Bridging cultural and cognitive perspectives on similarity reasoning

Alexandra Carstensen (abc@ucsd.edu)[1⋆], Chiara Saponaro (c.saponaro@campus.unimib.it)[2⋆], Michael C. Frank (mcfrank@stanford.edu)[3], and Caren M. Walker (carenwalker@ucsd.edu)[1]

[1]Department of Psychology, UC San Diego, [2]Department of Psychology, University of Milano-Bicocca, [3]Department of Psychology, Stanford University

Abstract

Is a cow more closely related to grass or to a chicken? Responses vary by culture and age, among other factors. Those from western societies (or independent-leaning regions within interdependent non-western societies) are more likely to endorse the taxonomic match, the chicken, over the thematic match, grass (Chiu, 1972; Talhelm et al., 2014). This preference has been documented – largely in western cultures – to increase over development (e.g., Smiley & Brown, 1979). While neither development nor culture occur independently of the other, comparisons across these areas are problematic. We address one potential barrier to comparing cultural and developmental research using this classic paradigm – stimulus format – and show that the use of text (versus image) stimuli can bias participants toward taxonomic responding in some contexts. We present stimuli designed for cross-cultural use with children and adults and document country, regional, and demographic variation across the US and Italy.

Keywords: similarity; taxonomic and thematic semantics; analytic and holistic reasoning; culture; relational reasoning

Introduction

The concept of *similarity* has been studied broadly within the frameworks of cognitive processing and development. It is fundamental to categorization and a linchpin of conceptual development in childhood, supporting abstract thought. A key paradigm for studying similarity is the taxonomic-thematic triads task, which has been used widely in cognitive, developmental, and cultural psychology research. Here, we take initial steps toward bridging these research traditions and methodologies by (1) investigating possible confounding effects of stimulus format (text vs. image) and (2) assessing cultural, regional, and demographic variation in similarity reasoning using this task.

Taxonomic and thematic perspectives on similarity have a long history within cognitive and developmental research. One of the earliest approaches to documenting similarity reasoning relied on open-ended sorting to identify common strategies in categorization, including analytic and holistic reasoning (e.g. Annett, 1959; Kagan, Moss, & Sigel, 1963). These foundational studies identified analytic reasoning as a focus on the similarity between object parts and properties, e.g., cows and chickens are related because they are both animals. In contrast, they define relational, or holistic similarity as motivated by the functional relationships between objects, e.g., cows and grass are related because cows eat grass. Kagan, Moss, & Sigel (1963) documented analytic and relational reasoning patterns across a variety of behavioral tasks, reported stable individual differences in reasoning style, and argued for links between reasoning style and a broad range of demographic and personality factors. In more recent work, approaches to similarity reasoning have been systematically contrasted in *triad tasks* that are designed to pit taxonomic and thematic matches against each other to evaluate participants' reliance on analytic versus holistic similarity. Specifically, two objects are assigned to the same taxonomic category if they share similar attributes, and they are assessed as sharing a thematic relationship if they co-occur in causal, spatial, and temporal contexts (Markman & Hutchinson, 1984), including culture-specific associations (Golonka & Estes, 2009; Wisniewski & Bassok, 1999), a somewhat broader view than in earlier accounts.

Preferences for taxonomic or thematic relations in similarity reasoning are also often cited as central tenants of analytic versus holistic thought and core to cross-cultural differences in cognition. For example, Chiu (1972) linked culture to analytic and holistic reasoning using a taxonomic-thematic triad task, showing that children in China were more likely than their peers in the US to select thematic matches and to highlight thematic similarities in justifying their choices. Ji, Zhang, & Nisbett (2004) expanded on this work, demonstrating that cross-cultural differences in similarity reasoning, while often confounded with language, cannot be fully explained as an effect of language. They showed that Chinese-English bilinguals (both Mandarin and Cantonese speakers) respond more thematically than English-speaking Americans, whether tested in Chinese or English, and in mainland China or the United States. Indeed, systematic variation in taxonomic and thematic reasoning exists between speakers of the same language and residents of the same country and has been linked to regional subcultures in Italy (Knight & Nisbett, 2007) and China (Talhelm et al., 2014), as well as a range of demographic factors in the US, including politics, gender, age, socioeconomic status, education, personality (extraversion, openness), and cognitive reflection (Talhelm et al., 2015). Despite extensive use of the triads task in cultural, cognitive, and developmental research, there is little consensus on how phenomena and mechanisms within each tradition relate across these literatures, resulting in a gap between these perspectives.

*The first two authors contributed equally to this work.

The taxonomic shift in similarity reasoning

Across several accounts, a key feature of cognitive development is the shift from a thematic to a taxonomic perspective on similarity relations (e.g. Annett, 1959; Denney, 1974; Denney & Ziobrowski, 1972; Gentner & Rattermann, 1991; Inhelder, Piaget, & Papert, 1964; Smiley & Brown, 1979). Specifically, studies have found evidence for a thematic preference in preschool and school-aged children (Bauer & Mandler, 1989; Greenfield & Scott, 1986; Scott, Serchuk, & Mundy, 1982; Walsh, Richardson, & Faulkner, 1993) and for a gradual shift toward increased taxonomic responding over middle childhood (Kagan, Moss, & Sigel, 1963; Smiley & Brown, 1979; cf. Greenfield & Scott, 1986; Walsh, Richardson, & Faulkner, 1993). This taxonomic shift has historically been explained as an effect of experience and education. That is, as children gain experience and formal education, they move from a basic conceptual organization, in which objects are represented according to their reciprocal relationship in the real world, to a formal and taxonomic organization, in which entities are represented according to their category membership (Smiley & Brown, 1979).

Critically, however, these age effects are often confounded with stimulus modality: children who are presented with images tend to respond more thematically, while adults who are presented with text tend to respond more taxonomically. Consistent with this potential confound, Lin & Murphy (2001) (Expt. 4) presented adults with a triad task using pictures (alongside text) and observed a thematic preference, a finding that ran opposite to their predictions. These results are consistent with adult performance in other sorting tasks using pictures as stimuli, with both nonliterate (Luria, 1976) and literate (Murphy, 2001) adults. In the latter study, participants tended to group items thematically, even when the instructions emphasized the taxonomic nature of the categories.

The present study

Despite considerable overlap in experimental methods and cognitive constructs, research on similarity in cultural and developmental psychology traditions remains largely distinct. Here, we take initial steps toward bridging these traditions by evaluating the comparability of experimental paradigms across areas. As previously noted, there is reason to suspect that the typical stimuli used within each tradition – images in developmental studies and text in cultural ones – may induce different modes of responding. If so, this presents a barrier to making meaningful comparisons across these literatures, and a potential confound to research documenting the taxonomic shift.

To our knowledge, there are no studies that have addressed the potential confounding effects of stimulus type by presenting the *same set of stimuli* in image *and* text format to participants from *one age group*. Here we provide such a test, with adults in the United States and Italy. To preview our results, we find that there is no effect of stimulus type in our US sample. However, given that US adults are considered to be among the most taxonomic-leaning across cultures (Nisbett, Peng, Choi, & Norenzayan, 2001), the Italian sample provides a stronger test of generality. Within the Italian sample, we do find evidence suggesting that stimulus type may act as a confounding factor in comparisons between studies with children and adults. In the first section of our analysis, we probe this and other issues of consistency and reliability in the triad task paradigm. In the second section, we explore variation between countries and regions, validating our task as a measure of variation within and across countries. Finally, we pursue a set of exploratory questions about demographic factors linked to variation in similarity reasoning within both cognitive and cultural frameworks.

Method

We assessed preferences in similarity judgments using a triad task that requires participants to choose between a taxonomic and thematic match for each cue item. All participants received the same set of triad stimuli, but we randomized the presentation format of each triad as image or text to assess the effect of stimulus format within subjects.

Our methods and confirmatory analyses were preregistered separately for the US (`https://aspredicted.org/iz4ui.pdf`) and Italian samples (`https://aspredicted.org/xt95z.pdf`). US and Italian participants were run as separate experiments. We consolidate them here for ease of presentation but separate the analyses for each. The web experiment and stimuli are available at `https://osf.io/9uve8/`.

Participants

We recruited 200 adult participants via Amazon Mechanical Turk, 100 native English speakers from the US and 100 native Italian speakers from Italy, 50 of whom came from northern Italian regions and 50 from southern.[1] All recruitment, consent, experiment, and demographic questionnaire text was presented in the test language (English in the US; Italian in Italy).

To ensure that participants followed our task instructions, we included 8 unambiguous attention check questions in our experiment (e.g., ear: eye/fan), with one on each of the study pages, and excluded any participants who made 2 or more errors across these questions, including a number of apparent bots in the US sample (28 excluded from the US sample, 5 IT). To minimize cultural influences external to the target

[1] We define northern and southern regions in Italy following the geographical criteria used by the Italian National Institute of Statistics (Istituto Nazionale di Statistica, 2021), grouping the Northwest and Northeast macroregions as northern Italy and the South and Islands macroregions as southern Italy. Amazon Mechanical Turk (AMT) recruitment is designed to be country-specific, but not region-specific. In half of our IT postings, we specified that participants must be from northern Italy in our AMT study title, and southern for the other half. At the end of the study, we asked participants whether they grew up in Italy, and if so, what region they grew up in. If participants reported growing up outside of Italy or in a central region (i.e., Tuscany, Lazio, Umbria or The Marches), we excluded their data and recruited replacement participants until we reached 50 participants each from north and south.

countries, we stated in recruitment materials that participants must be native speakers of the test language. We also included a question about native language and planned to exclude participants for noncompliance, but all participants in the two samples reported being native speakers of the relevant test language. We also excluded participants who had lived abroad for more than two years in a region (continent) with predominantly non-Western societies (5 US, 0 IT), and those who reported speaking any language predominantly spoken in non-Western societies with self-rated fluency of at least 6 out of 10 in speaking and comprehension (1 US, 10 IT). Multiple exclusion criteria applied to some participants, resulting in fewer unique exclusions than the total number of exclusion criteria met.

After exclusions, the final sample consisted of 70 participants from the US (mean age = 39.9y; 28F, 41M, 1 decline to state), 44 from northern Italy (mean age = 34y; 12F, 32M), and 44 from southern Italy (mean age = 29.7y; 20F, 24M).

Stimuli

We developed a novel stimulus set of 84 triads, each consisting of a single cue item and two match options, with one bearing a thematic relationship with the target and the other a taxonomic relationship (e.g., carrot: rabbit/peas; Figure 1). We collected and adapted stimuli from previous triad tasks (Ji, Zhang, & Nisbett, 2004; e.g., Markman & Hutchinson, 1984; Waxman & Namy, 1997) to create a stimulus set appropriate for children and adults across a range of cultural contexts that are compatible with both text and image presentation. As much as possible, we sought to avoid salient overlap between items within each triad, including lexical and phonological similarity in labels and shared shape, color, or style across images. We also ran a pilot study to ensure that our image stimuli were easily recognizable without labels and corresponded well to the assigned label.[2]

Following the same criteria for child-appropriateness and non-overlapping representations, we also collected, adapted, and produced filler triads for a total of 84 sets that did not follow the structure of test triads with contrasting taxonomic and thematic match options (e.g., sunset: day/night). Finally, we created 8 unambiguous attention check trials, in which a cue item was unambiguously more closely related to one of the two choice options, e.g., ear: eye/fan).

Figure 1: Example trials, test (left) and filler (right).

We included a range of demographic questions assessing factors that have been implicated in within-country variation in similarity reasoning, including languages spoken (Ji, Zhang, & Nisbett, 2004) age, gender, ethnicity, education, subjective socioeconomic status (Talhelm et al., 2015), and related factors like experience with international travel and (in the Italian sample) years living in northern, central, and southern Italian regions.

The task instructions, stimulus labels, and demographic questions were translated into Italian by a native Italian speaker fluent in English, backtranslated to English by an Italian-English bilingual blind to the original instructions, and this back translation was checked against the original instructions for accuracy, with corrections made in any cases where a closer translation could be identified.

Procedure

Each participant received half of the questions in image format and half as text, in two blocks. Block order (image/text) and the specific test and filler triads contained within each were randomized between subjects. As in previous work, participants read brief text instructions at the beginning of the study directing them to choose the option most closely related to the cue in each triad, with an example filler item in text format (i.e., today: yesterday/tomorrow). Each triad was presented vertically, with the cue item centered and options presented below (with left/right position randomized). Participants completed 8 pages of trials containing 21 filler and test items and 1 attention check each. An instruction line at the top of each page prompted participants to select the choice "most closely related" (Italian: "più strettamente collegata") to the cue item. After completing the triad portion of the task, participants received the demographic questions. The study took about 15 minutes to complete.

Results

The analysis script and all data are available at `https://osf.io/9uve8/`. Following our preregistration, all analyses are performed within-country with the exception of the model testing for differences between adults in the US and Italy. All models were implemented as Bayesian binomial logistic regressions using the BRMS package in R (Bürkner, 2018) with default (weakly informative normal) priors.

[2]To ensure that image and label stimuli were interpretable and corresponded to each other, we conducted a labeling manipulation check on Amazon Mechanical Turk. Ten participants labeled each image, and image-label pairs were only accepted when the most frequent term produced by participants matched the intended term. Image-label pairs that were labeled with high agreement using a term other than the target were updated to the most commonly elicited label. Images with poor labeling agreement were replaced and the labeling process repeated. If the most frequent label for an item was superordinate or ambiguous relative to the intended label, we favored specificity and maintained the subordinate label for the item (e.g., "frying pan" for "pan") as long as the more specific label was also produced by some participants. About 10% of the items were modified during this process; in general, labels elicited in the pilot largely matched the intended label, with most variation in naming coming from the use of synonyms or labels at different levels of specificity (e.g., "juice," "orange juice," "drink").

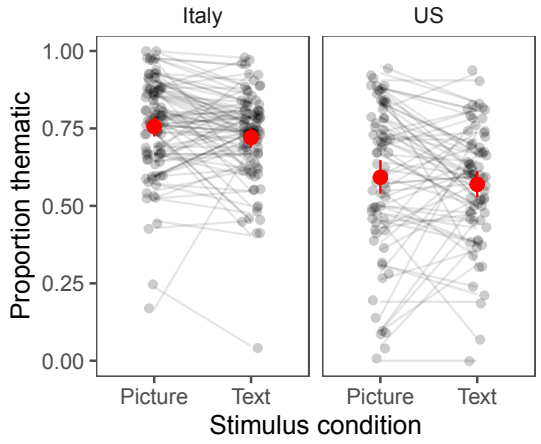

Figure 2: Proportion thematic judgments by country and condition. Individual black dots show participant means (connected by lines within participants). Red dots show condition means, and error bars show 95% confidence intervals.

Task characteristics and reliability

Do text stimuli bias toward taxonomic responding? To evaluate the hypothesis that the use of text stimuli partly contributes to taxonomic preferences observed in adults, we fit a regression predicting thematic responding on each trial as a function of stimulus type (image or text), modeled as a fixed effect. As random effects, we included intercepts for triads and subjects, as well as by-subject random slopes for the effect of stimulus format. Data for both the US and Italy are shown in Figure 2. Among participants from the US, we did not observe a reliable effect of stimulus type (β = -0.168, 95% Bayesian Credible Interval = [-0.534, 0.197]). We evaluated evidence for the null hypothesis, that stimulus type had no effect (on the US sample), using a Bayes Factor analysis to compare support for the test model relative to a null model that omitted the stimulus format term but was otherwise identical. The estimated Bayes Factor in favor of the test model over the null was 1.971, which does not meet our criterion of BF $>$ 3 or $<$ 1/3.

In addition to our preregistered confirmatory test of stimulus type, we also performed an exploratory test for fatigue effects that included an additional fixed effect of trial number, an interaction between stimulus type and trial number, and the same random effect structure as above. This model showed a very small effect of trial number with a credible interval whose lower bound was very close to zero (β = 0.0034, 95% CI = [0, 0.0068]), indicating a potentially small decrease in thematic responding across the experiment, and no other effects with credible intervals excluding zero.

We repeated both of the above regressions with data from our Italian sample, though in this case, we preregistered the second model, including a trial number term and interaction, as our main analysis. Both analyses show reliable effects of stimulus type among IT participants (from the preregistered model, text β = -0.516, 95% CI = [-1.037, 0.004]). Again, we observed a very small effect of trial number (β = -0.0034, 95% CI = [-0.0065, -0.0004]).

Sample	Condition	Intra-rater ICC	Inter-rater ICC
US	Pictures	0.37	0.92
US	Text	0.29	0.93
IT	Pictures	0.31	0.95
IT	Text	0.30	0.96

Table 1: ICCs for reliability within participants (intra-rater) and stimuli (inter-rater) for both samples and conditions.

Taken together, these analyses indicate that (1) using text stimuli (versus images) can lead participants to favor taxonomic over thematic matches in some contexts, highlighting a potentially confounding factor, but that (2) the biasing effect of text varies across populations, and (3) fatigue effects may be negligible even in studies with nearly 200 items.

Is responding reliable within individuals or stimuli? We used intraclass correlation coefficients (ICCs) to analyze the stability of participants' judgments across triads. Following the standard taxonomy for ICCs, we used a two-way, non-interactive model of the effects of stimulus and participant (model 2A from McGraw & Wong, 1996). Across both stimulus presentation conditions and both samples, we found a very consistent pattern (Table 1): while participants' judgments were variable across triads, inter-rater agreement about specific triads was quite high. Thus, it appears that specific triads had very reliable levels of taxonomic vs. thematic responding.

Measure validity and variation

Conceptual replication of Knight and Nisbett (2007) We designed our Italian sample to support a conceptual replication of previous work documenting cultural variation between northern and southern Italy. Regional effects are shown in Figure 3. Following our preregistration, we fit a model predicting thematic responding on each trial, with the region in which participants grew up (north or south) as a fixed effect, and as random effects, we included intercepts for triads and subjects, as well as by-subject random slopes for the effect of stimulus format. As predicted, we found a reliable effect of region (southern β = 0.521, 95% CI = [0.068, 0.965]), showing that participants from southern Italy were more likely to select thematic matches. This finding provides validation for our task as a measure of variation between cultures and extends the findings of Knight and Nisbett to our broader sample population, which includes adults convenience sampled from throughout northern and southern Italy (as compared to secondary school students from four schools).

We also tested for an effect of socioeconomic status (SES), and an interaction between SES and region within Italy. Knight & Nisbett (2007) observed a main effect of region, with southern Italians responding more thematically, and an

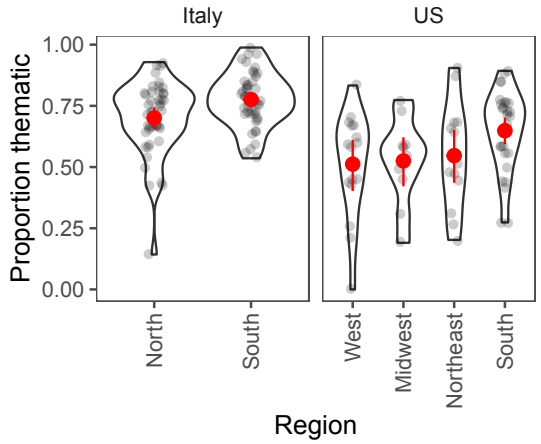

Figure 3: Proportion thematic judgments by region within each country. Plotting convention as above.

interaction with SES, such that this effect was especially strong for low-SES southern Italians, and weaker for their high-SES peers. We fit a model identical to that above, but with region, SES, and their interaction as fixed effects. This analysis did not identify any reliable effects. We tested evidence for the null hypothesis that SES has no effect using a Bayes Factor analysis. The null model in this analysis included only one fixed effect, region, and the test model included fixed effects of region and SES (though no interaction term, for closer comparability with the null model). This analysis estimated a Bayes Factor of 0.482 in favor of the test model over the null, which does not satisfy our criterion to accept the null, suggesting that our SES analysis is underpowered.

Finally, we explored effects of US region (west/midwest/northeast/south[3]) on thematic responding, with the same random effects structure as above. We found an effect of region, with participants who grew up in the south responding more thematically than those from the west (south $\beta = 0.379$, 95% CI = [-0.485, 1.297]).

Testing for US-Italy differences In addition to our extension of previous cross-cultural findings, we predicted a novel cross-cultural difference: that participants in the US would be more likely to select taxonomic matches than those in Italy. To test this prediction, we modeled thematic responding as a function of country (US/IT) as a fixed effect, with the same random effect structure as in the region models. Our regression bore out this prediction, finding a reliable effect of country (US $\beta = -0.99$, 95% CI = [-1.338, -0.664]) in the predicted direction.

Exploring demographic effects within cultures A broad range of demographic factors have been implicated in variation within both developmental and cross-cultural studies of taxonomic and thematic reasoning (e.g., Kagan, Moss, & Sigel, 1963; Talhelm et al., 2015), including gender, age, education, and SES. All of these analyses were not preregistered and should be considered exploratory; all will need to be replicated in future work.

Within our US and Italian samples we predicted thematic responding as a function of gender (M/F/other), age (in years, centered), education (coded ordinally), and SES (MacArthur ladder, centered), all as fixed effects in a large omnibus regression with the same random effect structure as in previous regressions. In the omnibus regression with US data, none of these effects held. In the Italian omnibus model, we observed an effect of SES ($\beta = 0.166$, 95% CI = [0.025, 0.31]) with higher SES participants responding more thematically [in contrast to previous findings; Knight & Nisbett (2007); Talhelm et al. (2015)].

Discussion

This study aimed to (1) investigate possible confounding effects of stimulus format (text or image) in a taxonomic-thematic triad task, and (2) assess cultural, regional, and demographic variation in similarity reasoning.

While we found no effect of stimulus type in the US sample, we did observe an effect in the Italian sample, where text stimuli biased participants toward more taxonomic responding compared with image stimuli. This finding presents a cautionary note for direct comparisons between adult and developmental data, as studies with adults tend to use text stimuli and find more taxonomic responding, while studies with children often use image stimuli and report relatively more thematic responding. As a result, stimuli may present a confound for developmental findings indicating a taxonomic shift. On the other hand, we observed a stimulus format effect for only one of the two populations we examined. Thus, it is unclear to what extent (if any) it undermines these findings. Despite this, the differential effect of stimulus format in our two samples provides another cautionary note: that task format can introduce confounds which interact with culture.

Future work should aim to discover whether this stimulus format effect is caused by, for example, reliance on word co-occurrence information (for text stimuli) that may differ from scene statistics (invoked in processing image stimuli), or perhaps a variation on Smiley & Brown's (1979) proposal, by which formal education increases taxonomic responding. Elaborating this view, it could be that text stimuli prime propositional knowledge about taxonomic relationships, or featural similarities between the cue item and taxonomic match, while image stimuli could induce more holistic reasoning, highlighting associative or ecological relationships.

We also found a small effect of trial number in both populations, suggesting mild fatigue effects. However, the small size of the effect suggests that many more trials can be included in triad tasks than is common in the literature (where roughly 20 trials is typical). We therefore recommend inclu-

[3]We operationalized US regions following the US Census Bureau's definitions (United States Census Bureau, 2013).

sion of additional trials and the use of filler stimuli to obscure the structure of test trials and reduce the influence of task demands when practicable.[4]

We evaluated the consistency of responses in our task within subjects and within triads, and found stable patterns across the text and image conditions, and across the two countries. While individual participants varied their responses, we observed high consistency across subjects in responses to individual stimuli, suggesting high reliability in the measure.

We validate our task as a measure of cultural variation through a conceptual replication of previous work showing regional differences between northern and southern Italy (Knight & Nisbett, 2007). Specifically, we show that participants who grew up in southern regions tend to respond more thematically than their counterparts from northern regions. Knight & Nisbett (2007) attribute the thematic preference in southern Italy to a tendency toward social interdependence in the south (compared to independence in the north), which may also be amplified by economic necessity. They argue that the increased interdependence in southern Italy promotes holistic processing – as in East Asia – resulting in reasoning that is less focused on taxonomic similarities and more on thematic. Following this prior work, we examined the effects of region and socioeconomic status (SES) within Italy. Contrary to previous findings, we did not observe an interaction between region and SES, though our study was likely underpowered to detect such an effect. Nonetheless, we do extend the previously observed regional differences to a broader, less controlled sample. While Knight and Nisbett's study tested participants from four Italian high schools, in Naples and Crotone (the south) and Milan and Monza (north), we recruited a broader sample of adults from the two regions, for a more naturalistic and perhaps more conservative test.

Although likely underpowered, our exploratory analyses of US regional variation and demographic effects within both cultures found an intriguing effect of region within the US and SES within the Italian sample. Within the US, participants who grew up in southern states were significantly more likely to select thematic matches than those who grew up in western states, perhaps reflecting effects of voluntary settlement (see, e.g., Kitayama, Varnum, & Sevincer, 2014), or differing levels of interdependence within US regional cultures more broadly. Future work is needed to examine these effects in a confirmatory manner. In Italy, higher SES was related to more thematic responding (contrasting with previous findings; Knight & Nisbett (2007); Talhelm et al. (2015)). The reliable main effect of SES in this omnibus regression also contrasts with the results of our confirmatory analysis, which tested for an interaction between region and SES in Italy and did not show a reliable main effect of SES or any interaction. This, combined with the modest effect estimate, suggests that the effect of SES in our sample is small and likely emerges in the omnibus regression as a result of statistically controlling for the other demographic variables included there.

Drawing on previous research showing that US adults are among the most analytic reasoners across cultures, including those in western Europe (Nisbett, Peng, Choi, & Norenzayan, 2001), we predicted a bias toward taxonomic responding in our US sample compared to Italy, which we found.[5] This novel finding is consistent with previous work and aligns with cross-cultural theory relating social independence to taxonomic reasoning.

This study has taken initial steps toward bridging between similarity research in cognitive and cultural traditions. In doing so, we engage with recent calls to foster cultural perspectives on development (Heiphetz & Oishi, 2021) and local, mechanism-focused explanations of cultural differences, i.e., closer to the level of individuals than countries (Miyamoto, 2013). To this end, we developed a stimulus set appropriate for children and validated its use as a measure of cross-cultural variation, providing a foundation for future work to explore behavior and mechanisms at the intersection of development and cross-cultural variation.

Many open questions remain for future research: Do cross-cultural differences in similarity reasoning extend beyond the taxonomic-thematic contrast to broader construals of similarity? Is the taxonomic shift evident in development across cultures, or culture-specific? Are the sources of country-level, regional, and demographic variation in adults relevant for children? Do common or distinct mechanisms underlie these effects? For example, Smiley & Brown (1979) found preferences for thematic responding in both preschoolers and elderly adults, attributing taxonomic responding to the influence of formal education, and concluding that adults drift further from this mode of responding as they age. Taken together with cross-cultural variation, it seems plausible that *western* education in particular may bias individuals toward taxonomic similarity, explaining both age and at least some of the culture effects. Within the cognitive development framework, future research should explore regional variation over development in school-aged children, evaluating the extent to which the taxonomic shift is typical of development versus specific to (regional) subcultures, or perhaps formal education in a western cultural context.

Going beyond effects of age and country-level culture, it is less clear that the western formal education view can account for regional variation within countries linked to social factors like independence or interdependence – or to demographic factors like gender and socioeconomic status. A remaining challenge for cognitive and cross-cultural psychology is to disentangle demographic and contextual factors at the level of psychological mechanisms, clarifying when it is or is not appropriate to view demographic variation as a type of cultural variation.

[4] In an early pilot study, we varied the ratio of test to filler items, and found that participants who received twice as many test as filler items were often aware of the taxonomic-thematic contrast among most items. However, few of the participants in the 1:1 pilot articulated this feature of the trial structure.

[5] NB: this taxonomic preference is relative, not absolute, as both populations show an overall preference for thematic matches.

Acknowledgements

We are very grateful to Nicky Sullivan for help with early stimulus design and piloting. We also thank members of the Early Learning and Cognition Lab at UC San Diego and the Language and Cognition Lab at Stanford University for their feedback and helpful comments, especially Khuyen Le and George Kachergis. This work was funded in part by NSF under grant SBE-2047581, a CAREER Award supporting CMW and AC, and by awards from the McDonnell Foundation and the Center for the Study of Language and Information at Stanford University supporting AC.

References

10 Annett, M. (1959). The classification of instances of four common class concepts by children and adults. *British Journal of Educational Psychology*, *29*(3), 223–236.

Bauer, P. J., & Mandler, J. M. (1989). Taxonomies and triads: Conceptual organization in one- to two-year-olds. *Cognitive Psychology*, *21*, 156–184.

Bürkner, P.-C. (2018). Advanced bayesian multilevel modeling with the r package brms. *R Journal*, *10*, 395–411. http://doi.org/10.32614/RJ-2018-017

Chiu, L.-H. or L. H. (1972). A cross-cultural comparison of cognitive styles in chinese and american children. *International Journal of Psychology - INT J PSYCHOL*, *7*, 235–242. http://doi.org/10.1080/00207597208246604

Denney, N. W. (1974). Evidence for developmental changes in categorization criteria for children and adults. *Human Development*, *17 1*, 41–53.

Denney, N. W., & Ziobrowski, M. (1972). Developmental changes in clustering criteria. *Journal of Experimental Child Psychology*, *13*, 275–282.

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. E. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 225–277). Cambridge University Press. http://doi.org/10.1017/CBO9780511983689.008

Golonka, S., & Estes, Z. (2009). Thematic relations affect similarity via commonalities. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*, 1454–64. http://doi.org/10.1037/a0017397

Greenfield, D., & Scott, M. S. (1986). Young children's preference for complementary pairs: Evidence against a shift to a taxonomic preference. *Developmental Psychology*, *22*, 19–21.

Heiphetz, L., & Oishi, S. (2021). Viewing development through the lens of culture: Integrating developmental and cultural psychology to better understand cognition and behavior. *Perspectives on Psychological Science*, *17*(1), 62–77. http://doi.org/10.1177/1745691620980725

Inhelder, B., Piaget, J., & Papert, D. (1964). *The early growth of logic in the child: Classification and seriation*. Routledge.

Istituto Nazionale di Statistica. (2021). Popolazione residente e dinamica demografica anno 2020. [Press release]. Retrieved from `https://www.istat.it/it/files//2021/12/CEN%20SIMENTO-E-DINAMICA-DEMOGRAFICA-2020.pdf`

Ji, L.-J., Zhang, Z., & Nisbett, R. (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*, 57–65. http://doi.org/10.1037/0022-3514.87.1.57

Kagan, J., Moss, H. A., & Sigel, I. E. (1963). Psychological significance of styles of conceptualization. *Monographs of the Society for Research in Child Development*, 73–112.

Kitayama, S., Varnum, M., & Sevincer, A. T. (2014). The frontier: Voluntary settlement and cultural change. *Culture Reexamined: Broadening Our Understanding of Social and Evolutionary Influences*, 93–127. http://doi.org/10.1037/14274-005

Knight, N., & Nisbett, R. (2007). Culture, class and cognition: Evidence from italy. *Journal of Cognition and Culture*, *7*(3-4), 283–291.

Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General*, *130*(1), 3.

Luria, A. R. (1976). *Cognitive development: Its cultural and social foundations*. Harvard university press.

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, *16*(1), 1–27.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*(1), 30.

Miyamoto, Y. (2013). Culture and analytic versus holistic cognition: Toward multilevel analyses of cultural influences. In *Advances in experimental social psychology* (Vol. 47, pp. 131–188). Elsevier.

Murphy, G. L. (2001). Causes of taxonomic sorting by adults: A test of the thematic-to-taxonomic shift. *Psychonomic Bulletin & Review*, *8*(4), 834–839.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, *108*(2), 291.

Scott, M. S., Serchuk, R., & Mundy, P. (1982). Taxonomic and complementary picture pairs: Ability in two to five-year-olds. *International Journal of Behavioral Development*, *5*(2), 243–256.

Smiley, S. S., & Brown, A. L. (1979). Conceptual preference for thematic or taxonomic relations: A nonmonotonic age trend from preschool to old age. *Journal of Experimental Child Psychology*, *28*(2), 249–257.

Talhelm, T., Haidt, J., Oishi, S., Zhang, X., Miao, F. F., & Chen, S. (2015). Liberals think more analytically (more "WEIRD") than conservatives. *Personality and Social Psychology Bulletin*, *41*(2), 250–267.

Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., & Kitayama, S. (2014). Large-scale psychological differences within china explained by rice versus wheat agri-

culture. *Science*, *344*(6184), 603–608.

United States Census Bureau. (2013). Statistical groupings of states and counties. [Press release]. Retrieved from https://www2.census.gov/geo/pdfs/reference/GARM/Ch6GARM.pdf

Walsh, M., Richardson, K., & Faulkner, D. (1993). Perceptual, thematic and taxonomic relations in children's mental representations: Responses to triads. *European Journal of Psychology of Education*, *8*(1), 85–102.

Waxman, S. R., & Namy, L. L. (1997). Challenging the notion of a thematic preference in young children. *Developmental Psychology*, *33*(3), 555.

Wisniewski, E. J., & Bassok, M. (1999). What makes a man similar to a tie? Stimulus compatibility with comparison and integration. *Cognitive Psychology*, *39*(3-4), 208–238.