9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Tasks Beyond Taxonomies: A Multidimensional Embedding Space of Team Tasks

Keywords: teams, groups, tasks, experimental design, generalizability

Extended Abstract

Teamwork is an essential component of success in many areas, but the specific skills and approaches that lead to a team's success vary greatly depending on the task at hand. A star soccer team does not have the same skills, approaches, and interaction patterns as a star quiz bowl team. Consequently, building generalizable theories about teams requires accounting for task characteristics as a moderator. Differences in tasks can reverse a team's outcomes [1] or attenuate the effects of an intervention [3]. In fact, task characteristics can explain as much as 60% of the variance in team behaviors [7]. Understanding team success and failure therefore requires an equally clear understanding of tasks [5].

Current methods to describe task characteristics generally use conceptual dimensions and categories (for example, "creativity tasks" [6] and "maximizing tasks" [8]), but different task frameworks often focus on different attributes. As a result, the field encounters an apples-to-oranges problem: when different frameworks give contradictory answers about how similar two tasks are, it is difficult to reconcile their insights and know what types of team behaviors to expect for a given task. Consider, for example, "Anagrams," the task of solving as many anagrams as possible; and "Brainstorming," the task of coming up with as many ideas as possible for improving a library. According to Steiner [8], these tasks are similar, since both involve doing an action as much as possible ("maximizing tasks"). But according to McGrath [6], the two tasks are quite different, as Anagrams is an "intellective task" (Type 3), while Brainstorming is a "creativity task" (Type 2). Thus, researchers might rightly wonder: will a finding from a study conducted using Anagrams generalize to a study using Brainstorming? What exactly qualifies two tasks as being sufficiently "similar" or "different?" And how many tasks should one test before being confident that an effect will generalize?

In this research, we integrate multiple taxonomies and typologies of group tasks into a single theoretical framework — creating a similarity score for tasks. Our framework is represented as a matrix, in which each column defines a task dimension and each row defines a task (Figure 1). We sourced 23 characteristics from five popular frameworks (McGrath 1984, Shaw 1963, Steiner 1966, Zigurs et al. 1999, and Laughlin and Ellis 1986), and we asked a group of trained raters to label 102 tasks along these dimensions. Each cell represents the mean binary rating given by 20+ raters for a given task and dimension, which we interpret as "the extent to which a given dimension applies to a task." For example, if "Writing Story" is not a maximizing task, the mean value along the "Maximizing" dimension should be close to 0; for the converse, the value should be close to 1. We then demonstrate, first, that this 102 x 23 matrix meaningfully encodes task differences, and second, that it can be useful for selecting tasks for studies, conducting adaptive experiments, and identifying boundary conditions for theories.

We call this multidimensional approach the "Task Space" on the grounds that every task is describable in our framework by some choice of parameters that maps to a unique point in the space. The Task Space answers the call to shift from categorical representations to a multidimensional method of measuring team constructs [5, 2, 4]. As with the example set by the

Three-Dimensional Team Scaling Model [4], our system creates "coordinates" for team tasks — a map that can help researchers better navigate their experimental decisions.

The Task Space is an important step towards building context-sensitive theories [5] and conducting *integrative experiments* [2]. One might wish to select tasks that are as different as possible in order to show that a theory has broad applicability, or to focus only on tasks containing a certain characteristic of interest. By identifying a task's coordinates in the Task Space — either by using our pre-labeled tasks or by applying a survey metric we provide — researchers can optimize their experimental designs based on their criteria of interest.

For example, suppose that someone is interested in whether a particular theory holds across the tasks "Advertisement Writing," "Writing Story," and "Typing Game." In Figure 1, we can see that "Advertisement Writing" and "Writing Story" (Far Left, Red) are similar, while "Typing Game" is different (Far Right, Green). If the theory holds only on the two red tasks and not on the green task, the researcher could hypothesize a boundary condition for their theory as a hyperplane separating the two groups; they could then test this hypothesis by studying tasks on either side of the hyperplane. Conversely, if the theory holds for all three tasks, the researcher may want to stress test the theory's generalizability by choosing a task that is as different as possible from the existing experiments.

Over time, as researchers test their theories using the Task Space, they may discover new tasks or task dimensions that are missing from the current representation. For example, if two tasks lie in close proximity in the existing task space and yet show qualitatively different team behaviors, the researcher may hypothesize that one or more differentiating dimensions are missing. Fortunately, the matrix format of the Task Space makes it easy to append a new row (for tasks) or a new column (for dimensions). We aim for the Task Space to evolve alongside our cumulative knowledge, serving as a versatile tool from designing one experiment to finding inspiration for the next.

References

[1] Abdullah Almaatouq, Mohammed Alsobay, Ming Yin, and Duncan J Watts. Task Complexity Moderates Group Synergy. *PNAS*, 118(36):e2101062118, 2021.

[2] Abdullah Almaatouq, Thomas L Griffiths, Jordan W Suchow, Mark E Whiting, James Evans, and Duncan J Watts. Beyond Playing 20 Questions with Nature: Integrative Experiment Design in the Social and Behavioral Sciences. *BBS*, pages 1–55, 2022.

[3] David M Herold. Improving the Performance Effectiveness of Groups Through a Task-Contingent Selection of Intervention Strategies. *AMR*, 3(2):315–325, 1978.

[4] John R Hollenbeck, Bianca Beersma, and Maartje E Schouten. Beyond Team Types and Taxonomies: A Dimensional Scaling Conceptualization for Team Description. *AMR*, 37(1):82–106, 2012.

[5] James R Larson Jr. *In Search of Synergy in Small Group Performance*. Psych. Press, 2010.

[6] Joseph Edward McGrath. *Groups: Interaction and Performance*, volume 14. Prentice-Hall Englewood Cliffs, NJ, 1984.

[7] Charles G Morris. Task Effects on Group Interaction. *JPSP*, 4(5):545, 1966.

[8] Ivan D Steiner. Models for Inferring Relationships Between Group Size and Potential Group Productivity. *Behavioral Science*, 11(4):273–283, 1966.

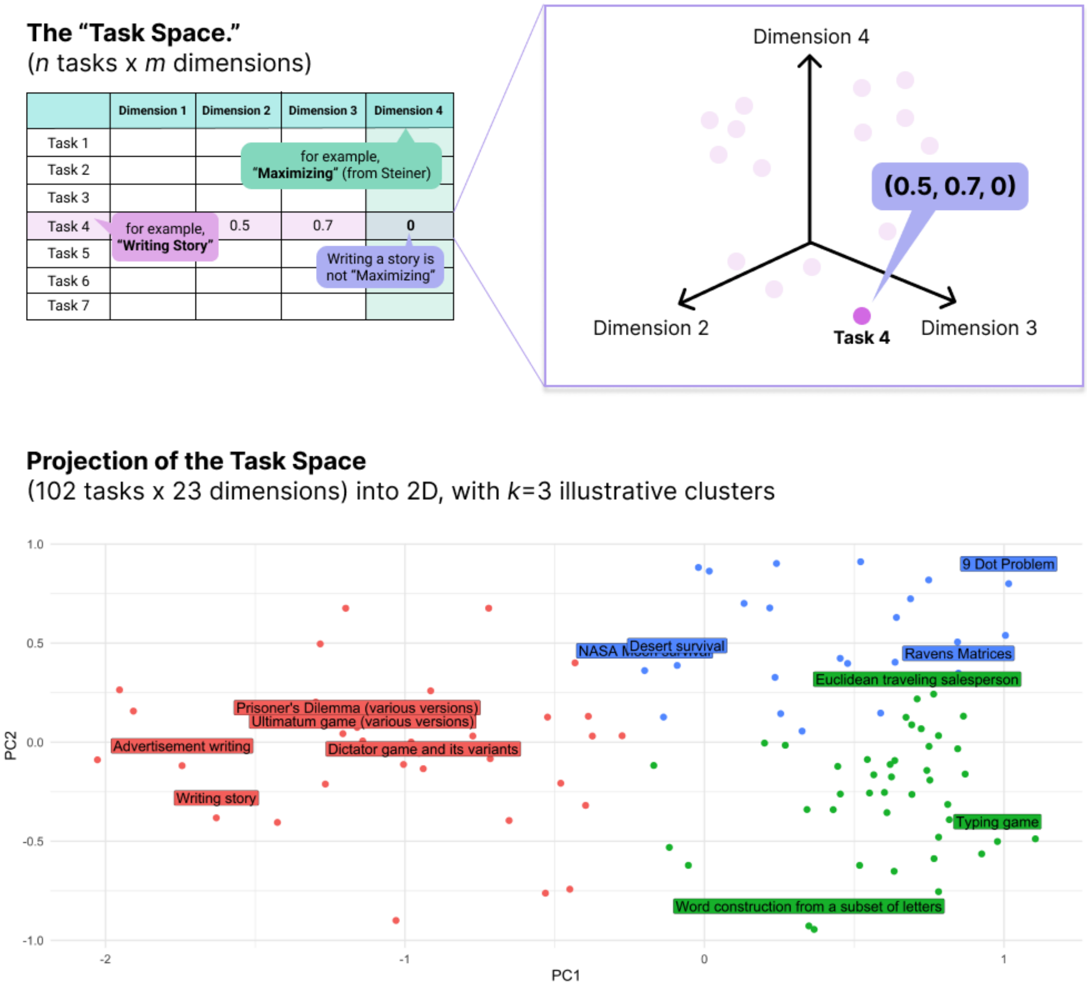

Figure 1: **(Top)** An illustration of the Task Space. The space can be represented as a matrix (Top Left), in which each row represents a task, and each column a task dimension. Each row, or task vector, can then be conceptualized as a point in a multidimensional space (Top Right).
(Bottom) A two-dimensional projection of empirical ratings for 102 tasks, sourced from a cross-section of social science domains, including the Collective Intelligence Task Battery (Woolley et al. 2010), collections of forecasting and prediction tasks (Silver, Mellers, and Tetlock 2021), and collections of "classic" tasks used in group studies (Lorge and Solomon 1960). Here, the x-axis is the first Principal Component, and the y-axis is the second Principal Component. Colors represent clusters generated using k-means, and are intended merely to be illustrative of ways in which tasks with similar dimensions can be grouped together. The true "clustering" of tasks will be determined empirically — based on whether teams performing certain tasks exhibit the same outcomes. We highlight the fact that the map captures semantically meaningful differences between tasks: for example, "Advertisement Writing" and "Writing story" are next to each other (Far Left, Red); "Prisoner's dilemma," "Ultimatum game," and "Dictator game" are in a similar region (Middle left, Red), and "NASA Moon Survival" and "Desert surval," which are two variants of the same game, are positioned so close together as to be overlapping (Middle Right, Blue). Meanwhile, "Word construction from a subset of letters" is positioned far from "9 Dot Problem," and the "Typing game" is more similar to "Word construction" than it is to "9 Dot Problem." These observations qualitatively demonstrate that the Task Space sensibly encodes task characteristics.