–15 and 27 are from the World Values Survey wording, available at <http://www.worldvaluessurvey.org>. Questions 28–31 are from the Ethical Position Questionnaire ([Forsyth, 1980](#)).

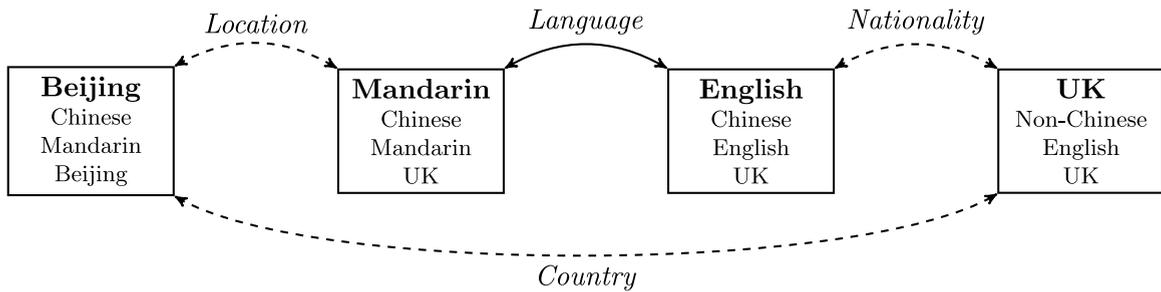


Fig. 1. Experimental design: Four groups and four effects. Note: The four groups are in bold, with the nationality, language and location of each group given directly below. The four effects are in italics, showing the identifying comparison which keeps other elements constant. The language effect is drawn with a solid line to indicate it is a causal estimate. The other three effects are observational estimates. The nationality requirement for the UK group was ‘non-Chinese’, with almost all (82/87) born in the UK.

UK. The two groups entered the same rooms, and each individual had an equal chance of turning over their paper to find a Chinese version or an English version of the survey.³ The use of standard survey questions meant we could often use standard wording for both languages. Where this was not possible, we carefully translated and back-translated questions, with help from two Chinese-English bilingual researchers. Subjects were recruited on the understanding that we were interested in Chinese students studying in the UK, with email addresses of the researchers (one Chinese, the other English) shown in the third paragraph. For these sessions, English and Chinese people were present, and were available to answer questions quietly. Experimenters were instructed to answer any questions in the language they were asked, though questions were rare.

The two other samples are non-randomised reference groups, made up of relevant convenience samples. *Beijing* comprises students at Beijing Normal University. It is difficult to establish the counterfactual cultural environment that Chinese students in the UK would have been exposed to if they had not studied in the UK. There is selection into studying in the UK which likely means these students are different, regardless of any exposure to a different country. However the two universities are very close, at least by some measures. The QS world rankings⁴ for 2017 placed UEA at 252 and Beijing Normal at 257, so the two universities are remarkably close in terms of global rankings. *UK* are other students in the UK, with no nationality requirement other than not being Chinese. These students were recruited using the experimental lab’s mailing list, and 94% of the resulting sample were born in the UK. This group is designed to resemble the UK student body, so it did not need to be exclusively British.

The experiment lasted 20–30 min, and subjects were paid between £2-7 or ¥20-70 according to their self-reported die-roll. This was approximately the same at the time of the experiments using the nominal exchange rate. In both universities, the minimum earnings afforded a hot drink on campus, while the maximum was sufficient to also get a hot meal.

2.1. Empirical approach to estimating the four effects

In order to test the theories discussed in Section 2.2, we first need to calculate the size and direction of differences in preferences between groups. To calculate these, consider the response y of individual i of group j for any given question:

$$y_{ij} = \alpha_j + \beta X_{ij} + \epsilon_{ij} \tag{1}$$

The group means are captured by α , controlling for a range of individual characteristics (captured by the vector X) and an error term (ϵ) describing individual variations. Simple comparisons between group means give rise to the effects of interest as in Fig. 1. We can calculate the average causal effect of language on preferences for a given question as $\Delta_{Language} = \alpha_{Mandarin} - \alpha_{English}$. The location and nationality of the two groups are the same, with the only difference (language) randomly assigned.

We provide non-causal estimates of three other effects of interest. The country effect is $\Delta_{Country} = \alpha_{Beijing} - \alpha_{UK}$. This is the standard observational effect which shows the difference in preferences between people in different countries, as has been calculated before (e.g. Falk et al., 2018; Gächter & Schulz, 2016; Henrich et al., 2001, 2010, 2006; Rieger et al., 2014; Roth et al., 1991; Vieider et al., 2015). The location effect is $\Delta_{Location} = \alpha_{Beijing} - \alpha_{Mandarin}$. The two groups both comprise Chinese students answering the survey in Mandarin, but they do so in different locations. While we are interested in the effect of location on preferences, our estimate is potentially confounded. Chinese students with UK-style preferences are more likely to study in the UK, which would tend to overestimate the size of the location effect. The nationality effect is $\Delta_{Nationality} = \alpha_{English} - \alpha_{UK}$. Both groups answer the survey in English in the UK, but their nationality varies. Note that this compares Chinese to non-Chinese students, as the latter are the relevant reference group for Chinese students. This observational estimate does not attempt to measure the causal impact of nationality on

³ We randomise at the individual level, rather than the session level. This ensures that all other ‘noise’ factors (such as the experimenters, room temperature, lights, and news on a given day) were identical for both treatment groups in the same session. As a result, the effect of the ‘noise’ would not confound with that of the treatment, i.e. the language of the survey.

⁴ Available at <https://www.topuniversities.com/university-rankings/world-university-rankings/>.

preferences, merely to describe any preference differentials between Chinese students in the UK and other (94% British) students in the UK. By definition the well-known country effect is the sum of the other three effects: $\Delta_{Country} = \Delta_{Location} + \Delta_{Language} + \Delta_{Nationality}$.

Using 19 questions means we have rich data, with a little under 6500 observations covering 340 people and nine topics. However, unless we control for the family-wise error rate, we could mistakenly report significant effects where there are none. To control for this, we use the free step-down resampling method for each of the four effects (Jones et al., 2019; Westfall & Young, 1993), which provides adjusted p values for the 19 times we calculate each effect.

2.2. Theoretical predictions regarding the four effects

Consider a simplified model from identity economics (Akerlof & Kranton, 2000; Benjamin et al., 2016, 2010), which can help organise relevant theories. A subject chooses x to maximise their utility:

$$U = -a \cdot (x - x_a)^2 - c \cdot (x - x_c)^2 - b \cdot (x - x_b)^2 \quad (2)$$

x_a is their own preferred option “in the absence of identity considerations” (Benjamin et al., 2010, p.1915), with any deviations from this incurring a disutility weighted by a . Identities are included, with x_c and x_b the actions prescribed for members of categories *Chinese* and *British* respectively. Deviating from these prescribed actions incurs disutilities. Subjects weigh the three disutilities, with $a \geq 0$, $c \geq 0$, $b \geq 0$ and $a + c + b = 1$.

In our experimental design two groups clearly have a single identity. The *Beijing* group does not have a British identity but can be expected to have a Chinese one ($b = 0$, $c \geq 0$). Conversely, the *UK* group may have a British identity but no Chinese one ($b \geq 0$, $c = 0$). What should we expect for the middle two groups? Classic assimilation theory sees “...the gradual disappearance of original cultural and behavioural patterns in favour of new ones” (Algan et al., 2012, p.4). This implies that over time we would expect Chinese students in the UK to have a higher b and lower c , meaning that the groups *Mandarin* and *English* should lie between the two reference groups, as found by others (Cameron et al., 2015).

The exact speed of assimilation in a given setting is an empirical question, but its direction is clear: there is an inexorable move towards the values and preferences of the new culture.⁵ Full assimilation is typically measured in decades or generations (Abramitzky et al., 2020; Alesina & Fuchs-Schündeln, 2007; Fernández, 2011), but there are reasons to think it may be quicker for our sample. Short-term travel, language proficiency, and education have all been found to predict greater assimilation (Bleakley & Chin, 2010; Cameron et al., 2015; Heine & Lehman, 2004), meaning our sample may see somewhat quicker assimilation than is typically measured.

We can test whether the theory of *slow assimilation* is accurate in the short run. Assimilation is seen as a slow process of closing the gap in country-level differences. The average time in the UK for our sample is just over a year, though it ranges from a few weeks to five years. For most of our international students, we should then expect relatively small differences in preferences by language or location: the largest differences in preferences should be by country. In the language of the model, the increase in b and decrease in c are expected to be small for Chinese students in the UK, so differences by nationality should be almost as large as country-level differences. The theory allows for small moves towards the preferences of the host country by location, as found by Heine and Lehman (2004) and Ji, Zhang, and Nisbett (2004), but differences by language are not easily incorporated. When language is discussed in the context of assimilation, it is normally in the context of proficiency (e.g. Bleakley & Chin, 2010). A large language effect would be difficult to reconcile with the overriding message of slow assimilation: the process takes significant amounts of time.⁶ The relative magnitudes are then $|\Delta_{country}| \geq |\Delta_{nationality}| \gg |\Delta_{location}| \geq |\Delta_{language}| \approx 0$. There are also implications for correlations between effects. Whilst assimilation is expected to be limited, the predicted direction of any changes are clear: any differences should be towards the UK sample. This means assimilation expects non-negative correlations between all effects, e.g. $cov(\Delta_{country}, \Delta_{location}) \geq 0$.

The process of one set of preferences replacing another need not be gradual. Frame-switching argues that rather than mixing two cultural influences in each decision, people could select between them (Hong et al., 2000). There is an equivalent idea in identity economics: “[w]hen an individual’s identity is associated with multiple social categories, the “situation” could determine, for example, which categories are most salient” (Akerlof & Kranton, 2000, p.731). Ji et al.’s (2004) evidence supports this view, as they find location and language to progressively move Chinese subjects away from a more Chinese way of categorising objects, and towards an American one. A bicultural understanding of frame switching or identity priming models has slightly different predictions from slow assimilation. These models predict the same direction of movement (e.g. $cov(\Delta_{country}, \Delta_{location}) \geq 0$), but with larger differences by location and/or language. In our context location or language could make a British identity more salient, increasing b and decreasing c , which would result in $|\Delta_{location}| > 0$, $|\Delta_{language}| > 0$ and a smaller difference by nationality.

However, whilst our Chinese subjects in the UK are bilingual, they are not bicultural (Soffietti, 1955). This simple observation changes Akerlof and Kranton’s (2000) and Hong et al.’s (2000) predictions discussed above. Hong et al. (2000) argues that we should only expect a frame to be active if it is available (or internalised). If subjects are monocultural, a given prime may deactivate one

⁵ Within the context of acculturation theory (Berry, 1997), this is discussed as *blending*. This process ‘splits the difference’ between x_c and x_b , with greater exposure to a British identity captured by an increase in b .

⁶ To be precise, slow assimilation theory does not explicitly state that the language of a survey will have no effect on preferences. Rather it is silent on the topic, despite discussing how language proficiency may affect the speed of assimilation. Therefore it is reasonable to summarise slow assimilation as expecting any language effects to be small. Note that any large language effects would also imply much of the empirical research into assimilation has a confounding factor as it typically uses the host-country language (e.g. Cameron et al., 2015).

Table 1
Predictions by theory.

Theory	Slow assimilation (with bicultural subjects)	Identity priming (with monocultural subjects)
Effect magnitudes	$ \Delta_{country} > \Delta_{nationality} $ $ \Delta_{nationality} > \Delta_{location} $ $ \Delta_{location} \geq \Delta_{language} $ $ \Delta_{language} \approx 0$	$ \Delta_{country} > 0$ $ \Delta_{nationality} > 0$ $ \Delta_{language} > 0$ $ \Delta_{location} \approx 0$
Covariances	Non-negative	Ambiguous

Note: Identity priming or frame switching models with bicultural subjects are not included above. Like slow assimilation, they predict $|\Delta_{country}|$ to be the largest effect and all covariances to be non-negative. They differ by allowing for larger differences by location and/or language. See text for further details. For clarity, ≈ 0 denotes the theory predicts small effects.

frame without activating another. Speaking English or moving to the UK could then reduce c without increasing b (the slack would be taken up by an increase in a). A similar logic applies in Akerlof and Kranton's (2000) model with monocultural subjects. It is common to test identity economics models by priming a subject, and comparing choices between primed and control groups (Benjamin et al., 2016, 2010; Chakravarty & Fonseca, 2014). This experimental design seeks to increase an identity's salience through priming, and so reveal the category's prescribed action. In our setting, if Chinese subjects in the UK are monocultural, the UK identity cannot be 'switched on' when the subject speaks English. This implies that speaking English reduces the pull of the Chinese identity, without increasing the pull of the UK one.

Identity priming with monocultural subjects makes distinct predictions. Because they only have one cultural frame internalised, merely moving location but doing the survey in Chinese is expected to have a limited effect on preferences. Therefore we expect small differences by location ($|\Delta_{location}| \approx 0$), notwithstanding selection into migration due to preference similarity. Language is expected to activate one frame in Mandarin but deactivate it in another, implying large language effects ($|\Delta_{language}| > 0$). There is no clear implication of the direction of this effect, as it depends upon the sign of $[x_a - x_c]$. Nationality effects are expected to be large ($|\Delta_{nationality}| > 0$), as this compares people with a frame (mostly British students in the UK, for whom $b > 0$) to those of a different background with a less salient identity (Chinese students in the UK speaking English). Likewise country effects are expected to be large as it compares people with different frames ($|\Delta_{country}| > 0$). There is no clear prediction about the relative size of differences by language, nationality and country. Nor does frame switching make clear predictions about correlations between effects, as it does not see influences as inherently cumulative.

Table 1 summarises the main theoretical predictions. Later, we consider alternative theories or explanations for our results. Whilst we cannot observe people who have been in the UK for longer than five years, in Section 3.3 we test whether time is a confounder. In Section 4 we consider alternative explanations for our language results. These include that language influences preferences through specific grammatical features, and that people respond differently in native and second languages. Our experiment is not designed to test these ideas. However, these ideas have recently gained prominence, and suggest alternative explanations for our results.

2.3. The use of stated preferences

We mostly use stated preference questions. A potential drawback is that unincentivised questions could provide less accurate answers than incentivised ones (see Bardsley et al. (2010) chapter 6, for a general discussion of the trade-offs, and Alós-Ferrer and Yechiam (2020) for the position of this journal's editors). This is perhaps mitigated by using ten questions from the Global Preference Survey. Falk, Becker, Dohmen, Huffman, and Sunde (2023) show that these survey questions provide answers which are reliable, parsimonious and cost-effective, relating closely to incentivised tasks on the same theme. Likewise, Falk et al. (2018) show that they predict real-world decisions at the individual level. For the World Values Survey questions we use, Johnson and Mislin (2012) show the unincentivised trust question predicts experimentally-measured trust. On Forsyth's (1980) ethics questions, an early study ($N = 80$) found no link between relativist moral attitudes and cheating behaviour (Forsyth & Berger, 1982). Rai and Holyoak (2013) find somewhat different results ($N = 120$), and argue that the die-rolling method we use may be more sensitive to detecting cheating than methods used by studies finding null results. For World Value Survey questions on whether different types of behaviour can be justified, and the relative moral attitudes, we test in Section 3.3 whether these questions can predict our own incentivised question on cheating.

The benefits of using stated preference questions include their speed and practicality (Falk et al., 2023). Using only incentivised questions would have meant a choice between having longer sessions or fewer questions. The former would have diluted the monetary incentive to recruit subjects, and potentially reduced the attention of those that did attend. The option to reduce the number of questions is also unattractive. We can measure the same preference domain several times, seeing whether any differences are robust. Rather than having one experiment on trust, we can see whether our results are similar in both trust questions. For some domains, it would be more difficult to administer an incentivised question. Time preferences would have required administering a delayed payment, and all two-player games would have required coordination. Other preference domains would have been harder to incentivise; we are able to ask about ethical values and opinions. In short, using some unincentivised questions provides a larger sample, a broader range of preference domains and a greater ability to see whether differences are robust.

Table 2
Summary statistics and balance tests.

Characteristics	Group				Balance tests			
	Beijing	Mandarin	English	UK	<i>p</i> for equality between			
	(1)	(2)	(3)	(4)	(1)&(2)	(2)&(3)	(3)&(4)	(1)&(4)
Female (%)	79.8	73.3	67.9	44.2	.65	.85	.016*	.000***
Urban (%)	80.0	96.6	89.3	60.5	.002**	.41	.001***	.02*
Age	21.2 (0.24)	23.2 (0.26)	23.1 (0.27)	20.2 (0.26)	.000***	.98	.000***	.01*
# People known in room	0.47 (0.16)	1.67 (0.18)	2.02 (0.19)	1.30 (0.18)	.000***	.85	.097	.000***
Religiosity	3.36 (0.24)	3.89 (0.26)	3.48 (0.26)	3.25 (0.26)	.42	.85	.78	.79
Socio-economic Status	4.91 (0.13)	5.92 (0.15)	5.72 (0.15)	5.80 (0.15)	.000***	.85	.78	.001***
Maths ability	4.99 (0.24)	4.72 (0.27)	5.20 (0.27)	4.32 (0.27)	.71	.85	.10	.14
N (min/max)	95/104	82/88	78/84	85/87	.71	.98	.37	.14

Note: Under the Group heading we present means, with standard errors in parentheses. Under the Balance Tests heading we provide *p* values of four hypotheses, relating to balance on observables. We use adjusted *p* values, controlling the family-wise error rate in each column using the *wyouth* package in Stata 16, using 10,000 iterations (Jones et al., 2019; Westfall & Young, 1993). ***, ** and * denote significance at 0.1, 1 and 5% levels respectively. The four tests relate to the later regressions that estimate the location, language, nationality and country effects. The last three questions are all out of 10 and self-perceptions, with higher numbers indicating having more of that trait (e.g. 10 = very religious). N min refers to the number of subjects who answered all questions, while max denotes those that answered at least one. The most skipped question was age, with 11 people declining to answer (as allowed by the instructions).

2.4. Sample details

In Table 2 we report summary statistics for control questions by group.⁷ We provide balance tests for the four comparisons we will use later, controlling the family-wise error rate for the eight variables. The experimental groups (*Mandarin* and *English*) are similar in each of the controls. Whilst this is unsurprising given the randomised nature of allocation to treatment, it is worth noting that there is not differential skipping of questions (see last line). That could indicate differential understanding, and so is reassuring. In the classic cross-country comparison, we see there are statistically significant differences in five of the control questions. The *Beijing* group is more female, poorer and knows fewer people in the same session, while the *UK* group is more gender-balanced, less urban and younger. These differences are standard in cross-country comparisons: if countries differ in their socio-economic status that is part of the comparison. The other observational comparisons also lack balance in observables, with significant differences in four or five control variables. We control for these observable characteristics, but all of the observational estimates could also have differences in unobservables. This underlines that the observational estimates are not causal, but describe the size of differences in preferences.

The characteristics of the experimental groups are of further interest. 147 subjects gave information on their length of stay in the UK, with an average of 57 weeks, a minimum of 3 weeks and a maximum of 5 years. Of these, 22 have been in the UK for less than 6 months, 79 for between 6 months and one year, 24 for between 1 and 2 years, and 22 for over 2 years. For proficiency in the language, 115/166 students stated they had intermediate English, with 39/166 stating an advanced level of English. All students needed to pass an English test to study in the UK and there is not differential skipping of questions, so we do not have concerns over the level of understanding. However, we return to language proficiency (and the length of time students have spent in the UK) in Section 3.3.

3. Results

To give a sense of the raw data, we show the average response by group in Table 3. We move on to the four effects later, but note here there are group-level differences. Of the 19 outcome questions, 15 are significantly different by group at the 5% level. The precise wording of each question can be found in the replication package (Clist, 2023a), with Table 3 ordering the questions by the eight themes. In the vast majority of cases there appears to be agreement between questions regarding the same topic. This is the case even when questions had reversed scales (see table note for details), and were not asked consecutively.

3.1. The magnitude of effects

We estimate the four effects using a series of regressions in the form of (1). *y* is the response to one of the 19 outcome questions, and determined by group means (α), a vector of controls (*X*) and an error term. Each effect is simply the difference between two

⁷ We aimed for a minimum of 80 students in each group. This could identify a given outcome effect of 0.21 standard deviations with power of 0.9 at the 5% significance level.

Table 3
Summary statistics for outcome questions, by group.

Q. Number & Short description	Beijing	Mandarin	English	UK	P value
Die-roll					
1. Reported roll (0–5)	3.15 (0.17)	3.20 (0.17)	3.31 (0.16)	3.40 (0.17)	0.742
Cheating is justifiable					
13. For government benefits	3.30 (0.20)	3.51 (0.22)	4.66 (0.24)	2.14 (0.14)	0.000***
14. For public transport	2.38 (0.21)	2.98 (0.32)	4.60 (0.30)	3.86 (0.24)	0.000***
15. For paying taxes	2.16 (0.16)	2.25 (0.19)	3.87 (0.30)	1.94 (0.14)	0.000***
Patience					
16. Self-assessment	6.84 (0.23)	6.91 (0.24)	6.42 (0.28)	6.95 (0.21)	0.446
24. (Free from) procrastination	6.07 (0.28)	6.16 (0.29)	6.13 (0.23)	6.61 (0.27)	0.471
Negative reciprocity					
17. Punish on own behalf	6.30 (0.21)	6.18 (0.24)	5.45 (0.28)	4.91 (0.23)	0.000***
18. Punish on others behalf	5.58 (0.20)	5.69 (0.20)	5.56 (0.24)	5.71 (0.18)	0.935
21. Revenge on own behalf	4.84 (0.19)	4.69 (0.23)	5.11 (0.27)	4.05 (0.24)	0.0182*
Altruism					
19. Self-assessment	6.97 (0.20)	7.41 (0.25)	6.15 (0.24)	7.41 (0.23)	0.000***
26. Hypothetical (/£1200)	281.0 (27.2)	366.5 (32.4)	463.6 (42.6)	169.8 (18.7)	0.000***
Positive reciprocity					
20. Self-assessment	8.43 (0.20)	9.06 (0.17)	8.46 (0.22)	8.28 (0.21)	0.0128*
25. Gift (/£40) responding to £30	32.2 (0.83)	33.0 (1.00)	26.4 (1.24)	19.71 (1.16)	0.000***
Trusting					
22. People have best intentions	7.09 (0.19)	7.11 (0.22)	5.81 (0.21)	5.19 (0.24)	0.000***
27. People are fair	6.90 (0.20)	6.72 (0.24)	5.65 (0.25)	4.70 (0.19)	0.000***
Ethics are flexible					
28. Ethics vary by situation & society	7.96 (0.20)	8.50 (0.15)	6.35 (0.20)	7.08 (0.22)	0.000***
29. Codification of ethics is bad	6.64 (0.24)	6.80 (0.24)	6.18 (0.17)	7.08 (0.18)	0.004**
30. Codification of ethics is impossible	7.28 (0.20)	7.40 (0.23)	6.23 (0.25)	7.07 (0.22)	0.003**
31. Morality of lies is context dependent	7.81 (0.21)	7.65 (0.22)	6.73 (0.26)	7.83 (0.18)	0.004**

Note: Higher numbers indicate higher values of the trait or agreement with the sentiment. Scales run 1–10 unless stated. For questions 16–22, 24 and 28–31 the questionnaire asked a question with 1 denoting most and 10 denoting least: these have been transformed. Question numbers allow the reader to find the precise wording of the question in the survey. Robust standard errors are provided in parentheses. P values come from tests for equal group means, operationalised using dummies for each group but no constant term, and testing whether the coefficients are equal in a regression on the outcome question. N varies between 360–363 for each row.

group means, as in Fig. 1. We standardise all outcome data, transforming each question's response to have a mean of 0 and a standard deviation of 1. This eases the interpretation of effect sizes, as they are simply differences in standard deviations. It also allows comparisons over different questions to be more natural.

We present the results in two ways. First, we visually show the four effects of interest (Fig. 2) and controls (Fig. 3). These are taken from 19 regressions (one per question) which pool all four samples together, using marginal effects to show the four effects of interest. This allows all data to be used in estimating the effects of controls. Second, we address multiple hypothesis testing concerns in Table 4. We adjust the p values for the 19 outcome questions for each of the four effects to control the false discovery rate (Jones et al., 2019; Westfall & Young, 1993).

The results make several points clear. First, let us consider the average magnitude of location, language, nationality and country effects. The location effect is the smallest of the four at 0.122. As data is centred this can be interpreted as *Beijing* and *Mandarin* differing by a little under one eighth of a standard deviation, on average. This is surprisingly small as the two samples differ in a variety of ways. Table 2 shows those studying in the UK are richer, older, and more likely to have an urban background. The regression controls for these observed differences, but there is not a single significant effect across the 19 questions (Table 4).

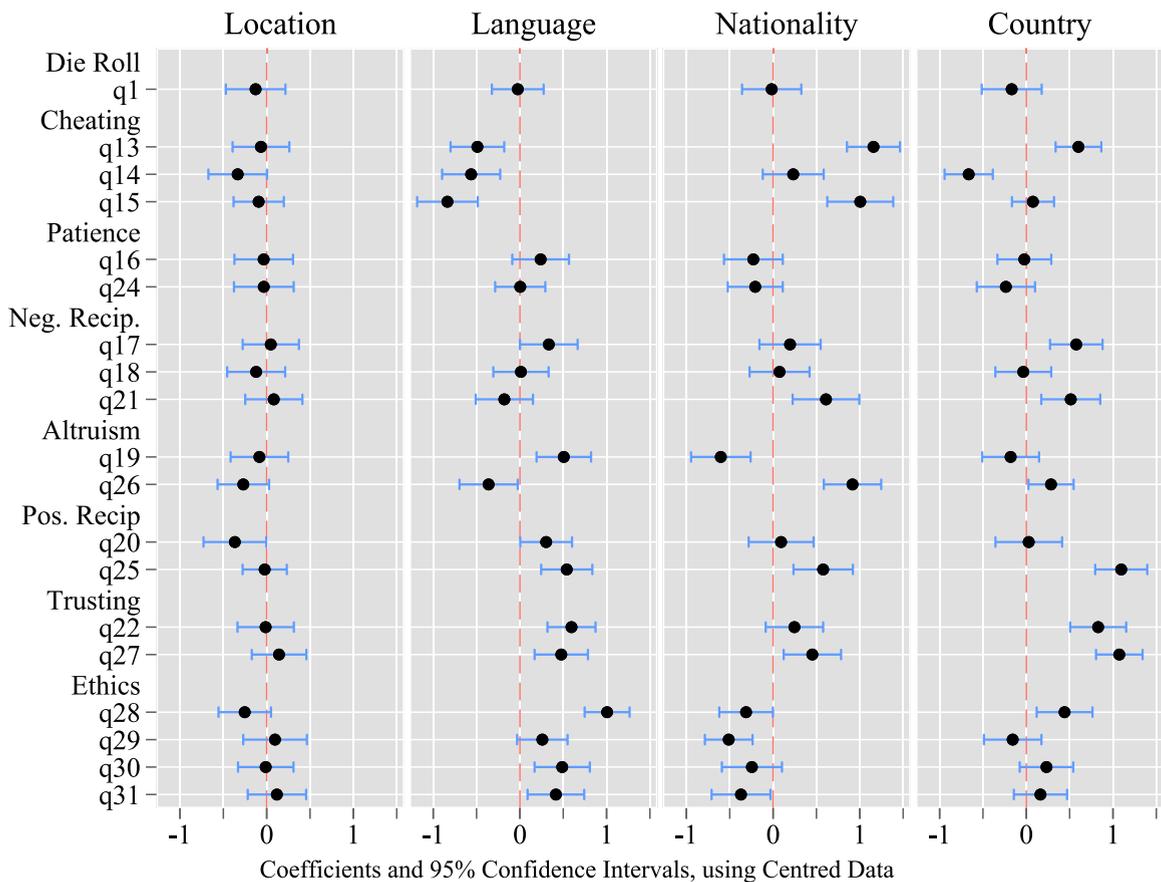


Fig. 2. The four effects, by outcome question. Note: Each effect is identified by comparing two groups, as shown in Fig. 1. Coefficients can be interpreted as differences in standard deviations, as the data is centred. The 95% confidence intervals come from robust standard errors. The effect in each case is being ‘more Chinese’.

The average absolute country effect is 0.388. Of the 19 questions, there are 7 significant effects at the 5% level after correcting for multiple comparisons (Table 4). These include differences in four of the nine topics. The average absolute language effect is slightly larger at 0.402, with 9 significant differences at the 5% level. This is perhaps surprising as the country effect varies language and several other factors. The language effect is also a pure causal estimate, given the experimental control. Moving to nationality, it has the largest average absolute effect: 0.424. In 7 cases it is significant at the 5% level. This means Chinese students in the UK differ by around half a standard deviation from other students in the UK when speaking English. Together, these results show that, on average, Chinese students in the UK are very similar to Chinese students in China when speaking Mandarin. However, they are not similar to other students in the UK when they speak English. These results are not compatible with *slow assimilation*, which expects $|Δ_{country}|$ to easily be the largest effect, which it is not. Evidence from the magnitude of effects are in line with identity priming and frame-switching models with monocultural subjects. Language activates associated cultural frames, but only if they have been internalised.

Second, moving to controls, we see much smaller average absolute effects. This is clear in Fig. 3, which requires a smaller x axis than Fig. 2. The average magnitude for controls across the 19 questions ranges from 0.04 (for socio-economic status) to 0.15 (for gender),⁸ compared to average magnitudes of 0.12–0.42 for the effects of interest. The average language effect is around three times larger than the average gender effect.

Third, whilst we are not principally interested in the preferences themselves, we find very similar country effects to others. This is despite differences in sample, time period and implementation.⁹ For the incentivised die-roll question, a large meta-analysis found the UK report to be 0.45 higher than China’s on our scale (Abeler, Nosenzo, & Raymond, 2019), meaning UK samples tend to lie for

⁸ Specifically, they are female 0.15, urban 0.11, age 0.05, the number of people in the room known 0.05, religious 0.07, socioeconomic status 0.04 and maths ability 0.07.

⁹ We use a student sample in 2017–18, and normally a 10 point scale. Falk et al. (2018) use an 11 point scale, a representative sample and collected data in 2012. The last World Values Survey to cover the UK and China was the fifth wave (2005–9). While it uses a 10 point scale, it uses a representative sample. Forsyth, O’Boyle, and McDaniel (2008) is a meta-analysis, which reports only summary statistics and mixes 5 and 9 point scales. It also averages 10

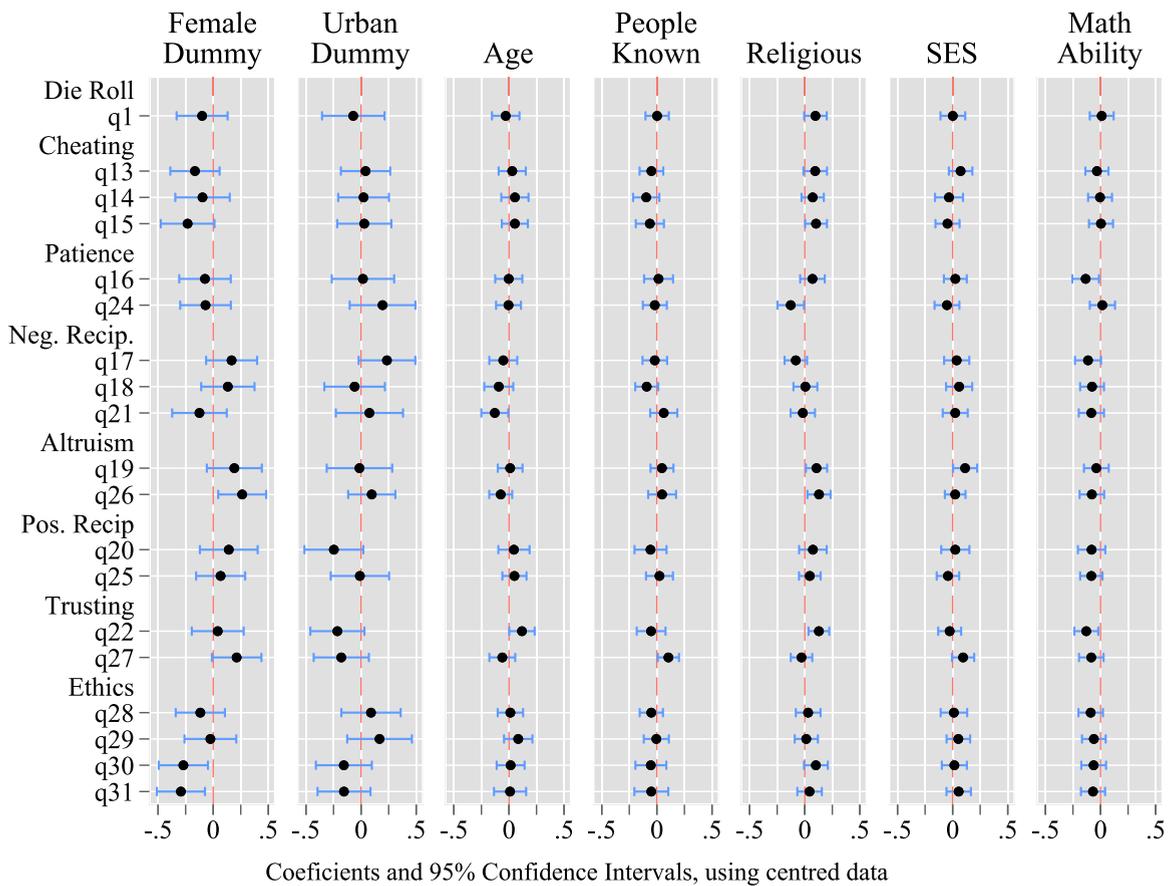


Fig. 3. Controls, by outcome question. Note: Coefficients can be interpreted as differences in standard deviations, as the data is centred. The 95% confidence intervals come from robust standard errors.

profit at a higher rate. We find a difference of 0.27 in the same direction (on a 0–5 scale). For Falk et al.’s (2018) questions (16–22 and 24–26), we find significantly higher negative reciprocity, positive reciprocity and trust for our Beijing sample. For patience and altruism there are no significant differences at the 5% level. Falk et al. (2018) finds very similar results: China has higher positive reciprocity (0.52 standard deviations) and trust (0.31 standard deviations), with no real difference in negative reciprocity (China is higher by 0.001 standard deviations). They also report slightly lower patience (-0.14) and higher altruism (0.48), with the latter the largest discrepancy. For the World Value Survey questions (13–15 and 27, from Inglehart et al., 2014) the fifth wave found country differences in standard deviations of 0.48, -0.17, -0.14 and 0.58 respectively. This echoes the mixed findings in our own results, including the apparent contradiction in questions 13 and 14. For the moral relativism questions (28–31, from Forsyth, 1980; Forsyth et al., 2008) report almost no difference between the UK and China. They report the UK has higher moral relativism by less than 0.04 standard deviations. In short, the country effects we find are very similar to previous work.

Fourth, significant effects on the same topic tend to agree. We find a total of 23 significant effects in Table 4 at the 5% level. In six cases there is only one significant effect for that topic. In 13 cases there is agreement: one trio and five pairs. There is disagreement in two pairs: the nationality effect for altruism, and the country effect for cheating. We cannot check the nationality effect as this has not been calculated before, but we can check country effects. As noted above this apparent contradiction has been found previously: the fifth wave of the World Values Survey found the same pattern for questions 13 and 14. This implies country differences on the justifiability of cheating on government benefits is different from that on public transport. These high levels of agreement by topic increase our confidence in the results.

3.2. The direction of effects

Having considered average effect magnitudes, we now turn to their direction. As summarised in Table 1 *slow assimilation* predicts all correlations will be non-negative, as Chinese students will gradually move towards UK style preferences with each new experience

questions, of which we use only 4. The average score for moral relativism for China is 0.687 (N = 1081, SD = 0.146) compared to the UK’s 0.690 (N = 289, SD = 0.035). To cautiously calculate the difference in standard deviations, we used the grand mean’s SD of 0.076 (N = 30,230).

Table 4
Testing the four effects, with multiple hypothesis corrections.

Effect:	Location Beijing		Language Mandarin		Nationality Chinese		Country China	
	Δ	p	Δ	p	Δ	p	Δ	p
Die-roll								
1	-0.128	1.000	-0.024	0.997	-0.016	0.951	-0.168	0.910
Cheating is justifiable								
13	-0.066	1.000	-0.490	0.028*	1.158	0.000***	0.602	0.000***
14	-0.334	0.601	-0.562	0.019*	0.232	0.755	-0.664	0.000***
15	-0.092	1.000	-0.836	0.000***	1.005	0.000***	0.077	0.947
Patience								
16	-0.034	1.000	0.239	0.546	-0.228	0.755	-0.023	0.995
24	-0.033	1.000	0.004	0.997	-0.206	0.755	-0.235	0.792
Negative reciprocity								
17	0.047	1.000	0.334	0.308	0.196	0.755	0.577	0.005**
18	-0.122	1.000	0.012	0.997	0.075	0.951	-0.035	0.995
21	0.082	1.000	-0.179	0.732	0.610	0.032*	0.513	0.050*
Altruism								
19	-0.084	1.000	0.507	0.026*	-0.603	0.012*	-0.180	0.910
26	-0.270	0.710	-0.361	0.273	0.916	0.000***	0.286	0.300
Positive reciprocity								
20	-0.367	0.575	0.303	0.308	0.093	0.951	0.029	0.995
25	-0.023	1.000	0.540	0.007**	0.578	0.019*	1.095	0.000***
Trusting								
22	-0.012	1.000	0.595	0.001***	0.246	0.747	0.829	0.000***
27	0.143	0.999	0.477	0.031*	0.453	0.094	1.073	0.000***
Ethics are flexible								
28	-0.253	0.794	1.006	0.000***	-0.312	0.387	0.442	0.096
29	0.096	1.000	0.259	0.390	-0.512	0.007**	-0.157	0.910
30	-0.011	1.000	0.488	0.032*	-0.244	0.755	0.233	0.748
31	0.118	1.000	0.415	0.122	-0.370	0.308	0.163	0.910
Average magnitude								
$ \bar{\Delta} $	0.122		0.402		0.424		0.388	

Note: Each effect is estimated using the full sample for each question. This allows a more precise estimate of the effects of controls, shown in Fig. 3. Each effect is identified by comparing two groups, as in Fig. 1. The Δ parameters can be interpreted as differences in standard deviations, as the data is standardised. The original regressions use robust standard errors. The p columns report adjusted p values from the `wyoutng` package in Stata 16 (Jones et al., 2019; Westfall & Young, 1993). This controls for the family-wise error rate for each effect, i.e. each effect is treated as one family of hypotheses. We use the recommended 10,000 iterations, with ***, ** and * denoting significance at 0.1, 1 and 5% levels respectively. $|\bar{\Delta}|$ denotes the average absolute effect in standard deviations. Questions 15 and 29 have a sample of 338. Questions 13, 14, 19, 21, 22, 24 and 26 have a sample of 339. The rest (1, 16, 17, 18, 20, 25, 27, 28, 30 and 31) have a sample of 340. This gives a total sample of 6449.

or exposure to British culture. Models of *identity priming* with monocultural subjects do not make predictions on the correlations of effects, stating only that subjects may simply be less influenced by the Chinese identity when speaking English. In Fig. 4 we plot all six relationships alongside correlation coefficients.

Fig. 4 shows country effects are positively correlated with each other effect, albeit never significantly so. These relationships are noisy, so looking only at one outcome variable could mislead. For example, the results from question 14 could be presented in isolation as compelling evidence that Chinese students in the UK adopt Chinese preferences in Mandarin and British preferences in English. That would echo Ji et al.'s (2004) findings in support of bicultural *identity priming* models. Likewise question 21 could be presented in isolation as evidence that Chinese students do not assimilate at all, with nationality and country differences very similar. That would echo very *slow assimilation* findings (Abramitzky et al., 2020; Alesina & Fuchs-Schündeln, 2007; Bisin & Verdier, 2011). Looking at a range of preferences, which is possible because we measure stated preferences, neither theory fits the overall picture.

The effects are not additional, as whilst location, language and nationality effects are positively correlated with country effects, they are not positively correlated with each other. This contradicts *slow assimilation* and bicultural *identity priming* models. Fig. 4's bottom row shows two negative correlations, including the only significant correlation. Questions 13 and 15 are useful extreme examples to explain the negative correlation between nationality and language effects. Table 4 shows the respective language effects (-0.49 and -0.84) and the nationality effects (1.2 and 1). In other words, the causal effect of speaking Mandarin for Chinese students in the UK is to substantially decrease how acceptable they find cheating. Comparing students in the UK, Chinese students speaking English find cheating substantially more acceptable than the (mainly British) reference group. This restates previous findings: Chinese students in the UK speaking Mandarin are similar to their compatriots in Beijing. For questions 13 and 15, Table 3 shows they are indeed remarkably similar. In English they are not similar to the Beijing or UK group.

These results fit models of *identity priming* or *frame switching* with monocultural subjects. An internalised Chinese cultural frame can be activated by speaking Mandarin, but this identity is less salient when speaking English. International students have not internalised a UK identity, and so it cannot be activated. This theory explains why the location effect is small, as moving country

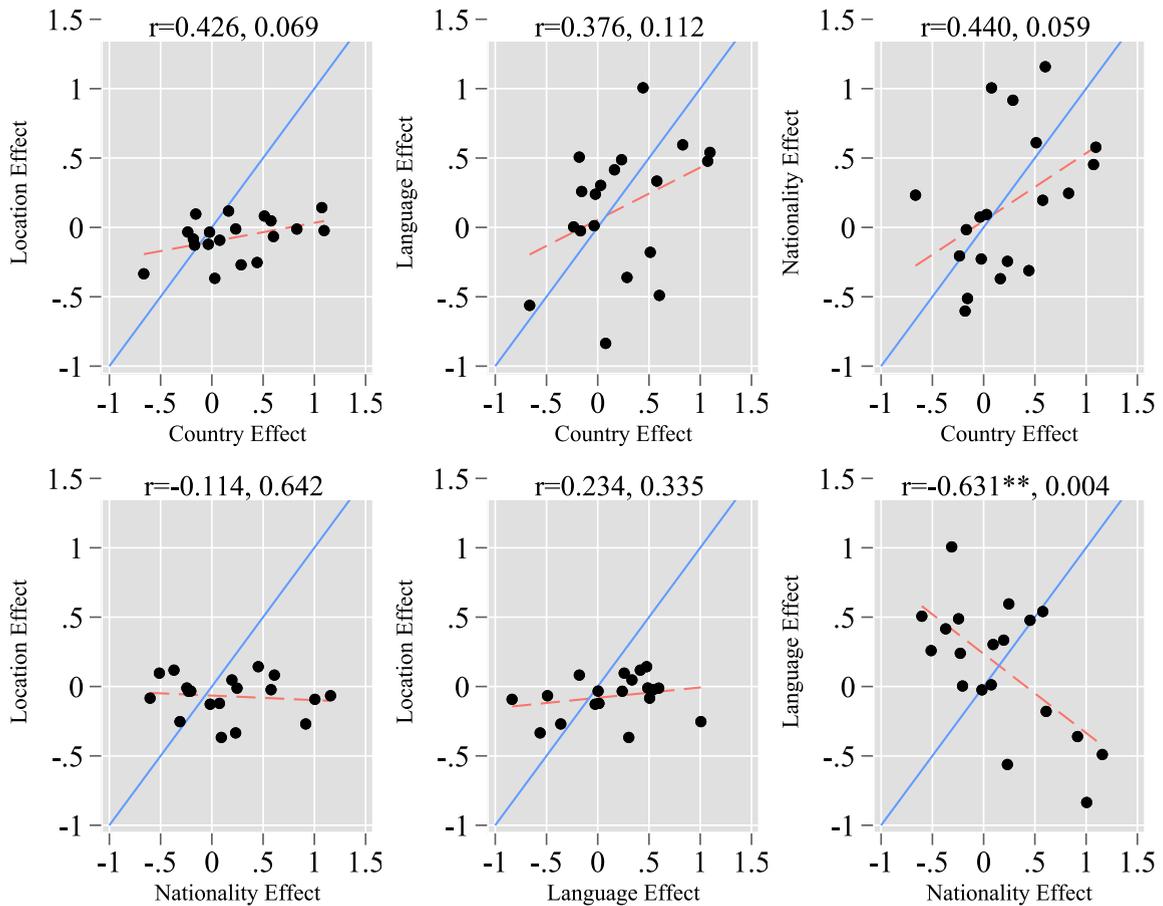


Fig. 4. Correlations between effects. Note: 45° lines are solid, while lines of best fit are dashed. Each dot represents one of the 19 outcome questions. At the top of each scatter plot is Pearson's correlation coefficient r , with the associated p value. ***, ** and * denote significance at 0.1, 1 and 5% levels respectively.

does not remove a cultural frame. It also explains why language and nationality effects can be large but not additional. Speaking English weakens the Chinese cultural frame, but does not replace it. In this sense Chinese students in the UK speaking English are less bound by identity-based concerns (Akerlof & Kranton, 2000; Fehr & Hoff, 2011). In the language of (2), speaking English sees c reduce, b remain unchanged, and a increase. In our sample international students do not learn foreign preferences, but learn to neglect older ones.

3.3. The robustness of effects

We present three tests of robustness. The first controls for the amount of time a student has spent in the UK. The second selectively drops individuals with weaker language skills. The third examines whether our stated preference questions can predict the observed level of cheating.

'Time in country' is often found to be an important influence on assimilation (Alesina & Fuchs-Schündeln, 2007; Bisin & Verdier, 2011). For example, Abramitzky et al.'s (2020) data summary shows the speed of assimilation in name choice is very close to linear over 20 years. Our sample allows only one-quarter of that time horizon, as we are focused on relatively short-term migrants. Rather we can see any changes over the first months and years, and whether controlling for time in the UK affects our core results.

Fig. 5 shows three coefficients from regressions which augment the standard regression specification with (a) a variable capturing *time in the UK* and (b) an interaction term for *time in the UK* and whether the survey was conducted in Mandarin. We only use Chinese students in the UK for these regressions to keep a consistent set of controls for all subjects. Missing data on *time in the UK* reduces the sample to 136 or 137 for each the 19 regressions. Half of these have spent less than 40 weeks in the UK, with a mean of 58 weeks and a standard deviation of 49.

The first panel shows the effects of speaking Mandarin on preferences, which closely resemble our main results (see Table 4 and Fig. 2). Our language results are unaffected by controlling for the time a student has spent in the UK. The second panel shows that *time in the UK* has a small average effect on preferences. A Chinese student spending around one extra year in the UK has preferences

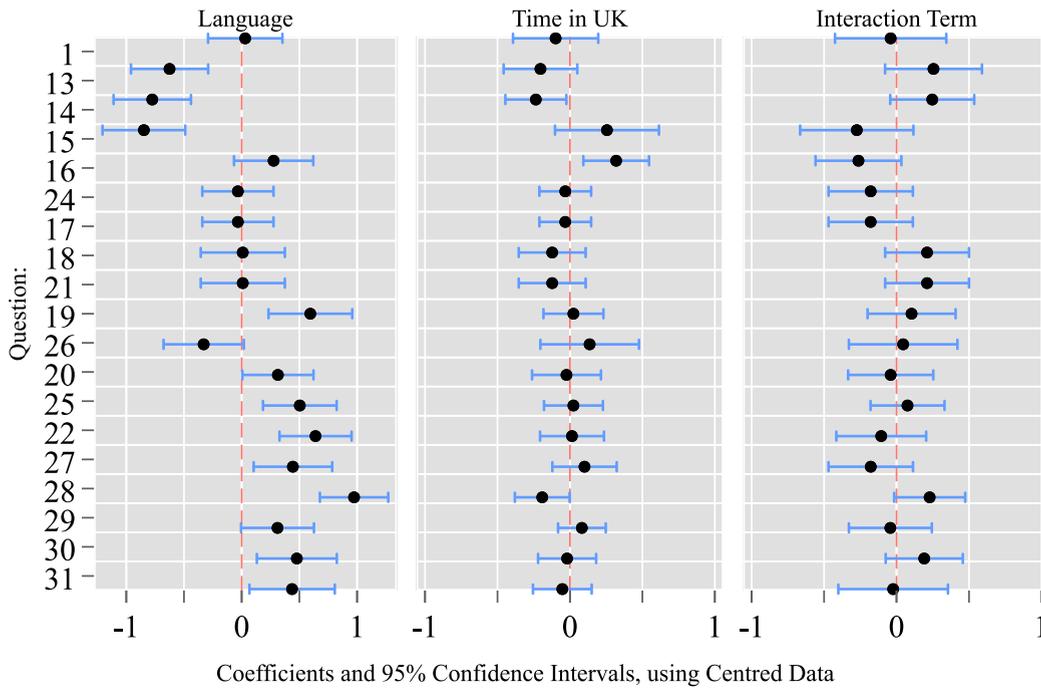


Fig. 5. The effect of time in the UK, For Chinese students in the UK.

that differ in average magnitude by around 0.12 standard deviations (in an uncertain direction). In one case the effect is significant at the 1% level (question 16), and in two further cases at the 5% level (questions 14 and 28). However, the significant effects do not survive using the Holland method to control the family-wise error rate for the 19 questions asked (the lowest q value is 0.11). The third panel displays the interaction term, measuring the additional effect of speaking Chinese and time in the UK. None of these approach significance.

These results are consistent with a small effect of time on assimilation found by others (Abramitzky et al., 2020; Alesina & Fuchs-Schündeln, 2007; Bisin & Verdier, 2011). As this was not the main focus of our study, we cannot distinguish between a small effect and a null result, especially for any interaction effects. However, we do not find large jumps in preferences over the first few years a Chinese student spends in the UK, and controlling for this factor does not affect our main results.

For our second robustness check, we exclude students with lower language proficiency in order to see whether the conclusions still hold. Table 5 shows the set of results for four samples. Column (1) is the full sample, provided for comparison. Column (2) drops 8 students who were born in Hong Kong, but were included in the English or Mandarin samples. They responded to a call for ‘Chinese students’, but it is possible that including them could introduce bias or noise into the results. Column (3) drops 56 students from the English group, who report speaking less than ‘advanced’ English. Column (4) additionally drops 70 students from the Mandarin group for the same reason. It is possible that a differential level of English could influence the results, even if students have passed University entrance requirements.

Despite dropping large amounts of data, Table 5 shows that the conclusions are unchanged. The average magnitude of the four effects of interest are very similar. There is a little more movement in the correlations between effects, but significant correlations appear robustly estimated. This shows that the results are robust to rather drastic changes in sample composition and size: the third column drops 15% of the total sample, the fourth drops 35%.

Our third robustness check investigates the validity of using stated preferences. We use 18 such questions in our survey, and one revealed preference question. The benefit of measuring stated preferences is that we get a sense of average effect sizes across a range of domains, and reduce the likelihood of cherry-picking a misrepresentative finding. A possible downside is that people may pay less attention, or not be incentivised to report accurately. We can test this concern by seeing whether relevant stated preferences are correlated with the measured revealed preference. We pay subjects on the basis of their reported die-roll, so a higher number signals either good luck or cheating. We also ask seven questions covering the domains of cheating and ethics. This allows us to test whether the stated preferences can help predict a subject’s reported die-roll: a tough test as it measures if subjects are honest about their feelings regarding dishonesty. In Table 6 we present seven bivariate regressions where the (non-standardised) die-roll is the dependent variable, and the (standardised) independent variable is a question on cheating or ethics. As we conduct multiple tests of this hypothesis, we use a step-down procedure to adjust the p values for the 7 tests, reporting the resulting q values.

Table 6 shows two significant effects at the 5% level, two at around 0.1 which are insignificant, and three which are both small and insignificant. The significant predictors are questions 14 and 31, where a one standard deviation higher answer predicts an

Table 5
Summary of results for four samples.

Sample	(1)	(2)	(3)	(4)
Average magnitude of the four effects				
Location	0.122	0.132	0.125	0.158
Language	0.402	0.402	0.424	0.447
Nationality	0.424	0.429	0.468	0.461
Country	0.388	0.392	0.391	0.397
Correlations & Their significance				
Location & Language	0.234	0.209	0.196	-0.110
<i>p</i>	0.335	0.391	0.422	0.655
Location & Nationality	-0.114	-0.157	-0.042	-0.065
<i>p</i>	0.642	0.521	0.865	0.792
Language & Nationality	-0.631	-0.605	-0.696	-0.622
<i>p</i>	0.004**	0.006**	0.001**	0.004**
Location & Country	0.426	0.355	0.455	0.129
<i>p</i>	0.069	0.136	0.050	0.597
Language & Country	0.376	0.393	0.312	0.393
<i>p</i>	0.112	0.096	0.193	0.096
Nationality & Country	0.440	0.451	0.434	0.429
<i>p</i>	0.059	0.053	0.063	0.067
N	6449	6297	5445	4211

Note: See text for details on samples. Sample 4 drops 126 students in the Mandarin and English groups because they self-report language proficiency as basic (1), beginner (10) or intermediate (115). ***, ** and * denote significance at 0.1, 1 and 5% levels respectively.

Table 6
Are stated preferences for cheating and ethics related to revealed cheating?.

Independent variable:	Cheating questions			Ethics questions			
	13	14	15	28	29	30	31
β	0.100	0.209	-0.022	-0.004	0.012	0.119	0.191
<i>p</i>	0.213	0.007**	0.792	0.959	0.888	0.175	0.026*
<i>q</i>	0.618	0.045*	0.991	0.991	0.991	0.618	0.147

Note: Each column represents one bivariate regression, where a constant term was included but not reported. The dependent variable in each case is question 1, measured 0–5, where a higher number captures a higher chance of having cheated. The independent variables are all standardised, so β measures the effect of a one standard deviation increase in the independent variable on question 1. The *q* value uses the Holland method to control the family-wise error rate for the seven tests included, using a step-down method. For both *p* and *q*, ***, ** and * denote significance at 0.1, 1 and 5% levels respectively. *N* varies between 360 and 363.

average reported die-roll of 0.2 higher. These two questions are plausibly related to the die-roll: question 14 asks subjects how justifiable it is to cheat on public transport, while question 31 asks whether a lie's morality is context-dependent. Of the three cheating questions, cheating on a transport fare is the lowest stakes and most related to the die-roll task, and it is only this question that is significant after the multiple-testing correction.

The effect size is reasonably large. For comparison, we found a country effect of 0.25 with higher cheating in the UK. Abeler et al.'s (2019) meta-analysis found a country difference in the same direction (of 0.45, translated to our 0–5 scale) and that men report an average of 0.15 more. These results show that our stated preferences are capturing something real. They are able to predict revealed preference outcomes in our setting, as well as having been shown reliable in other settings.

4. Alternative explanations of language effects

Our main interest in language effects is in how they relate to other differences, both in magnitude and direction. This helps us understand the short-run dynamics of preference assimilation for international students, and test between assimilation and identity-priming models. However, there are competing reasons why we might find language effects, alternative mechanisms and different ways of understanding the results. Below we start by discussing two possible confounds of our language effects: certain linguistic features could shape preferences, or native languages could influence choices differently from second languages. Our experiment was not designed to answer these questions, but we can reconsider our evidence in the light of alternative explanations. We then discuss a possible mechanism for effects, and alternative ways of understanding differences in measured preferences.

4.1. Neo-Whorfian

Recent economic research has been influenced by Benjamin Whorf's old assertion that grammar influences preferences. Here, we consider whether our results are robust to neo-Whorfian ideas. A simple test for this is whether the language effect is correlated

with the country effect. If language determines preferences, the experimental group should resemble the Beijing group in Mandarin and the UK group in English. As shown in Fig. 4 (upper middle panel), that is not the case.

A more nuanced test of neo-Whorfian ideas is to examine differences only where the theory predicts them. Economists are interested in four main areas that relate to how languages encode gender, politeness, the self and time. We cannot test the effect of ‘gendered languages’ (Jakiela & Ozier, 2018), as we do not measure relevant attitudes or outcomes. Nor can we test the idea that politeness distinctions mark a deeper respect of social hierarchy (Tabellini, 2008), because neither Mandarin nor English requires politeness distinctions (Davis & Abdurazokzoda, 2016; Kashima & Kashima, 1998). We can consider three neo-Whorfian predictions regarding trust and time.

Pronouns, such as you or I, are required in some languages but optional in others. Kashima and Kashima (1998) argued people speaking languages that allow ‘pronoun drop’ will be less egotistical. Tabellini (2008) tested this idea, finding that speakers of languages which allow pronoun drop have lower trust/respect, which in turn affects governance. More recent work has tested correlations with a variety of other outcomes (Feldmann, 2019; Licht, Goldschmidt, & Schwartz, 2007) or used internal differences in the language (He, Riyanto, Tanaka, & Yamada, 2020). We focus on Tabellini’s (2008) original test, as we measure trust in a similar way. While Tabellini (2008) uses the wording from the sixth wave of the World Values Survey, our question 27 uses wording from the first five waves. English requires the use of the pronoun, whereas Mandarin does not (Davis & Abdurazokzoda, 2016; Kashima & Kashima, 1998), so Tabellini’s (2008) theory predicts greater trust in English. This would mean negative language and country effects for question 27. We can also follow Falk et al. (2018) in using their trust question (our question 22) to test the theory. Our results are the opposite of Tabellini’s (2008) predictions. For question 27, Table 4 shows that the language and country effects are positive and (respectively) significant at the 5% and 1% levels. For question 22, Table 4 shows that the language and country effects are positive and significant at the 1% level. Both trust questions have significant differences in the opposite direction to a leading neo-Whorfian prediction.

On time, Chen (2013) argues that strong ‘future time reference’ languages (including English) make a larger distinction between the present and future than weak ‘future time reference’ languages (including Mandarin). A number of more recent papers have also considered this relationship (Angerer et al., 2021; Chen et al., 2019; Falk et al., 2018; Roberts et al., 2015; Sutter et al., 2018). The original (Chen, 2013) article has a clear, testable prediction in our experiment: positive language and country effects for the patience questions. (Note, however, that Galor et al. (2020) use data which predicts no differences, as Mandarin and English are both classified as having a ‘periphrastic future tense’.) Our results do not support the linguistic-savings hypothesis. Table 4 shows patience is one of only two themes not to have any significant effects.

Chen and He (2021) test a novel hypothesis, linking within-language variations in the inclusion of pronouns to time preferences. They find that dropping a pronoun in the description of a discounting task led to more patient choices, using 135 Taiwanese students. This could perhaps imply we should see more patient choices in Mandarin than English in our results; the same prediction as from the future time reference theory. As mentioned above, we do not. Hence our results are not in line with Chen and He’s (2021) hypothesis or results.

We do not wish to over-interpret our results. They do not falsify a general relationship, one classification predicts no relationship, and it comprises merely two languages (cf. Herz, Huber, Maillard-Bjedov, & Tyahlo, 2021; Sutter et al., 2018). However, our specific results do not support neo-Whorfian predictions for the three cases we can test, and in one case strongly contradicts it. It is clear that our results are not driven by a neo-Whorfian confound.

4.2. The Foreign Language Effect (FLE)

Another potential confound of our language results is that English is a second language for international students. Results from psychology suggest that people may behave differently in a second language than their native language, perhaps because people are less emotional in their second language. Three recent meta-analyses focus on differences in two established domains: moral decision-making and risk-aversion, generally finding a robust effect which is attenuated by language similarity (Circi et al., 2021; Del Maschio et al., 2022; Stankovic et al., 2022). Effects have been found in other domains, though these are less well established and/or relate to more distant domains.¹⁰ These studies do not neatly map onto our own experiment, as they mainly deal with moral decisions where there is a clear utilitarian choice, such as the footbridge or switch dilemmas. The typical finding is that people are less biased in a foreign language, but this does not have a clear implication in our study. Our own ethical questions relate to whether ethical codes are flexible and context-dependent, whilst the die-roll and cheating questions relate to ethical standards. Looking at the size and significance of language effects in our results, we do not see a noticeable pattern in questions which are more related to moral decisions, implying our results are not driven by a foreign language effect.

Mostly closely related to our own experimental results, Alempaki, Doğan, and Yang (2021) examine a potential FLE for lying behaviour and associated norms for people in Germany and China, finding no overall effects on behaviour.¹¹ Our incentivised measure of cheating (question 1) also found no significant sample differences. For norms and appropriateness, they report social

¹⁰ For example, Díaz-Lago and Matute (2019) find that bilinguals are less susceptible to causality bias in their second language across two experiments with 80 native Spanish speakers and 36 native English speakers. Gao, Zika, Rogers, and Thierry (2015) find the ‘hot hand’ effect disappears when feedback is given in a second language, using 16 Chinese-English bilinguals.

¹¹ One experiment in China found fewer lies in English than in Mandarin (by 32 percentage points, $p = .006$, $N = 92$). In another experiment in China they found no difference (6.5 percentage points, $p = .566$, $N = 74$), nor in an experiment in Germany between German and English (9 percentage points, $p = .414$, $N = 64$).

norms are stricter in one's native language. Some of our results are similar, as speaking Mandarin decreases the perceived justification of cheating in three domains (questions 13–15). The FLE could indeed confound these results. However, for question 28 there is a larger effect in the opposite direction: students are more likely to say ethics are flexible if speaking their native language. Further, [Alempaki et al. \(2021\)](#) did not find differential norms between German and English for Germans. This is difficult to reconcile with a FLE on cheating, perhaps pointing to a role for cultural associations. The domains in which FLE might be important is also unclear, and so this appears to warrant further research. Specifically, the Foreign Language Effect needs to be disentangled from the associated norms and frames of different languages, perhaps by using a broad spread of preferences and languages.

4.3. Language as a coordination device

The identity economics model in Section 2.2 discusses the pull of norms: the idea that choices can be affected by the norms of a given category, and that category's salience at the time of the decision. It is also possible that language acts as a coordination device for some questions, influencing choices by altering people's expectations of others' behaviour (see [Clist, 2023b](#), and the references therein). For example, a preference for cooperation could be contingent upon the expected cooperation level of others. If language alters our expectation of others' behaviour, then the language effect could be explained not by an inherently greater desire for cooperation in one language, or by an elicitation bias, but by higher expectations of cooperation in that language. [Li \(2017\)](#) explores this hypothesis, finding a language difference in strategic interaction games, but no effect for non-strategic elicitations of social preferences. He interprets this as evidence for an expectation-based explanation, at least among his 64 Hong Kong subjects. In other settings, [Clist \(2023b\)](#) discusses cases where language effects on cooperation are not accompanied by different expectations. In our own results, we find large effects for trusting behaviour, with some significant results for positive reciprocity and altruism, but no significant effects for negative reciprocity. The large effects for trusting behaviour is broadly compatible with an expectations-based mechanism, but this mechanism cannot explain our results more broadly.

4.4. Are treatment effects evidence of different preferences or biases?

We have shown elicited preferences differ in a sizeable and significant fashion, by language, country and nationality. Our main contribution is to show that these language differences are present across a wide range of preference domains (where previous research mostly examined one domain), and to show that the language differences are not unambiguously in the direction of country differences, and so cannot be explained using a standard model of assimilation. We have also discussed possible confounds and mechanisms for these effects, with none providing a compelling explanation for our results.

A reader may wonder whether the differences in elicited preferences are better understood as differences in 'true' preferences, or merely biases related to elicitation.¹² Different schools of thought within behavioural economics and psychology would understand our results in quite different ways (for a textbook treatment, see chapter 2, [Bardsley et al., 2010](#)). [Stigler and Becker \(1977\)](#) argue that people's tastes do not change easily, nor do they differ meaningfully between people. Any change in behaviour should not be understood as a change in tastes or preferences, but in the ability to turn a given good into utility. [Cowen \(1989, p.134\)](#) responds critically: "[a]ny attempt... to salvage the assumption of constant tastes... only pushes the problem back one step".

[Stigler and Becker's \(1977\)](#) argument is echoed in later discussions of the discovered preference hypothesis ([Binmore, 1999](#); [Plott, 1996](#); [Smith, 1994](#)). This posits that people have rational and consistent preferences: they merely need sufficient deliberation, incentives and experience to 'discover' their own preference. In this view one would expect our language differences to essentially be the result of elicitation bias; if subjects had more time, incentive or deliberation in each language, effects should disappear. (Our location and nationality effects would be understood as reflecting different underlying preferences, in line with the discovered preference hypothesis but not [Stigler and Becker's](#) view.)

There are dissenting voices. Proponents of the constructed preference hypothesis may argue that our language effects are not the result of different biases, but of different preferences in each language (see [Bruni & Sugden, 2007](#), and the references therein). Essentially it would imply preferences are endogenous to the language in which they are elicited.¹³

Our data does not allow us to say whether language differences in elicited preferences are the result of elicitation bias, or fundamentally different 'true' preferences, or a combination of the two. Rather, we can simply assert that there is a difference, and the direction of this difference is best understood using *identity priming*, and not a model of assimilation. Future work would need to provide the necessary conditions to see whether language effects change with greater experience, deliberation or incentive.

¹² Our thanks to the editor for suggesting this point. There are several large literatures on related issues. See [Sugden \(2022\)](#) for a critical view on the attempt to separate 'true preferences' from 'biases' in the context of stated preference surveys. For an empirical test of [Stigler and Becker's \(1977\)](#) theory (discussed in the text), see [Dasgupta, Gangadharan, Maitra, and Mani \(2017\)](#). On elicitation and preference stability, see [Harrison's \(2015\)](#) critical review of a popular book, which provides a brief overview of preference consistency and the difficulties of elicitation in a domain we do not measure (risk). For more recent work on biases and true preferences, see [Alós-Ferrer, Fehr, and Garagnani \(2023\)](#) in another domain we do not consider: transitive preferences, which finds a substantial minority of choices are due to nontransitive choices, which cannot be explained by noise or elicitation bias.

¹³ We have included Binmore in the discovered preference hypothesis school, though he argues that the discovered preference could be path-dependent, and so the language in which the process of discovery occurs could colour the outcome (see [Bruni & Sugden, 2007](#)).

5. Conclusion

We measure a wide range of (mostly stated) preferences for four groups. Simple comparisons provide observational estimates of how they differ by country, location, and nationality. We estimate one causal effect, finding that randomising the language of the survey leads to average differences of 0.4 standard deviations. This is slightly larger than cross-country differences, and does not appear to be caused by Whorfian effects or the Foreign Language Effect. Rather the large language effects appear to be due to language acting as a trigger for identity-based preferences (Akerlof & Kranton, 2000; Fehr & Hoff, 2011) where those identities have been absorbed.

The nationality effect is the largest of the four: Chinese students in the UK speaking English differ by an average of 0.42 standard deviations from other students in the UK. Prima facie large differences by nationality appear to concur with previous research, which finds assimilation is slow (Abramitzky et al., 2020; Alesina & Fuchs-Schündeln, 2007; Bisin & Verdier, 2011; Fernández, 2011). However, the underlying model of this research is that there are two cultural influences. As time passes, people's attachment to old influences weaken, gradually displaced by new ones (Verdier, Manning, Bisin, & Algan, 2012). By using international students, we observe the dynamics of assimilation for a relevant and receptive population over the first few years. Our results do not provide support for slow assimilation, as students do not simply move towards UK-style preferences. In English, Chinese students studying in the UK do not resemble their British or Beijing-based counterparts.

Our results have two implications for understanding different determinants of elicited preferences. First, we cannot gain a complete picture of migrants' preferences in only one language. In English, UK-based Chinese students' responses are far from those of their Chinese counterparts. Yet when asked in Mandarin there is no significant difference between them and the Beijing sample. This implies future research into assimilation would benefit by measuring preferences in two languages. Second, the pull identity exerts on preferences can be turned down by language, not only increased (Akerlof & Kranton, 2000; Benjamin et al., 2016, 2010). This implies a middle step in assimilation. People may shake off their old preferences and in the long run adopt new ones, as argued elsewhere. However, between abandoning old preferences and adopting new ones, they may simply be relatively free of identity-based preferences. This result is in line with dynamic constructivism with monocultural subjects (Clist & Verschoor, 2017; Hong et al., 2000). For people that are bicultural, switching languages may switch cultural frames. For more short-term migrants, a new language simply deactivates older cultural influences.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Everything needed to replicate all of our results, and the experimental scripts, are available at <https://doi.org/10.7910/DVN/Z9CANV>.

Acknowledgements

We received helpful comments from Sean Roberts, Toman Barsbai, Arjan Verschoor, Mike Brock, Joshua Hill, Gabriele Restelli, Daniel J. Benjamin, the editor Carlos Alós-Ferrer and two anonymous referees. We would also like to thank attendants at talks for CBESS and BEDERG. We received excellent research assistance from Ines Ferreira, Mingpei Li, Han Lin, Borja Perez-Viana Martinez, Huizhong Wang, and Heran Zhang.

Funding

This research was partially supported by a Senior Research Fellow Scheme (Grant No. SRFS2122-4H01) from the Research Grant Council of Hong Kong SAR government and a research grant from Beijing Normal University, China awarded to Ying-Yi Hong, and a grant from DEVCo (UEA) awarded to Paul Clist.

Compliance with Ethical Standards

The research was accepted by the International Development Research Ethics Committee at the School of International Development, project code: 1998285A.

References

- Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4), 1115–1153.
- Abramitzky, R., Boustan, L., & Eriksson, K. (2020). Do immigrants assimilate more slowly today than in the past? *American Economic Review: Insights*, 2(1), 125–141.

- Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *Quarterly Journal of Economics*, 115(3), 715–753.
- Alempaki, D., Doğan, G., & Yang, Y. (2021). Lying in a foreign language? *Journal of Economic Behaviour and Organization*, 185, 946–961.
- Alesina, A., & Fuchs-Schündeln, N. (2007). Goodbye Lenin (or not?): The effect of communism on people's preferences. *American Economic Review*, 97(4), 1507–1528.
- Algan, Y., Bisin, A., & Verdier, T. (2012). Introduction: Perspectives on cultural integration of immigrants. In *Cultural integration of immigrants in Europe* (pp. 1–48). Oxford: Oxford University Press.
- Alós-Ferrer, C., Fehr, E., & Garagnani, M. (2023). *Identifying nontransitive preferences: Working Paper Series, 415*, University of Zurich Department of Economics.
- Alós-Ferrer, C., & Yechiam, E. (2020). At the eve of the 40th anniversary of the Journal of Economic Psychology: Standards, practices, and challenges. *Journal of Economic Psychology*, 80, Article 102309.
- Angerer, S., Glätzle-Rützler, D., Lergetporer, P., & Sutter, M. (2021). The effects of language on patience: an experimental replication study of the linguistic-savings hypothesis in Austria. *Journal of the Economic Science Association*, 7(1), 88–97.
- Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., & Sugden, R. (2010). *Experimental economics: Rethinking the rules*. Princeton University Press.
- Benjamin, D. J., Choi, J. J., & Fisher, G. (2016). Religious identity and economic behavior. *The Review of Economics and Statistics*, 98(4), 617–637.
- Benjamin, D. J., Choi, J. J., & Strickland, A. J. (2010). Social identity and preferences. *American Economic Review*, 100(4), 1913–1928.
- Berry, J. W. (1997). Immigration, acculturation, and adaptation. *Applied Psychology*, 46(1), 5–34.
- Binmore, K. (1999). Why experiment in economics? *The Economic Journal*, 109(453), 16–24.
- Bisin, A., & Verdier, T. (2011). The economics of cultural transmission and socialization. In *Handbook of social economics, Vol. 1* (pp. 339–416). Elsevier.
- Bleakley, H., & Chin, A. (2010). Age at arrival, English proficiency, and social assimilation among US immigrants. *American Economic Journal: Applied Economics*, 2(1), 165–192.
- Bruni, L., & Sugden, R. (2007). The road not taken: how psychology was removed from economics, and how it might be brought back. *The Economic Journal*, 117(516), 146–173.
- Cameron, L., Erkal, N., Gangadharan, L., & Zhang, M. (2015). Cultural integration: Experimental evidence of convergence in immigrants' preferences. *Journal of Economic Behaviour and Organization*, 111, 38–58.
- Chakravarty, S., & Fonseca, M. A. (2014). The effect of social fragmentation on public good provision: an experimental study. *Journal of Behavioral and Experimental Economics*, 53, 1–9.
- Chen, M. K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review*, 103(2), 690–731.
- Chen, J. I., & He, T.-S. (2021). Discounting from a distance: The effect of pronoun drop on intertemporal decisions. *Journal of Economic Psychology*, 87, Article 102454.
- 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.
- Circi, R., Gatti, D., Russo, V., & Vecchi, T. (2021). The foreign language effect on decision-making: A meta-analysis. *Psychonomic Bulletin & Review*, 28, 1131?1141.
- Clist, P. (2023a). [Dataset] replication data for: Do international students learn foreign preferences? The interplay of language, identity and assimilation. *Harvard Dataverse*, <http://dx.doi.org/10.7910/DVN/Z9CANV>.
- Clist, P. (2023b). Language and cooperation. In B. Kebede (Ed.), *Encyclopedia of experimental social science*. Elgar, volume forthcoming, available at: <https://paulclit.github.io/>.
- Clist, P., & Verschoor, A. (2017). Multilingualism and public goods provision: An experiment in two languages in Uganda. *Journal of Development Economics*, 129, 47–57.
- Cowen, T. (1989). Are all tastes constant and identical?: A critique of Stigler and Becker. *Journal of Economic Behaviour and Organization*, 11(1), 127–135.
- Dasgupta, U., Gangadharan, L., Maitra, P., & Mani, S. (2017). Searching for preference stability in a state dependent world. *Journal of Economic Psychology*, 62, 17–32.
- Davis, L. S., & Abdurazokzoda, F. (2016). Language, culture and institutions: Evidence from a new linguistic dataset. *Journal of Comparative Economics*, 44(3), 541–561.
- Del Maschio, N., Crespi, F., Peressotti, F., Abutalebi, J., & Sulpizio, S. (2022). A meta-analysis of the Foreign Language Effect. *Bilingualism: Language and Cognition*, 25(4), 617–630.
- Díaz-Lago, M., & Matute, H. (2019). Thinking in a Foreign language reduces the causality bias. *Quarterly Journal of Experimental Psychology*, 72(1), 41–51.
- Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences. *Quarterly Journal of Economics*, 133(4), 1645–1692.
- Falk, A., Becker, A., Dohmen, T., Huffman, D., & Sunde, U. (2023). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science*.
- Fehr, E., & Hoff, K. (2011). Introduction: Tastes, castes and culture: the influence of society on preferences. *The Economic Journal*, 556(121), F396–F412.
- Feldmann, H. (2019). Do linguistic structures affect human capital? The case of pronoun drop. *Kyklos*, 72(1), 29–54.
- Fernández, R. (2011). Does culture matter? In *Handbook of social economics, Vol. 1* (pp. 481–510). Elsevier.
- Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3), 525–547.
- Forsyth, D. R. (1980). A taxonomy of ethical ideologies. *Journal of Personality and Social Psychology*, 39(1), 175.
- Forsyth, D. R., & Berger, R. E. (1982). The effects of ethical ideology on moral behavior. *The Journal of Social Psychology*, 117(1), 53–56.
- Forsyth, D. R., O'Boyle, E. H., & McDaniel, M. A. (2008). East meets West: A meta-analytic investigation of cultural variations in idealism and relativism. *Journal of Business Ethics*, 83(4), 813–833.
- Gächter, S., & Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature*, 531(7595), 496–499.
- Galor, O., & Özak, Ö. (2016). The agricultural origins of time preference. *American Economic Review*, 106(10), 3064–3103.
- Galor, O., Özak, Ö., & Sarid, A. (2020). Linguistic traits and human capital formation. *American Economic Review: Papers and Proceedings*, 110, 309–313.
- Gao, S., Zika, O., Rogers, R. D., & Thierry, G. (2015). Second language feedback abolishes the 'hot hand' effect during even-probability gambling. *Journal of Neuroscience*, 35(15), 5983–5989.
- Harrison, G. W. (2015). Book review, *Risky Curves: On the empirical failure of Expected Utility*, Friedman Daniel, Isaac R. Mark, James Duncan, Sunder Shyam. *Journal of Economic Psychology*, 48, 121–125.
- He, T.-S., Riyanto, Y. E., Tanaka, S. C., & Yamada, K. (2020). Pronoun drop and prosocial behavior: Experimental evidence from Japan. *Journal of the Economic Science Association*, 1–13.
- Heine, S. J., & Lehman, D. R. (2004). Move the body, change the self: Acculturative effects on the self-concept. In *Psychological foundations of culture, Vol. 8* (pp. 305–331).
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., et al. (2001). In search of homo economicus: behavioral experiments in 15 small-scale societies. *American Economic Review*, 91(2), 73–78.
- Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., et al. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science*, 327(5972), 1480–1484.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., et al. (2006). Costly punishment across human societies. *Science*, 312(5781), 1767–1770.

- Herz, H., Huber, M., Maillard-Bjedov, T., & Tyahlo, S. (2021). Time preferences across language groups: Evidence on intertemporal choices from the Swiss language border. *The Economic Journal*, 131(639), 2920–2954.
- Hong, Y.-y., Morris, M. W., Chiu, C.-y., & Benet-Martinez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American Psychologist*, 55(7), 709–720.
- Hugh-Jones, D. (2016). Honesty, beliefs about honesty, and economic growth in 15 countries. *Journal of Economic Behaviour and Organization*, 127, 99–114.
- Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., et al. (2014). *World values survey: Round five—country-pooled datafile 2005–2008*. Madrid: JD Systems Institute.
- Jakiela, P., & Ozier, O. (2018). *Gendered language: Policy Research Working Paper, 8464*, World Bank.
- Ji, L.-J., Zhang, Z., & Nisbett, R. E. (2004). Is it culture or is it language? Examination of language effects in cross-cultural research on categorization. *Journal of Personality and Social Psychology*, 87(1), 57–65.
- Johnson, N. D., & Mislin, A. (2012). How much should we trust the World Values Survey trust question? *Economics Letters*, 116(2), 210–212.
- Jones, D., Molitor, D., & Reif, J. (2019). What do workplace wellness programs do? Evidence from the Illinois workplace wellness study. *Quarterly Journal of Economics*, 134(4), 1747–1791.
- Kashima, E. S., & Kashima, Y. (1998). Culture and language: The case of cultural dimensions and personal pronoun use. *Journal of Cross-Cultural Psychology*, 29(3), 461–486.
- Lambarraa, F., & Riener, G. (2015). On the norms of charitable giving in Islam: Two field experiments in Morocco. *Journal of Economic Behaviour and Organization*, 118, 69–84.
- Li, K. K. (2017). How does language affect decision-making in social interactions and decision biases? *Journal of Economic Psychology*, 61, 15–28.
- Licht, A. N., Goldschmidt, C., & Schwartz, S. H. (2007). Culture rules: The foundations of the rule of law and other norms of governance. *Journal of Comparative Economics*, 35(4), 659–688.
- Plott, C. (1996). Rational individual behaviour in markets and social choice processes: The discovered preference hypothesis. In K. Arrow, E. Colombatto, M. Perlman, & C. Schmidt (Eds.), *The rational foundations of economic behaviour* (pp. 225–250). Basingstoke: International Economic Association and Macmillan.
- Rai, T. S., & Holyoak, K. J. (2013). Exposure to moral relativism compromises moral behavior. *Journal of Experimental Social Psychology*, 49(6), 995–1001.
- Rieger, M. O., Wang, M., & Hens, T. (2014). Risk preferences around the world. *Management Science*, 61(3), 637–648.
- Roberts, S. G., Winters, J., & Chen, K. (2015). Future tense and economic decisions: Controlling for cultural evolution. *PLoS One*, 10(7).
- Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M., & Zamir, S. (1991). Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. *The American Economic Review*, 81(5), 1068–1095.
- Smith, V. L. (1994). Economics in the laboratory. *Journal of Economic Perspectives*, 8(1), 113–131.
- Soffiatti, J. P. (1955). Bilingualism and biculturalism. *Journal of Educational Psychology*, 46(4), 275–277.
- Stankovic, M., Biedermann, B., & Hamamura, T. (2022). Not all bilinguals are the same: A meta-analysis of the moral Foreign Language Effect. *Brain and Language*, 227, Article 105082.
- Stigler, G. J., & Becker, G. S. (1977). De gustibus non est disputandum. *The American Economic Review*, 67(2), 76–90.
- Sugden, R. (2022). Debiasing or regularisation? Two interpretations of the concept of ‘true preference’ in behavioural economics. *Theory and Decision*, 92(3–4), 765–784.
- Sutter, M., Angerer, S., Glätzle-Rützler, D., & Lergetporer, P. (2018). Language group differences in time preferences: Evidence from primary school children in a bilingual city. *European Economic Review*, 106, 21–34.
- Tabellini, G. (2008). Institutions and culture. *Journal of the European Economic Association*, 6(2–3), 255–294.
- UNESCO Institute for Statistics (2020). *Education statistics*. available at <http://data.uis.unesco.org/Index.aspx?queryid=172>.
- Verdier, T., Manning, A., Bisin, A., & Algan, Y. (2012). *Cultural integration of immigrants in Europe*. Oxford University Press.
- Vieider, F. M., Lefebvre, M., Bouhouicha, R., Chmura, T., Hakimov, R., Krawczyk, M., et al. (2015). Common components of risk and uncertainty attitudes across contexts and domains: Evidence from 30 countries. *Journal of the European Economic Association*, 13(3), 421–452.
- Westfall, P. H., & Young, S. S. (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*, Vol. 279. John Wiley & Sons.