An Empirical Exploration of Diversity Perception

Vassilis Markos[1,*], Loizos Michael[1,2,*]

[1]Open University of Cyprus (OUC), Nicosia, Cyprus
[2]CYENS Center of Excellence, Nicosia, Cyprus

Abstract

As our societies grow more complex and machine-mediated interactions become a prevalent means of communication, properly handling diversity comes as a necessity. This requires a quantification of diversity sensitive to how individuals perceive it. With this aim, we review existing popular measures of diversity, and examine their ability to capture human perceptions of diversity in a variety of cases, demonstrating their insufficiency in many of them. Moreover, we also present a draft exploration of factors that possibly affect individual perception in those cases.

Keywords

diversity metrics, diversity perception, rankings

1. Introduction

From retrieving search results through a search engine to automatically assessing and ranking job applications, diversity rises as a natural desideratum, enhancing fairness and boosting user confidence and trust. Thus, developing a sound metric of diversity, taking individual perceptions into account, is of crucial importance. Indeed, it is often the case that people act based on what they *think* is real versus what it might *be* real [1]. In particular, when asked to describe why and how people within a working team might be different, they tend to provide a plethora of context-specific differences as reasons [2]; a tendency illustrating the importance of *perceived* over *actual* reality. Also, Danbold and Unzueta [3] demonstrate how one's belonging to a certain group affects their perception of diversity. Similar cases of the effects of individual attributes on one's perception of diversity are demonstrated in [4], providing further evidence that whether a collection of items is diverse or not does not depend solely on properties of the collection itself.

With this work we aim to provide some preliminary results of our analysis on how people do perceive diversity under various circumstances. Our aim is not to provide a new definition of perceived diversity, but to observe and determine some factors that influence individual perception of diversity and the way they do so. In particular, we focus on special cases (e.g., rankings) and how people understand diversity in those settings.

HHAI-WS 2023: Workshops at the Second International Conference on Hybrid Human-Artificial Intelligence (HHAI), June 26—27, 2023, Munich, Germany

*Corresponding author.
vasileios.markos@st.ouc.ac.cy (V. Markos); loizos@ouc.ac.cy (L. Michael)

2. Background and Methodology

Most of the many definitions of diversity conceptualize it as dispersion across some dimension(s) (e.g., gender in workplace [5, 6] or species abundance in natural habitats [7]), often discriminating between group- and individual-level diversity [1], resulting to two typologies [1, 8]: (i) *Organizational Demography* (OD), which considers diversity at the collective level, as a distribution of certain individual features across a group/unit [9]. (ii) *Relational Demography* (RD), which considers diversity at both collective and individual levels, as an individual's distance from a group [1]. Under OD, diversity emerges as a group property [10], while within RD diversity is considered as a cross-level property of an individual within a group [11]. Regarding the defining attributes of diversity there are [8]: (i) *Mono-Attribute* approaches, viewing diversity over a single (group of) attribute(s) and; (ii) *Multiple-Attribute* approaches, involving several attributes. Mostly within the latter, one finds works exploring aspects of diversity perception [8]. In our context *perceived diversity* "might be defined as members' awareness of differences" [6], on the basis that the differentiated understanding of one's differences with others vastly affects their perception of diversity [12]. Moreover, individual expectations in certain settings affect perception of group differences [13] and, hence, one's views of diversity [14, 15]. In [3, 4, 16], it is argued that contextual and subjective factors, like group membership and structure, impact one's identity and, hence, their perception of diversity.

The simplest diversity metric, *richness*, counts the number of different classes C_i within a population P [17]. *Entropy*, H, [18] measures disorder within a group through log-weighted relative abundance, p_i, of each class C_i, $H = -\sum_{i=1}^{n} p_i \log p_i$. *Simpson's Index*, λ, [19] computes diversity as the approximate probability of belonging to the same class, $\lambda = \sum_{i=1}^{n} p_i^2$. *Berger-Parker Index*, bp, [20] utilizes the relative abundance of the largest class, $bp = \max_{i=1,\ldots,n} p_i$. If r_i denotes the rank of class C_i, a common ranked diversity metric is *Hall and Tideman's*, TH *Index* [21], $TH = {}^{1}/{2\sum_{i=1}^{n} r_i p_i - 1}$ (a variation of Simpson's Index embedding class ranks). Other approaches to measure diversity include using some dispersion measures (e.g., gender / race diversity [22, 23]) and qualitative measures within the specific context and purpose [24]. For more, see, e.g., [25].

We examined five variables: (a) ***Population awareness***, k, i.e., whether the sampling population was known; (b) ***Population ranking***, p, i.e., whether the sampling population was ranked; (c) ***Sample ranking***, s, i.e., whether the sample was ranked; (d) ***User involvement***, u, i.e., whether participants were assigned to a class; (e) ***Observed / Constructed diversity***, o, i.e., whether participants were asked to observe and assess or construct diversity. We encode each condition with a 5-letter string (e.g., kPsuo, where upper-case means the condition was controlled). We created 104 ordered pairs of conditions differing in exactly one dimension, collecting responses from 1040 crowd-workers, each compensated with 0.50$. Each was shown two groups (one per condition) of 1 training and 10 actual items. Consistency was controlled by including two identical items per group. Throughout the study, classes were denoted by randomly assigned colors; populations contained 24 elements and samples 12; unranked ensembles were represented as hollow pie-charts; ranked ones as linear arrays with order verbally indicated; user involvement was mentioned in task description and the participant's class color was shown; diversity estimates in observation tasks were provided through a slider; samples in construction tasks were sampled by a given population, or a list of available classes.

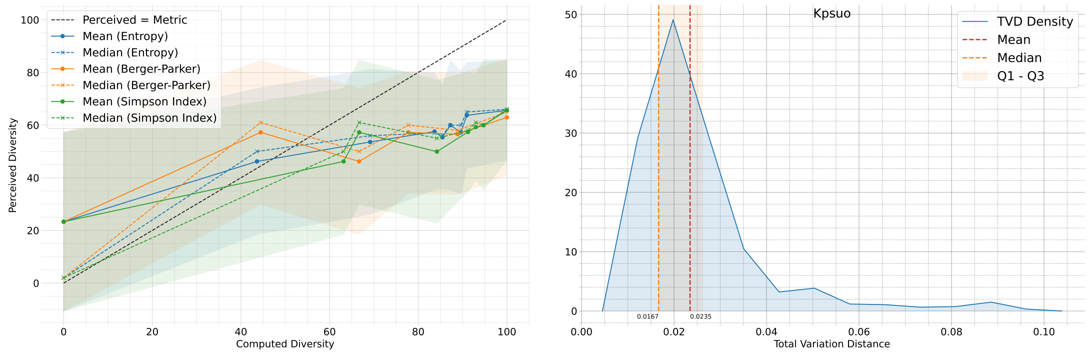


Figure 1: [Left] Diversity in kpsuO as determined by participants vs normalized metrcs (shaded area indicates ±1 std). [Right] TVD of population and sample distributions (Kpsuo).

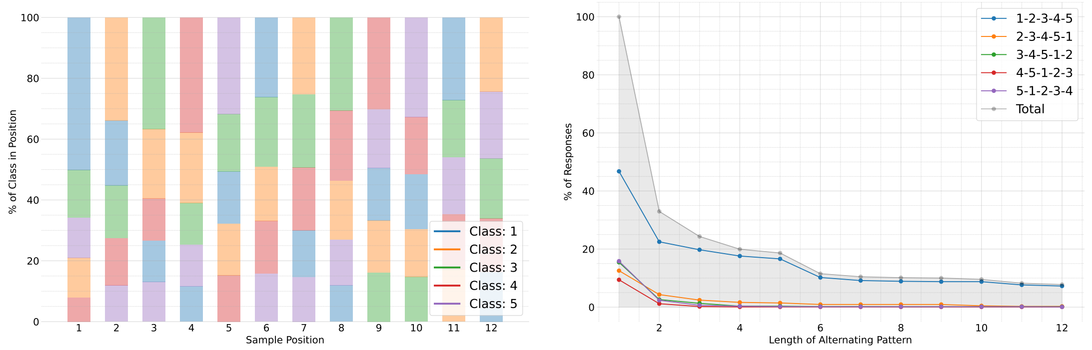
3. Empirical Results and Analysis

Since observation and construction tasks are structurally different, inconsistency in observation tasks was measured as the distance between the participants' provided estimations of diversity, while in construction tasks inconsistency between unranked samples was measured as the Manhattan distance between the corresponding class distributions, and in ranked ones as their Hamming distance. Responses on identical items were at most 17.5% inconsistent, thus none was excluded from further analysis. In simple settings, such as observing (kpsuO) or constructing (kpsuo) an unranked sample, usual diversity metrics (Entropy, Simpson's, and Berger-Parker Indices) align with individual perception, (Fig. 1, left), even if understating both extremes. While there are cases where participants constructed minimally diverse samples, these can be interpreted as training error[1]. Hence, in simple cases, typical diversity metrics properly capture perceived diversity.

Total Variation Distance (TVD) between population / sample distributions[2] across all responses in Kpsuo is skewed towards lower values ($g_1 = 2.60$, Fig. 1, right), implying an effect of population on sample distribution. Also, class participation was found to skew perceived diversity towards that class, echoing past results [3]. Namely, in kpsUo class distributions were significantly different (Kruskal-Wallis test, $H(4) = 869.06$, $p < .001$) with participants' class significantly preferred both at an aggregate (aggregate class means: 2.04, 2.19, **3.79**, 2.12, 1.85) and individual level, as in 78.08% of responses the participant's class was the most abundant. Notably, in all cases, we observed a slight under-representation of the fifth class (pairwise Wilcoxon test, $p < .04$, BH p-value adjustment), possibly attributed to most participants coming from left-to-right reading countries and sampling 12 elements from 5 classes being inherently unbalanced.

Regarding rankings, (Fig. 2a), there is an alternating pattern following the order classes were presented (kpSuo). In Figure 2b we present the percentage of responses that respected some cyclic permutation of 1-2-3-4-5. Notably, most participants that did follow a pattern, chose to

[1]Regarding samples of one class, 7 participants provided 1 or 2 while 4 of them provided more than 7 such responses, corresponding in most cases to training tasks.
[2]TVD in this discrete case coincides with half the Manhattan distance of the two distributions.



(a) Distribution of participants class choices per ranked sample position in kpSuo.

(b) Percentage of responses that adopted cyclic permutations of the presented class order in kpSuo.

Figure 2: Results regarding perceived diversity in the case of constructed rankings.

adopt the order classes appeared (left-to-right). There is a considerable drop after position 5 (i.e., the class number) hinting towards the first part of a ranking being more important when it comes to diversity in this setting. These results resemble similar assumptions of previous work in ranking diversification [26]. We examined whether participants considered ranked samples with an alternating part as more diverse than others in observation tasks, however we found no significant difference (normalized Mann-Whitney $U = 0.73$, $p = 0.23$), even if one focuses on alternating patterns that were presented to participants after they had observed a non-alternating pattern at the same iteration, ($U = 0.93$, $p = 0.18$). This imbalance between observed and constructed diversity in rankings can be interpreted by structural differences of the two tasks, since observation tasks required significantly less time compared to construction tasks ($U = 30.71$, $p < 10^{-5}$, $f = 0.84$, for kpSuo vs. kpSuO) which implies different levels of elaboration.

4. Discussion and Future Work

Even if analysis of all gathered data is yet to be completed, widely used diversity metrics have been found inadequate at capturing perceived diversity in complex settings. Thus, we expect that further analyses will unveil more sophisticated patterns of behavior. While this work focused on providing some preliminary results on how people conceptualize diversity, another related problem is how such results can help formulating an informed metric of diversity. More precisely, utilizing previous work, it would be interesting to examine up to what extent individual diversity perceptions can be captured by using existing explainable human-machine interaction protocols, like Interactive Machine Learning [27] or Machine Coaching [28]. Another limitation of this work is that we did not utilize realistic settings, e.g., workplace scenarios, a direction worth exploring, for several of the observed effects may vanish or be amplified in such cases. Similarly, considering multi-dimensional items might also affect positively or negatively any observed trends.

Privacy Notice: No data of private nature were collected and participants were required to complete a statement of informed consent prior to taking the survey.

Acknowledgments

This work was supported by funding from the EU's Horizon 2020 Research and Innovation Programme under grant agreements no. 739578 and no. 823783, and from the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation, and Digital Policy.

References

[1] E. Hobman, P. Bordia, C. Gallois, Consequences of Feeling Dissimilar from Others in a Work Team, Journal of Business and Psychology 17 (2003) 301–325. doi:10.1023/A:1022837207241.

[2] B. Meyer, M. Shemla, C. C. Schermuly, Social category salience moderates the effect of diversity faultlines on information elaboration, Small Group Research 42 (2011) 257–282. URL: https://doi.org/10.1177/1046496411398396. doi:10.1177/1046496411398396. arXiv:https://doi.org/10.1177/1046496411398396.

[3] F. Danbold, M. M. Unzueta, Drawing the diversity line: Numerical thresholds of diversity vary by group status., Journal of personality and social psychology (2019).

[4] A. Jaiswal, L. Dyaram, Diversity: a matter of reality or perception?, International Journal of Organizational Analysis 26 (2018) 00–00. doi:10.1108/IJOA-01-2018-1324.

[5] D. van Knippenberg, M. C. Schippers, Work group diversity, Annual Review of Psychology 58 (2007) 515–541. URL: https://doi.org/10.1146/annurev.psych.58.110405.085546. doi:10.1146/annurev.psych.58.110405.085546. arXiv:https://doi.org/10.1146/annurev.psych.58.110405.085546, pMID: 16903805.

[6] M. Shemla, B. Meyer, L. Greer, K. A. Jehn, A review of perceived diversity in teams: Does how members perceive their team's composition affect team processes and outcomes?, Journal of Organizational Behavior 37 (2014). doi:10.1002/job.1957.

[7] H. Tuomisto, A Consistent Terminology for Quantifying Species Diversity? Yes, It Does Exist, Oecologia 164 (2010) 853–860. URL: https://doi.org/10.1007/s00442-010-1812-0. doi:10.1007/s00442-010-1812-0.

[8] J. Qin, N. Muenjohn, P. Chhetri, A Review of Diversity Conceptualizations: Variety, Trends, and a Framework, Human Resource Development Review 13 (2014) 133–157. doi:10.1177/1534484313492329.

[9] J. Pfeffer, Organizational demography: Implications for management, California Management Review 28 (1985) 67–81. URL: https://doi.org/10.2307/41165170. doi:10.2307/41165170. arXiv:https://doi.org/10.2307/41165170.

[10] J. Pfeffer, Organizational Demography., Research in organizational behavior (1983).

[11] C. B. Goldberg, Relational Demography and Similarity-Attraction in Interview Assessments and Subsequent Offer Decisions: Are we Missing Something?, Group & Organization

Management 30 (2005) 597–624. URL: https://doi.org/10.1177/1059601104267661. doi:10.1177/1059601104267661. arXiv:https://doi.org/10.1177/1059601104267661.
[12] S. M. Thatcher, K. A. Jehn, E. Zanutto, Cracks in Diversity Research: The Effects of Diversity Faultlines on Conflict and Performance, Group Decision and Negotiation 12 (2003) 217–241. URL: https://ideas.repec.org/a/spr/grdene/v12y2003i3d10.1023_a1023325406946.html. doi:10.1023/A:1023325406946.
[13] M. A. Maltarich, S. M. B. Thatcher, D. J. Schepker, J. Park, Perceived Faultlines in Group Dynamics: An Individual-Level Perspective, Small Group Research 52 (2021) 565–599. URL: https://doi.org/10.1177/1046496420986870. doi:10.1177/1046496420986870. arXiv:https://doi.org/10.1177/1046496420986870.
[14] K. A. Jehn, K. Bezrukova, The faultline activation process and the effects of activated faultlines on coalition formation, conflict, and group outcomes, Organizational Behavior and Human Decision Processes 112 (2010) 24–42. URL: https://www.sciencedirect.com/science/article/pii/S0749597809001095. doi:https://doi.org/10.1016/j.obhdp.2009.11.008.
[15] S. M. B. Thatcher, P. C. Patel, Group Faultlines: A Review, Integration, and Guide to Future Research, Journal of Management 38 (2012) 969–1009. URL: https://doi.org/10.1177/0149206311426187. doi:10.1177/0149206311426187. arXiv:https://doi.org/10.1177/0149206311426187.
[16] S. Shrivastava, J. Gregory, Exploring the antecedents of perceived diversity, Journal of Management & Organization 15 (2009) 526–542. doi:10.5172/jmo.15.4.526.
[17] R. Colwell, Biodiversity: Concepts, Patterns, and Measurement, Princeton Univ. Press, 2009, pp. 257–263. doi:10.1515/9781400833023.257.
[18] C. E. Shannon, A Mathematical Theory of Communication, The Bell System Technical Journal 27 (1948) 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.
[19] E. H. Simpson, Measurement of Diversity, Nature 163 (1949) 688–688. URL: https://doi.org/10.1038/163688a0. doi:10.1038/163688a0.
[20] W. H. Berger, F. L. Parker, Diversity of Planktonic Foraminifera in Deep-Sea Sediments, Science 168 (1970) 1345–1347. URL: https://www.science.org/doi/abs/10.1126/science.168.3937.1345. doi:10.1126/science.168.3937.1345. arXiv:https://www.science.org/doi/pdf/10.1126/science.168.3937.1345.
[21] M. Hall, N. Tideman, Measures of concentration, Journal of the American Statistical Association 62 (1967) 162–168.
[22] T. V.;Gümüs, The Diversity Perception and The Attitudes of Employees: A Study on Human Resource Professionals and Hotel Workers, Ankara Üniversitesi SBF Dergisi 65 (2010) 219–237. doi:10.1501/SBFder_0000002161.
[23] L. T. Phillips, M. L. Slepian, B. L. Hughes, Perceiving groups: The people perception of diversity and hierarchy., Journal of Personality and Social Psychology 114 (2018) 766–785. URL: https://doi.org/10.1037/pspi0000120. doi:10.1037/pspi0000120.
[24] R. Balter, J. Chow, Y. Jin, What diversity metrics are best used to track and improve employee diversity? [executive summary], 2014.
[25] D. G. Mcdonald, J. Dimmick, The conceptualization and measurement of diversity, Communication Research 30 (2003) 60–79. URL: https://doi.org/10.1177/0093650202239026. doi:10.1177/0093650202239026. arXiv:https://doi.org/10.1177/0093650202239026.
[26] V. Markos, M. Loizos, Post-hoc Diversity-aware Curation of Rankings, in: Proceedings of

the 14th International Conference on Agents and Artificial Intelligence (ICAART 2022), volume 2, 2022, pp. 323–334. doi:10.5220/0010839600003116.
[27] J. A. Fails, D. R. Olsen, Interactive machine learning, in: Proceedings of the 8th International Conference on Intelligent User Interfaces, IUI '03, Association for Computing Machinery, New York, NY, USA, 2003, p. 39–45. URL: https://doi.org/10.1145/604045.604056. doi:10.1145/604045.604056.
[28] L. Michael, Machine Coaching, in: Proceedings of the IJCAI 2019 Workshop on Explainable Artificial Intelligence (XAI @ IJCAI 2019), 2019, pp. 80–86.

A. Materials

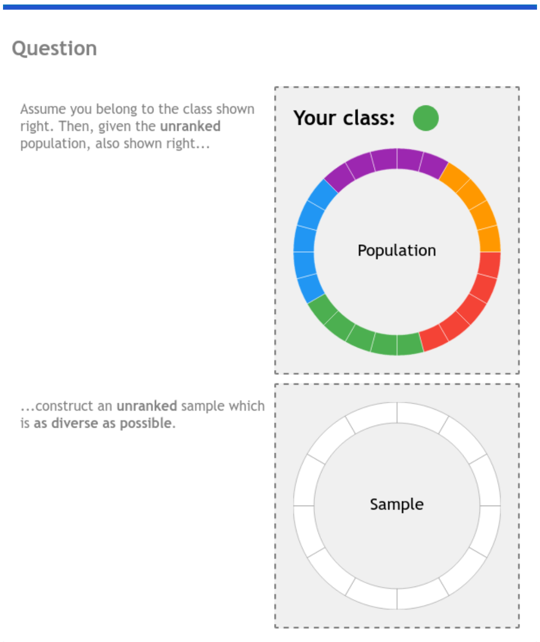
In Figure 3 we showcase indicative materials used for some examined conditions that demonstrate how all five examined variables were mapped to visual and/or textual items. Regarding observation tasks, participants were required to drag the slider at least once in order to proceed with the next task, in order to ensure user involvement.

B. Population Awareness

As we discussed in Section 3, there seems to be a significant correlation between the underlying population and sample distributions, as illustrated in Figure 4a (right). Similar plots for all four conditions where the underlying unranked population was known in construction tasks are shown in Figure 4. As one may observe, there seems to be a similar effect in all five cases in total, suggesting that participants, in general, took the distribution of the underlying population into account when constructing their samples.

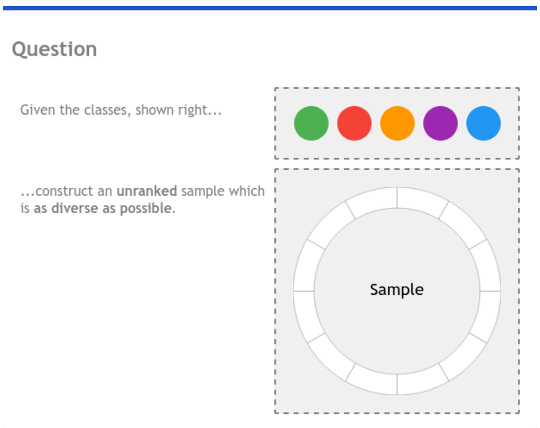
C. Order Effects

As discussed in Section 2, we examined a total of 104 ordered pairs of conditions. We chose to manipulate order between conditions to study any effect it might have on perceived diversity. For instance, it might be the case that how participants judge sample diversity in kpsuO is affected by whether they have first seen items of KpsuO, i.e., whether they are aware that the underlying population might play a role in their judgment, even if they are not informed about it. In Figure 5 we present the distributions of all observation task responses with respect to order of appearance. Single letters indicate which variable was manipulated in each pair while highlighted plots correspond to cases where a significant order effect appears (normalized Mann-Whitney $U > 2.09$, $p < .05$ in all highlighted cases). Since there seem to be cases where the order of appearance plays a significant role in observation tasks, there is need for further analysis.

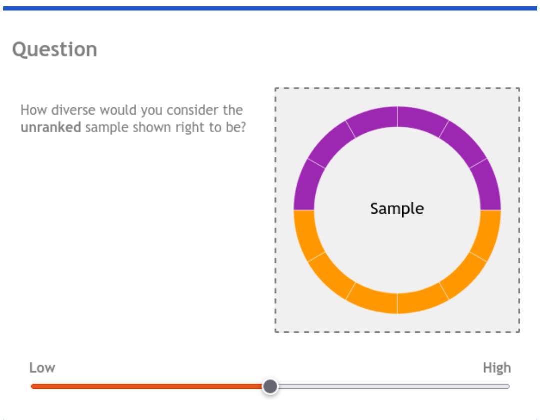


(a) Sample material for KpsUo.

(b) Sample material for KpSuO.

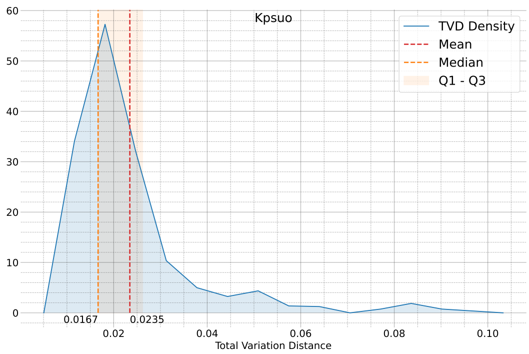


(c) Sample material for kpsuo.

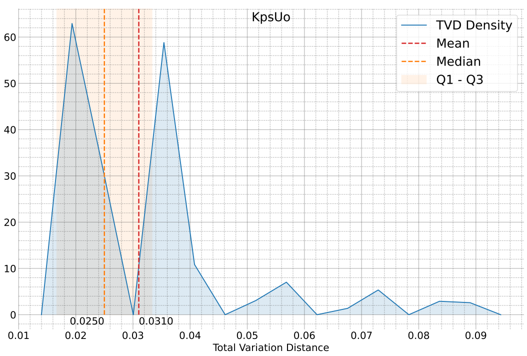


(d) Sample material for kpsuO.

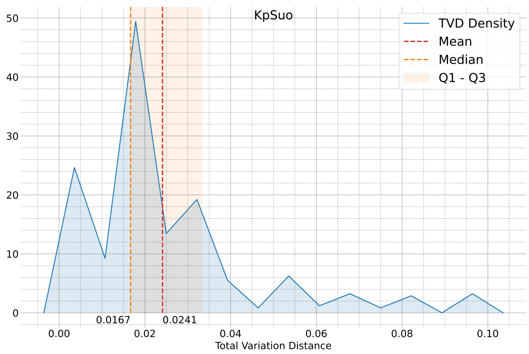
Figure 3: Materials for four conditions used throughout this study.



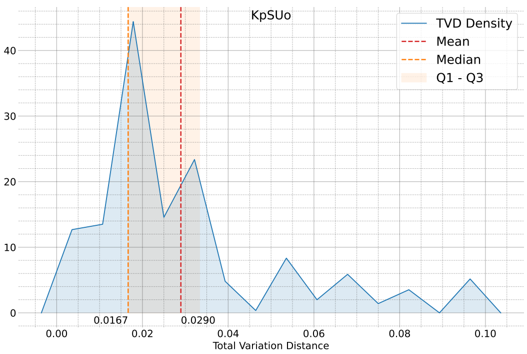
(a) Kpsuo, $g_1 = 2.60$.



(b) KpsUo, $g_1 = 1.75$



(c) KpSuo, $g_1 = 1.54$.



(d) KpSUo, $g_1 = 1.36$.

Figure 4: TVD distributions for construction tasks where the underlying unranked population was known. In all cases there appears to be a significant correlation between the sample and population distributions, judging by skewness ($g_1 > 1.0$ in all four cases).

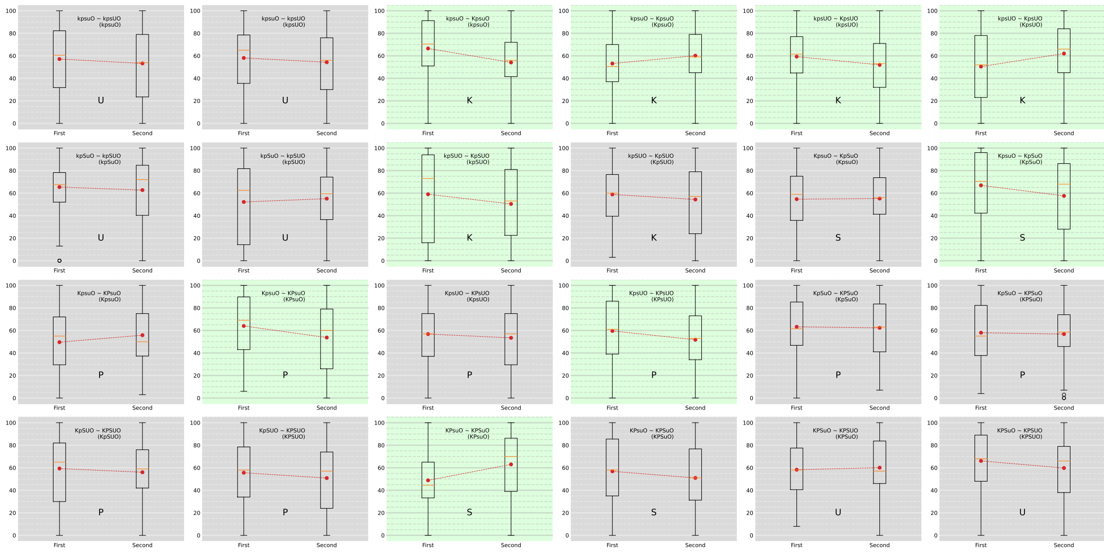


Figure 5: Response distributions in observation tasks per pair with respect to order of appearance. Single letters indicate the switching variable, while highlighted plots correspond to statistically significant differences in response distributions ($p < .05$).