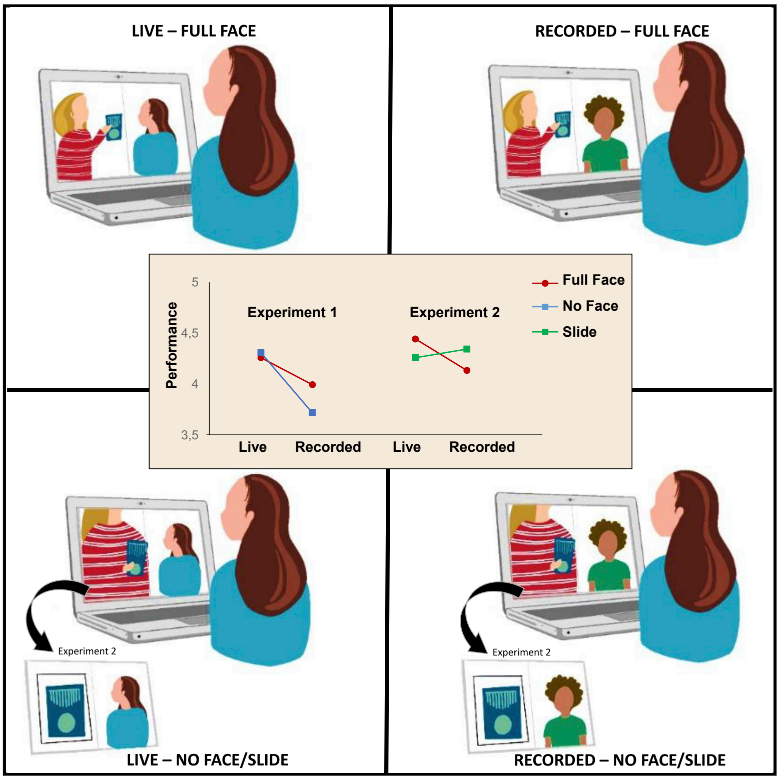
Report

Current Biology

Social interaction is a catalyst for adult human learning in online contexts

Graphical abstract



Highlights

- Social learning is characterized by a contingent student-teacher exchange
- People learn better in live, interactive video calls compared to yoked recorded videos
- In live, interactive teaching, seeing the face of the teacher improves learning
- In recorded teaching, seeing a slide is more beneficial for learning

Authors

Sara De Felice, Gabriella Vigliocco, Antonia F. de C. Hamilton

Correspondence

sara.felice.16@ucl.ac.uk

In brief

De Felice et al. find that in online sessions, live, interactive teaching improves learning compared to pre-recorded teaching, and this effect is sustained over a week. Viewing the teacher's face improves learning specifically during live interactions. These results demonstrate that rich social interaction enhances learning in an online context.

De Felice et al., 2021, Current Biology *31*, 1–7
November 8, 2021
https://doi.org/10.1016/j.cub.2021.08.045

CellPress

Current Biology

CellPress

Report

Social interaction is a catalyst for adult human learning in online contexts

Sara De Felice,[1,3,4,*] Gabriella Vigliocco,[2] and Antonia F. de C. Hamilton[1]
[1]Institute of Cognitive Neuroscience, University College London, Queen Square, London WC1N 3AZ, UK
[2]Department of Experimental Psychology, University College London, Bedford Way, London WC1H 0AP, UK
[3]Twitter: @Sara_De_Felice
[4]Lead contact
*Correspondence: sara.felice.16@ucl.ac.uk
https://doi.org/10.1016/j.cub.2021.08.045

SUMMARY

Human learning is highly social.[1–3] Advances in technology have increasingly moved learning online, and the recent coronavirus disease 2019 (COVID-19) pandemic has accelerated this trend. Online learning can vary in terms of how "socially" the material is presented (e.g., live or recorded), but there are limited data on which is most effective, with the majority of studies conducted on children[4–8] and inconclusive results on adults.[9,10] Here, we examine how young adults (aged 18–35) learn information about unknown objects, systematically varying the social contingency (live versus recorded lecture) and social richness (viewing the teacher's face, hands, or slides) of the learning episodes. Recall was tested immediately and after 1 week. Experiment 1 (n = 24) showed better learning for live presentation and a full view of the teacher (hands and face). Experiment 2 (n = 27; pre-registered) replicated the live-presentation advantage. Both experiments showed an interaction between social contingency and social richness: the presence of social cues affected learning differently depending on whether teaching was interactive or not. Live social interaction with a full view of the teacher's face provided the optimal setting for learning new factual information. However, during observational learning, social cues may be more cognitively demanding[11] and/or distracting,[12–14] resulting in less learning from rich social information if there is no interactivity. We suggest that being part of a genuine social interaction catalyzes learning, possibly via mechanisms of joint attention,[15] common ground,[16] or (inter-) active discussion, and as such, interactive learning benefits from rich social settings.[17,18]

RESULTS

Learning new information is critical to human survival and often occurs in social contexts. However, the majority of research on learning examines either asocial learning (the student is alone) or observational learning (the student watches another individual but does not interact). Learning as part of a live social interaction has been shown to be particularly valuable in infants[3] but has rarely been systematically studied in adults. Here, we examine adult social learning—and in particular the process of acquiring novel information and factual knowledge—to gain a better understanding of how social factors impact on learning and what cognitive processes may support this, in order to advance both education and research across psychology and neuroscience.

Social learning refers to any learning happening between two or more individuals. Observational learning[19] involves acquisition of information through passive exposure to the material (e.g., learning from a pre-recorded video). In contrast, interaction-based learning[1] requires mutual feedback between student and teacher (e.g., learning in live conversations).[20] In observational learning, we learn *from* others, while in interaction-based learning we learn *with* others. These forms of social learning mainly differ on the basis of social contingency, that is, the bi-directional exchange during an interaction between two or more people, where each person can initiate an action and/or directly react to their partner (mutual feedback). Contingent interactions are cognitively demanding[11] and could impact on learning in different ways. Interaction might impair learning by increasing cognitive load and/or fear of being evaluated poorly by the interlocutor.[21] Alternatively, socially contingent teaching might boost learning, as seen in children,[3] but not always in adults.[9,10]

A second important factor in social learning is social richness, that is, the type (and quantity) of social information available from one's partner. Information could be presented in a variety of formats, including by video,[22] multimedia characters,[23] recorded slides,[24] or podcasts.[25] Previous studies have not systematically examined social richness as a contributing factor in learning. As with social contingency, the relationship between social richness and learning could go in either direction. Rich social features could increase cognitive load[11,26] and/or distract learners.[17] Alternatively, social cues, such as eye-gaze[18] and gestures of a teacher,[27] could benefit learning by facilitating the coordination and "attunement" between student and teacher[28] via mechanisms of joint attention and social engagement.[29–32]

Here, we report a direct—and, to our knowledge, the first—investigation of different (online) social learning contexts in adults. We present two experiments conducted during the coronavirus disease 2019 (COVID-19) pandemic, where online

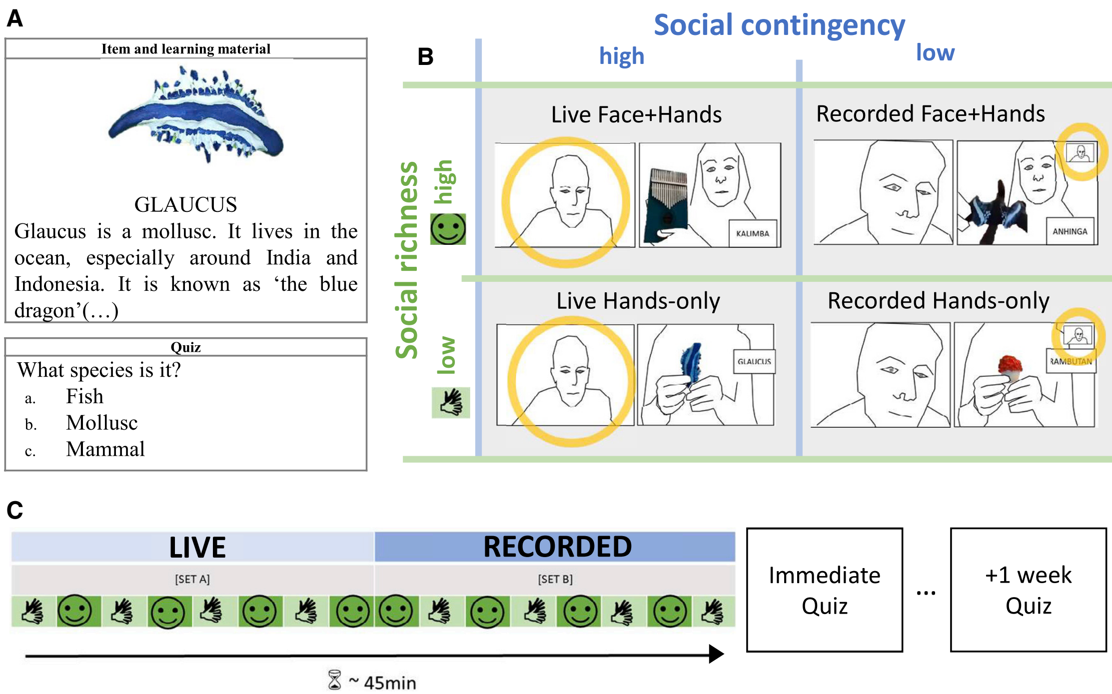


Figure 1. Experimental paradigm

(A) Sample item. Participants learned facts about unusual items in a 2-min structured conversation.

(B) Design for experiment 1. In a 2 × 2 factorial design, participants were taught about items with high or low social contingency and high or low social richness. In each screenshot, the participant is circled in yellow. In the recorded sessions, participant learned from a recorded video of a previous participant, so that recorded session of participant 2 was the live session of participant 1 (the recorded session of participant 3 was the live session of participant 2 and so on). Experiment 2 used the same design with the "hands only" conditions replaced with "slides" showing only the item.

(C) Experiment timeline. In each 45-min session, participants learned about 4 items in each of the 4 conditions. The order of conditions, sets, and trials were counterbalanced. Learning was tested with 80 computer quiz questions (5 per item) administered immediately after the learning session and again 1 week later. Full information and question sets for all items are available online at https://files.osf.io/v1/resources/rm2zp/providers/osfstorage/60fe9c0e317620013237db38?direct&version=1.

learning has become widespread. Our aim is to better understand what key components of social interaction support adult human learning in an online context and whether these play a cumulative beneficial effect when employed together. Both experiments use a 2 × 2 factorial design, where participants learn novel information over a video call in four teaching formats, differing on the basis of social contingency (live versus recorded) and social richness (more or less visual social cues; Figure 1). Verbal information about the object of learning was matched across all conditions, and recorded conditions were yoked to the live conditions, allowing us to focus on how live interaction and visual cues impact on learning. Learning performance—as measured via a multiple-choice quiz—was assessed immediately after teaching and 1 week later.

Experiment 1 (n = 24 participants) investigated the difference in learning performance between interactive learning and observational learning (social contingency factor), with either full-face (and hands) view of the teacher or a limited view of the hands only (social richness factor). Figure 2A illustrates the main effects. There was a main effect of time: not surprisingly, participants recalled more things straight after they learned them compared to a week later, independently of the learning conditions ($F_{(1, 23)} = 25.81$; $p < 0.0001$; $\eta^2 = 0.53$). There was also a main effect of social contingency ($F_{(1, 23)} = 33.34$; $p < 0.0001$; $\eta^2 = 0.59$): participants remembered more things learned during live teaching (compared to pre-recorded videos), irrespective of when they were tested and of whether the teacher's face was visible during teaching. There was no main effect of social richness ($F_{(1, 23)} = 1.28$; $p = 0.27$; $\eta^2 = 0.05$). However, we found an interaction effect between social contingency and social richness ($F_{(1, 23)} = 6.28$; $p = 0.017$; $\eta^2 = 0.22$; Figure 2B). To interpret this interaction, given that the same pattern of results has been observed at both times, we collapsed across the factor time and considered the social contingency and social richness factors (Table S2). While there was no difference in the live condition, in the recorded condition, recall was significantly better for material learned when the teacher's face was fully visible compared to when only the hands were presented ($t_{(23)} = 2.15$; $p = 0.04$). In addition, both post hoc comparisons for the social contingency factor (live face versus recorded face [$t_{(23)} = 2.99$; $p = 0.007$] and live hands versus recorded hands [$t_{(23)} = 5.61$; $p = 0.001$]) showed that performance in the live conditions was significantly higher. These results suggest that being engaged in a socially contingent interaction boosts learning and that socially rich cues may also be relevant. Similar results were found using a multi-level logistic regression model (the code and output of the model are available at https://osf.io/rm2zp/?view_only=6f5b20c6392e4bf48bf69681b2e42f34).

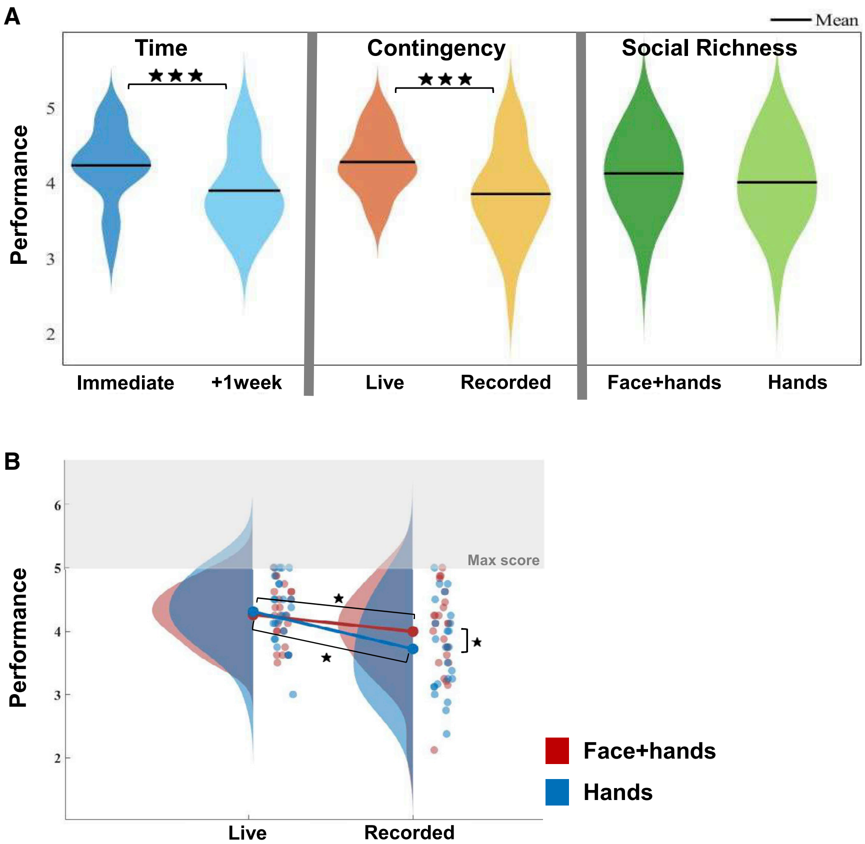


Figure 2. Results for experiment 1

(A) Main effects of time, contingency, and view on learning performance. Significant effects of time and contingency were found; ***p < 0.0001.

(B) Interaction effects. There was an interaction effect between contingency and richness; $F_{(1, 23)} = 6.28$; $p = 0.017$; $\eta 2 = 0.22$. Significant difference as measured by paired t test *$p < 0.05$ is shown. Here, performance has been averaged across time (immediate and +1 week test). The violin plots show the probability density function (Kernel density estimation), which can go beyond the smallest and largest data points. For clarity, we have indicated the max possible score on performance (x axis). Dots are showing individual scores.

See also Tables S1 and S2.

Experiment 2 (n = 27 participants) was a pre-registered extension of experiment 1. Here, we repeated the conditions with a full view of the teacher's hands and face (both live and recorded), but instead of the hands-only view, we included a condition where information was presented in slides to provide a stronger distinction in social richness. We found a main effect of time ($F_{(1, 26)} = 30.68$; $p < 0.0001$; $\eta^2 = 0.54$; Figure 3A). However, we did not find a main effect of social contingency ($F_{(1, 26)} = 1.67$; $p = 0.21$; $\eta^2 = 0.06$) or of social richness ($F_{(1, 26)} = 0.04$; $p = 0.84$; $\eta^2 = 0.002$). Importantly, we replicated the interaction effect between social contingency and social richness ($F_{(1, 26)} = 5.28$; $p = 0.03$; $\eta^2 = 0.16$; Figure 3B).

To interpret this interaction, given that the same pattern of results has been observed at both times, we collapsed across the factor time and considered the social contingency and social richness factors (Table S2). In the face condition, results from experiment 2 replicated those of experiment 1: when the teacher's face was visible, learning from a live, interactive session was more effective than learning via a recorded video ($t_{(26)} = 2.45$; $p = 0.02$). Additionally, in the live condition, exposure to face might lead to more learning than slides ($t_{(26)} = 1.77$; $p = 0.09$), while the opposite was observed in the recorded condition ($t_{(26)} = -1.87$; $p = 0.07$; see Table S2 for details). In other words, seeing the teacher's face seems to be advantageous, specifically when learning was interactive, while during observational learning, a slide presentation seems more beneficial. Similar results were found using a multi-level logistic regression model (the code and output of the model are available at https://osf.io/rm2zp/?view_only=6f5b20c6392e4bf48bf69681b2e42f34).

DISCUSSION

Understanding how learning is affected by social interaction is important for education and training in many contexts. This has become even more important during the COVID-19 pandemic, where social contact has been constrained across all domains of our lives. We investigated which social factors modulate how adults learn new concepts online. In two experiments, we manipulated social contingency (whether teaching happens through a live interaction or via a recorded video) and social richness (the extent to which the teaching context is rich in social cues, e.g., seeing the teacher's face or just a slide) and measured learning immediately after the teaching session and a week later.

Findings from both experiments point to two main conclusions: first, in the case of full-face view, interaction-based learning is more effective than observational learning. Both our studies showed that, when the teacher's face was fully visible, playing an active role in the interaction improves learning over yoked observation of the same sessions. Second, visual social cues impact learning differently depending on whether learning is interactive or observational (Figure 4): both studies showed a strong interaction effect between social contingency and social richness. To our knowledge, this is the first study showing that rich social cues specifically improve interactive, but not observational, learning.

We discuss first the impact of social contingency on learning from sessions when teacher's full face was visible. The social contingency contrast was directly replicated in both studies (red lines on Figure 4): interactive learning (live video call) resulted in better performance compared to observational learning (recorded video). These data are consistent with previous work on children, which have emphasized the benefits of social contingency for learning. Social connections with a teacher (e.g., parent versus stranger)[33,34] and social contingency[4,6,34,35] significantly improve learning in a variety of contexts.[7,36,37] Previous work on adults had more mixed results. A majority of studies found no difference between interaction-based learning

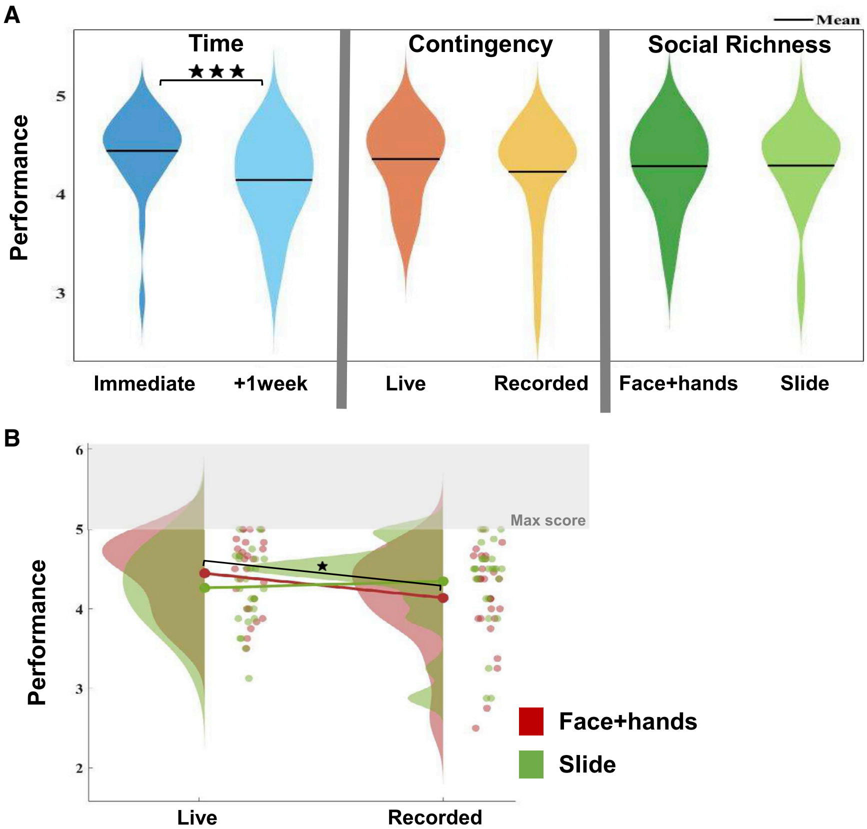


Figure 3. Results for experiment 2

(A) Main effects. A significant effect of time was found; ***p < 0.0001.

(B) Interaction effects. There was an interaction effect between contingency and richness; $F_{(1, 26)} = 5.28$; $p = 0.03$; $\eta 2 = 0.16$. Significant difference as measured by paired t test *$p < 0.05$ is shown. Here, performance has been averaged across time (immediate and +1 week test). The violin plots show the probability density function (Kernel density estimation), which can go beyond the smallest and largest data points. For clarity, we have indicated the max possible score on performance (x axis). Dots are showing individual scores.

See also Tables S1 and S2.

and observational learning.[10,24,25,38] However, these did not control for exposure time (e.g., recorded material could be replayed multiple times, whereas the live session was only played once) and did not specifically manipulate how interactive the teaching session was. Direct comparison of interaction-based with observational learning found a significant improvement in learning during interactive teaching.[9,22,39] These studies, however, failed to control for factors beyond interactivity (e.g., attending a class versus watching a video of one teacher speaking to the camera presents a number of differences beyond interactivity per se). Our work goes beyond previous studies by using a carefully controlled video-call method, which allows interactivity during live learning (participants were free to interrupt, ask questions, etc.) but also a yoked control for recorded sessions. Here, participants observed the previous participant, while the same exact information as the interactive sessions was available (overall across participants). Therefore, our results are in line with previous studies and furthermore can specifically support the conclusion that interactivity is the factor that enhances human learning in social contexts. Together with our pre-registered replication (experiment 2), this makes our results robust and relevant. The key role played by interactivity in social learning raises the question of which aspects of the interaction contributed the most.[40,41] Although a systematic analysis of verbal and non-verbal behaviors observed during the sessions is beyond the scope of this paper, we do not believe that performance could be driven by differences in participants' active engagement (e.g., clarifications requested): for each item, the researcher (teacher) ensured that two repetitions were given consistently in each session (see STAR Methods for an example).

Our second important finding across both experiments is the interaction effect between social contingency and social richness (Figure 4). Although it seems sensible to think that the format in which information is delivered (slides, video, podcast, etc.) could impact learning, to our knowledge, no other study has directly investigated this. The fact that the social richness of a learning context influences learning differently when students engage in a social interaction or just observe one suggests that different cognitive mechanisms may support interactive and observational learning. When participants take part in interactive learning with a full-face view of their teacher, they may engage in either joint attention,[15] common ground,[16] shared intentionality,[42] or all these processes together in order to attune with the teacher.[43] This attunement may allow information to be shared more effectively.[3,44] Rich visual cues may enable stronger attunement by providing more information about the interaction partner's gaze and mental states.[17,18] If this interpretation is correct, this may explain the results of experiment 1, where more socialness (more contingency and more richness) leads to better learning, and also for the replication in experiment 2, when learning from full-face stimuli was better for interactive conditions.

However, in experiment 2, learning was also good for the recorded-slides condition. In this observational learning, the learner is passively decoding an interaction between two external actors. Previous studies suggest that being an observer of a social interaction is more cognitively demanding than actively engaging in that interaction[11] and social cues may become distracting.[12–14] In our study, the fact that, during the recorded videos, participants were presented with the view of another participant as well as the teacher could have possibly contributed to diverge attention away from the learning material, making it a possible explanation for worse learning performance in this condition. Instead, a slide may help to focus the attention on the learning content, compared to a "socially rich" view (experiment 2), whereas decoding a social situation where only the hands are visible may be particularly hard (cognitively

Current Biology
Report

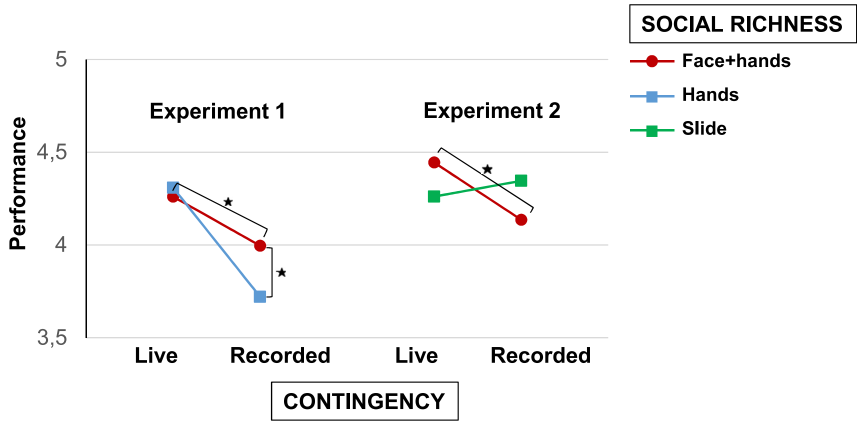


Figure 4. Summary of experiments 1 and 2

Performance has been averaged across time (immediate and +1 week test). * $p < .05$. For experiment 1, we found a main effect of contingency ($F_{(1,\ 23)} = 33.34$; $p < 0.0001$; $\eta 2 = 0.59$) and an interaction effect between contingency and social richness ($F_{(1,\ 23)} = 6.28$; $p = 0.017$; $\eta 2 = 0.22$). Paired-sample t tests revealed significant differences between recorded face+hands and recorded hands condition ($t_{(23)} = 2.15$; $p = 0.04$), live face+hands and recorded face+hands ($t_{(23)} = 2.99$; $p = 0.007$), and live hands and recorded hands ($t_{(23)} = 5.61$; $p = 0.001$). For experiment 2, we found an interaction effect between contingency and social richness ($F_{(1,\ 26)} = 5.28$; $p = 0.03$; $\eta 2 = 0.16$). Paired-sample t tests revealed significant differences between live face+hands and recorded face+hands ($t_{(26)} = 2.45$; $p = 0.02$). We also observed some trends for live face+hands versus live slide ($t_{(26)} = 1.77$; $p = 0.09$) and for recorded face+hands versus recorded slide ($t_{(26)} = -1.87$; $p = 0.07$). See also Tables S1 and S2.

demanding), given its atypicality (experiment 1). Note that the differences between interactive and recorded conditions cannot be driven by the stimuli, which are matched in our yoked design, or by audience effects,[45] as the teacher was online in all conditions (and participants were aware of it). Our claim that different mechanisms are engaged in interactive versus observational learning is compatible with the idea that being part of a social interaction engages different neural and cognitive mechanisms compared to observation.[46,47]

We use the term "social contingency" here to refer to our live-teaching condition, but we acknowledge that this goes beyond a simple time-dependent exchange and includes rich and complex behavioral dynamics, with bi-directional responses and original input between two or more people. This is not driven by a single cognitive mechanism but rather a series of cognitive processes (e.g., attention, motivation, back channeling, monitoring, and language) that may be absent in a non-interactive situation. It is hard to separate individual components because live interaction cannot be easily deconstructed.[48] Future studies using virtual reality might be able to do so[49] by experimentally manipulating which aspects of interaction are most important to learning.

The present work employed a naturalistic task that aimed to realistically recreate the student-teacher interaction online. However, in real-world education, teaching usually occurs in bigger groups, leading to two important considerations: first, in the context of a classroom, the teacher does not engage directly with each and every student throughout the whole session. It remains unknown how our results generalize to a one-to-many situation, like a lecture. Previous work comparing video lectures with face-to-face teaching suggests that the live teaching advantage generalizes to the context of a classroom.[9,39] However, remote video-call and face-to-face teaching may still involve different social dynamics. Video-call interfaces can suffer from time lags, video distortions, and a lack of mutual eye contact. It may be that the video-call context accentuates both the sense of engagement and the sense of disengagement, depending on whether a given student feels the teacher is directly interacting with them. Recently, an informal survey run across a large professional network revealed that, during Zoom calls, only about 27% of the 4,671 respondents reported to pay attention, while the rest either engaged in other activities or found it hard not to zone out, confessing to remain alert only to their name being called.[50] The catalyst role of social interaction may be even more impactful in online teaching, as attention and engagement are fundamental pre-requisites to successfully acquire new information.

Second, learning in a classroom environment implies learning *in the presence* of others (this being either offline in the same room or online in the same Zoom call): the mere presence of peers could modulate arousal, attentional, and motivational processes,[51] which in turn could either significantly improve learning[52] or make it harder.[53] Given that our design only involves one student, we do not know how our results generalize to such peer-group effects.

In conclusion, we have shown that social interaction acts as a catalyst to support learning and improves information transfer across people, and as such, it benefits from aspects that makes social interaction complex and rich. These findings contribute to our understanding of human learning: they point at the importance of interaction-based learning over observational learning and at social richness in the context of interaction-based learning, where social cues may support the student-teacher effort of achieving a shared understanding and co-creating knowledge. Future work can dissect the features of interaction that correlate with learning and identify ways to optimize learning in real-life educational contexts.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.08.045.

ACKNOWLEDGMENTS

S.D.F. was supported by Leverhulme award DS-2017-026. G.V. was supported by European Research Council Advanced Grant (ECOLANG, 743035) and Royal Society Wolfson Research Merit Award (WRM\R3\170016).

AUTHOR CONTRIBUTIONS

Conceptualization, S.D.F., G.V., and A.F.d.C.H.; methodology, S.D.F. and A.F.d.C.H.; formal analysis, S.D.F.; investigation, S.D.F.; writing – original draft, S.D.F.; writing – review & editing, S.D.F., G.V., and A.F.d.C.H.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We worked to ensure that the study questionnaires were prepared in an inclusive way. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list.

Received: May 14, 2021
Revised: July 28, 2021
Accepted: August 17, 2021
Published: September 14, 2021

REFERENCES

1. Shamay-Tsoory, S.G. (2021). Brains that fire together wire together: interbrain plasticity underlies learning in social interactions. Neuroscientist. https://doi.org/10.1177/1073858421996682.
2. Meltzoff, A.N., Kuhl, P.K., Movellan, J., and Sejnowski, T.J. (2009). Foundations for a new science of learning. Science *325*, 284–288.
3. Kuhl, P.K. (2007). Is speech learning 'gated' by the social brain? Dev. Sci. *10*, 110–120.
4. Myers, L.J., LeWitt, R.B., Gallo, R.E., and Maselli, N.M. (2017). Baby FaceTime: can toddlers learn from online video chat? Dev. Sci. *20*, e12430.
5. Roseberry, S., Hirsh-Pasek, K., Parish-Morris, J., and Golinkoff, R.M. (2009). Live action: can young children learn verbs from video? Child Dev. *80*, 1360–1375.
6. Roseberry, S., Hirsh-Pasek, K., and Golinkoff, R.M. (2014). Skype me! Socially contingent interactions help toddlers learn language. Child Dev. *85*, 956–970.
7. Troseth, G.L., Saylor, M.M., and Archer, A.H. (2006). Young children's use of video as a source of socially relevant information. Child Dev. *77*, 786–799.
8. Richert, R.A., Robb, M.B., and Smith, E.I. (2011). Media as social partners: the social nature of young children's learning from screen media. Child Dev. *82*, 82–95.
9. Ramlogan, S., Raman, V., and Sweet, J. (2014). A comparison of two forms of teaching instruction: video vs. live lecture for education in clinical periodontology. Eur. J. Dent. Educ. *18*, 31–38.
10. Solomon, D.J., Ferenchick, G.S., Laird-Fick, H.S., and Kavanaugh, K. (2004). A randomized trial comparing digital and live lecture formats [ISRCTN40455708. BMC Med. Educ. *4*, 27.
11. Kourtis, D., Jacob, P., Sebanz, N., Sperber, D., and Knoblich, G. (2020). Making sense of human interaction benefits from communicative cues. Sci. Rep. *10*, 18135.
12. Phillips, L.H., Tunstall, M., and Channon, S. (2007). Exploring the role of working memory in dynamic social cue decoding using dual task methodology. J. Nonverb. Behav. *31*, 137–152.
13. Friedman, R.S., and Förster, J. (2010). Implicit affective cues and attentional tuning: an integrative review. Psychol. Bull. *136*, 875–893.
14. Kirkorian, H.L., Choi, K., and Pempek, T.A. (2016). Toddlers ' word learning from contingent and noncontingent video on touch screens. Child Dev. *87*, 405–413.
15. Schertz, H.H., Odom, S.L., Baggett, K.M., and Sideris, J.H. (2013). Effects of joint attention mediated learning for toddlers with autism spectrum disorders: an initial randomized controlled study. Early Child. Res. Q. *28*, 249–258.
16. Bohn, M., Tessler, M.H., and Frank, M.C. (2019). Integrating common ground and informativeness in pragmatic word learning. PsyArXiv. https://doi.org/10.31234/osf.io/cbx46.
17. Kajopoulos, J., Cheng, G., Kise, K., Müller, H.J., and Wykowska, A. (2021). Focusing on the face or getting distracted by social signals? The effect of distracting gestures on attentional focus in natural interaction. Psychol. Res. *85*, 491–502.
18. Marotta, A., Lupiáñez, J., Martella, D., and Casagrande, M. (2012). Eye gaze versus arrows as spatial cues: two qualitatively different modes of attentional selection. J. Exp. Psychol. Hum. Percept. Perform. *38*, 326–335.
19. Laland, K.N., and Rendell, L. (2019). Social learning: theory. In Encyclopedia of Animal Behavior, , J.C. Choe, ed. (Elsevier), pp. 380–386.
20. Morgan, T.J.H., Uomini, N.T., Rendell, L.E., Chouinard-Thuly, L., Street, S.E., Lewis, H.M., Cross, C.P., Evans, C., Kearney, R., de la Torre, I., et al. (2015). Experimental evidence for the co-evolution of hominin toolmaking teaching and language. Nat. Commun. *6*, 6029.
21. Hertel, P.T., Brozovich, F., Joormann, J., and Gotlib, I.H. (2008). Biases in interpretation and memory in generalized social phobia. J. Abnorm. Psychol. *117*, 278–288.
22. Zhang, D., Zhou, L., Briggs, R.O., and Nunamaker, J.F. (2006). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. Inf. Manage. *43*, 15–27.
23. Kwok, K., Ghrear, S., Li, V., Haddock, T., Coleman, P., and Birch, S.A.J. (2016). Children can learn new facts equally well from interactive media versus face to face instruction. Front. Psychol. *7*, 1603.
24. Vaccani, J.P., Javidnia, H., and Humphrey-Murto, S. (2016). The effectiveness of webcast compared to live lectures as a teaching tool in medical school. Med. Teach. *38*, 59–63.
25. Schreiber, B.E., Fukuta, J., and Gordon, F. (2010). Live lecture versus video podcast in undergraduate medical education: a randomised controlled trial. BMC Med. Educ. *10*, 68.
26. Moore, C., and Barresi, J. (2017). The role of second-person information in the development of social understanding. Front. Psychol. *8*, 1667.
27. Wakefield, E., Novack, M.A., Congdon, E.L., Franconeri, S., and Goldin-Meadow, S. (2018). Gesture helps learners learn, but not merely by guiding their visual attention. Dev. Sci. *21*, e12664.
28. Mundy, P., and Newell, L. (2007). Attention, joint attention, and social cognition. Curr. Dir. Psychol. Sci. *16*, 269–274.

29. Kasari, C., Paparella, T., Freeman, S., and Jahromi, L.B. (2008). Language outcome in autism: randomized comparison of joint attention and play interventions. J. Consult. Clin. Psychol. *76*, 125–137.

30. Saito, D.N., Tanabe, H.C., Izuma, K., Hayashi, M.J., Morito, Y., Komeda, H., Uchiyama, H., Kosaka, H., Okazawa, H., Fujibayashi, Y., and Sadato, N. (2010). "Stay tuned": inter-individual neural synchronization during mutual gaze and joint attention. Front. Integr. Nueurosci. *4*, 127.

31. Kawai, N. (2011). Attentional shift by eye gaze requires joint attention: Eye gaze cues are unique to shift attention. Jpn. Psychol. Res. *53*, 292–301.

32. Hoehl, S., Fairhurst, M., and Schirmer, A. (2021). Interactional synchrony: signals, mechanisms and benefits. Soc. Cogn. Affect. Neurosci. *16*, 5–18.

33. Kuhl, P.K., Tsao, F.M., and Liu, H.M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. Proc. Natl. Acad. Sci. USA *100*, 9096–9101.

34. Lauricella, A.R., Gola, A.A.H., and Calvert, S.L. (2011). Toddlers' learning from socially meaningful video characters. Media Psychol. *14*, 216–232.

35. O'Doherty, K., Troseth, G.L., Shimpi, P.M., Goldenberg, E., Akhtar, N., and Saylor, M.M. (2011). Third-party social interaction and word learning from video. Child Dev. *82*, 902–915.

36. Goldstein, M.H., and Schwade, J.A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. Psychol. Sci. *19*, 515–523.

37. Calvert, S.L., Strong, B.L., Jacobs, E.L., and Conger, E.E. (2007). Interaction and participation for Young Hispanic and Caucasian Girls' and Boys' learning of media content. Media Psychol. *9*, 431–445.

38. Davis, J., Crabb, S., Rogers, E., Zamora, J., and Khan, K. (2008). Computer-based teaching is as good as face to face lecture-based teaching of evidence based medicine: a randomized controlled trial. Med. Teach. *30*, 302–307.

39. John, J.R., Hemasundar, H.P.A., and Rajendran, G. (2016). A collation of two forms of demonstration: in-person versus video demonstration which gains the upper hand? Res. Rev. J. Dent. Sci. *4*, 23–29.

40. Dale, R., Fusaroli, R., Duran, N.D., and Richardson, D.C. (2013). The Self-Organization of Human Interaction (Elsevier).

41. Dideriksen, C., Christiansen, M.H., Tylén, K., Dingemanse, M., and Fusaroli, R. (2020). Building common ground: quantifying the interplay of mechanisms that promote understanding in conversations. PsyArXiv. https://doi.org/10.31234/osf.io/a5r74.

42. Sabbagh, M.A., and Baldwin, D.A. (2001). Learning words from knowledgeable versus ignorant speakers: links between preschoolers' theory of mind and semantic development. Child Dev. *72*, 1054–1070.

43. Schmitz, M. (2015). Joint attention and understanding others. Synth. Philos. *29*, 235–251.

44. Hu, J., Qi, S., Becker, B., Luo, L., Gao, S., Gong, Q., Hurlemann, R., and Kendrick, K.M. (2015). Oxytocin selectively facilitates learning with social feedback and increases activity and functional connectivity in emotional memory and reward processing regions. Hum. Brain Mapp. *36*, 2132–2146.

45. Hamilton, A.F.C., and Lind, F. (2016). Audience effects: what can they tell us about social neuroscience, theory of mind and autism? Cult. Brain *4*, 159–177.

46. Seuren, L.M., Wherton, J., Greenhalgh, T., and Shaw, S.E. (2021). Whose turn is it anyway? Latency and the organization of turn-taking in video-mediated interaction. J. Pragmatics *172*, 63–78.

47. Rice, K., and Redcay, E. (2016). Interaction matters: a perceived social partner alters the neural processing of human speech. Neuroimage *129*, 480–488.

48. De Jaegher, H., Di Paolo, E., and Gallagher, S. (2010). Can social interaction constitute social cognition? Trends Cogn. Sci. *14*, 441–447.

49. Pan, X., and Hamilton, A.F.C. (2018). Why and how to use virtual reality to study human social interaction: the challenges of exploring a new research landscape. Br. J. Psychol. *109*, 395–417.

50. Riccobono, F. (2020). Blind polls: Zoom fatigue. https://www.teamblind.com/blog/index.php/2020/07/30/blind-polls-zoom-fatigue/.

51. Guerin, B. (1986). Mere presence effects in humans: a review. J. Exp. Soc. Psychol. *22*, 38–77.

52. Lytle, S.R., Garcia-Sierra, A., and Kuhl, P.K. (2018). Two are better than one: infant language learning from video improves in the presence of peers. Proc. Natl. Acad. Sci. USA *115*, 9859–9866.

53. Skuballa, I.T., Xu, K.M., and Jarodzka, H. (2019). The impact of co-actors on cognitive load: when the mere presence of others makes learning more difficult. Comput. Human Behav. *101*, 30–41.

54. Prolific (2014). Prolific. https://prolific.co/.

55. Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., and Evershed, J. (2020). Gorilla in our MIDST: An online behavioral experiment builder. Behav Res *52*, 388–407.

56. The MathWorks (2020). MATLAB (The MathWorks).

57. RStudio Team (2020). RStudio: Integrated Development for R. RStudio (PBC).

58. IBM Corporation (2020). IBM SPSS Statistics for Windows, Version 27.0 (IBM).

59. Vigliocco, G., Motamedi, Y., Murgiano, M., Wonnacott, E., Marshall, C., Milán-Maillo, I., and Perniss, P. (2019). Onomatopoeia, gestures, actions and words: how do caregivers use multimodal cues in their communication to children? Proc. 41st Annu. Conf. Cogn. Sci. Soc. *41*, 1171–1177.

60. Al-Rukban, M.O. (2006). Guidelines for the construction of multiple choice questions tests. J. Family Community Med. *13*, 125–133.

61. Liebowitz, M.R. (1987). Social phobia. Mod. Probl. Pharmacopsychiatry *22*, 141–173.

62. Jolliffe, D., and Farrington, D.P. (2006). Development and validation of the basic empathy scale. J. Adolesc. *29*, 589–611.

Please cite this article in press as: De Felice et al., Social interaction is a catalyst for adult human learning in online contexts, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.08.045

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models/subjects		
Young adults: see "Experimental model and subject details" for details.	Online recruitment portal[54]	N/A
Data and code availability		
Raw data and analysis code	This paper	https://osf.io/rm2zp/?view_only=6f5b20c6392e4bf48bf69681b2e42f34 - https://doi.org/10.17605/OSF.IO/RM2ZP
Software and algorithms		
Gorilla Experiment Builder	Anwyl-Irvine et al.[55]	https://www.gorilla.sc
MATLAB R2020a	The MathWorks[56]	https://www.mathworks.com/products/matlab.html
Prolific Academic	Prolific[54]	https://www.prolific.co
R and Rstudio	RStudio Team[57]	https://www.rstudio.com
SPSS	IBM SPSS Statistics for Windows, Version 27.0[58]	https://www.ibm.com/uk-en/products/spss-statistics

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Sara De Felice (sara.felice.16@ucl.ac.uk).

Materials availability

Full information about the experimental items is available at the public repository Open Science Framework https://osf.io/rm2zp/?view_only=6f5b20c6392e4bf48bf69681b2e42f34.

Data and code availability

Original data have been deposited at the Open Science Framework https://osf.io/rm2zp/?view_only=6f5b20c6392e4bf48bf69681b2e42f34 (https://doi.org/10.17605/OSF.IO/RM2ZP). All original code has been deposited at the Open Science Framework https://osf.io/rm2zp/?view_only=6f5b20c6392e4bf48bf69681b2e42f34 (https://doi.org/10.17605/OSF.IO/RM2ZP). Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

This study was approved by the UCL ethics committee. All participants gave consent to take part, and a separate optional consent to share the video recordings of their session with others. Some people chose not to consent to video sharing but were still able to complete the learning task. Experiment 2 was pre-registered on the 21st December 2020 on OSF (De Felice & Hamilton, 2020 10.17605/OSF.IO/NXS37).

Experiment 1

43 participants took part in the study. Data from the first 13 participants formed our pilot study (not reported here). Of the remaining 30 participants, 6 participants were excluded due to poor videocall quality (N = 2, we only accepted subjects who reported 4 and above on a 1(poor)-5(excellent) videocall quality scale), inattention (N = 1), not completing the one-week after test (N = 1), revisiting the material during the week-gap (N = 2). The final sample (N = 24, 11 female) included in the analysis had a mean age of 27.29 (SD = 4.28, range 19-35 years). They were either native English speakers (45.83%) or reported to be regularly speaking English since at least more than 5 years.

Experiment 2

We used the software program G*Power to conduct a power analysis. From experiment 1, we used the minimum effect size of interest of $\eta^2 = 0.05$ (effect size F = 0.22, Social contingency contrast) and a correlation among repeated-measure of 0.66, aiming for 0.95

power at 0.05 alpha error probability. The power analyses indicated a sample size of 20 people. We recruited 30 to ensure our sample to account for data loss due to post hoc exclusion (see Experiment 1 sample and data pre-processing for exclusion criteria).

30 participants took part in the study. Overall, 3 participants were excluded due to either poor video-call quality (N = 1), or speaking English since less than 5 years (N = 2). The final sample (N = 27, 14 female) included in the analysis had a mean age of 25.23 (SD = 5.04, range 19-35 years). 37.03% of the sample reported to be native English speakers (the rest reported to be regularly speaking English since at least more than 5 years). All participants completed all the steps of the study.

METHOD DETAILS

Sample recruitment

Sample was recruited via Prolific (https://www.prolific.co) [2020]. To be included in the study, participants had to be aged 18-35 (inclusive); be fluent in English (having spoken English regularly for at least the past 5-10 years); giving consent to have their camera and microphone on as well as being recorded for the whole duration of the experiment. In addition to these criteria, participants could only take part in experiment 2 if they did not took part in experiment 1. Participants were paid £7.50 for the first hour of the experiment, and then a further £5 when they completed the learning quiz a week later.

Material

Two learning sets were created, each including eight items, two from each of the following four categories: animals, food, ancient objects and musical instruments. Item selection started from a pool used in Vigliocco et al.[59] Final items were selected on the basis of an initial pilot (N = 15) run face-to-face before the covid-19 pandemic. The 16 selected items were considered highly unlikely to be known by the general population. Wherever possible, models for each item were bought online, when not found these were hand-made in ceramic and acrylic, ensuring high resemblance to the real item. Learning material and quiz were adapted for this experiment based on a pilot study conducted online (N = 13, Subject Details for Experiment 1). For the learning material, a descriptive paragraph was created for each item, made of 5 core pieces of information (e.g., *where is the item from? what does the name mean?* etc) plus two or three extra curiosities to make it more challenging (these were not tested). For the quiz, there were five multiple choice questions (each testing memory for one of the five core pieces of information): each question had three options (the correct one, a misleading one and a completely wrong one; see Al-Rukban[60]). See Figure 1A for an example of the learning material and quiz (full information and question sets for all items are available online at https://files.osf.io/v1/resources/rm2zp/providers/osfstorage/60fe9c0e317620013237db38?direct&version=1).

Design

This study aimed to investigate whether i) actively participating and ii) seeing the face of the teacher was beneficial during a virtual learning session (compared to learning passively from a recorded video without seeing the face of the teacher). A 2x2 within-subject design was used to look at the influence of the two factors of interest on learning performance: social contingency (live versus recorded video-call) and social richness (face versus hands-only view). There were four conditions: live_face, live_hands, recorded_face, recorded_hands (Figures 1B and 1C). Where possible (depending on participant's consents) the recorded session for a given participant was made using a recording of the live session for the participant immediately before them. Each live session was never used more than three times as a recorded session (i.e., the same recorded session was shown to a maximum of three participants).

A typical session would run as follows: the live interactive teaching session had 8 trials with 8 different items and then the recorded teaching session had 8 trials with 8 different items. The 8 live trials alternated between a face+hands view and a hands-only view, and similarly the 8 recorded trials alternated in view (Figure 1D). The order of the live/recorded conditions, the set of items assigned to each condition and the starting point for the alternating viewpoint trials was counterbalanced over participants. Thus, each of the 16 items appeared in either live or recorded and either face or hands condition equally often. Also, we controlled for the list position of each item (that is, whether an item was presented as first or last), so that each item appeared fairly equally on each of the 16 list positions. Learning performance was tested twice: immediately after the experimental session and one week later.

The learning context in which participants learned about a given item was manipulated. We had three binary factors: i) Contingency (live or recorded); ii) View (face+hands or only hands); iii) Time (learning quiz delivered immediately after the experiment and one week after). Our main outcome variable was performance on the learning quiz (5 multiple-choice questions x 16 items): Performance = SumCorrectAnswers/TotalTrials (note that number of total trials could change across participants depending on whether some data points were excluded – see Data Pre-processing section).

We also collected data on: pre-knowledge of the experimental items; psychometric questionnaires on social anxiety (24 items)[61] and basic empathy scale (20 items);[62] video-call quality (1-5 scale, 1 = very poor to 5 = excellent); enjoyment rating (general and separately for live and recorded session); source memory of the learned items ('in which context did you learn this? Live/Recorded call; Face/Slide view') and review of any experimental item during the one week-delay period in between quizzes (see Procedure section).

For experiment 2, we replicated the same design as in experiment 1, with the only difference being the contrast in the social richness factor: here, we compare exposure to teacher's face to presentation of PowerPoint slide (instead of the 'Hands' condition as in experiment 1). In the slide condition, participants were presented with a slide with white-background, the name of the item placed on the top-center of the screen, and three pictures of the item taken from different perspectives. During the slide presentations, the teacher used the mouse cursor to point to the item or parts of it on the slide. This allows the slide conditions to maintain some aspect of attention/joint attention without any visible face. The order of conditions and trials were the same as for experiment 1.

Procedure

For both experiments, procedure involved four main parts: invitation on Prolific, the videocall (main experiment), completion of the immediate learning quiz and completion of the same quiz a week later. First, participants responded to our advert on Prolific (https://www.prolific.co) [accessed July 2020], when they were directed to Gorilla Experiment Builder (https://www.gorilla.sc) to complete consent form and demographic variables. If meeting the inclusion criteria for the study, they were invited to arrange a videocall with the researcher. A zoom link was sent via the prolific chat: this ensured complete anonymity.

Second, the researcher introduced herself and made sure the participant could see/hear well. Participants were then asked to make sure the zoom window was in full-screen mode and that gallery-view was selected. The researcher gave oral instructions always in the same way (alternating only the order of instructions for live and recorded session depending on the participant): "*the aim of this experiment is for you to learn information about a bunch of different items including animals, food, musical instruments and ancient objects. You will learn about these in slightly different context: for the first half of the experiment we will be chatting over this live call. I will be showing a model of the item and tell several facts about it. When I have finished, you can interact with me, ask questions about the item and I can repeat any information you may have missed. You are very welcome to interact with me as much as you want. We will have 2 mins per item, then we will move to the next item. For the second part of the experiment, I will share my screen and play a video of a previous participant who did the same study before you. Your task is always the same: try to learn as many facts as you can about the items you will hear of, as after the experiment, you will be asked to complete a quiz to test your learning. Please do not take any notes while we go through the items: just listen and try to see what you can remember. Also, you will notice that sometimes I will adjust my screen like this [lowering down the camera so that only hands would be visible], this is just part of the experiment. Do not worry if it feels there is a lot of information: this is meant to be challenging. Hope you can just have fun listening to these different items and learn new things! Is it all clear?*" Participants had the opportunity to ask questions at this point. Before starting the experiment, participants pre-knowledge on the items were tested by reading each items aloud "*Have you ever heard of any of these items before?*." If an item was known, it was still included in the experimental session but it was noted and excluded from the analysis. The experiment then started with either the live session followed by the recorded session or vice versa (order was counterbalanced). For each trial, the name of the item was presented on the bottom-left side of the screen via a clip-holder, printed in capitals in black ink over white background. This was always visible throughout the whole duration of the trail and in all conditions. Trials alternated between face and hands condition. For each trial, after the description of the item, two prompts were included (e.g., "Do you remember what the name means"? and "Can you recall where it comes from"? – the researcher would give the correct answer if participants could not recall it). The researcher would omit prompting if the participant asked for repetition themselves, to ensure each session would have equal number of prompts/repetitions. The full session lasted approximately 45 mins (16 trials of 2 min each plus some time for instructions and debriefing, Figure 1D).

Third, at the end of the learning session, participants were redirected to prolific, where they could access a link to complete the learning quiz (immediate performance) in Gorilla. At this point, we also collected information about the video-call quality and measures of social anxiety[61] and empathy[62]. This part lasted about 10 mins.

Fourth and finally, exactly one week after they completed the videocall and immediate learning quiz, participants were given access to a new study on Prolific. Those who wanted to participate responded through Prolific and were directed to the delay quiz on Gorilla. At this point, in addition to their learning performance, we also collected information on source memory of the learned items ('*in which context did you learn this? Live/Recorded call; Face/Slide view*') and whether they reviewed of any experimental item during the one week-delay period. This part lasted about 5-10 mins.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data pre-processing

Single trials (i.e., 'item') were excluded from the whole dataset based on the following criteria: i) participant reported pre-knowledge of the item before the experiment; ii) the connection was temporarily bad for one or two trials (but overall good enough to keep the participant as a whole); iii) the information presented by the teacher during the learning phase was somehow inaccurate, misleading or incomplete.

In addition, single trials were excluded from the delayed performance only if participant reported to having revisited the item in any form (telling a friend about it, reading/googling about it) during the one-week gap between immediate and delayed learning quiz.

Performance was calculated out of 5 questions per item, based on the valid trials: Performance (/5) = Sum Correct Answers / Total Trials.

Data analysis

We used SPSS to run a 2x2x2 factorial ANOVAs to test the difference in learning performance between Call (Live versus Recorded video-call) and View (Face versus Hands-only for experiment 1 and Slide for experiment 2) and Time (immediate versus delay recall). Sample size, Means and SD are reported for both experiments in Table S1. Statistical tests, p values and Confidence Intervals are reported for all contrasts for both experiments in Table S2.