Journal of Cognitive Psychology

ISSN: 2044-5911 (Print) 2044-592X (Online) Journal homepage: https://www.tandfonline.com/loi/pecp21

The spatial mapping of concepts in English and Mandarin

Qiong Wu, Evan Kidd & Stephanie C. Goodhew

To cite this article: Qiong Wu, Evan Kidd & Stephanie C. Goodhew (2019): The spatial mapping of concepts in English and Mandarin, Journal of Cognitive Psychology, DOI: 10.1080/20445911.2019.1663354

To link to this article: https://doi.org/10.1080/20445911.2019.1663354

Published online: 06 Sep 2019.

JOURNAL OF COGNITIVE PSYCHOLOGY
https://doi.org/10.1080/20445911.2019.1663354

The spatial mapping of concepts in English and Mandarin

Qiong Wu[a], Evan Kidd[a,b,c] and Stephanie C. Goodhew[a]

[a]Research School of Psychology, The Australian National University, Canberra, Australia; [b]ARC Centre of Excellence for the Dynamics of Language, Canberra, Australia; [c]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

ABSTRACT

English speakers have been shown to map abstract concepts in space, which occurs on both the vertical and horizontal dimensions. For example, words such as God are associated with up and right spatial locations, and words such as *Satan* with down and left. If the tendency to map concepts in space is a universal property of human cognition, then it is likely that such mappings may be at least partly culturally-specific, since many concepts are themselves language-specific and therefore cultural conventions. Here we investigated whether Mandarin speakers report spatial mapping of concepts, and how these mappings compare with English speakers (i.e. are words with the same meaning associated with the same spatial locations). Across two studies, results showed that both native English and Mandarin speakers reported spatial mapping of concepts, and that the distribution of mappings was highly similar for the two groups. Theoretical implications are discussed.

ARTICLE HISTORY
Received 14 December 2018
Accepted 27 August 2019

KEYWORDS
Language; space; concepts; English; Mandarin

English speakers often invoke spatial locations when conceptualising and communicating both concrete and abstract concepts. The mapping of concrete concepts conforms to the typical spatial location of the relevant object in the environment (Estes, Verges, & Barsalou, 2008; Zwaan & Yaxley, 2003). For example, sky is associated with *up* and grass *down*, and we typically see the sky above us and the grass below us. Abstract concepts, which have no physical or tangible locations in space, also take on these mappings (Lakoff & Johnson, 1999). For example, positive emotions are described as related to *up* (such as *cheer up*), while negative emotions are often represented by expressions related to *down* (such as feeling *down in the dumps*). Consistent with these linguistic patterns, English speakers have been found to associate positive terms such as *happy* with *up*, and negative terms such as *sad* with *down* (Goodhew, McGaw, & Kidd, 2014; Gozli, Chasteen, & Pratt, 2013). Furthermore, these associations between abstract concepts and space are not limited to the vertical dimension, but are also observed on the horizontal dimension. For example, the word *God* is associated with both *up* and *right*, whereas *Devil* is associated with both *down* and *left* (Chasteen, Burdzy, & Pratt, 2010). While exactly how abstract concepts inherit spatial affordances is the subject of debate, the language use framework suggests that patterns of language use may influence how abstract concepts are spatial-mapped (Goodhew et al., 2014). Thus, it is possible that people who use different languages emerging from distinct cultural backgrounds could have different patterns of spatial mapping of concepts.

Studies that have reported on the spatial mapping of concepts have typically examined the phenomenon in languages which share some common cultural heritage. This includes those that belong to Indo-European language family, such as English (Crawford, Cohn, & Kim, 2014; Goodhew et al., 2014; Zwaan & Yaxley, 2003), German (Dudschig, De la Vega, & Kaup, 2015), and Spanish (Ouellet, Santiago, Funes, & Lupiáñez, 2010; Santiago, Lupianez, Perez, & Funes, 2007). While these studies have not made direct cross-cultural comparisons, they have shown the presence of the spatial mapping of concepts in these languages, and the pattern appears to be similar across them. The present study, in contrast, directly compared

CONTACT Stephanie C. Goodhew Stephanie.Goodhew@anu.edu.au

English versus Mandarin speakers in their spatial mapping of concepts. Mandarin is a branch of the Sino-Tibetan language family and speakers of its languages constitute a distinct cultural group from speakers of English. This means that examining the spatial mapping of concepts in this group has the potential to a provide unique insight into how culture as encoded in language may shape this phenomenon. There is some preliminary evidence for the presence of the spatial mapping of time-based concepts in Mandarin (Boroditsky, Fuhrman, & McCormick, 2011; Zhou & Fan, 2015). However, to our knowledge, this is the first study to systematically investigate the spatial mapping of a broad array of concept words in this language.

There are several possible ways in which the spatial mapping of concepts could play out across language and cultures. Firstly, the phenomenon could be invariant, such that the same concepts are mapped similarly. Alternatively, while the general tendency to locate concepts in space might be a universal tendency, the specific associations formed might be specific to particular linguistic groups, especially those that speak genetically distinct languages that are spoken in different cultures. There is some existing evidence for universality, including similarities in how English, Arabic, and Hebrew speakers represent concepts such as quantity and preference (Tversky, Kugelmass, & Winter, 1991). In contrast, there is also some existing evidence for this language-specificity in relation to how some concepts are mapped in space. In particular, English speakers have been found to arrange temporal sequences from left to right, whereas Hebrew and Arabic speakers have been found to arrange them right to left, consistent with reading direction in each of the language (Fuhrman & Boroditsky, 2010; Tversky et al., 1991). Similarly, English, Hebrew, and Arabic-reading participants have been found to differ with respect to their spatial representation of number (Shaki, Fischer, & Petrusic, 2009), and Italian and Arab participants to differ in how the spatial sequence of simple sentences in a sentence-picture matching task (Maass & Russo, 2003). However, the spatial mapping of abstract concepts has yet to be investigated more systematically or comprehensively across a diverse array of abstract concepts, and across the two very distinct languages, namely English and Mandarin. This is what was done here.

Abstract concepts were chosen as the focus of the study because we reasoned that it is more likely that these would show differences between the linguistic groups, if there are differences to be found. This is because such associations are not constrained by physical realities of the world, and therefore have greater opportunity to be shaped by cultural factors. In particular, the language-use hypothesis (Goodhew et al., 2014) predicts that language use patterns, which may be specific to particular languages, should play a role in determining how abstract concepts are mapped in space. In contrast, concrete words which usually refer to prototypical objects are more likely to show uniformity across cultures (e.g. physically *sky* is up and *grass* is down, irrespective of linguistic background), assuming that they may at least in part be shaped by sensorimotor factors. This does not preclude a role for language in shaping the mappings for concrete words, however, the mappings for abstract concept words are likely most sensitive to revealing this influence. Accordingly, here we present two studies comparing the spatial mapping of concepts in English and Mandarin across a wide array of abstract concepts.

Study 1

The purpose of Study 1 was to compare and contrast the similarities and differences between English and Mandarin speakers in the how they associate concepts with space. The spatial mapping of 151 concept words was assessed via questionnaire, and the distribution of ratings was then compared between the two groups. A two-alternative forced-choice rating scale was used, in which participants were required to choose between two alternative spatial associations on the vertical (i.e. up/down) and horizontal (i.e. left/right) dimensions. They also rated the words for their valence (positive/negative).

We conducted several analyses to assess the nature of the spatial mappings both within each language groups, as well as between the two language groups. Firstly, we examined the correlation between the three rated dimensions within each language. Previous research has suggested that items that are associated with *up* also tend to be associated with *right* (and correspondingly *down* with *left*) (Chasteen et al., 2010). Examining the correlations between the vertical and horizontal dimensions provides insight into whether this pattern is also observed here, in both language groups. Given the scoring (described below in the Method section), a positive correlation signifies relationships between *up* and *right* on the one hand, and *down* and *left* on the other. Furthermore, we assessed the correlation between valence and each of the vertical and horizontal dimensions. Here, positive correlations indicate that the *up* and *right* are associated with *positive*, and *down* and *left* with *negative*. Moreover, comparing whether these correlations differ between the two language groups reveals whether the pattern of associations are similar or different between them. Additionally, correlating the ratings for each item provided by the English versus Mandarin language group indicates whether the rank ordering of the items is similar or different between the two, as will examining the distributions of ratings for both language groups. Finally, we also examined which individual items attract the most extreme ratings on each dimension, as well as the most similar and the most discrepant ratings between the two groups on each dimension and offer some discussion of why this occurred.

Method

Participants

All participants were students at The Australian National University. All provided informed consent, and the study protocol was approved by ANU's Human Research Ethics Committee. Native English and Mandarin speakers were recruited. A total of 280 participants volunteered in exchange for course credit. Theirs average age was 20.8 years (SD = 2.7). The sample contained 73 males and 161 females.

Materials and procedure

Participants completed the questionnaire online via Qualtrics. First, after reading the information sheet, participants were asked to provide explicit consent to participate. Then, initial questions asked about participants' demographic information (age, gender, country of birth, and first language). If participants selected Chinese[1] as their first language, then they were instructed on how to change the default setting of the survey into modern simplified Chinese, and to explicitly confirm compliance with this instruction. For the reasons articulated in Note 1, while the recruitment material stated *Chinese*, this group reflected predominately if not exclusively individuals whose native language was *Mandarin*. Hereafter, therefore, this grouped will be referred to as the *Mandarin* language group.

In the main part of the questionnaire, participants were presented with 151 words that were selected from the Conceptual Cueing Database (Goodhew & Kidd, 2016). The words were selected on the basis that (a) the list was compromised of predominately abstract concept words, as it seems likely that abstract concept spatial mappings have a greater chance of differing across languages than concrete word mappings, and (b) Qualtrics' automatic translation from English into simplified Chinese was appropriate (as confirmed by the first author, a native Mandarin speaker). The word lists can be seen here: https://osf.io/qd8xr/. Participants were instructed to rate each word across three dimensions: vertical (up / down), horizontal (left / right), and valence (positive / negative). Specifically, they were asked to indicate whether each of the words associated with up or down, then left or right, and then positive or negative. They were told that there were no correct or incorrect answers, and instead they were encouraged to respond according to their first thought. Each word occupied a row, and there were three distinct columns, one for the vertical dimension, one for the horizontal dimension, and one for valence, each subdivided into the two possible response options. The first two dimensions provided information about the spatial mappings, and the valence question provided insight into whether the word was conceptualised the same way in the two languages. The questions were two-alternative response options with no neutral point. This was done to avoid potential response-criteria differences between the groups. The logic here was that if an item has no clear spatial association, then ratings should approximate 50% of each response option. Participants could leave items blank, in which case the questionnaire prompted them to complete the items, but this was not a requirement for completing the questionnaire.

Results & discussion

Participant details

Four participants were removed from analysis because they nominated a language other than Mandarin or English as their native language. Eight participants were removed because they did not answer affirmatively to altering the language of presentation when instructed, and 34 participants were excluded due to excessive missing data (>10 items were not answered), leaving a final sample of 234 participants. Of these, 110 identified

Mandarin as their native language, and 124 identified English as their native language. Of the 110 Mandarin speakers, 107 identified China as their country of birth (3 Other). Of the 124 English speakers, 101 identified Australia as their country of birth (23 Other).

Response scoring

Responses were scored as 1 for *Up, Right,* and *Positive,* and 0 for *Down, Left,* and *Negative*. An average rating for each item was then calculated by taking the mean response across participants in each group. Any cells where there was missing data (≤10 per participant) were simply omitted from the computation of averages for each condition. Raw data can be found here: https://osf.io/yht4q/.

English inter-dimensions analysis

Next, we sought to establish the relationships between the three dimensions (vertical, horizontal, and valence) for the English speakers. To do so, we calculated the correlation between the mean rating for each of the 151 items by the English speakers on each dimension with each other dimension. This revealed strong positive correlations (all correlation coefficients reported in the results section for this and the following study are Spearman's rho) between the three dimensions: vertical and horizontal ratings ($r = .86$, $p < .001$), vertical and valence ($r = .89$, $p < .001$), and horizontal and valence ($r = .84$, $p < .001$). In other words, there was a high degree of correspondence between the ratings on the dimensions amongst English speakers.

The positive correlation between the vertical and horizontal dimensions means that the items that were rated as more strongly associated with *up*, were also more strongly associated with *right* than *left,* whereas items more strongly associated with *down* were also more strongly associated with *left* than *right*. The positive correlation between the vertical and valence dimension indicates that the more positively-rated an item was, the stronger its association with *up* rather than *down*, and vice versa for negative-rated items. The positive correlation between the horizontal and valence dimension implies that items that attracted stronger *right* ratings also received more positive rather than negative ratings, whereas items that attracted stronger *left* ratings received more negative than positive ratings. These associations are consistent with previous work with English speakers showing behavioural associations between positive-valence religious words (e.g. *God*) and *up* and *right* spatial dimensions, and associations between negative-valence religious words (e.g. *Devil*) and *down* and *left* (Chasteen et al., 2010). The association between the directions on the horizontal plane and valence has long been a feature of European Judeo-Christian culture, where, for instance, Jesus "sat at the right hand of God" and the Latin (and modern Italian) word for left is *sinistra*, which in English is synonymous with *evil* (i.e. sinister) (Casasanto & Chrysikou, 2011).

The correlations for the different dimensions all appear quite similar to one another. Diedenhofen and Musch's (2015) comparison of correlations (two-tailed test) indicated that the correlation between the vertical and horizontal dimensions ($r = .86$) was not reliably different from that between the vertical and valence dimensions ($r = .89$), $p > .05$[2], 95% CI for difference = [−.07 to +.01]. In contrast, the correlation between the horizontal and valence dimensions ($r = .84$) was reliably different from (i.e. lower than) the correlation between the vertical and valence dimensions, $p < .05$, CI = [+.01 to +.09]. The correlation between vertical and horizontal was not found to be reliably different from the correlation between the horizontal and valence dimension ($p > .05$), CI = [−.02 to .06]. In summary, the single strongest association was the one between the vertical and valence dimensions. That is, the strongest tendency was for words that received consistently stronger *up* ratings to also receive stronger *positive* ratings, whereas items that received more *down* ratings also received stronger *negative* ratings.

From Figures 1–3, it appears as though the spatial mappings may be more extreme for the vertical dimension than the horizontal dimension. The mode is an appropriate measure of central tendency to illustrate where the bulk of the distribution sits. However, it is clear that the distributions are bimodal rather than unimodal. Therefore, to test this, we calculated the mode of each half of the distribution (i.e. computed the mode for the scores <0.5, and then the mode for the scores >0.5) for the vertical and horizontal dimension for English speakers. This provides insight into the extremeness of the spatial mapping of concepts, because if the mode is more extreme (i.e. further from the midpoint of the response scale, 0.5), then it suggests that the spatial mapping of concepts is more extreme. In contrast, if the mode if closer to the centre of the response scale, then it suggests that

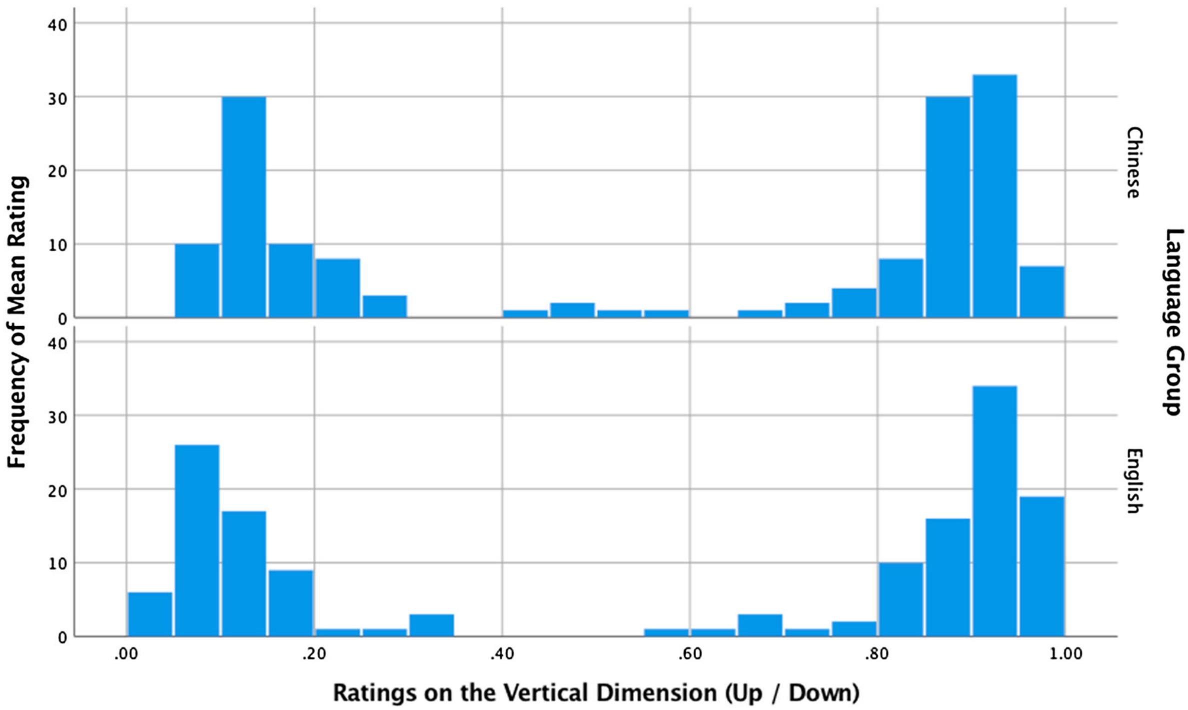

Figure 1. Distribution plots of the frequency of items with given mean scores on the vertical dimension in Study 1 (0 = down, 1 = up). Mandarin speakers upper panel, English speakers lower panel.

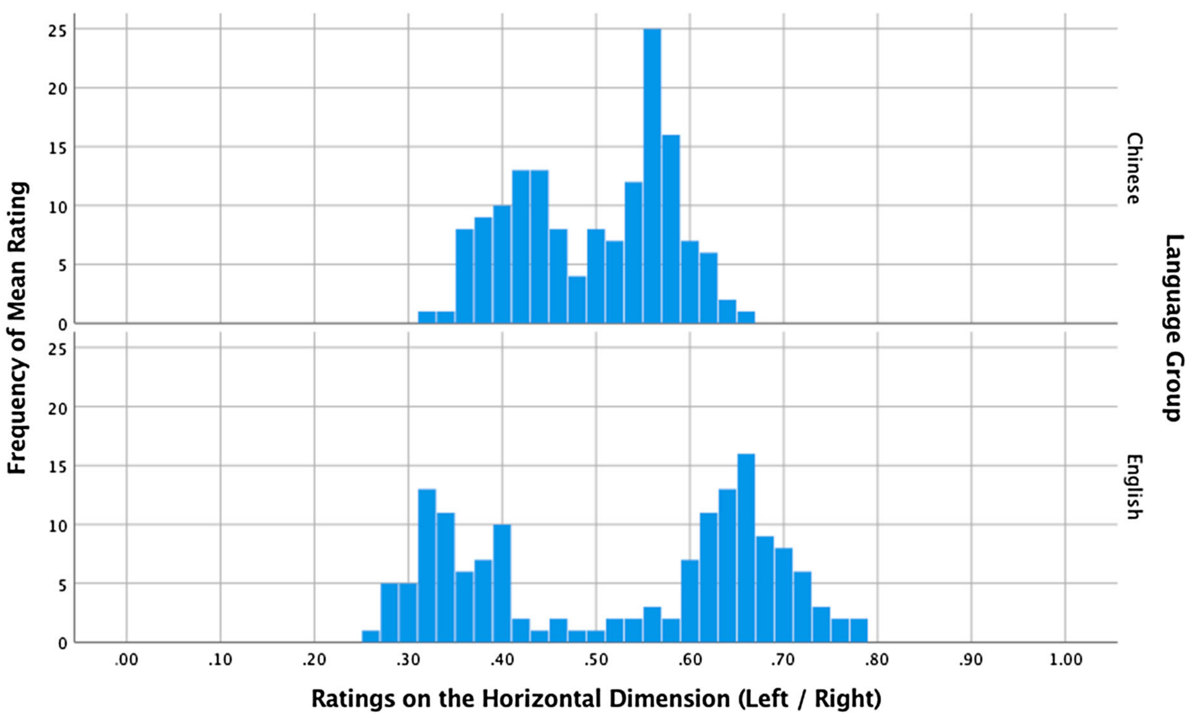

Figure 2. Distribution plots of the frequency of items with given mean scores on the horizontal dimension in Study 1 (0 = left, 1 = right). Mandarin speakers upper panel, English speakers lower panel.

the spatial mapping of concepts is less extreme. For the vertical ratings, the mode for the lower part of the distribution was .09, whereas the mode for the higher scores was .93. For the horizontal dimension, the mode for the lower scores was 0.32 (i.e. closer to the mid-point than that for vertical), and for the

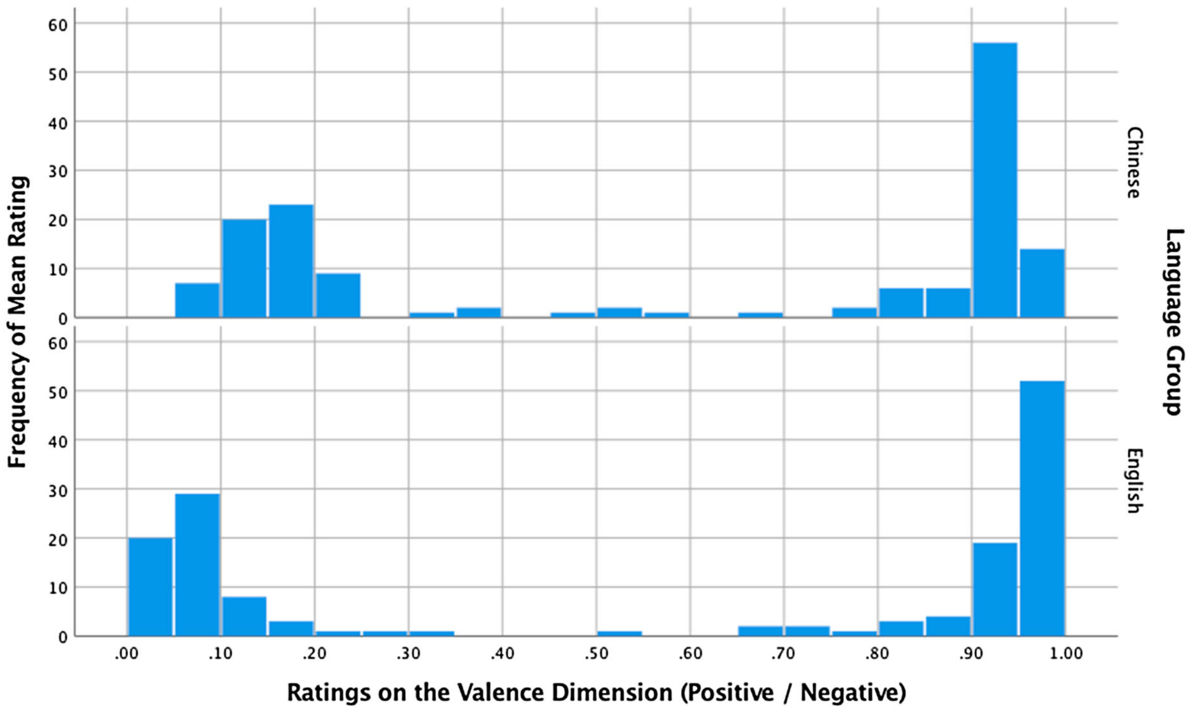

Figure 3. Distribution plots of the frequency of items with given mean scores on the valence dimension in Study 1 (0 = negative, 1 = positive). Mandarin speakers upper panel, English speakers lower panel.

higher scores was .65 (i.e. closer to the mid-point than for vertical). This suggests that there is more pronounced spatial mapping of concepts along the vertical than the horizontal dimension for English speakers.[3]

Mandarin inter-dimensions analysis

We next sought to assess the relationships between the three dimensions (vertical, horizontal, and valence) for the Mandarin speakers. As we saw for the English speakers, this revealed strong positive correlations between the three dimensions: vertical and horizontal ratings ($r = .75$, $p < .001$), vertical and valence ($r = .90$, $p < .001$), and horizontal and valence ($r = .78$, $p < .001$). This means that words that received strong *up* associations also tended to receive strong *right* associations, and that items that were rated strongly as either *up* or *right* were also more likely to be rated as *positive* rather than *negative*. In other words, there was a high degree of correspondence across the dimensions in their ratings amongst Mandarin speakers, although the strongest correspondence was that between the vertical and valence dimensions. This was confirmed via Diedenhofen and Musch's (2015) comparison of correlations, which indicated that the correlation between the vertical and valence dimensions ($r = .90$) was greater than the correlation between the vertical and horizontal dimensions ($r = .75$), $p < .001$, 95% CI for the difference between the correlations = [+.10 to +.22], and also greater than the correlation between the horizontal and valence dimensions ($r = .78$), $p < .001$, CI = [+.07 to +.19], whereas the correlation between vertical and horizontal did not reliably differ from the correlation between horizontal and valence ($p > .05$), CI = [−.08 to +.02]. This means that the most pronounced pattern of association was such that items that received greater *up* (versus *down*) ratings, also received greater *positive* (versus *negative*) ratings.

From Figures 1–3, it appears that the spatial mappings may be more extreme for the vertical dimension than the horizontal dimension. To test this, we calculated the mode of each half of the distribution (i.e. computed the mode for the scores <0.5, and then the mode for the scores >0.5) for the vertical and horizontal dimension for the Mandarin group. For the vertical ratings, the mode for the lower scores was .13 whereas the mode for the higher scores was .90. For the horizontal dimension, the mode for the lower scores was .41 (i.e. much closer to the mid-point than that for vertical), and for the higher scores was .55 (i.e. much closer to the mid-point than for vertical). This indicates that there is more pronounced spatial mapping of concepts along the vertical than the horizontal dimension for Mandarin speakers.

Cross-language distribution comparisons

The patterns of distribution about the frequency of the items receiving particular scores that were averaged across participants for each language group and each dimension are displayed in Figures 1–3. The average score for each item (i.e. each word) for both groups (English, Mandarin) for each dimension (vertical, horizontal, and valence) was then compared via a series of Mann–Whitney U tests. That is, the dependent variable was the score for each item on a given dimension, and the independent variable was language group. This test assesses whether the distribution of scores is similar or different between the two groups. These revealed similar patterns of distribution for the two cultural groups on the valence and vertical dimensions ($p = .884$, $d = .02$[4] and $p = .655$, $d = .05$) while the distribution of the two groups' ratings differed on the horizontal dimension ($p = .020$, $d = .27$). Examination of Figure 2 indicates that this is because the scores tend to be less extreme in the Mandarin group on the horizontal dimension, which suggests that this group have weaker concept-space mappings on this dimension.

Furthermore, we ran a series of rank-order correlations to determine the degree to which the two cultural groups rated the individual items in a similar manner. Specifically, we calculated the correlation between the mean rating for each of the 151 items by the English speakers on a given dimension with the mean rating for each of the 151 items by the Mandarin speakers on that same dimension. The results revealed significant associations for all three dimensions, although the correlations were higher for vertical ($r = .87$, $p < .001$) and the valence dimension ($r = .82$, $p < .001$) when compared to the horizontal dimension ($r = .74$, $p < .001$). Critically, according to Diedenhofen and Musch's (2015) comparison of correlations, the vertical and horizontal dimension correlations were different ($p < .001$, 95% CI for difference = [+.07 to +.21]). This is converging evidence that the two groups were more similar in their ratings along the vertical dimension than the horizontal dimension. Furthermore, the groups had a higher correlation with one another on the valence dimension than the horizontal dimension ($p < .05$, 95% CI for difference = [+.01 to +.15]), as well as a higher correlation on the vertical dimension than the valence dimension ($p < .05$, CI for difference = [+.01 to +.10]).

In the previous section, we examined the correlations across dimensions within each language group. Here we sought to establish whether there were differences between the groups in the extent to which these dimensions correlated with one another. To do this, we once again used Diedenhofen and Musch's (2015) comparison of correlations function. This indicated that the correlation between the vertical and horizontal dimensions was greater for English speakers ($r = .86$) than for Mandarin speakers ($r = .75$), $p = .006$, 95% CI = [+.03 to +.20]. In contrast, neither the correlation between the vertical and valence dimensions ($p = .665$, CI = [−.06 to +.04]), nor the correlation between the horizontal and valence dimensions ($p = .131$, CI = [−.02 to +.14]) was different for the two groups.

Item analysis

Here we examined which particular items produced the most extreme mean ratings on each dimension in each language group, and also identified the items that had the smallest and largest discrepancy in how the two groups rated them.

English group item analysis.. On the vertical dimension, there were 30 items that scored less than 0.1 (i.e. most strongly *down*). Six items scored at or below .05, and these were: *Bitter, Bleak, Ashamed, Ugly, Dread,* and *Dead*. At the other end of the spectrum, there were 49 items that scored greater than 0.9 (i.e. most strongly *up*). There were 20 items that scored 0.95 or greater, and these were: *Achieve, Inspired, Highlight, Cheer, Celebrate, Complete, Positive, Admire, Happy, Height, Friend, Active, Champion, Ambitious, Hero, Light, Laugh, Halo, Bright,* and *Radiant*.

On the horizontal dimension, no items received a score of less than 0.1 (i.e. most strongly *left*). The 11 leftmost items had scores of 0.3 or below ($M = .28$, $SD = .01$). (Note that we used 11 rather than 10 due to tied scores for values to two decimal places). These were: *Lost, Bitter, Aimless, Bleak, Delay, Insomnia, Unfair, Ashamed, Depressed, Weasel,* and *Theft*. Similarly, no items received a score of greater than 0.1 (i.e. most strongly *right*). The fact that no items received the most extreme left/right scores (whereas items did for the vertical dimension) is consistent with the mode-based analysis indicating the less pronounced spatial mapping of concepts on this dimension. The 11 rightmost items received a score of .72 or greater ($M = .74$, $SD = .02$). These were: *Achieve, Brave, Friday, Active, Reliable, Celebrate, Champion, Hero, Holiday, Respect,* and *Leader*.

On the valence dimension, 47 items score below 0.1 (i.e. most strongly negative). 20 items scored .05 or below, which were: *Diseased, Hostile, Despair, Devil, Depressed, Rude, Enemy, Cruel, Crime, Doom, Cheat, Dread, Hate, Dead, Negative, Unhappy, Fraud, Bleak, Insolent,* and *Defeat*. At the other end, 74 items scored greater than 0.9 (i.e. most strongly positive). Of these, 56 items scored 0.95 or greater. The 26 most positive items (i.e. which received a score .98 or greater) were: *Achieve, Active, Admire, Peace, Cheer, Celebrate, Highlight, Beauty, Belong, Brave, Hero, Love, Laugh, Polite, Champion, Holiday, Positive, Loyal, Sincere, Reliable, Quality, Talented, Grateful, Generous, Clean,* and *Radiant*.

Mandarin group item analysis.. On the vertical dimension, the 10 items that had a mean rating score of less than 0.1 (i.e. most strongly *down*) were: *Depressed, Cheat, Theft, Ashamed, Crime, Despair, Diseased, Dread, Fearful,* and *Obnoxious*. At the other end of the spectrum, there were many more items that scored greater than 0.9 (i.e. most strongly *up*): 32. There were 10 items that received a score 0.95 or greater, and these were: *Friend, Cheer, Celebrate, Achieve, Studious, Positive, Inspired, Happy, Grateful,* and *Champion*.

On the horizontal dimension, no items scored below .1 (i.e. most strongly *left*). The 10 leftmost items received a score of .36 or less, and consisted of: *Argue, Guilty, Negative, Stingy, Bleak, Gloomy, Frustrated, Hate, Obnoxious, Neurotic*. The mean score for these items was .35 ($SD = .02$). Similarly, no items received a score greater than 0.9 (i.e. most strongly *right*). The thirteen items that received a score of 0.6 or greater were: *Ambitious, Mature, Summer, Rich, Righteous, Hero, Holiday, Lucky, Smile, Satisfying, Reliable, Complete,* and *Business* (mean score = .61, $SD = .02$). This is consistent with the mode-based analysis indicating the less pronounced spatial mapping of concepts on this dimension.

On the valence dimension, 7 items received a mean score less than .1 (i.e. most strongly negative). These were: *Cruel, Steal, Crime, Theft, Diseased, Unfair,* and *Contempt*. Sixty-six items score greater than 0.9 (i.e. most strongly positive). The 26 items that scored .95 or greater were: *Achieve, Smile, Happy, Studious, Light, Brave, Friday, Active, Hero, Respect, Love, Justice, Admire, Complete, Sweet, Wise, Cheer, Satisfying, Rich, Bright, Lucky, Neat, Inspired, Laugh, Polite,* and *Earnest*.

Comparative item analysis.. The following analyses were conducted in order to provide a deeper understanding of the similarities and differences between the English and Mandarin language groups in how concepts are associated with physical space. Specifically, we calculated a *discrepancy score*, by subtracting the mean rating for the Mandarin group from the mean rating for the English group for a given dimension. A larger absolute value discrepancy score therefore indicates a greater discrepancy between the how the two groups rated the item, whereas a discrepancy score of zero indicates the absence of a discrepancy. Positive discrepancy scores indicate higher ratings by the English group, whereas negative discrepancy scores indicate higher ratings by the Mandarin group.

The items with the smallest and largest discrepancy scores on the vertical dimension are displayed in Table 1. We sought to identify five items in each range (positive, neutral, negative), however, where there are tied scores, more items were identified.

The basis of some of the discrepant ratings between the two groups appears to make sense given their respective cultures. For example, the Mandarin-language group consisted of predominately Chinese-born individuals, for whom *red* has particular political and cultural significance. Given the general tendency in both groups for positive

Table 1. Items with the largest (positive and negative) discrepancy scores on the **vertical** dimension, and items with zero discrepancy scores (i.e. no difference between the two groups) in Study 1.

Item	Discrepancy Score	English Rating	Mandarin Rating
Defeat	−.42	.06	.48
Argue	−.39	.15	.54
Subordinate	−.25	.17	.42
Red	−.24	.59	.83
Bleak	−.20	.03	.23
Fearful	0	.09	.09
Liar	0	.10	.10
Vulgar	0	.11	.11
Neurotic	0	.17	.17
Enraged	0	.23	.23
Crime	0	.09	.09
Home	0	.85	.85
Sweet	0	.93	.93
Friend	0	.96	.96
Contempt	.18	.31	.13
Belong	.18	.90	.72
Proud	.26	.85	.59
Funny	.44	.91	.47
Radiant	.70	.95	.25

items to have stronger associations with *up*, it is therefore logical that this item would receive a higher *up* rating for the Mandarin versus the English-speaking group. In contrast, many predominately English-speaking cultures have what would typically be described as an *individualistic* rather than a *collectivist* culture (Imada, 2012). Consistent with this, the English speakers tended to rate items that could be considered to be more typically used to describe individuals' traits (such as *radiant, proud* and *funny*) as more strongly associated with *up*, which is also consistent with the tendency for Westerners but not East Asians to use adjectives to describes people as having stable dispositions, where East Asians typically contextualise actions in contexts (Kashima, Kashima, & Kidd, 2014; Kashima, Kashima, Kim, & Gelfand, 2006). *Radiant* was particularly polarising, with the English speakers rating it near maximum in association with *up,* whereas Mandarin speakers rated it convincingly *down*. In contrast, the English and Mandarin groups provided very similar high *up* ratings for words such as *friend* and *sweet*, and similarly extreme *down* ratings for *fearful*, *crime*, *liar,* and *vulgar*.

The items with smallest and largest discrepancy scores on the horizontal dimension are displayed in Table 2. We aimed to shortlist the most extreme five items, and five with the highest agreement. However, eight items are included with negative discrepancy scores due to tied scores (i.e. could not isolate the top five). Four items are listed with zero discrepancy because these were the only items that had zero scores, whereas multiple items had scores of −.01 and +0.01.

Table 2. Items with the largest (positive and negative) discrepancy scores on the **horizontal** dimension, and items with zero discrepancy scores (i.e. no difference between the two groups) in Study 1.

Item	Discrepancy Score	English Rating	Mandarin Rating
Defeat	−.16	.33	.49
Doom	−.15	.31	.46
Bitter	−.14	.27	.41
Insomnia	−.14	.28	.42
Lazy	−.14	.32	.46
Aimless	−.14	.27	.41
Ashamed	−.14	.29	.43
Delay	−.14	.28	.42
Worried	0	.38	.38
Divorce	0	.41	.41
Pompous	0	.43	.43
Ethical	0	.56	.56
Achieve	.22	.77	.55
Celebrate	.25	.74	.49
Funny	.25	.69	.44
Radiant	.27	.68	.41
Brave	.28	.77	.49

Similar to the ratings on the vertical dimension, the English-language group showed stronger *right* ratings to traits such as *radiant, funny*, and *brave,* as well as to the words *achieve* and *celebrate*. Furthermore, the English-language group also gave clearer *left* ratings to personal traits such as *aimless* and *lazy*. Notably, however, the discrepancies on this dimension were less pronounced than either the vertical or valence dimensions, consistent with the observation that words tended to be less clearly associated left/right than they did up/down or positive/negative.

The items with smallest and largest discrepancy scores on the valence dimension are displayed in Table 3.

Red had a much stronger positive association for the Mandarin group, whereas the English group were relatively neutral about it. This most likely reflects from the same reasons as discussed for its discrepant ratings on the vertical dimension. Some of the discrepantly-rated items on the valence dimension can be linked back to a collectivist/individualist cultural difference. For example, *subordinate* was rated as more negative for the English group, whereas if anything it was slightly positive for the Mandarin group (i.e. >.5). Again, individual personality traits such as *funny* were rated far more positively by the English group than the Mandarin group, whereas words such as *mediocre* attracted much more negative ratings by the English than the Mandarin group. Interestingly, *defeat* was another large-discrepancy item, it was approaching the most extreme negative rating for the English group, whereas it had a neutral mid-point rating for the Mandarin group. One possible explanation for this

Table 3. Items with the largest (positive and negative) discrepancy scores on the **valence** dimension, and items with zero discrepancy scores (i.e. no difference between the two groups) in Study 1.

Item	Discrepancy Score	English Rating	Mandarin Rating
Defeat	−.45	.05	.50
Red	−.34	.54	.88
Subordinate	−.31	.24	.55
Argue	−.30	.08	.38
Mediocre	−.23	.12	.35
Weekend	0	.94	.94
Light	0	.96	.96
Speak	0	.89	.89
Happy	0	.97	.97
First	0	.94	.94
Belong	.19	.99	.80
Contempt	.22	.31	.09
Proud	.28	.82	.54
Funny	.46	.94	.48
Radiant	.73	.98	.25

discrepancy is that the word may have been interpreted differently in the two languages. That is, in English, the word *defeat* can refer to the noun (i.e. an instance of being defeated), or it could refer to the verb (i.e. the action of defeating – thereby winning). These two different interpretations of defeat would appear to have quite different valences. In Mandarin, this word also has these dual meanings. However, it is possible that the two language groups differed in the extent to which they favoured one of these interpretations of the word, and this is the source of the discrepant ratings between the two language groups. *Argue* was also far more negative for the English versus Mandarin group. In contrast, both groups felt similarly positive about words such as *weekend, light,* and *happy*.

Summary

In summary, Study 1 revealed several key findings. First and foremost, there was clearly the presence of the spatial mapping of concepts in both the English- and Mandarin-language participants: the distributions of responses to the items were bimodal, with the modes typically separated from the mid-point score, which we assume would indicate no systematic spatial mapping. Furthermore, there was remarkable similarity in how the two groups rated the set of items across the three dimensions of vertical, horizontal, and valence. Both groups had stronger spatial mapping of concepts on the vertical than the horizontal spatial dimension. However, the groups differed on the horizontal dimension. In particular, the spatial mapping of concepts appeared less pronounced for the Mandarin group than the English group on the horizontal dimension.

Study 2

The results from Study 1 suggested significant overlap in how English- and Mandarin-speakers conceptualise concepts in space. In Study 2 we attempted to replicate these effects using a more sensitive item scale. Specifically, instead of a forced choice scale we used a graded Likert-scale, which may reveal more subtle differences across dimensions and groups, if they exist. The spatial mapping of 151 concept words was again assessed via questionnaire, and the distribution of ratings was then compared between the two groups.

Method

Participants

All participants were students at The Australian National University. All provided informed consent, and the protocol was approved by ANU's Human Research Ethics Committee. Native English and Mandarin speakers were recruited.[5] A total of 332 participants volunteered in exchange for course credit. Their average age was 20.6 years (SD = 4.9). The sample contained 169 males and 159 females. (Note that these values do not include four participants who declined to report their demographics).

Materials and procedure

As per Study 1, participants completed the questionnaire online via Qualtrics. First, the information sheet was displayed to participants, and a forced-response (yes/no) question about whether they consented to participate. Following this, initial questions asked about participants' demographic information (age, gender, country of birth, and first language). If participants selected Mandarin as their first language, then they were instructed on how to change the default setting of the survey into modern simplified Mandarin, and to explicitly confirm compliance with this instruction.

In the main part of the questionnaire, participants were presented with the same 151 words as Study 1. Participants were instructed to rate each word across three dimensions: vertical (up / down), horizontal (left / right), and valence (positive / negative). The response scale for the vertical dimension consisted of: *Strongly Down, Somewhat Down, Neither Up nor Down (Neutral), Somewhat Up, and Strongly Up*. The response scale for the horizontal dimension consisted of: *Strongly Left, Somewhat Left, Neither Left nor Right (Neutral), Somewhat Right, Strongly Right*. Finally, the response scale for the valence dimension consisted of: *Strongly Negative, Somewhat Negative, Neither Positive nor Negative (Neutral), Somewhat Positive, and Strongly Positive*. The to-be-rated words were shown only in English or Mandarin (depending on the participant's language).[6]

Results & discussion

Participant details

Fourteen participants were removed from analysis because they nominated a language other than

Mandarin or English as their native language, and 62 participants were excluded due to excessive missing data (>10 items were not answered), leaving a final sample of 256 participants. Of these, 125 identified Mandarin as their native language, and 131 identified English as their native language. Of the 125 Mandarin speakers, 117 identified China as their country of birth (2 Australia, 6 Other). Of the 131 English speakers, 107 identified Australia as their country of birth (2 China, 22 Other).

Response scoring

Responses on the vertical dimension were scored as 1 for *Strongly Down,* 2 for *Somewhat Down,* 3 for *Neutral*, 4 for *Somewhat Up*, and 5 for *Strongly Up*. Responses on the horizontal dimension were scored as 1 for *Strongly Left,* 2 for *Somewhat Left*, 3 for *Neutral*, 4 for *Somewhat Right*, and 5 for *Strongly Right*. Responses on the valence dimension were scored as 1 for *Strongly Negative*, 2 for *Somewhat Negative,* 3 for *Neutral*, 4 for *Somewhat Positive,* and 5 for *Strongly Positive*. An average rating for each item was then calculated by taking the mean response across participants in each group. Any cells where there was missing data (≤10 per participant) were simply omitted from the computation of averages for each condition.

English inter-dimensions analysis

Next, we sought to establish the relationships between the three dimensions (vertical, horizontal, and valence) for the English speakers. This revealed very high positive correlations between the three dimensions: vertical and horizontal ratings ($r = .96$, $p < .001$), vertical and valence ($r = .98$, $p < .001$), and horizontal and valence ($r = .95$, $p < .001$). In other words, there was a high degree of correspondence between the dimensions in their ratings amongst English speakers. In particular, there was a very strong tendency such that as items received stronger *up* ratings they also received stronger *positive* ratings, closely followed by a strong tendency for items receiving clear *up* ratings to also receive clear *right* ratings, and then also for items receiving stronger *right* ratings to also receive stronger *positive* ratings. Diedenhofen and Musch's (2015) comparison of correlations indicated that the correlation between the vertical and horizontal ratings ($r = .96$) was less than the correlation between the vertical and valence ratings ($r = .98$), ($p < .001$, 95% CI for the difference = [−.03 to −.01]), and that the correlation between vertical and valence ($r = .98$) was greater than the correlation between the horizontal and valence dimensions ($r = .95$), ($p < .001$, CI = [+.02 to +.05]). Similarly, the correlation between the vertical and horizontal ratings ($r = .96$) was greater than that between the horizontal and valence dimensions ($r = .95$), ($p < .036$, CI = [+.001 to +.022]). In other words, the strongest observed correlation was that between the vertical and valence dimension ratings. That is, while all of the relationships were strong, the strongest was the tendency for words associated with *up* to also be associated with *positive*, and correspondingly for words associated with *down* to also be associated with *negative*.

From Figures 4–6, it appears as though the spatial mappings may be more extreme for the vertical dimension than the horizontal dimension. The mode is an appropriate measure of central tendency to illustrate where the bulk of the distribution sits. As in Study 1, the distributions are bimodal. Therefore, following our analyses in Study 1, we calculated the mode of each half of the distribution (i.e. computed the mode for the scores <3, and then the mode for the scores >3) for the vertical and horizontal dimension for English speakers. This provides insight into the extremeness of the spatial mapping of concepts, because if the mode is more extreme (i.e. further from the mid-point of the response scale, 3), then it suggests that the spatial mapping of concepts is more extreme. In contrast, if the mode if closer to the centre of the response scale, then it suggests that the spatial mapping of concepts is less extreme. For the vertical ratings, the mode for the lower scores was 1.69, whereas the mode for the higher scores was 4.02. For the horizontal dimension, the mode for the lower scores was 2.36 (i.e. closer to the mid-point than that for vertical), and for the higher scores was 3.65 (i.e. closer to the mid-point than for vertical). This suggests that there is more pronounced spatial mapping of concepts along the vertical than the horizontal dimension for English speakers.

Mandarin inter-dimensions analysis

Here, we sought to establish the relationships between the three dimensions (vertical, horizontal, and valence) for the Mandarin speakers. As we saw for the English speakers, this revealed very high positive correlations between the three dimensions: vertical and horizontal ratings ($r = .98$, $p < .001$), vertical and valence ($r = .99$, $p < .001$), and horizontal and valence ($r = .98$, $p < .001$). In other words, there

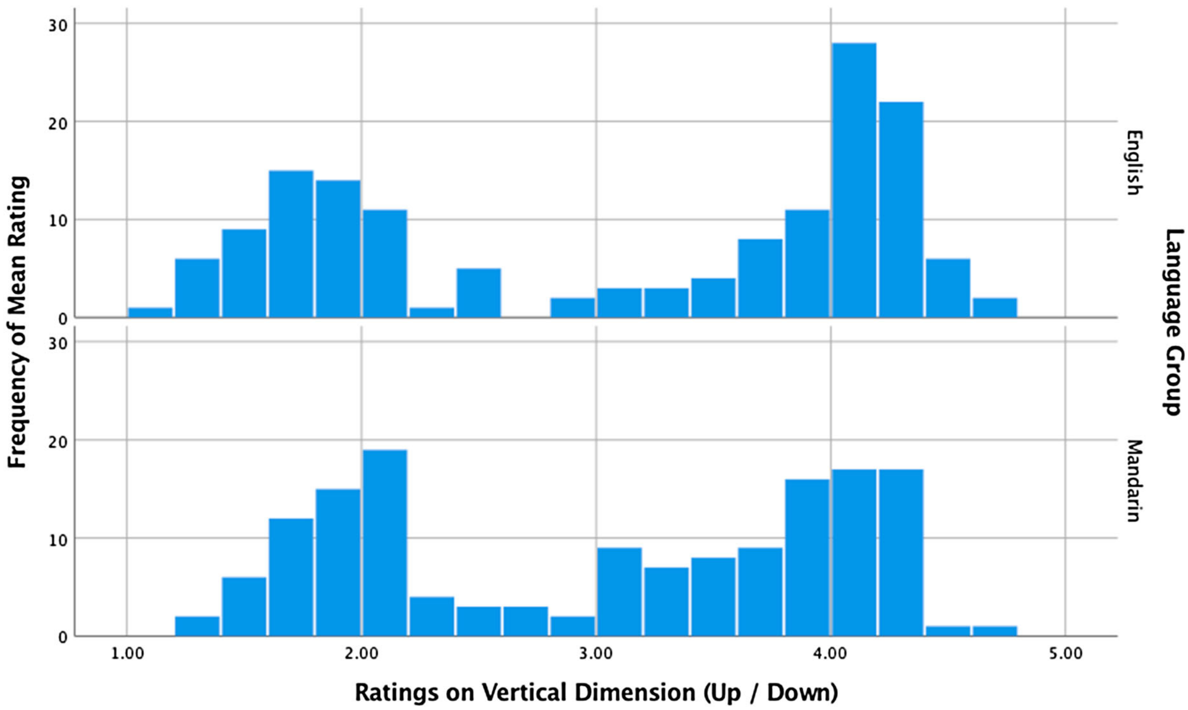

Figure 4. Distribution plots of the frequency of items with certain scores for the vertical dimension in Study 2 (1 = strongly down, 5 = strongly up). English speakers upper panel, Mandarin speakers lower panel.

was a high degree of correspondence between the dimensions in their ratings amongst Mandarin speakers. In particular, there were clear tendencies such that items associated with up were also associated with right, items associated with vertical were also associated with positive, and items associated

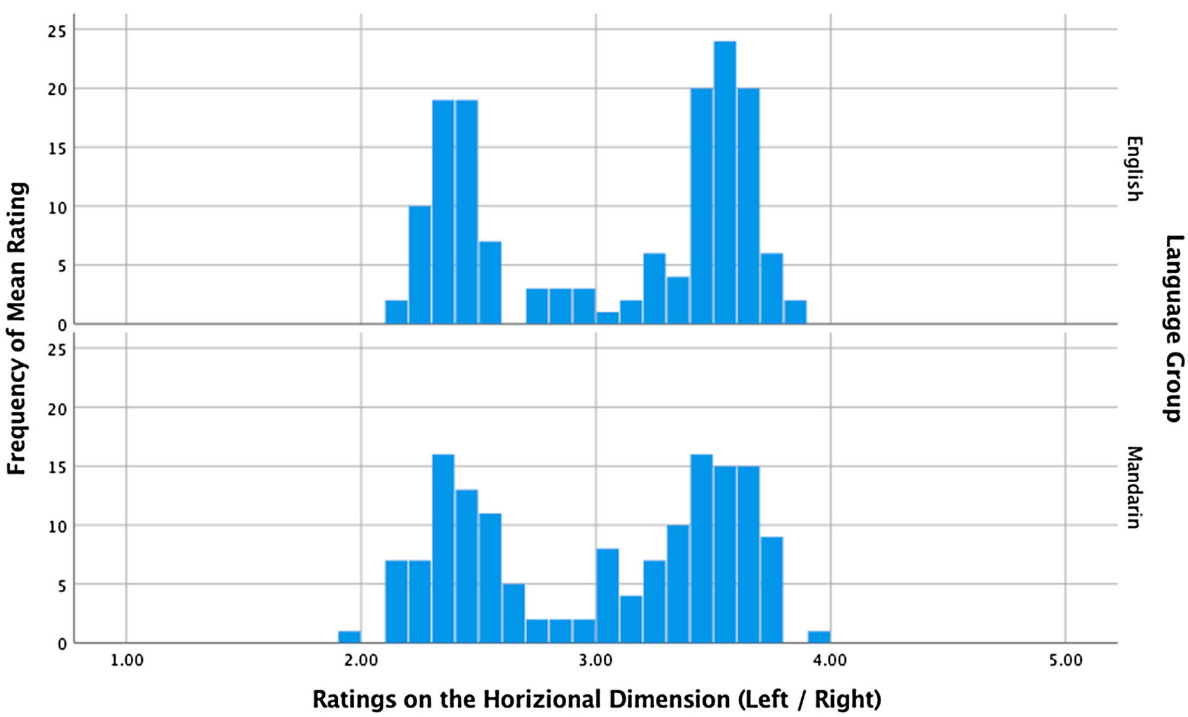

Figure 5. Distribution plots of the frequency of items with certain scores for the horizontal dimension in Study 2 (1 = strongly left, 5 = strongly right). English speakers upper panel, Mandarin speakers lower panel.

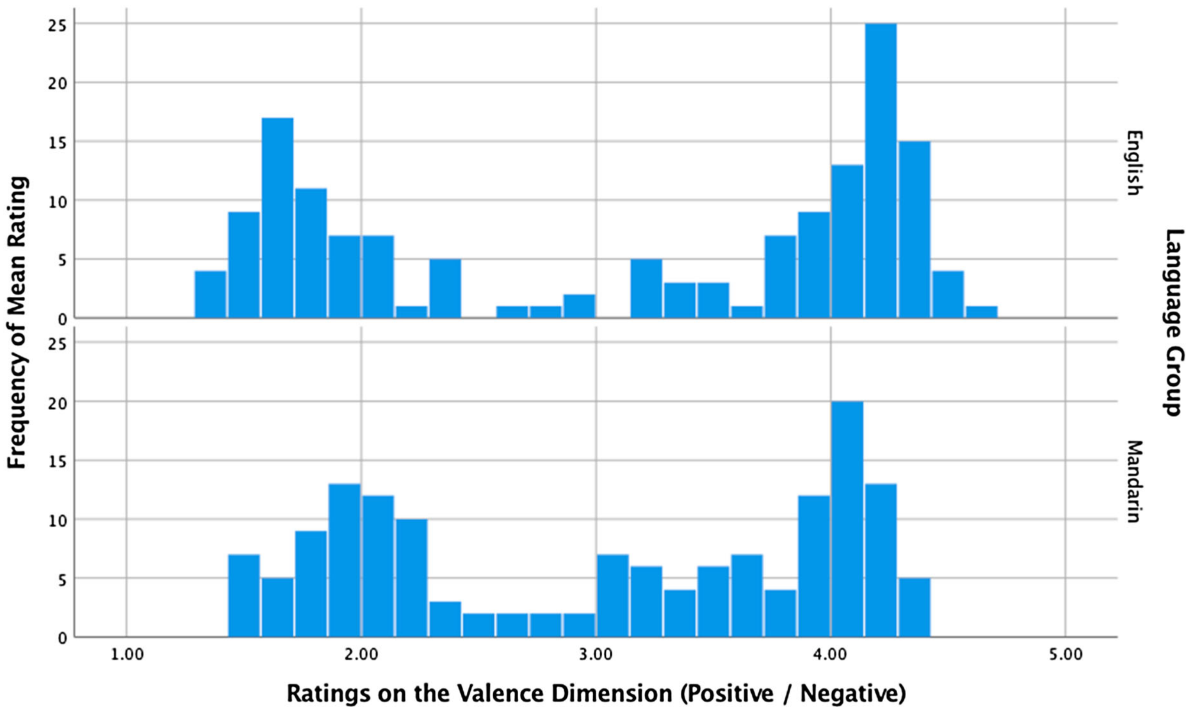

Figure 6. Distribution plots of the frequency of items with certain scores for the valence dimension in Study 2 (1 = strongly negative, 5 = strongly positive). English speakers upper panel, Mandarin speakers lower panel.

with right were also associated with positive. Clearly, the good is up/right and bad is down/left associations are not unique to English. Comparison of correlations analysis revealed that the correlation between vertical and valence was greater than either of the other two correlations ($p < .001$, 95% CI for the difference = [+.01 to +.02]), which did not differ from one another numerically at two decimal places. This means that as per the English language group, the correlation between the vertical and valence dimensions was the strongest observed correlation. That is, while all of the associations were strong, the strongest was the tendency for items associated with up (versus down) to be associated with positive (versus negative).

From Figures 4–6, it is evident that the spatial mappings may be more extreme for the vertical dimension than the horizontal dimension. To test this, we calculated the mode of each half of the distribution (i.e. computed the mode for the scores <3, and then the mode for the scores >3) for the vertical and horizontal dimension for Mandarin speakers. For the vertical ratings, the mode for the lower scores was 2.07 whereas the mode for the higher scores was 3.99. For the horizontal dimension, the mode for the lower scores was 2.37 (i.e. closer to the mid-point than that for vertical), and for the higher scores was 3.67 (i.e. closer to the mid-point than for vertical). This suggests that for Mandarin speakers, there is more pronounced spatial mapping of concepts along the vertical than the horizontal dimension.

Cross-language distribution comparisons

The patterns of distribution about the frequency of the items receiving particular scores that were averaged across participants for each language group and each dimension are displayed in Figures 4–6. A series of Mann–Whitney U tests revealed similar patterns of distribution for the two cultural groups on all three dimensions (vertical, horizontal, and valence), ($p = .467$, $d = .08$, $p = .457$, $d = .09$, and $p = .330$, $d = .11$, respectively).

Furthermore, we ran a series of rank-order correlations to determine the degree to which the two cultural groups rated the individual items in a similar manner. The results revealed significant associations for all three dimensions, with similar high correlations for the vertical ($r = .88$, $p < .001$), horizontal dimension ($r = .87$, $p < .001$) and the valence dimension ($r = .89$, $p < .001$). According to Diedenhofen and Musch's (2015) comparison of correlations, the vertical and horizontal dimension correlations were indistinguishable ($p = .714$, 95% CI for the difference = [−.04 to +.07]), as were the vertical and valence dimensions ($p = .691$, CI = [−.06 to

+.04]), and the horizontal and valence dimensions ($p = .445$, CI = [−.07 to +.03]). In other words, this analysis indicated that the three different dimensions were associated to similar extents when comparing the two language groups.

Whereas the previous analyses did not reveal any evidence for language-group differences in this study, comparing the magnitude of the correlations between the dimensions across the languages did reveal evidence for differences. In particular, Diedenhofen and Musch's (2015) comparison of correlations function indicated that the correlation between the vertical and horizontal dimensions was greater for Mandarin speakers ($r = .98$) than for English speakers ($r = .96$, $p < .001$, 95% CI of the difference between the correlations = [+.01 to +.03]). Similarly, the correlation between the vertical and valence dimensions was larger for the Mandarin group ($r = .99$) than for the English group ($r = .98$), $p < .001$, CI = [+.01 to +.02]. Finally, the correlation between the horizontal and valence dimensions was also greater for the Mandarin group ($r = .98$) than for the English group ($r = .95$), $p < .001$, CI = [+.02 to +.05]. Altogether, this means that the Mandarin group demonstrated stronger correlations between all dimensions than the English group.

Item analysis

Here we examined which particular items produced the most extreme mean ratings on each dimension in each language group, and also identified the items that had the smallest and largest discrepancy in how the two groups rated them.

English group item analysis.. On the vertical dimension, there were 13 items that scored below 1.5 (i.e. most strongly *down*). These were: *Depressed, Despair, Dead, Cruel, Doom, Devil, Defeat, Ashamed, Diseased, Satan, Dread, Negative,* and *Hate*. At the other end of the spectrum, there were three items that scored 4.5 or greater (i.e. most strongly *up*). These were *Achieve, Champion,* and *Happy*. On the horizontal dimension, no items scored below 1.5 (i.e. most strongly *left*). Similarly, no items scored above 4.5 (i.e. most strongly *right*). This is consistent with the mode-based analysis indicating that the spatial mapping of concepts was less pronounced on the horizontal than the vertical dimension.

The 10 most left-rated items were: *Depressed, Diseased, Dead, Satan, Despair, Devil, Doom, Obnoxious, Cruel,* and *Insane*. These items had mean ratings ranging between 2.16 and 2.28 (Mean = 2.24, SD = .04). The strong down and left associations for items such as *Devil* and *Satan* is consistent with previous behavioural research (Chasteen et al., 2010). The 10 items most right rated items were: *Achieve, Loyal, Champion, Active, Victory, Ambitious, Positive, Brave, Love,* and *Respect*. These items had mean ratings ranging between 3.68 and 3.89 (Mean = 3.76, SD = .06). On the valence dimension, 10 items scored below 1.5 (i.e. most strongly *negative*): *Depressed, Dead, Despair, Doom, Diseased, Devil, Negative, Dread, Hate, Cruel*. Two items scored above 4.5 (i.e. most strongly *positive*): *Achieve* and *Positive*.

Mandarin group item analysis.. On the vertical dimension, five items had a mean rating below 1.5 (i.e. most strongly *down*). These were: *Despair, Cheat, Cruel, Crime,* and *Doom*. One item scored above 4.5 (i.e. most strongly *up*): *Champion*. On the horizontal dimension, no items had a score below 1.5, and no items had a score above 4.5. This supports the mode-based analysis in highlighting that the spatial mapping of concepts is less pronounced on the horizontal than the vertical dimension.

The 10 most left-rated items were: *Despair, Crime, Dread, Cheat, Cruel, Doom, Dead, Danger, Devil,* and *Guilty*. Their mean ratings ranged between 1.96 and 2.24 (Mean = 2.14, SD = .08). The 10 most right-rated items were: *Champion, Lucky, Righteous, Victory, Love, Genius, Loyal, Dream, Justice,* and *Sincere*. The mean ratings for these items ranged between 3.70 and 3.94 (Mean = 3.76, SD = .07). On the valence dimension, four items scored below 1.5 (i.e. most strongly *negative*): *Doom, Despair, Cruel,* and *Dread*. No items scored above 4.5 (i.e. most strongly *positive*). However, the 10 most positive scoring items were: *Champion, Love, Loyal, Honesty, Lucky, Righteous, Respect, Happy, Dream,* and *Justice*. These items' scores ranged between 4.24 and 4.38 (Mean = 4.30, SD = .05).

Comparative item analysis.. Here, discrepancy scores were calculated by subtracting the English rating from the Mandarin rating, such that a larger positive value indicates a higher rating by the Mandarin speakers. The discrepancy scores for the most and least discrepant items on the vertical dimension can be seen in Table 4. This includes the five most extreme discrepancy scores in each direction (i.e. higher ratings by English/Mandarin), as well as five that had zero or almost zero discrepancy scores.

The results of the item analysis in Study 2 shares many similarities with that of Study

Table 4. Items with the largest (positive and negative) discrepancy scores on the **vertical** dimension, and items with zero discrepancy scores (i.e. no difference between the two groups) in Study 2.

Item	Discrepancy Score	English Rating	Mandarin Rating
Radiant	−1.91	4.40	2.49
Funny	−1.34	4.21	2.86
Contempt	−1.19	2.82	1.62
Belong	−0.94	4.04	3.10
Achieve	−0.93	4.71	3.78
Stingy	−0.01	2.05	2.05
Mediocre	0	2.50	2.50
Steam	0	3.05	3.05
Neurotic	0	2.07	2.07
Theft	0.01	1.67	1.68
Earnest	0.55	3.64	4.19
Depressed	0.62	1.17	1.79
Negative	0.65	1.48	2.13
Argue	0.68	2.12	2.80
Defeat	1.31	1.37	2.68

1. Individualistic personality traits such as *radiant* and *funny* garnered stronger *up* associations amongst English than Mandarin speakers. *Defeat* was also differentially rated by the two groups, with strong *down* associations for the English group, versus modestly *up* associations for the Mandarin group. Similarly, *argue* was slightly *down* for the English group, but slightly *up* for the Mandarin group. Conversely, both groups had equivalent associations with *down* for words such as *theft, stingy,* and *neurotic,* and equivalent associations with *up* for *steam*. Somewhat surprisingly, *mediocre* received a mid-point score (2.5) from both groups, indicative of neither clear associations with *up* nor *down*.

The discrepancy scores for the most and least discrepant items on the horizontal dimension can be seen in Table 5.

Once again, *Radiant* attracted contrasting ratings for the English and Mandarin language groups. English speakers clearly had a *right* association for this item, whereas the Mandarin speakers' ratings hovered just above the neutral mid-point. While both groups gave *funny*, *achieve*, and *belong* scores indicative of *right* associations, this was more pronounced for the English group. *Contempt* was rated as slightly *right* versus slightly *left* for the English and Mandarin groups respectively. While both groups gave *right* associations for *righteous, young,* and *sincere*, this was more pronounced for the Mandarin group. The Mandarin group gave *argue* and *defeat* ratings of *right*, whereas these were *neutral* and *left* respectively for the English group. Both groups rated the items *reliable, home, victory,* and *pretty* as virtually equivalently *right*, whereas both groups gave a modest *left* rating for *vain*. Some of these results concur with a collectivist versus individualistic framework, such as the scores for *radiant* and *funny*.

The discrepancy scores for the most and least discrepant items on the valence dimension can be seen in Table 6.

English speakers rated *radiant, funny, belong, quality, and purpose* as much more strongly positive than did Mandarin speakers, however, both groups gave them positive ratings. *Argue* and *defeat* garnered positive associations in Mandarin, but negative ones in English. *Enraged, negative,* and *enemy* received negative ratings from both groups, but to a greater extent for English versus Mandarin. In contrast, both groups equivalently rated *danger, fearful*, and *dread* as negative, and *lucky* and *champion* as positive.

Table 5. Items with the largest (positive and negative) discrepancy scores on the **horizontal** dimension, and items with zero discrepancy scores (i.e. no difference between the two groups) in Study 2.

Item	Discrepancy Score	English Rating	Mandarin Rating
Radiant	−1.10	3.65	2.54
Funny	−0.84	3.57	2.74
Contempt	−0.57	2.96	2.39
Achieve	−0.52	3.89	3.37
Belong	−0.47	3.56	3.10
Reliable	−0.01	3.66	3.65
Home	−0.01	3.56	3.55
Victory	0	3.76	3.75
Pretty	0.01	3.55	3.56
Vain	0.01	2.42	2.43
Righteous	0.23	3.56	3.79
Young	0.24	3.31	3.55
Sincere	0.25	3.45	3.70
Argue	0.42	2.50	2.93
Defeat	0.50	2.31	2.81

Table 6. Items with the largest (positive and negative) discrepancy scores on the **valence** dimension, and items with zero discrepancy scores (i.e. no difference between the two groups) in Study 2.

Item	Discrepancy Score	English Rating	Mandarin Rating
Radiant	−1.90	4.32	2.42
Funny	−1.56	4.27	2.71
Belong	−1.01	4.19	3.18
Quality	−0.99	4.24	3.25
Purpose	−0.93	4.15	3.22
Danger	0	1.58	1.58
Fearful	0	2.11	2.10
Lucky	0	4.31	4.30
Dread	0	1.47	1.47
Champion	0	4.38	4.38
Enraged	0.54	1.66	2.19
Negative	0.60	1.47	2.06
Enemy	0.60	1.74	2.34
Argue	0.75	2.06	2.82
Defeat	1.15	1.56	2.71

Summary

In summary, both the English and Mandarin language groups showed patterns indicative of concepts associated with physical space dimensions when a more graded response scale was used. Furthermore, there were considerable similarities between the two groups in their responses. Unlike in Study 1 where there were language-group differences in the extent to which concepts were spatially mapped on the horizontal dimension, here in Study 2, both groups demonstrated this to equivalent extents. That said, notably, as in the previous study, for both groups the mappings on the horizontal dimension were less pronounced than for the vertical dimension. Both groups showed strong correlations among the three dimensions, with the strongest between the valence and vertical dimensions, indicating groupings of up/positive and down/negative dimensions. However, these associations were stronger for the Mandarin than the English group.

General discussion

Across two studies we found that Mandarin speakers exhibit spatial mapping of concepts, as has been shown in other languages (Crawford et al., 2014; Dudschig et al., 2015; Fuhrman & Boroditsky, 2010; Maass & Russo, 2003; Ouellet et al., 2010; Santiago et al., 2007; Shaki et al., 2009). This extends previous research showing spatial mapping of time-based concepts in Mandarin (Boroditsky et al., 2011), revealing that this tendency occurs for a wide variety of abstract concepts. This supports the notion that spatial mapping is not just limited to European cultures, and adds weight to the suggestion that tendency to ground concepts in space may be *in general* a universal phenomenon. Of course, it remains possible that other untested languages do not share this property. However, given that English and Mandarin languages are linguistically and culturally unrelated, the present results provide tentative support for its more universal presence.

The most resounding result from the two studies was a remarkable degree of similarity between the two language groups in the associations with space that concepts had, as well as the strength of these associations. One very consistent and striking finding was that the identical set of concept words produced stronger spatial mapping on the vertical than the horizontal dimension. Other findings in cognitive psychology indicate that human are more efficient at making up/down judgements than left/right ones (Franklin & Tversky, 1990). This likely stems from the fact that many natural visual scenes are more similar about a left/right axis than an up/down one. The embodied cognition framework (Barsalou, 1999; Gallese & Lakoff, 2005; Lakoff & Johnson, 1999) proposes that humans draw on the same mechanisms for both cognition (including concept representation) as they do for perception and action. Given the observed perceptual asymmetries that favour the vertical dimension, therefore, the present finding that humans are more likely to use the vertical than the horizontal dimension in association with concept words, is consistent with the embodied cognition framework.

According to the embodied cognition framework, the tendency to have reduced spatial mapping of concepts on the horizontal dimension is consistent with the notion that the physical world (i.e. natural scenes) are typically more informationally-diverse along the vertical than horizontal axis. Furthermore, our interactions with the world are more physically determined on an up/down axis than a left/right one. That is, when in motion, a person could decide to walk to either the left or to the right (i.e. same action along horizontal dimension), whereas moving up or down in space would require quite distinct actions (e.g. crouching versus jumping). These factors are then reflected in the human tendency to have superior up/down than left/right perceptual acuity, which is therefore also reflected in greater vertical than horizontal associations with abstract concepts. Interestingly, researchers appear to have known this intuitively, as it is far more common to see assessments of the spatial mapping of concepts on the vertical (e.g. Dudschig, Souman, Lachmair, de la Vega, & Kaup, 2013; Goodhew et al., 2014; Gozli, Chasteen, et al., 2013; Gozli, Chow, Chasteen, & Pratt, 2013; Louwerse & Jeuniaux, 2010; Setic & Domijan, 2007) versus horizontal (Chasteen et al., 2010) dimension.

Previous research has suggested that an individuals' handedness (i.e. whether their dominant hand is right or left) can determine which side they associate positive versus negative concepts with. That is, with English-speaking participants, right-handers tend to associate rightward space with positive concepts and leftward space with negative concepts, whereas left-handers tend to associate leftward space with positive concepts and rightward space with negative concepts (Casasanto, 2009). It is possible that heterogeneity due to a latent individual-difference factor such as this explains why spatial mappings were less clearly pronounced on the horizontal dimension. We suspect that it has more to do with the inherent perceptual confusability of the left-right dimension, however, this would be interesting to examine in future research.

While there were considerable similarities in the spatial mapping of concepts in English and Mandarin speakers, there was some evidence of differences on the horizontal dimension. In particular, the results of Study 1 suggested that English speakers tended to rate concepts more extremely on the horizontal dimension than did the Mandarin speakers. This finding is consistent with the semantic associations between space and concepts (i.e. *right* means correct) and right-is-good belief in English (Chasteen et al., 2010). In contrast, the Mandarin speakers spatialised concepts on the horizontal to a lesser degree. One possible reason for this difference is that horizontal mappings seem to be less consistent in Mandarin. For instance, left represents either "superior" or "inferior" in different Mandarin idiomatic expressions (i.e. *zuo'zun'you'bei,* meaning lit. "left is superior and right is inferior" versus *zuo'-bei'you'zun*, meaning lit. "left is inferior and right is superior"). Contrary metaphors associating left and right locations also exist in seating manner for official occasions. In some dynasties, people who have higher reputations are placed at the seats on the left side of the host's seat, while in other dynasties, people who are placed at the right side of the host's seat are more respected. Consistent with this, in Study 1 an item such as *leader* was rated as more strongly associated with right for the English speakers (.72) than the Mandarin speakers (.53), for whom it approximated the midpoint. This indicates that there may be cross-cultural differences in the manifestation of the spatial mapping of concepts. However, this result was specific to Study 1 and did not occur in Study 2. It is not immediately clear why the two studies differed in this regard, and therefore we can only say that there is suggestive, not conclusive evidence of differences in spatial mappings between the two languages.

Another theoretical explanation that has been offered for the spatial mapping of concepts is the *language use hypothesis* (Goodhew et al., 2014). This account suggests that observed spatial mappings (e.g. between *happy* and *up*) both result from, and are reflected in, language use patterns. Multiple demonstrations of English language-use statistics predicting observed spatial mappings (both explicit and behavioural) has supported this hypothesis (Goodhew et al., 2014; Louwerse, 2008; Louwerse & Jeuniaux, 2010). Furthermore, language-use statistics have also been found to explain variance in other systematic associations, such as those between concepts and colours (e.g. *happy* and *yellow*) (Goodhew & Kidd, 2017). Given the large degree of overlap between the spatial mapping of concepts in English and Mandarin observed in the present work, this framework would predict similar language-use biases (e.g. between *happy* and *up*) in Mandarin as those already documented in English. This would be an interesting avenue for future research to test this prediction. However, it will likely await improved tools. For example, in previous research in English we have used the Google Ngram tool (Michel et al., 2011), which aggregates millions of digitised books to examine for language use patterns, such as the frequency of individual words, or word pairs (e.g. *happy-up*). While Google Ngram provides comprehensive data on English, and it does offer a Mandarin-language search, this appears to be far more limited. For example, while a search for "happy up" co-occurrences and "happy down" co-occurrences yields clear data for English, showing that the former occurs far more often than the latter, the same search in simplified Mandarin yields an error message indicative of no available data.

Can language use patterns account for the reduced spatial mappings on the horizontal versus vertical dimension? Again, an answer to this for Mandarin speakers awaits new tools. Notably, however, even if language-use patterns did mirror the observed spatial mappings in this study, it still does not disentangle causality. That is, it does not distinguish between the possibility that language-use actively *shaped* the conceptual representation, versus language-use merely *reflected* systematic patterns already embedded in the cognitive architecture. If the latter, then it leaves open the possibility for another explanatory mechanism, such as the embodied cognition framework. That is, the language use framework is not mutually exclusive with the embodied cognition framework, but instead both mechanisms could co-occur.

Finally, the present research focussed on *explicit* ratings of the spatial mapping of concepts. That is, we outright asked participants to indicate the extent to which they considered concepts mapped in space. This approach relies on the premise that participants have conscious access to their cognitive representations of concepts. There is compelling evidence for this assumption. In particular, there are strong correlations between participants' explicit reports of spatial mapping of concepts and how behaviour is implicitly affected by these spatial mapping of concepts, for example, in conceptual cueing paradigms (Goodhew & Kidd, 2016). In conceptual cueing paradigms, participants are presented with a concept word (a conceptual "cue") in the centre of the screen, and then a spatially-offset target appears (e.g. a letter that participants have to identify, or a stimulus whose presence they detect). Conceptual cueing occurs when response efficiency to this target is influenced by the meaning of the word. For example, conceptual cueing would be reflected in participants' facilitated responses to the target when it physically appears in an upper location of the screen following the word *happy*, and impaired responses to the same target following the word *sad*, and vice versa for a lower-location target (e.g. Chasteen et al., 2010; Estes et al., 2008; Goodhew et al., 2014; Gozli, Chasteen, et al., 2013). The fact that explicit reports of spatial mappings bear considerable resemblance to those that produce conceptual cueing indicates that participants do indeed have conscious access to the spatial mapping of concepts.

In a similar vein, in recent work, we have found evidence that participants' explicit associations between concepts and *colour* (Goodhew & Kidd, 2017) also correspond to those associations that produce congruency effects on implicit behavioural responses. This is further evidence that participants do indeed have explicit access to how concepts are represented, and in particular, how they are related to physical dimensions such as space, thereby validating the approach we adopted here. However, it would be useful for future research to show that the associations identified here in the explicit ratings have a corresponding behavioural manifestation both English and Mandarin speakers. We anticipate that the explicit ratings provided in this study (including the full list of rated items for each language group) will be useful in guiding word selection for such endeavours.

In conclusion, the present study definitively revealed that explicit spatial mapping of concepts also existed in Mandarin, as in English. Similar patterns of spatial mapping were displayed on the vertical dimension, whereas there was some evidence of differences on the horizontal dimension. These results suggest that while the tendency to map concepts in space may be a universal phenomenon, there may be context-specificity in the manifestation of these mappings, such that not all languages with different cultural backgrounds manifest the same mappings. At the same time, the most clear-cut results were that the spatial mapping of concepts robustly occurs in both languages, and there was considerable similarity in how native speakers of these two very different languages mapped concepts in space. Furthermore, both groups showed demonstrably reduced spatial mapping of concepts on the horizontal dimension relative to the vertical dimension in both studies. This likely reflects the physical and psychological realities of how humans perceive and interact with the world around them, consistent with the embodied cognition framework.

Notes

1. In Study 1, our participants self-identified as Chinese-speaking but we did not ask which specific Chinese language(s) they spoke. The test was conducted in Modern Simplified Chinese (MSC), which is most typically used to represent spoken Mandarin. However, since other Chinese languages also use MSC (e.g. Cantonese spoken in in Guangzhou province), we cannot rule out that some of our participants also spoke other varieties. However, in Study 2, participants were specifically recruited on the basis of *Mandarin* being their native language, see note #5.
2. Note that depending on the type of comparison (e.g. different or same groups, overlapping correlations or not), the comparison of correlations function provides output from a varying number of different statistical tests to compare the two correlations (e.g. Pearson and Fillon's *z*, Hotelling's *t*, etc). Where there were multiple provided, while the output of these tests was very similar, they were not identical. Therefore, all *p* values reported reflect the results of *all* of the tests provided. For example, if we say $p < .001$, then this means that *all* of the tests provided by this function indicated $p < .001$. Where they were slightly different (e.g. $p = .286$ versus $p = .287$), we instead report summary statistics that capture all of the *p*-values (e.g. $p > .05$). Also note that the for the *size of the sample*, we used the number of items feeding into the correlation (151), as this best represents the correlation sample.
3. While it would be useful to be able to statistically compare the modes across the dimensions, we do not know of a statistical test that provides this

information for modes, rather than those that operates on means (for parametric tests) or medians (non-parametric tests).

4. Effect sizes for the Mann-Whitney *U*-test were calculated via https://www.psychometrica.de/effect_size.html.
5. This differs from Study 1, where we recruited "Chinese-speakers". While the vast majority of participants in Study 1 were likely native Mandarin speakers because of our use of simplified Mandarin script (which is only used for Mandarin), we removed any possibility that participants in Study 2 spoke other languages native to mainland China (e.g. Cantonese, spoken predominantly in Hong Kong, Macau, and Guangzhou region of mainland China) by recruiting only Mandarin speakers.
6. The instructions and response labels were shown in both English and Mandarin.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by an Australian Research Council (ARC) Future Fellowship (FT170100021) awarded to S.C.G.

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. doi:10.1017/S0140525X99002149

Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently? *Cognition*, *118*(1), 123–129. doi:10.1016/j.cognition.2010.09.010

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General*, *138*(3), 351–367. doi:10.1037/a0015854

Casasanto, D., & Chrysikou, E. G. (2011). When left is "right": Motor fluency shapes abstract concepts. *Psychological Science*, *22*(4), 419–422. doi:10.1177/0956797611401755

Chasteen, A. L., Burdzy, D. C., & Pratt, J. (2010). Thinking of god moves attention. *Neuropsychologia*, *48*(2), 627–630. doi:10.1016/j.neuropsychologia.2009.09.029

Crawford, L. E., Cohn, S. M., & Kim, A. B. (2014). "Good is up" is not always better: A memory advantage for words in metaphor-incompatible locations. *Plos One*, *9*(9), e108269. doi:10.1371/journal.pone.0108269

Diedenhofen, B., & Musch, J. (2015). Cocor: A comprehensive solution for the statistical comparison of correlations. *Plos One*, *10*(4), e0121945. doi:10.1371/journal.pone.0121945

Dudschig, C., De la Vega, I., & Kaup, B. (2015). What's up? Emotion-specific activation of vertical space during language processing. *Acta Psychologica*, *156*, 143–155. doi:10.1016/j.actpsy.2014.09.015

Dudschig, C., Souman, J., Lachmair, M., de la Vega, I., & Kaup, B. (2013). Reading "sun" and looking up: The influence of language on saccadic eye movements in the vertical dimension. *Plos One*, *8*(2), e56872. doi:10.1371/journal.pone.0056872

Estes, Z., Verges, M., & Barsalou, L. W. (2008). Head up, foot down: Object words orient attention to the objects' typical location. *Psychological Science*, *19*(2), 93–97. doi:10.1111/j.1467-9280.2008.02051.x

Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, *119*(1), 63–76. doi:10.1037/0096-3445.119.1.63

Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cognitive Science*, *34*(8), 1430–1451. doi:10.1111/j.1551-6709.2010.01105.x

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*(3/4), 455–479. doi:10.1080/02643290442000310

Goodhew, S. C., & Kidd, E. (2016). The conceptual cueing database: Rated items for the study of the interaction between language and attention. *Behavior Research Methods*, *48*, 1004–1007. doi:10.3758/s13428-015-0625-9

Goodhew, S. C., & Kidd, E. (2017). Language use statistics and prototypical grapheme colours predict synaesthetes' and non-synaesthetes' word-colour associations. *Acta Psychologica*, *173*, 73–86. doi:10.1016/j.actpsy.2016.12.008

Goodhew, S. C., McGaw, B., & Kidd, E. (2014). Why *is* the sunny side always up? Explaining the spatial mapping of concepts by language use. *Psychonomic Bulletin & Review*, *21*(5), 1287–1293. doi:10.3758/s13423-014-0593-6

Gozli, D. G., Chasteen, A. L., & Pratt, J. (2013). The cost and benefit of implicit conceptual cues for visual attention. *Journal of Experimental Psychology: General*, *142*, 1028–1046. doi:10.1037/a0030362

Gozli, D. G., Chow, A., Chasteen, A. L., & Pratt, J. (2013). Valence and vertical space: Saccade trajectory deviations reveal metaphorical spatial activation. *Visual Cognition*, *21*(5), 628–646. doi:10.1080/13506285.2013.815680

Imada, T. (2012). Cultural narratives of individualism and collectivism: A content analysis of textbook stories in the United States and Japan. *Journal of Cross-Cultural Psychology*, *43*(4), 576–591. doi:10.1177/0022022110383312

Kashima, Y., Kashima, E., & Kidd, E. (2014). Language and culture. In T. M. Holtgraves (Ed.), *The Oxford handbook of language and social psychology* (pp. 46–61). Oxford: Oxford University Press.

Kashima, Y., Kashima, E. S., Kim, U., & Gelfand, M. (2006). Describing the social world: How is a person, a group, and a relationship described in the East and the West? *Journal of Experimental Social Psychology*, *42*(3), 388–396. doi:10.1016/j.jesp.2005.05.004

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh*. New York, NY: Basic Books.

Louwerse, M. M. (2008). Embodied relations are encoded in language. *Psychonomic Bulletin & Review*, *15*(4), 838–844. doi:10.3758/PBR.15.4.838

Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition*, *114*(1), 96–104. doi:10.1016/j.cognition.2009.09.002

Maass, A., & Russo, A. (2003). Directional bias in the mental representation of spatial events: Nature or culture? *Psychological Science*, *14*(4), 296–301. doi:10.1111/1467-9280.14421

Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., … Google Books, T. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182. doi:10.1126/science.1199644

Ouellet, M., Santiago, J., Funes, M. J., & Lupiáñez, J. (2010). Thinking about the future moves attention to the right. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 17–24. doi:10.1037/a0017176

Santiago, J., Lupianez, J., Perez, E., & Funes, M. J. (2007). Time (also) flies from left to right. *Psychonomic Bulletin & Review*, *14*(3), 512–516. doi:10.3758/BF03194099

Setic, M., & Domijan, D. (2007). The influence of vertical spatial orientation on property verification. *Language and Cognitive Processes*, *22*(2), 297–312. doi:10.1080/01690960600732430

Shaki, S., Fischer, M. H., & Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychonomic Bulletin & Review*, *16*(2), 328–331. doi:10.3758/pbr.16.2.328

Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology*, *23*(4), 515–557. doi:10.1016/0010-0285(91)90005-9

Zhou, Y., & Fan, Y. (2015). A contrastive study of temporal-spatial metaphor between Chinese and Americans. *Theory and Practice in Language Studies*, *5*(1), 119. doi:10.17507/tpls.0501.16

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin & Review*, *10*(4), 954–958. doi:10.3758/BF03196557

Appendixes

Appendix A

A screen shot of the questionnaire in Study 1. (All words and headers would appear in Mandarin for Mandarin-language participants).

Delay	○	○	○	○	○	○
Depressed	○	○	○	○	○	○
Despair	○	○	○	○	○	○
Devil	○	○	○	○	○	○
Diseased	○	○	○	○	○	○
Divorce	○	○	○	○	○	○
Doom	○	○	○	○	○	○
Dread	○	○	○	○	○	○
Dream	○	○	○	○	○	○
	Up	Down	Left	Right	Positive	Negative
Earnest	○	○	○	○	○	○
Enemy	○	○	○	○	○	○
Enraged	○	○	○	○	○	○
Ethical	○	○	○	○	○	○
Faith	○	○	○	○	○	○
Fearful	○	○	○	○	○	○
First	○	○	○	○	○	○
Foolish	○	○	○	○	○	○
Fraud	○	○	○	○	○	○
Free	○	○	○	○	○	○
	Up	Down	Left	Right	Positive	Negative
Friday	○	○	○	○	○	○
Friend	○	○	○	○	○	○
Frustrated	○	○	○	○	○	○
Funny	○	○	○	○	○	○
Generous	○	○	○	○	○	○
Genius	○	○	○	○	○	○
Gentle	○	○	○	○	○	○
Gloomy	○	○	○	○	○	○
God	○	○	○	○	○	○

Appendix B

A screen shot of the questionnaire in Study 2.

	Strongly Left 非常"左"	Somewhat Left 有一点"左"	Neither Left Nor Right (Neutral) (中性)	Somewhat Right 有一点"右"	Strongly Right 非常"右"
Achieve	○	○	○	○	○
Active	○	○	○	○	○
Admire	○	○	○	○	○
Agile	○	○	○	○	○
Aimless	○	○	○	○	○
Altitude	○	○	○	○	○
Ambitious	○	○	○	○	○
Argue	○	○	○	○	○
Ashamed	○	○	○	○	○
Atmosphere	○	○	○	○	○
Beauty	○	○	○	○	○
Belong	○	○	○	○	○
Bitter	○	○	○	○	○
Bleak	○	○	○	○	○
Brave	○	○	○	○	○
Bright	○	○	○	○	○
Business	○	○	○	○	○
Celebrate	○	○	○	○	○
Champion	○	○	○	○	○
Cheat	○	○	○	○	○
Cheer	○	○	○	○	○
Clean	○	○	○	○	○
Clumsy	○	○	○	○	○
Complete	○	○	○	○	○
Contempt	○	○	○	○	○
Creator	○	○	○	○	○
Crime	○	○	○	○	○
Cruel	○	○	○	○	○