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

Sara De Felice[1]; Gabriella Vigliocco[2], Antonia F. de C. Hamilton[1]

1. Institute of Cognitive Neuroscience, University College London, London, WC1N 3AZ, UK
2. Department of Experimental Psychology, University College London, London, WC1H 0AP, UK

SUMMARY

Human learning is highly social[1–3]. Advances in technology have increasingly moved learning online and the recent Covid-19 pandemic has accelerated this trend. Online learning can vary in terms of how 'socially' the material is presented (e.g. live/recorded), but there is 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 vs recorded lecture) and social richness (viewing the teacher's face/hands/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 & 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[15] and/or distracting[16–18], resulting in less learning from rich social information if there is no interactivity. We suggest that being part of a genuine social interaction catalyses learning, possibly via mechanisms of joint attention[11], common ground[12] or (inter)-active discussion, and as such, interactive learning benefits from rich social settings[13,14].

Key words: social learning, human learning, online schooling, social interaction, face-to-face interaction, education, psychology, social cues, joint attention, two-person neuroscience.

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 [15] 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 [15,26] and/or distract learners [13]. Alternatively, social cues such as eye-gaze [14] 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 Covid-19 pandemic where online 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 2x2-factorial design, where participants learn novel information over a video-call in four teaching formats, differing on the basis of social contingency (live vs 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 one week later.

---------------------------------------FIGURE 1 ABOUT HERE--------------------------------------

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<.0001, η^2=.53). There was also a main effect of *Social contingency* ($F_{(1, 23)}$=33.34, p<.0001, η^2=.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=.27, η^2 =.05). However, we found an interaction effect between *Social contingency* and *Social richness* ($F_{(1, 23)}$=6.28, p=.017, η^2=.22; Figure 2B). To interpret this interaction, given that the same pattern of results have been observed at both times – we collapsed across the factor *Time* and considered the *Social contingency* and *Social richness* factors (Appendix Table 3). 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=.04). In addition, both post-hoc comparisons for the social contingency factor (live-face vs recorded-face $t_{(23)}$=2.99, p=.007, and live-hands vs recorded-hands $t_{(23)}$=5.61, p=.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 (see SuppInfo).

--------------------------------------FIGURE 2 ABOUT HERE-------------------------------------

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<.0001, η^2=.54; Figure 3A). However, we did not find a main effect of *Social contingency* ($F_{(1, 26)}$=1.67, p=.21, η^2=.06) or of *Social richness* ($F_{(1, 26)}$=.04, p=.84, η^2=.002). Importantly, we replicated the interaction effect between *Social contingency* and *Social richness* ($F_{(1, 26)}$=5.28, p=.03, η^2=.16; Figure 3B). Similar results were found using a multi-level logistic regression model (see SuppInfo).

--------------------------------------FIGURE 3 ABOUT HERE---------------------------------------

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 (Appendix Table 3). 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=.02). Additionally, in the live condition exposure to face has a trend to an advantage over slides ($t_{(26)}$= 1.77, p=.09), while the opposite was observed in the recorded condition ($t_{(26)}$=-1.87, p=.07). 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.

--------------------------------------FIGURE 4 ABOUT HERE-------------------------------------

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, interaction-based learning is more effective than observational learning for full-face view situations. Both our studies showed that during a full-face view student-teaching exchange, playing an active role in the interaction improves learning over yoked observation of the same sessions. Second, visual social cues impact on learning differently depending on whether learning is interactive or observational (Figure 4): both studies show 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 Fig 4): interactive learning (live video-call) resulted in better performance compared to observational learning (recorded video). This data is consistent with previous work on children, which have emphasised the benefits of social contingency for learning. Social connections with a teacher (e.g. parent vs 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 and observational learning (Davis et al., 2008; Schreiber et al., 2010; Solomon et al., 2004; Vaccani et al., 2016). However, these did not control for exposure time (e.g. recorded material could be replayed multiple times while the live session was only played once) and did not specifically manipulated 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 vs watching a video of one teacher speaking to the camera present 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 which present the same exact information as the interactive sessions (overall across participants) without interactivity. Therefore, our results are in line with previous studies, and furthermore can specifically support the conclusion that interactivity may be 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[45,46]. While a systematic analysis of verbal and non-verbal behaviours 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 (Fig 4). While 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 [11], common ground (Bohn, Tessler & Frank, 2019), shared intentionality (Sabbagh & Baldwin, 2005) or all these processes together in order to attune with the teacher [40]. This attunement may allow information to be shared more effectively [3,41]. Rich visual cues may enable stronger attunement by providing more information about the interaction partner's gaze and mental states [13,14]. 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 [15] and social cues may become distracting [16–18]. Therefore, a slide may help to focus the attention on the learning content, compared to a 'socially rich' view (Experiment 2), while decoding a social situation where only the hands are visible may be particularly hard (cognitively 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, nor by audience effects [42], 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 engage different neural and cognitive mechanisms compared to observation [43,44].

It is worth noting that interactivity is interpreted here as the social mechanism through which we engage in a given situation (e.g. learning). Such mechanism should not be regarded as a mere motor and mental action, but rather as a series of cognitive processes (e.g. attention, motivation, back-channelling, monitoring, language) that may be absent in a non-interactive situation. It is hard to separate individual components because live interaction cannot be easily deconstructed [47]. Future studies using virtual reality might be able to do so [48], by experimentally manipulating which aspects of interaction are most important to learning.

The present work employed a naturalistic task which aimed to realistically recreate the student-teacher interaction online. However, in real-world education, teaching usually occurs in bigger groups. This gives rise 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 generalise 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 generalises 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 4671 respondents reported to pay attention, while the rest either engaged in other activities or found it hard not to zoning out, confessing to remain alert only to their name being called (Blind, 2020). This could mean that the catalyst role of social interaction may be even more impactful in online teaching, as attention and engagement is a fundamental pre-requisite 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[49], which in turn could either significantly improve learning[50], or making it harder[51].

Educators and teachers should consider carefully the teaching formats they employ depending on the tools they have available and their audience, in order to improve student engagement and optimise learning. Interactive learning requires a two-way connection between student and teacher: therefore, encouraging students to switch webcam on and directly prompting them with questions and feedback may be particularly impactful in the context of virtual schooling.

In conclusion, we have shown that social interaction can act as a catalyst to support learning and improve information-transfer across people, and 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 optimise learning in real-life educational contexts.

STAR METHOD

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

This study did not generate any new unique materials.

Data and code availability

All raw data reported in this paper will be shared by the lead contact upon request. This paper does not report original code. 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 η^2 =.05 (effect size F = .22, Social contingency contrast) and a correlation among repeated measure of .66, aiming for .95 power at .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 (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 payed £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 [52]. 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, 2006). See Figure 1.A for an example of the learning material and quiz. Full information and question sets for all items are reported in the Supplementary Material.

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 vs recorded video-call) and social richness (face vs hands-only view). There were four conditions: live_face, live_hands, recorded_face, recorded_hands (Figure 1.B and 1.C). 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 1.D). 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 in results).

We also collected data on: pre-knowledge of the experimental items; psychometric questionnaires on social anxiety (24 items, Liebowitz, 1987) and basic empathy scale (20 items, Jolliffe & Farrington, 2006); 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).

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-centre 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 (www.prolific.co) [accessed July 2020], when they were directed to Gorilla Experiment Builder (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 1.D).

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 (Liebowitz, 1987) and empathy (Jolliffe & Farrington, 2006). 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) = SumCorrectAnswers/TotalTrials.

Data analysis

We used SPSS to run a 2x2x2 factorial ANOVAs to test the difference in learning performance between Call (Live vs Recorded video-call) and View (Face vs Hands-only for experiment 1 and Slide for experiment 2) and Time (immediate vs delay recall). Sample size, Means and SD are reported for both experiments in Table 2. Statistical tests, p-values and Confidence Intervals are reported for all contrasts for both experiments in Table 3.

We also run a Logistic Regression Model using the glmer function in the lme4 package in R (ADD REF). The model was built with the three factors of interest (Time, Social Contingency and Social Richness) as predictors, and question (nested per item) and participant as intercepts: Performance ~ 1 + Time*Contingency*Richness + (1 | Item/Question) + (1 | Participant). The outcome of the regression confirmed the pattern of results found with the factorial ANOVA. For experiment 1, we found that both Time ($p<.0001$) and Contingency ($p<.0001$) were good predictors of Performance: delay test was associated with a decreased of performance of .298 unit ($R^2= -.298 \pm .047$), and the recorded teaching condition was associated with worse learning ($R^2= -.346 \pm .046$). The model also revealed an interaction effect between Time and Contingency ($R^2=.098 \pm .046$, $p<.01$). For experiment 2, we found that both Time ($p<.0001$) and Contingency ($p<.01$) were good predictors of Performance: delay test was associated with a decreased of performance of .317 unit ($R^2= -.317 \pm .051$), and the recorded teaching condition was associated with worse learning ($R^2= -.118 \pm .051$). The model also revealed an interaction effect between Time and Contingency ($R^2=.195 \pm .052$, $p<.0001$).

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Gorilla Experiment Builder	Anwyl-Irvine, Massonnié, Flitton, Kirkham & Evershed, 2018	www.gorilla.sc
MATLAB R2020a	MATLAB (2020). Natick, Massachusetts: The MathWorks Inc.	https://www.mathworks.com/products/matlab.html
Prolific Academic	Prolific (2014). Oxford, UK	www.prolific.co
R and RStudio	RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA	http://www.rstudio.com/

SPSS	IBM Corp. Released 2020. IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY: IBM Corp	https://www.ibm.com/uk-en/products/spss-statistics

Author Contributions

Conceptualization, S.D.F., G.V. and A.H.; Methodology, S.D.F. and A.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.H.

References

1. Shamay-Tsoory, S. G. Brains that Fire Together Wire Together: Interbrain Plasticity Underlies Learning in Social Interactions. *Neuroscientist* (2021) doi:10.1177/1073858421996682.
2. Meltzoff, A., Kuhl, P., Movellan, J. & Sejnowski, T. J. Foundations for a New Science of Learning -- Meltzoff et al. 325 (5938): 284 -- Science. *Science (80-.).* **325**, 284–288 (2009).
3. Kuhl, P. K. Is speech learning 'gated' by the social brain? *Dev. Sci.* **10**, 110–120 (2007).
4. Myers, L. J., LeWitt, R. B., Gallo, R. E. & Maselli, N. M. Baby FaceTime: can toddlers learn from online video chat? *Dev. Sci.* **20**, (2017).
5. Roseberry, S., Hirsh-Pasek, K., Parish-Morris, J. & Golinkoff, R. M. Live action: Can young children learn verbs from video? *Child Dev.* **80**, 1360–1375 (2009).
6. Roseberry, S., Hirsh-Pasek, K. & Golinkoff, R. M. Skype Me! Socially Contingent Interactions Help Toddlers Learn Language. *Child Dev.* **85**, 956–970 (2014).
7. Troseth, G. L., Saylor, M. M. & Archer, A. H. Young children's use of video as a source of socially relevant information. *Child Dev.* **77**, 786–799 (2006).
8. Richert, R. A., Robb, M. B. & Smith, E. I. Media as Social Partners: The Social Nature of Young Children's Learning From Screen Media. *Child Dev.* **82**, 82–95 (2011).
9. Ramlogan, S., Raman, V. & Sweet, J. A comparison of two forms of teaching instruction: video vs. live lecture for education in clinical periodontology. 31–38 (2013).
10. Solomon, D. J., Ferenchick, G. S., Laird-Fick, H. S. & Kavanaugh, K. A randomized trial comparing digital and live lecture formats ISRCTN40455708. *BMC Med. Educ.* **4**, 1–6 (2004).

11. Schertz, H. H., Odom, S. L., Baggett, K. M. & Sideris, J. H. Effects of Joint Attention Mediated Learning for toddlers with autism spectrum disorders : An initial randomized controlled study. *Early Child. Res. Q.* **28**, 249–258 (2013).
12. Bohn, M., Tessler, M. H. & Frank, M. Integrating Common Ground and Informativeness in Pragmatic Word Learning. (2019) doi:10.31234/osf.io/cbx46.
13. Kajopoulos, J., Cheng, G., Kise, K., Müller, H. J. & Wykowska, A. 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 (2020).
14. Marotta, A., Lupiáñez, J., Martella, D. & Casagrande, M. Eye gaze versus arrows as spatial cues: Two qualitatively different modes of attentional selection. *J. Exp. Psychol. Hum. Percept. Perform.* **38**, 326–335 (2012).
15. Kourtis, D., Jacob, P., Sebanz, N., Sperber, D. & Knoblich, G. Making sense of human interaction benefits from communicative cues. *Sci. Rep.* **10**, 1–8 (2020).
16. Phillips, L. H., Tunstall, Æ. M. & Channon, Æ. S. Exploring the Role of Working Memory in Dynamic Social Cue Decoding Using Dual Task Methodology. 137–152 (2007) doi:10.1007/s10919-007-0026-6.
17. Friedman, R. S. & Förster, J. Implicit Affective Cues and Attentional Tuning : An Integrative Review. **136**, 875–893 (2010).
18. Kirkorian, H. L. Hollins Digital Commons Toddlers ' Word Learning from Contingent and Non-Contingent Video on Touchscreens. 405–413 (2016).
19. Bandura, A. Social learning: Theory. *Encyclopedia of Animal Behavior* 380–386 (2019) doi:10.1016/B978-0-12-813251-7.00057-2.
20. Morgan, T. J. H. *et al.* Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nat. Commun.* **6**, 4–11 (2015).
21. Hertel, P. T., Brozovich, F., Joormann, J. & Gotlib, I. H. Biases in Interpretation and Memory in Generalized Social Phobia. *J. Abnorm. Psychol.* **117**, 278–288 (2008).
22. Zhang, D., Zhou, L., Briggs, R. O. & Nunamaker, J. F. Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Inf. Manag.* **43**, 15–27 (2006).
23. Kwok, K. *et al.* Children can learn new facts equally well from interactive media versus face to face instruction. *Front. Psychol.* **7**, 1–10 (2016).
24. Vaccani, J. P., Javidnia, H. & Humphrey-Murto, S. The effectiveness of webcast compared to live lectures as a teaching tool in medical school. *Med. Teach.* **38**, 59–63 (2016).

25. Schreiber, B. E., Junaid, F. & Fabiana, G. Live lecture versus video podcast in undergraduate medical education. *Med. Educ.* **10**, 1–6 (2010).
26. Moore, C. & Barresi, J. The role of second-person information in the development of social understanding. *Front. Psychol.* **8**, 1–15 (2017).
27. Wakefield, E., Novack, M. A., Congdon, E. L., Franconeri, S. & Goldin-Meadow, S. Gesture helps learners learn, but not merely by guiding their visual attention. *Dev. Sci.* **21**, 1–12 (2018).
28. Mundy, P. & Newell, L. Attention, joint attention & social cognition. *Soc. Cogn.* **16**, 269–274 (2007).
29. Kasari, C., Paparella, T., Freeman, S. & Jahromi, L. B. Language Outcome in Autism: Randomized Comparison of Joint Attention and Play Interventions. *J. Consult. Clin. Psychol.* **76**, 125–137 (2008).
30. Saito, D. N. *et al.* Stay tuned: Inter-individual neural synchronization during mutual gaze and joint attention. *Front. Integr. Neurosci.* **4**, 1–12 (2010).
31. Kawai, N. Attentional shift by eye gaze requires joint attention: Eye gaze cues are unique to shift attention. *Jpn. Psychol. Res.* **53**, 292–301 (2011).
32. Hoehl, S., Fairhurst, M. & Schirmer, A. Interactional synchrony: signals, mechanisms and benefits. *Soc. Cogn. Affect. Neurosci.* 5–18 (2020) doi:10.1093/scan/nsaa024.
33. Kuhl, P. K., Tsao, F. M. & Liu, H. M. Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9096–9101 (2003).
34. Lauricella, A. R., Gola, A. A. H. & Calvert, S. L. Toddlers' learning from socially meaningful video characters. *Media Psychol.* **14**, 216–232 (2011).
35. O'Doherty, K. *et al.* Third-Party Social Interaction and Word Learning From Video. *Child Dev.* **82**, 902–915 (2011).
36. Goldstein, M. H. & Schwade, J. A. Social feedback to infants' babbling facilitates rapid phonological learning. *Psychol. Sci.* **19**, 515–523 (2008).
37. Calvert, S. L., Strong, B. L., Jacobs, E. L. & Conger, E. E. Interaction and participation for Young Hispanic and Caucasian Girls' and Boys' learning of media content. *Media Psychol.* **9**, 431–445 (2007).
38. Davis, J., Crabb, S., Rogers, E., Zamora, J. & Khan, K. 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 (2008).
39. John, J. R., Priya, H., Hemasundar, A. & Rajendran, G. A Collation of Two Forms of

Demonstration : In-person Versus Video Demon- stration which Gains the Upper Hand ? **4**, 23–29 (2016).

40. Schmitz, M. Joint attention and understanding others. *Synth. Philos.* **29**, 235–251 (2014).
41. Hu, J. *et al.* 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 (2015).
42. Hamilton, A. F. de C. & Lind, F. Audience effects: what can they tell us about social neuroscience, theory of mind and autism? *Cult. Brain* **4**, 159–177 (2016).
43. Seuren, L. M., Wherton, J., Greenhalgh, T. & Shaw, S. E. Whose turn is it anyway ? Latency and the organization of turn-taking in video-mediated interaction. *J. Pragmat.* **172**, 63–78 (2021).
44. Rice, K. & Redcay, E. Interaction matters: A perceived social partner alters the neural processing of human speech. *Neuroimage* **129**, 480–488 (2016).
45. Dale, R., Fusaroli, R., Duran, N. D. & Richardson, D. C. *The Self-Organization of human interaction*. *Psychology of Learning and Motivation - Advances in Research and Theory* vol. 59 (Elsevier, 2013).
46. Dideriksen, C., Christiansen, M. H., Kristian, T., Dingemanse, M. & Fusaroli, R. Building common ground : Quantifying the interplay of mechanisms that promote understanding in conversations Building common ground : Quantifying the interplay of mechanisms that promote understanding in conversations. 1–156 (2020).
47. De Jaegher, H., Di Paolo, E. & Gallagher, S. Can social interaction constitute social cognition? *Trends Cogn. Sci.* **14**, 441–447 (2010).
48. Pan, X. & Hamilton, A. F. d. C. 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 (2018).
49. Guerin, B. Mere presence effects in humans: A review. *J. Exp. Soc. Psychol.* **22**, 38–77 (1986).
50. Lytle, S. R., Garcia-Sierra, A. & Kuhl, P. K. Two are better than one: Infant language learning from video improves in the presence of peers. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 9859–9866 (2018).
51. Skuballa, I. T., Xu, K. M. & Jarodzka, H. The impact of Co-actors on cognitive load: When the mere presence of others makes learning more difficult. *Comput. Human Behav.* **101**, 30–41 (2019).
52. Vigliocco, G. *et al.* 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 (2019).

53. Al-Rukban, M. O. Guidelines for the construction of multiple choice questions tests. *J. Family Community Med.* **13**, 125–33 (2006).