CoBlok: Collaborative Performance in Virtual Reality and Face-to-Face

Valtteri Wikström
valtteri.wikstrom@helsinki.fi
Cognitive Brain Research Unit,
University of Helsinki
Helsinki, Finland

Silja Martikainen
Cognitive Brain Research Unit,
University of Helsinki
Helsinki, Finland

Mari Falcon
Cognitive Brain Research Unit,
University of Helsinki
Helsinki, Finland

Niina Seittenranta
Cognitive Brain Research Unit,
University of Helsinki
Helsinki, Finland

Pyry Heikkinen
Cognitive Brain Research Unit,
University of Helsinki
Helsinki, Finland

Katri Saarikivi
Cognitive Brain Research Unit,
University of Helsinki
Helsinki, Finland

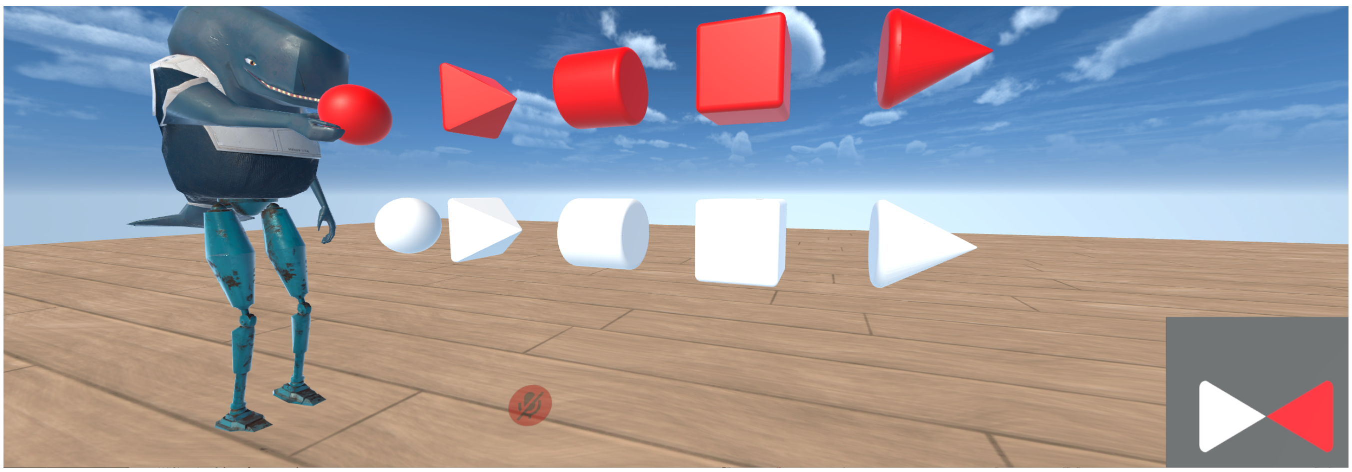


Figure 1: A screenshot of the CoBlok task in VR. In the center are the ten blocks used in the task, and on the bottom right is a puzzle card.

ABSTRACT

Computer-mediated communication is being adopted in work and personal life around the world. Measuring collaborative performance would be useful for evaluating and optimizing social computing applications, but there is a lack of methods for it. For this purpose, we have developed a collaborative block design task, which requires collaboration and depends on simple and abstract rules. The task is about constructing three-dimensional structures out of primitive shapes based on cards representing two-dimensional flat projections of the complete structure. The task is presented as a physical version, which can be manufactured using a 3D printer and a laser cutter, as well as a virtual version which is released as an open source VRChat world created in Unity, and as task components which can be imported into any virtual 3D environment. The task can be used for evaluating systems and augmentations for their fitness for collaboration, as well as to investigate other phenomena which seem to be linked with better cooperation, such as inter-individual and inter-brain synchronization.

CHI '22 Extended Abstracts, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9156-6/22/04.
https://doi.org/10.1145/3491101.3519883

CCS CONCEPTS

• **Human-centered computing → Collaborative and social computing design and evaluation methods**; *Empirical studies in collaborative and social computing*; **Virtual reality**; Mixed / augmented reality.

KEYWORDS

collaboration, cooperation, social computing, computer-mediated communication, metaverse

ACM Reference Format:
Valtteri Wikström, Silja Martikainen, Mari Falcon, Niina Seittenranta, Pyry Heikkinen, and Katri Saarikivi. 2022. CoBlok: Collaborative Performance in Virtual Reality and Face-to-Face. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '22 Extended Abstracts), April 29-May 5, 2022, New Orleans, LA, USA*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3491101.3519883

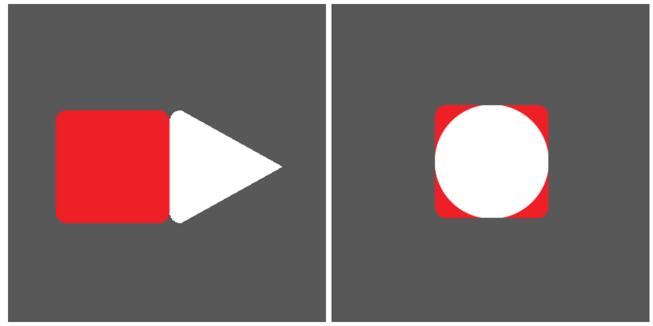
1 INTRODUCTION

Social computing technology is evolving rapidly. Increasing computing power and networking speeds, new materials, along with advanced sensing devices, audiovisual and haptic displays makes it possible to imagine a vast amount of possibilities for its design and development. The possibilities seem almost endless, as we do not need to limit ourselves to replicating or imitating face to face when creating virtual and augmented social worlds, as even physical laws such as gravity can be changed, replaced with others, or omitted altogether, things previously perceivable can be hidden and otherwise hidden information can be made perceivable. The enormity of this creative space seems formidable, and evaluating different applications is hard. One way to approach evaluation of social technology, especially for productivity applications, is to quantify the outputs of collaboration as a measure of collaborative success. Here, we present the latest version of a task created to evaluate collaborative performance in virtual and augmented environments.

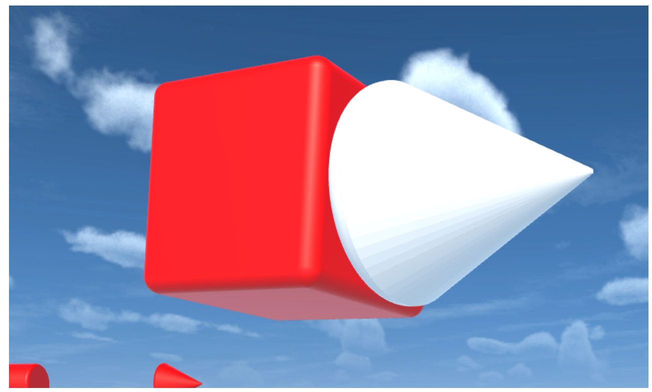
CoBlok is a collaborative block design task for assessing pair performance, which can be used in 3D environments as well as with physical blocks in face to face reality [20]. The task is designed to be symmetrical in regards to participant roles, to be based on simple and abstract rules making it quick to learn and easy to translate, and to require collaboration and communication to finish. The task appears to be engaging for participants in a similar way as the original block design test [13] which is widely used for measuring individual visuospatial reasoning abilities, and performance in the CoBlok task can be controlled by measuring participants individual visuospatial skill [20]. On the McGrath Group Task Circumplex [14], it represents a task of Type 3 (task with a correct answer), with aspects of Type 5 (cognitive conflict) and Type 8 (performance) [20]. Compared to many previous tasks used for collaborative performance measurement (such as in [6, 8, 18]), this task cannot be completed alone or without input from all participants, and compared to hidden profile tasks (such as in [2, 5, 12]), which were originally made to study group decision making bias, it does not rely on participants studying large amounts of verbal information, and it is specifically designed for quantifying performance in the form of puzzle completion times [20].

1.1 Task rules

The CoBlok task consists of puzzles, in which two participants receive cards, which depict different points of view of the same structure. The participants need to reconcile the different points of view to complete the structure. The structure is created by using a subset of 10 different blocks of the same dimensions, consisting of cube, sphere, cylinder, pyramid, and cone shapes in two different colours, red and white (see Figure 1 for the blocks). These blocks have been chosen, so that each shape resembles at least one other shape, when made into a flat projection from an orthogonal angle. Flat projection means in this case that perspective and depth information is omitted, depicting the shapes based on their outlines and colour. This makes it possible to create structures, which can not be unambiguously identified from one puzzle card depicting a flat projection.



(a) Puzzle card A **(b) Puzzle card B**



(c) 3D structure

Figure 2: Puzzle 2.2

Each puzzle consists of two cards and each participant receives one of the two cards, which they can look at, but they are not allowed to show their card to the other participant. Figure 2 shows two puzzle cards next to the 3D structure corresponding to them. Instead, the participants have to communicate with each other and build a structure which is consistent with both of their puzzle cards. The participants task is to figure out, as quickly and accurately as possible, the configuration of blocks that matches both of their puzzle cards. They are allowed to communicate freely, but without any external tools, such as pen and paper. The orientation of the cards do not matter, it is up to the participants to choose from which perspective the structure is presented in their card. The blocks can be occluded both completely and partially by another block being placed in front, but no blocks are completely occluded in both puzzle cards.

2 TASK DEVELOPMENT

An observation was made with the original 14-block version of the task, that especially tasks involving the double-sized blocks: long cuboid and long cylinder, were prone to confuse participants and produce multiple possible solutions. This was undesirable, as the number of blocks needed to complete each task was otherwise correlated with task difficulty, but this relationship became inconsistent when one block could sometimes be replaced with two other blocks in a puzzle. Additionally, some participants seemed to have trouble to remember all of the seven different shapes. Because of

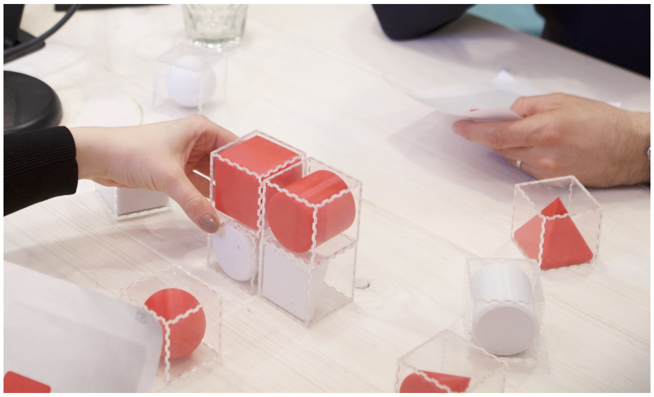


Figure 3: Gameplay with the physical version of the CoBlok task.

this, the long blocks were removed in the new version, resulting in five different shapes, in two different colours, for a total of 10 blocks (see Figure 1). To make the flat two-dimensional projections of the blocks easier to perceive, and to distinguish where one block ends and another one begins, the blocks were rounded in the corners for both the three-dimensional shapes and the two-dimensional projections.

With the modifications made to the task, the puzzles used in the original version were otherwise kept the same, but puzzles including the long blocks were replaced with similar puzzles using only the 10 blocks. Out of the ten puzzles in the original study, three were replaced for the new version, while seven puzzles remain the same. We tested the physical version of the updated task with 11 pairs for set 1 and 17 pairs for set 2 to obtain completion times with the new version. The order of the puzzles was kept consistent for all pairs, going from puzzle 1 to puzzle 5. The results, shown in Table 1, indicate that the difficulty rises gradually for both puzzle sets, with set 2 being more challenging.

2.1 Physical version

The physical version of the blocks were manufactured using a fused filament fabrication 3D printer in red and white PLA in dimensions of 50 mm x 50 mm x 50 mm. To make it possible to stack the blocks in any configuration, they were enclosed in transparent acrylic plastic boxes with the same inner dimensions. The boxes were designed with a free online tool [4] and manufactured with a laser cutter out of 2 mm thick acrylic. The puzzle cards were printed on regular printing paper with a color laser printer. Figure 3 shows gameplay with the physical version.

2.2 Virtual Reality version

We implemented the task as a World in the popular social VR application VRChat, which allows creating custom scenarios with Unity and a node-based programming language called Udon [7]. VRChat manages user avatars and voice transmission, and Udon contains functions for network synchronization, handling interactions, etc. The 10 blocks for the task were implemented as network synchronized (VRC Object Sync) and manipulable (VRC Pickup) kinematic game objects, meaning that they can be moved and rotated and they maintain the same transformations across all users, and they do not observe gravity or other forces, instead staying suspended in the air when released. A head-up display (HUD) was implemented as a local game object, which follows the users movements with a custom Udon script, being visible in the field-of-view. Buttons were added for choosing puzzle cards A or B and for choosing between ten puzzles. Custom Udon scripts were made so the buttons for puzzle cards toggle the texture of the HUD locally between the two cards belonging to the current puzzle, and the buttons for choosing puzzles send custom network events, which change the two active HUD textures based on the puzzle, and reset the positions of the blocks for all users.

The source code and components of the VRChat world are released as open source[1]. In this same repository, the files for 3D printing and laser cutting the physical version can also be found.

3 DEMONSTRATION AT CHI 2022

For virtual conference participants, we will demonstrate the version of the task created as a World in VRChat [7]. This world can be accessed either with a supported virtual reality headset, or on the desktop using mouse and keyboard. The virtual world will be available also outside of the interactivity time slot, and instructions to access the task will be provided on Discord.

At the onsite conference in New Orleans, we will present the physical version of the task for conference participants to try in a face to face setting. This version is also possible to manufacture remotely using a 3D printer for the blocks and a laser cutter for creating the acrylic enclosures, equipment found at digital fabrication laboratories, as well as many maker and hack spaces.

4 DISCUSSION

Our social abilities have evolved in "meat reality", and our understanding of the social brain can give us direction in the development of collaboration technology. The development of social cognition begins in the form of primary intersubjectivity: the intuitive, preverbal processes which makes it possible for a baby to perceive the gaze and facial, vocal and bodily expression of the caregiver. Interaction Theory and the Primary Interaction Thesis proposes that primary intersubjectivity is fundamental for understanding social interaction throughout the lifespan, that most situations do not require the drawing inferences or simulation, but instead rely on our ability to directly perceive the intentions and emotions of others from their expressions [10, 16]. Primary intersubjectivity has a strong connection to ideas about direct social perception and embodied thought from phenomenological philosophy, and offers a perhaps more compelling explanation than simulation for the existence of the mirror neuron system and inter-brain synchronization during social interaction [10]. Studies have found links between increased inter-brain and inter-individual synchronization and better cooperation, among them evidence has started to accumulate about the connection between collaborative task performance and synchronization [17, 19]. Inter-brain and inter-individual synchronization, along with different types of collaborative tasks, such

[1] https://www.github.com/coblok/coblok

Puzzle	1.1	2.1	1.2*	2.2	1.3	2.3*	1.4*	2.4	1.5*	2.5
Mean time (s)	28.4	96.9	64.6	54.5	95.5	145.9	159.3	199.8	158.2	252.0
Std (s)	8.6	69.8	24.2	36.4	50.8	116.0	97.1	122.1	168.2	136.5

Table 1: Mean completion times and standard deviations for correctly completed puzzles for 11 pairs (set 1) and 17 pairs (set 2). The puzzles which were different from the earlier version are marked with a star (*).

as the one presented here, offer ways to indirectly measure the constituents of successful cooperation.

It makes intuitive sense that computer-mediated communication and interaction is efficient in verbal, explicit communication, but more limited in non-verbal and implicit information transmission capability necessary for engaging primary intersubjectivity and direct perception, and it is deemed crucial to develop the sense of presence in distance collaboration [3]. Faster bandwidth and decreased latency, together with improved cameras and microphones certainly helps in transmitting the more delicate cues which engage the natural mechanisms for mindreading, but videotelephony can not easily overcome some of its inherent limitations regarding the lack of shared context. Sharing the same physical space makes it possible, for example, to read gaze cues, experience smells, touch, and actually use the space: to sense proximity, movement, and orientation. Over the years, a lot of hype has surrounded the idea of a shared virtual social world, the Metaverse [1]. This idea was popularized already in the 1980's by Gibson's novel *Necromancer* [11], implemented over the years for example in the form of the application VRChat which we have used in this study, and most recently brought into the spotlight as the new central strategy of the tech giant formerly known as Facebook, now Meta [15].

An important part of the envisioned near future development of virtual and augmented reality is to add sensors and to create algorithms for the sharing of gazes and facial expressions, using hands instead of controllers, and in other ways bringing the virtual space closer to the real one [15]. Meanwhile, since virtual interaction makes it possible to add information which would not be available in the real world, forward-thinking applications are utilizing biosignals for communication [9]. Biological information is in some sense inherently familiar to us, as we are able to perceive our own bodily state, and we can observe others' states, but sensors and displays makes it possible to bring this information into a new form, and to use signal processing to bring out hidden features. The forms which this information can take are virtually endless, calling for evaluation methods to be able to drive iterative development of prototypes.

Measuring of the non-verbal, phenomenal component, which by its nature escapes definitions, seems difficult, especially considering the vast space of possibilities for the development of virtual, multimodal interactive applications. We believe that controlled testing of collaborative performance, together with studying inter-brain and inter-individual synchronization, can offer ways to tap into the *je ne sais quoi* of social interaction: by testing augmentations and new iterations of prototypes we can drive their development for improving communication without needing to completely deconstruct human connection and the experience of presence.

REFERENCES

[1] Jeremy Bailenson. 2018. Bringing social back to the network. In *Experience on demand: What virtual reality is, how it works, and what it can do.* WW Norton & Company, New York, NY, USA.

[2] Aruna D. Balakrishnan, Susan R. Fussell, and Sara Kiesler. 2008. Do visualizations improve synchronous remote collaboration?. In *Proceedings of the 2008 SIGCHI Conference on Human Factors in Computing Systems (CHI '08).* ACM, New York, NY, USA, 1227–1236. https://doi.org/10.1145/1357054.1357246

[3] Merja Bauters, Jana Pejoska, Eva Durall, Katri Saarikivi, Valtteri Wikström, Mari Falcon, and Silja Martikainen. 2021. Are you there? Presence in collaborative distance work. *Human Technology* 17, 3 (Dec. 2021), 261–293. https://doi.org/10.14254/1795-6889.2021.17-3.5

[4] Rahul Bhargava. 2010. *BoxDesigner.* Northeastern University, MIT Media Lab, and Connection Lab. https://boxdesigner.connectionlab.org/ Accessed: 8 Jan 2022.

[5] Irina R. Brich, Inga M. Bause, Friedrich W. Hesse, and Ann-Katrin Wesslein. 2019. Working memory affine technological support functions improve decision performance. *Computers in Human Behavior* 92 (2019), 238–249. https://doi.org/10.1016/j.chb.2018.11.014

[6] Prerna Chikersal, Maria Tomprou, Young Ji Kim, Anita W. Woolley, and Laura Dabbish. 2017. Deep Structures of Collaboration. In *Proceedings of the 2017 ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW '17).* ACM, New York, NY, USA, 873–888. https://doi.org/10.1145/2998181.2998250

[7] VRChat documentation. 2022. *Worlds.* VRChat. https://docs.vrchat.com/docs/worlds Accessed: 8 Jan 2022.

[8] David Engel, Anita W. Woolley, Lisa X. Jing, Christopher F. Chabris, and Thomas W. Malone. 2014. Reading the Mind in the Eyes or Reading between the Lines? Theory of Mind Predicts Collective Intelligence Equally Well Online and Face-To-Face. *PLoS ONE* 9, 12 (December 2014), e115212. https://doi.org/10.1371/journal.pone.0115212

[9] Milou A. Feijt, Joyce H.D.M. Westerink, Yvonne A.W. De Kort, and Wijnand A. IJsselsteijn. 2021. Sharing biosignals: An analysis of the experiential and communication properties of interpersonal psychophysiology. *Human–Computer Interaction* 0, 0 (2021), 1–30. https://doi.org/10.1080/07370024.2021.1913164

[10] Shaun Gallagher. 2008. Direct perception in the intersubjective context. *Consciousness and Cognition* 17, 2 (2008), 535–543. https://doi.org/10.1016/j.concog.2008.03.003

[11] William Gibson. 1984. *Neuromancer.* Ace Books, New York, NY, USA.

[12] Martin D. Hassell and John L. Cotton. 2017. Some things are better left unseen: Toward more effective communication and team performance in video-mediated interactions. *Computers in Human Behavior* 73 (2017), 200–208. https://doi.org/10.1016/j.chb.2017.03.039

[13] Samuel C. Kohs. 1920. The Block-Design Tests. *Journal of Experimental Psychology* 3, 5 (1920), 357–376. https://doi.org/10.1037/h0074466

[14] J. E. McGrath. 1984. A Typology of Tasks. In *Groups: Interaction and Performance.* Prentice-Hall, Englewood Cliffs, NJ, USA, 53–66.

[15] Meta. 2021. *The Metaverse and How We'll Build It Together – Connect 2021.* Youtube. https://www.youtube.com/watch?v=Uvufun6xer8 Accessed: 8 Jan 2022.

[16] Henrike Moll, Ellyn Pueschel, Qianhui Ni, and Alexandra Little. 2021. Sharing Experiences in Infancy: From Primary Intersubjectivity to Shared Intentionality. *Frontiers in Psychology* 12 (2021), 2437. https://doi.org/10.3389/fpsyg.2021.667679

[17] Diego A. Reinero, Suzanne Dikker, and Jay J. Van Bavel. 2020. Inter-brain synchrony in teams predicts collective performance. *Social Cognitive and Affective Neuroscience* 16, 1-2 (09 2020), 43–57. https://doi.org/10.1093/scan/nsaa135

[18] Fazilat Siddiq and Ronny Scherer. 2017. Revealing the processes of students' interaction with a novel collaborative problem solving task: An in-depth analysis of think-aloud protocols. *Computers in Human Behavior* 76 (2017), 509–525. https://doi.org/10.1016/j.chb.2017.08.007

[19] Caroline Szymanski, Ana Pesquita, Allison A. Brennan, Dionysios Perdikis, James T. Enns, Timothy R. Brick, Viktor Müller, and Ulman Lindenberger. 2017. Teams on the same wavelength perform better: Inter-brain phase synchronization constitutes a neural substrate for social facilitation. *NeuroImage* 152 (2017), 425 – 436. https://doi.org/10.1016/j.neuroimage.2017.03.013

[20] Valtteri Wikström, Silja Martikainen, Mari Falcon, Juha Ruistola, and Katri Saarikivi. 2020. Collaborative block design task for assessing pair performance in virtual reality and reality. *Heliyon* 6, 9 (2020), e04823. https://doi.org/10.1016/j.heliyon.2020.e04823