# Neural and Cognitive Mechanisms of Real-World Interaction during Adult Learning

Sara De Felice

Institute of Cognitive Neuroscience
University College London (UCL)

Thesis submitted to UCL for the Degree of Doctor in Philosophy

May 2023

I, Sara De Felice confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed: [redacted]

Date: 24th May 2023

# Abstract

The goal of this thesis is to understand what makes a social interaction successful, and whether it supports learning of conceptual knowledge. Crucially, it distinguishes learning *via* the social from learning *about* the social, and asks the question of how *social interaction* supports declarative processing of *non*-social material. In doing so, it priorities ecological validity: all experiments involve relatively unconstrained teacher-learner interaction, and learning material resembled documentary-like content. The first half of the thesis shows a series of studies on how adults learn in online contexts (Study 1 and 2): Study 1 presents two online experiments, where social contingency (i.e. being part of a live interaction vs observing a pre-recorded one) and social cues (i.e. teacher's webcam on vs off vs showing a slide only) were manipulated. Results showed that learning in live interaction was associated with the best performance, and live social interaction with a full view of the teacher provided the optimal setting for learning, while seeing a slide had greater benefit during recorded sessions specifically. Study 2 replicates the live-learning advantage across three experiments and a large sample of adults with Autistic Spectrum Condition (ASC). The second half of this thesis (Study 3 and 4) investigates face-to-face interaction, using functional Near-Infrared Spectroscopy (fNIRS) hyperscanning and wavelet transform coherence (WTC) analysis, to measure brain synchrony in naturalistic interactions. Study 3 tests the hypothesis that being in the same room and engaging in conversation affects people's brain response to later novel stimuli. Study 4 asks whether teacher-student brain synchrony can be a marker of learning success and, if so, how it is modulated by social behaviours. Findings reveal a complex dynamic between neural responses and behavioural metrics, in particular mutual gaze and joint attention. Results are discussed in the frame of the mutual-prediction hypothesis, and advocate for a multi-modal investigation of social learning to fully understand its underlying cognitive mechanisms. Overall, this work advances the current understanding of naturalistic social interaction and has theoretical implications for cognitive models of information exchange and mutual

prediction, as well as practical significance for educational policies. The novel multi-modal and highly ecological approach used in this thesis makes this work an important example for real-world second person social neuroscience.

# Impact Statement

People have been learning from and with each other since we have record of human activity. Despite changes in the modern educational system, it remains similar to its ancestral versions in one feature at its core: it is *social*. Yet, we know very little about the cognitive and neural mechanisms that support human learning in social interaction, especially with regards to real-world contexts. This thesis employed highly ecological designs and multimodal paradigms to answer questions about what social contexts better support acquisition of new knowledge, and how social interaction facilitates information transfer between teacher (sender) and learner (receiver). To investigate these questions, this work makes use of both online platforms as well as wearable functional near-infrared spectroscopy (fNIRS) to record brain activity from two people simultaneously as they interact.

This thesis reports four main findings. First, in online contexts, learning in live (contingent) social interaction is more effective than learning from pre-recorded videos. Second, social interaction supports learning in adults with Autistic Spectrum Condition (ASC) too, more than non-social teaching and as much as it does in neurotypicals. Third, face-to-face social interaction synchronises brain systems involved in shared-understanding and common-ground during co-watching of naturalistic movies, beyond what would be expected by watching the same movies alone. Forth, learner-teacher brain synchrony predicts learning performance. Crucially however this is not a linear dynamic, but it is modulated by behavioural mechanisms of joint attention and mutual gaze between teacher and learner.

These findings advance the current understanding of the neurocognitive mechanisms engaged as people learning from and with others. Specifically, they show that neurocognitive mechanisms involved in interactive learning are distinct from mechanisms involved in non-interactive learning. In addition, they support the notion that learning *via* social interaction is distinct from learning *about* the social world: social cognition may support information transfer between people even in the context of non-social knowledge. This distinction may be particularly relevant in our understanding of neuro-diverse populations including ASC, and inform both cognitive models and experimental practices in the study of educational neuroscience. Lastly, this work emphasise the need for more ecologically valid investigations, with long-term implications for educational policies.

## UCL Research Paper Declaration Form (1/4)

This form references the doctoral candidate’s own published work(s)

1. For a research manuscript that has already been published (if not yet published, please skip to section 2).
   a) What is the title of the manuscript?

Learning *from* others is good, *with* others is better: the role of social interaction in human acquisition of new knowledge

   b) Please include a link to or doi for the work

https://doi.org/10.1098/rstb.2021.0357

   c) Where was the work published?

Philosophical Transaction of the Royal Society B

   d) Who published the work? (e.g. OUP)

The Royal Society Publishing

   e) When was the work published?

26 December 2022

   f) List the manuscript’s authors in the order they appear on the publication

De Felice, S.; Hamilton, A.; Ponari, M.; Vigliocco G.

   g) Was the work peer reviewed?

Yes

   h) Have you retained the copyright?

Yes

i) Was an earlier form of the manuscript uploaded to a preprint server? (e.g. medRxiv). If 'Yes', please give a link or doi)

No

If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

☒ I acknowledge permission of the publisher named under **1d** to include in this thesis portions of the publication named as included in **1c**.

2. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)

a) What is the current title of the manuscript?

Click or tap here to enter text.

b) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv; if 'Yes', please give a link or doi)

Click or tap here to enter text.

c) Where is the work intended to be published? (e.g. journal names)

Click or tap here to enter text.

d) List the manuscript's authors in the intended authorship order

Click or tap here to enter text.

e) Stage of publication (e.g. in submission)

Click or tap here to enter text.

3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4)

S.D.F.: conceptualization, writing—original draft, writing—review and editing; A.F.C.H.: conceptualization, supervision, writing—review and editing; M.P.: conceptualization, writing—review and editing; G.V.: conceptualization, supervision, writing—review and editing.

4. In which chapter(s) of your thesis can this material be found?

Chapter 1 (general Introduction)

**e-Signatures confirming that the information above is accurate** (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

Candidate

Sara De Felice

Date:

19th April 2023

Supervisor/ Senior Author (where appropriate)

Antonia Hamilton

Gabriella Vigliocco

Marta Ponari

Date

19th April 2023

# UCL Research Paper Declaration Form (2/4)

This form references the doctoral candidate's own published work(s)

1. For a research manuscript that has already been published (if not yet published, please skip to section 2)
    a) What is the title of the manuscript?

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

    b) Please include a link to or doi for the work

https://doi.org/10.1016/j.cub.2021.08.045

    c) Where was the work published?

Current Biology

    d) Who published the work? (e.g. OUP)

CellPress

    e) When was the work published?

14 September 2021

    f) List the manuscript's authors in the order they appear on the publication

De Felice, S.; Vigliocco, G.; Hamilton, A.

    g) Was the work peer reviewed?

yes

    h) Have you retained the copyright?

Yes

i) Was an earlier form of the manuscript uploaded to a preprint server? (e.g. medRxiv). If ‘Yes’, please give a link or doi)

No

If ‘No’, please seek permission from the relevant publisher and check the box next to the below statement:

☒ I acknowledge permission of the publisher named under **1d** to include in this thesis portions of the publication named as included in **1c**.

2. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)

a) What is the current title of the manuscript?

Click or tap here to enter text.

b) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv; if ‘Yes’, please give a link or doi)

Click or tap here to enter text.

c) Where is the work intended to be published? (e.g. journal names)

Click or tap here to enter text.

d) List the manuscript’s authors in the intended authorship order

Click or tap here to enter text.

e) Stage of publication (e.g. in submission)

Click or tap here to enter text.

3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4)

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

4. In which chapter(s) of your thesis can this material be found?

Chapter 2

**e-Signatures confirming that the information above is accurate** (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

Candidate

Sara De Felice

Date:

19th April 2023

Supervisor/ Senior Author (where appropriate)

Antonia Hamilton

Gabriella Vigliocco

Date

19th April 2023

## UCL Research Paper Declaration Form (3/4)

This form references the doctoral candidate’s own published work(s)

5. For a research manuscript that has already been published (if not yet published, please skip to section 2)
   - j) What is the title of the manuscript?
   - k) Please include a link to or doi for the work
   - l) Where was the work published?
   - m) Who published the work? (e.g. OUP)
   - n) When was the work published?
   - o) List the manuscript’s authors in the order they appear on the publication
   - p) Was the work peer reviewed?
   - q) Have you retained the copyright?
   - r) Was an earlier form of the manuscript uploaded to a preprint server? (e.g. medRxiv). If ‘Yes’, please give a link or doi)

If ‘No’, please seek permission from the relevant publisher and check the box next to the below statement:

☐ I acknowledge permission of the publisher named under **1d** to include in this thesis portions of the publication named as included in **1c**.

6. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)

   - f) What is the current title of the manuscript?
     Adults with Autistic Spectrum Condition benefit and enjoy learning via social interaction as much as neurotypical adults do
   - g) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv; if ‘Yes’, please give a link or doi)

n/a

h) Where is the work intended to be published? (e.g. journal names)

Psychological Science, Molecular Autism

i) List the manuscript's authors in the intended authorship order

De Felice, S., Hatilova, A., Trojan, F., Tsui, I. & Hamilton, A.

j) Stage of publication (e.g. in submission)

In submission

7. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4)

   S.D.F. – Conceptualization, methodology, formal analysis, investigation, writing original draft, writing review and editing; A. Hatilova – data collection, writing review and editing; F.T. – data collection, writing review and editing; I.T. – data collection, writing review and editing; A. Hamilton – Conceptualization, methodology, writing – review & editing.

8. In which chapter(s) of your thesis can this material be found?

Chapter 3

**e-Signatures confirming that the information above is accurate** (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

Candidate

Sara De Felice

Date:

19th April 2023

Supervisor/ Senior Author (where appropriate)

Antonia Hamilton

Date

19th April 2023

# UCL Research Paper Declaration Form (4/4)

This form references the doctoral candidate’s own published work(s)

9. For a research manuscript that has already been published (if not yet published, please skip to section 2)
   s) What is the title of the manuscript?
   t) Please include a link to or doi for the work
   u) Where was the work published?
   v) Who published the work? (e.g. OUP)
   w) When was the work published?
   x) List the manuscript’s authors in the order they appear on the publication
   y) Was the work peer reviewed?
   z) Have you retained the copyright?
   aa) Was an earlier form of the manuscript uploaded to a preprint server? (e.g. medRxiv). If ‘Yes’, please give a link or doi)

If ‘No’, please seek permission from the relevant publisher and check the box next to the below statement:

☐ I acknowledge permission of the publisher named under **1d** to include in this thesis portions of the publication named as included in **1c**.

10. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)

    k) What is the current title of the manuscript?

Social interaction increases brain synchrony during co-watching of novel movies

   l) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv; if ‘Yes’, please give a link or doi)

10.31234/osf.io/ruhmc

m) Where is the work intended to be published? (e.g. journal names)

Potential journals include Brain Imaging and Behavior, Current Opinion in Behavioral Sciences, Cognitive Neuroscience

n) List the manuscript's authors in the intended authorship order

De Felice, S., Hakim, U., Gunasekara, N., Pinti, P., Tachtsidis, I., & Hamilton, A.

o) Stage of publication (e.g. in submission)

In submission

11. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4)

    S.D.F. – Conceptualization, methodology, data collection, formal analysis, investigation, writing original draft, writing review and editing; U.H. – data collection, writing review and editing; N.G. – data collection, writing review and editing; P.P. – methodology, writing review and editing; I.T. – methodology, writing review and editing; A.F.d.C.H. – Conceptualization, methodology, writing – review & editing.
12. In which chapter(s) of your thesis can this material be found?

Chapter 4

**e-Signatures confirming that the information above is accurate** (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

Candidate

Sara De Felice

Date:

19th April 2023

Supervisor/ Senior Author (where appropriate)

Antonia Hamilton

Date

19th April 2023

# Acknowledgments

My PhD journey taught me about social interaction in ways that go far beyond the scientific results published in this thesis. None of the work reported here could have been possible without the social network of people that, in one form or another, shared this journey with me.

My deepest thanks go to my primary supervisor, Prof Antonia Hamilton, for making this PhD project an incredibly positive experience. I always felt better after every meeting with her, and learned an enormous amount of skills, techniques, and insightful ideas about social neuroscience from her (and *with* her, see experimental chapters! ☺).

I am immensely grateful to my other supervisor, Prof Gabriella Vigliocco, for believing in me when I interviewed for the Ecological Study of the Brain DTP program over four years ago, and for continuing to support me ever since, teaching me about science with wisdom, intelligence, and empathy.

I want to thank both Antonia and Gabriella, for always making me feel valued as a scientist and cared for as a person, and for always having a positive word on both sunny and rainy days. The most important lesson they both taught me has been that of leading an environment where scientific rigour, care, hard work, compassion and fun all co-exist in a perfect equilibrium. Growing next to these two exceptional scientists got me closer to the type of scientist I want to be. I feel lucky to have had the privilege to work with them and will foster their mentorship for the years to come.

My sincere thanks also go to past and present lab colleagues, collaborators, and mentors who, throughout the years, have improved the quality of this thesis and my understanding of many aspects of this work. In particular, I want to thank Isla Jones, Paula Wicher, Dwaynica Greaves, Ruihan Wu, Gal Rozic, Dr Roser Canigueral, Dr Paola Pinti, Uzair Hakim, Natalie Gunasekara, Prof Ilias Taschtsidis and Prof Joy Hirsch. Special thanks to the whole Ecological Brain team, made of outstanding academics, who helped me understand the value of real-world neuroscience. I am also grateful to the ICN community, IT team and staff, who create a unique work environment that anyone would be lucky to be part of.

Thanks also to the Leverhulme Trust for funding my PhD, and to the EPS and Bogue Fellowship for supporting my research visit at New York University (NYU).

A special thought goes to London, for being a home to me and to so many amazing people whom I feel lucky to call my friends. Thanks to them and also to those friends far from London and yet so close to my heart, for making these last few years so memorable. There is not space here to mention you all, but you know how much I love you.

Thanks also to my family for supporting me since always, for giving me the life opportunities that got me here today. It is a privilege to come from a loving and smart home, and I am grateful for that. Thanks to Andrea, who started and ended this journey with me, and for being together even when we were not. I couldn't ask for a more fun, loving, smart and supportive partner.

Last but not least, I want to thank my mum: I realised that despite I spent four solid years studying this subject, she had already taught me everything I needed to know about social interaction – or at least everything that truly matters. This thesis is dedicated to her. Wherever you are, I hope you are proud.

# Table of Contents

# Table of Figures

# Table of Tables

# Abbreviations

ASC = Autistic Spectrum Condition

dlPFC = dorso-lateral Pre-Frontal Cortex

EEG = Electro-Encephalography

fMRI = functional Magnetic Resonance Imaging

fNIRS = functional Infra-Red Spectroscopy

ISC = Inter-Subject Correlation

JA = Joint Attention

MG = Mutual Gaze

NT = Neuro-Typical

SPL = Superior Parietal Lobe

tACS = transcranial Alternating Current Stimulation

TPJ = Temporo-Parietal Junction

vPMC = ventral Pre-Motor Cortex

# 1. Chapter 1

This chapter has been adapted from the opinion piece published on 26th December 2022, full reference is:

De Felice, S.; Hamilton, A.; Ponari, M.; Vigliocco, G. (2022) Learning from others is good, with others is better: the role of social interaction in human acquisition of new knowledge *Phil. Trans. R. Soc.* http://doi.org/10.1098/rstb.2021.0357

## Introduction

Communities where experts would transfer knowledge to novices have existed in one form or another since we have records of human activity. Despite changes in the modern educational system, it remains similar to its ancestral versions in one feature at its core: it is social. Yet, we know very little about the cognitive mechanisms that support human learning in social interaction. Throughout our life, we acquire new information and form new conceptual representations largely in social contexts: for example, babies learn from their caregivers at home, pupils learn from teachers at school and by sharing their experiences with other students. In the same way, adult learning typically occurs in social contexts and in relation to peers, colleagues at work and/or mentors. Researchers in anthropology and sociology (e.g., Schegloff, 2007; Sacks, Schegloff & Jefferson 1974), as well as in developmental psychology (e.g., Bruner, 1957, 1978; Nomikou et al., 2016; Rohfling et al., 2016; Vygotsky, 1978) have emphasised in their work the importance of social interaction for learning and for development. However, cognitive psychology and neuroscience has traditionally studied cognition at the individual level.

The 'single-brain' approach (Gazzaniga, Ivry & Mangun, 2002) studies brain and cognition using experimental designs involving a sample of participants (children or adults) completing a given task individually, and then makes inferences about how the brain works more generally.  It is only in the last decade, that cognitive neuroscientists have begun to move to a 'second-person neuroscience' approach (Redcay & Schilbach, 2019) which studies cognitive processes in interaction, including the back-and-forth dynamics between two or more people. This thesis applies these ideas to the case of learning new concepts and knowledge.

I first set out the theoretical framework for this thesis, providing some definitions and introducing some ideas relevant to the study of human learning and for the experimental chapters (Chapter 2, 3, 4 and 5). I then move to review the evidence in children and adults, showing what we know so far about how learning benefits from social interaction across the lifespan. I also present neuroimaging studies to identify the neural signature of social interactive learning. After the literature review, I identify the possible cognitive mechanisms subserving interactive learning. I then highlight some methodological and theoretical issues that emerged from the literature review. In doing so, I describe how my work attempts to deal with some of these issues and present the rationale for the four experiments described in the following chapters.

## 1.1 Theories and definitions

The goal of this section is to introduce key concepts and set some definitions which would be useful throughout the thesis. As this work looks at both behavioural/cognitive *and* neural mechanisms of social learning, this section provides some relevant definitions on these two aspects respectively.

### 1.1.1 *Theories and definitions of social learning*

Human learning refers to any form of acquisition of new knowledge and skills by an individual. One can learn new information alone, e.g. memorising events via reading a history book. However, often learning occurs with and from other people. When such learning occurs via transmission of information across members of a social group, it is defined as *social learning* (Tomasello, 2004). Importantly, there are many ways in which learning can be social, depending on the role that the social agent(s) has in the learning process of a given individual. Therefore, the term social learning is a broad term that refers to any form of learning – such as motor, verbal and knowledge-based learning – via any form of social context, including observation of others (Cross, 2011; Cross et al., 2009), imitation (Bandura, 2019) and interactive learning (Figure 1.1).

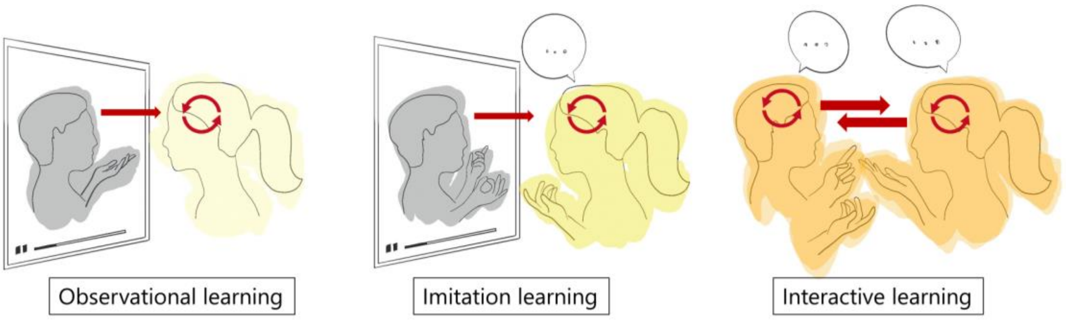


**Figure 1.1 Schematic of three types of social learning.**

Arrows indicate information flow between teacher and learner in three types of social learning. From left to right: Observational learning. Learner attends to information that flows from teacher to learner. Imitation learning. Learner attends to information that flows from teacher to learner and repeats/imitates the teacher. Interactive learning. Teacher and learner engage in social interaction and exchange reciprocal social signals. Importantly, information flows back and forth from teacher to learner.

Observational learning refers to learning via attending to someone else's actions and/or listening to them delivering information. Imitation refers to copying someone else (Bandura, 2019). Observational learning and imitation differ as observational learning only requires attention to the teacher, without immediate replication of their actions/words, while imitation learning involves observation plus active performance. Both of these can arise when a person watches a video of another person's actions or speech, with no interaction between the watcher and the video, and thus both involve the one-way transmission of information from teacher to student without interaction, where the learner is confined to the role of a receiver. In contrast, in interactive learning, both the teacher and the student are concurrently engaged in the learning process and they can both take full turns during the interaction. While interactive learning can vary in terms of how interactive any given contexts is (Rogoff, 1998; Rogoff et al., 2007), with variations even within single episodes (e.g. sessions), by definition any given interaction draws in all agents as contributors: all can participate in the interaction in both explicit (e.g. verbal feedback) and implicit (e.g. body language) forms.

Within cognitive models, interactive learning is defined by a two-way exchange of signals that includes subtle but critical reciprocity from student to teacher (Figure 1). These signals could indicate understanding (or lack thereof) as well as attentiveness (or inattentiveness) and thus allow the teacher to tune their lesson to the student. Here, I refer to reciprocity as any reaction fed-back during an real-time exchange which would inform the interlocutor(s) about the quality of the exchange (e.g. nodding for understanding, frowning for confusion etc), and thus possibly allow for a (re)direction of behaviour(s), as well as opportunities for the learner to elaborate what is being discussed. Thus, learning in interaction requires mutual feedback between a student (or learner, who is acquiring new knowledge) and a teacher (who is providing new information). Importantly, independently of the type of social context, for (social) learning to occur there must be an enduring change in the learner's action and/or knowledge as a consequence of either observing, imitating or interacting with others (Ramsey et al., 2021).

Such categories have been developed over the last decades especially in the context of action/motor learning, and their applicability to knowledge-based learning may be less obvious, although still useful to draw some conceptual distinctions. I employ these categories here to draw the distinction between learning *from* others (imitation and observational learning) and learning *with* others (interactive learning). There are only a few studies which directly contrast different types of learning. Examples include comparisons of physical or observational learning (Cross et al., 2017), and comparisons of sequence learning from imitation or verbal instruction (Renner et al., 2018) or of observational versus interactive learning (Matheson et al., 2013). Pioneering work from developmental psychologist Vygotsky (1962,1978) had long argued for a key role for the environment, especially the social environment in learning and development. His sociocultural theory of cognitive development views conceptual learning as an intrinsic social process. A number of other researchers in developmental psychology has also emphasised the importance of social interaction in cognitive and linguistic development (e.g., Rohfling et al., 2020).

### 1.1.2 *Theories and definitions of brain-to-brain dynamics*

The study of brain-to-brain dynamics – i.e. the relationship across multiple brains' neural responses – has received growing interest over the last few decades. This section will introduce some of the ideas emerging from this literature and define some terms to describe different aspects of brain-to-brain dynamics. In particular, it will consider two definitions, namely *inter-subject correlation* and *brain synchrony* (see figure 1.2), in relation to four main questions: a) what is the temporal relationship between multiple signals?; b) what does this relationship reflect?; c) when are multiple brain responses measured?; and d) how is the relationship quantified? (Table 1.1)

**Table 1.1 Characteristics of inter-subject correlation and brain-to-brain synchrony to study inter-brain dynamics. Also see Figure 1.2. a. Note that this list is not meant to be exhaustive and here we include just a few examples of the most popular metrics used in the literature.**

| | Inter-Subject Correlation | Brain-to-Brain Synchrony |
|---|---|---|
| **Question a:** what is the temporal relationship between multiple signals? | Symmetric mirroring: how two or more signals align or dis-align over time (i.e. A=B) | Dynamic coupling: how two or more signals co-vary in frequency and time (i.e. A⇔B) |
| **Question b**: what does the relationship between multiple signals believed to reflect? | Similarity in individuals' cognitive processing of either external stimuli or internal mental states | Complex social interaction, signal exchange and communication |
| **Question c:** what are the temporal characteristics of data collection, i.e. when is brain response of | Separate sessions | Simultaneously (with physical co-presence) |

| | | |
|---|---|---|
| participant A measured in relation to participant B? | | |
| **Question d:** what measure is used to quantify the brain-to-brain relationship? [a] | Pearson’s correlation | Wavelet Coherence Transform;<br>Circular correlation;<br>Granger causality |

*Inter-subject correlation* (ISC) quantifies the consistency of stimulus-driven responses among individual brains. In other words, it looks at how *aligned* (or *dis-aligned*) two or more brains are in response to different stimuli and/or environments. This has been first studied by Hasson et al. (2004) in a pivotal functional Magnetic-Resonance Imaging (fMRI) experiment, where five different individuals showed similar neural responses in occipital, parietal and temporal areas during free watching of a movie. Crucially, each individual brain was scanned individually. Since Hasson et al.’s study, a number of other investigations replicated the findings of ISC during natural viewing (Jääskeläinen et al., 2008; Lahnakoski et al., 2014; Parkinson et al., 2018), and extended them to interpretation of narratives (M. Nguyen et al., 2019; Yeshurun et al., 2017), speech comprehension (Wilson et al., 2008) and even reflection on socio-political issues (Dieffenbach et al., 2020).

Recently, Madsen and Parra (2022) cleverly demonstrated that ISC is the result of effective cognitive processing: they presented participants with informative videos in an attentive and distracted condition, while measuring their neural activity via EEG, as well as heart rate, gaze position, pupil size, breathing and head movement. ISC (i.e. *between*-subjects) emerged only for those signals that exhibit a robust *within*-subject brain–body connection. Within-subject brain-body connection gives a measure of how engaged an individual is to a given experience. By showing that ISC was specifically present in those instances when there was a high within-subject brain-body connection, the authors demonstrated that ISC is linked to the degree of individual participant’s attunement to the stimulus. Consistent with this interpretation, the strength of correlation changed with attentional state and predicted subsequent recall of information presented in the videos.

Therefore, ISC or neural alignment reflect the temporal similarity between two (or more) brain signals (question a, Table 1.1), resulting from common cognitive processing of an external stimulus (question b). It is obtained by scanning single brains one at a time (question c) and is usually quantified via correlation methods (question d). Importantly, neural alignment reveals the similarity of different brains to how we process the world around us, and although it has been shown to be modulated by inter-personal dynamics (e.g. Parkinson et al., 2018), it does not tell us anything about real-time interactive minds.

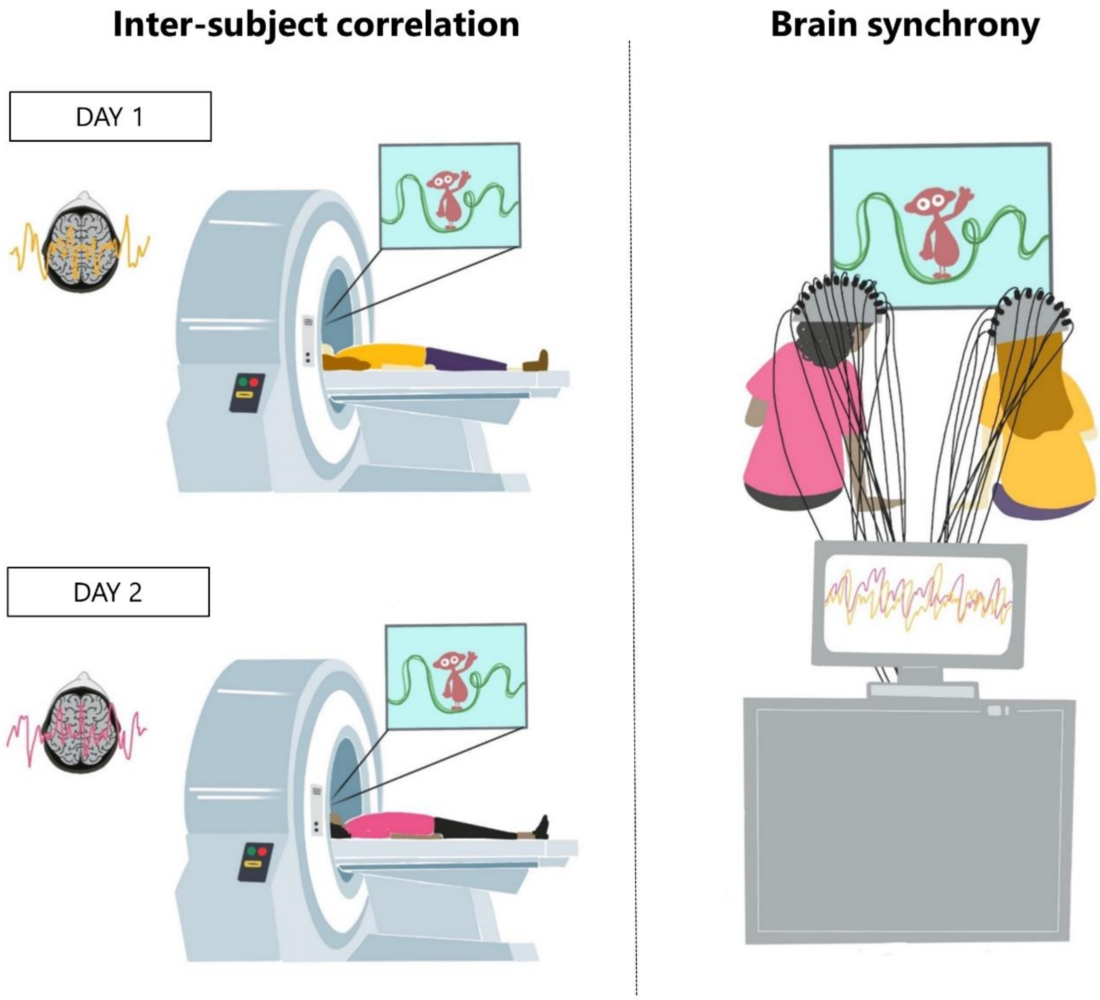


**Figure 1.2 Inter-Subject Correlation vs Brain Synchrony**

*Left:* on day 1, one participant is watching a cartoon while receiving a brain scan. On day 2, another participant is undergoing the same procedure, watching the same cartoon and receiving a brain scan. The neural responses from both participants are than compared. The similarity across neural responses to the same stimulus (e.g. the cartoon) collected from individual brains and different sessions is computed as inter-subject correlation. *Right:* On the same day, two participants watch a cartoon together, in real time, next to each other. Their brains response is recorded simultaneously, and the coherence between the two signals is computed as brain-to-brain synchrony. Also see Table 1.1.

In contrast, *brain-to-brain coherence* (or *synchrony*) reflects the degree of real-time co-variance across multiple interactive brains. As interactions are complex, signals from multiple interactive brains may not necessarily 'mirror' each other, but would instead 'couple' to each other dynamically over time (Hasson & Frith, 2016). The first study to measure neural activity from multiple interacting brains simultaneously has been Cui et al. (2011). In their pioneering hyperscanning functional near-infrared spectroscopy (fNIRS, see section 1.4.3 for a description of this method) study, they measured brain activity from two people as they engaged in either a cooperative or competitive task. Findings revealed that while an individual time-series analysis did not show task-specific neural patterns, an inter-brain coherence analysis did, with increased coherence emerging specifically during cooperative task blocks. This study was the first to provide evidence that collecting and analysing data from multiple interactive people can be more informative than single-person approaches, particularly in the study of social cognition. From this first pioneer study, many more have explored the potential of hyperscanning to shed light on real-time social dynamics in a variety of contexts (Babiloni & Astolfi, 2014a; Cañigueral et al., 2021b; Hirsch et al., 2018; Konvalinka et al., 2023; Lomoriello et al., 2022; T. Nguyen et al., 2020; Y. Pan et al., 2020; hyperscanning literature on social cognition is reviewed in details in section 1.2.3 and discussed in terms of social learning in section 1.3.3).

Therefore, brain-to-brain coherence or synchrony gives a measure of complex dynamic coupling, beyond simple alignment (question a, Table 1.1). This is believed to reflect real-time social dynamics (question b) and is achieved by measuring multiple brains simultaneously via hyperscanning (question c). As it reflects a complex signals dynamic, analysis tools used to compute brain synchrony include methods that can encompass not just temporal information about the signal, but also frequency components (e.g. wavelet transform coherence analysis, Grinsted et al., 2004) and directionality of the relationship (e.g. Granger causality, Granger, 1969) (question d).

The distinction between *neural alignment* (or ISC) and *brain-to-brain coherence* is relevant as it has implications for building cognitive models of how people process reality and interact with others (see Figure 1.2). In fact, neural alignment gives information on how similarly people's brains respond in a given context, or in other words how closely different brains experience and interpret the world. As such, it still involves the study of cognition within a *single-brain* framework. Instead, brain-to-brain coherence goes beyond similarities across *solo* brains, to grasp complex dynamics *between interactive brains*, as they continuously and mutually adapt over their interaction.

The shift towards a second-person neuroscience (Redcay & Schilbach, 2019) allows theories to go beyond single-brain models and consider the mutual-influence between different social agents. The *embodied mutual prediction framework* (Hamilton, 2020; Kingsbury et al., 2019) arises from such an experimental shift, and make clear predictions about complex social dynamics. Importantly, despite the fact that brain synchrony is more informative than ISC about social dynamics, it cannot exhaustively explain the mechanisms underlying a successful (or unsuccessful) social interaction. The mutual-prediction hypothesis explicitly acknowledges that interactive brains exist *within* interactive bodies and, in line with this premise, interprets brain synchrony as the result of auditory, visual and motor signals exchange between social agents. Therefore, interrogating inter-brain dynamics under different contexts and distinguishing between ISC and brain synchrony is valuable as it consequently leads to both experimental and theoretical renovation.

It is worth mentioning at this point that ISC and brain synchrony are not completely separate concepts and some overlaps between the two exist, in that brain synchrony *includes and extend from* ISC. In fact, when two (or more) people interact, a number of elements are processed by all parts involved which are not directly linked to any aspects of the interaction per se (e.g. the visual elements of the room, the sounds, the topic of the conversation etc). This would inevitably drive some degree of ISC. Therefore, brain synchrony would always include signal reflecting similarity in the way people process the world, while also including mutual-predictions about interactive agents.

Note that other terms have also been used in the literature to describe various forms of brain-to-brain dynamics, including brainwave entrainment, neural homophily and possibly more. However, for the sake of clarity, here we limit our terminology to *ISC/neural alignment* and *brain-to-brain coherence/synchrony* specifically within the definitions provided in this section (also see Figure 1.2). Any other terms is avoided to minimise confusion.

## 1.2 Review of the evidence

This section reports evidence from studies on children and adults to illustrate what we know so far about how humans learn in social contexts, and also reviews neuroimaging studies of interactive brains. As outlined in section 1.1.1, the primary focus of this work is studying the mechanisms of learning *via social* interaction *about non-social* conceptual knowledge (i.e. long-term memory for facts and concepts). Review of the evidence and discussion of the literature will however extend to other forms (e.g., single words, motor learning) when relevant.

### 1.2.1 *Social interaction for children learning*

Social interaction is crucial for optimal cognitive and brain development (Goswami, 2006; Kuhl, 2007; Meltzoff et al., 2009). This statement is relatively uncontroversial and is supported by a large body of literature (for the latest systematic review on the topic see Ilyka et al., 2021). Taking together the results from the 55 studies included in their review, Ilyka and colleagues concluded that an optimal development of cognitive functions – as measured via heterogeneous neuropsychological test batteries – and brain – as measured via structural and functional analysis of selected regions and networks – is contingent on child-caregiver interaction during the child's first years of life. Out of 55 relevant publications included in their systematic review, only six looked at both child and caregiver and how the dyadic interaction impacted cognitive and brain development (Beckwith & Parmelee, 2016; Elsabbagh et al., 2012; Gartstein et al., 2020; Jones et al., 2004; Perone & Gartstein, 2019; Pratt et al., 2019). Results coming from such an approach point to the importance of 'sensitivity' and 'reciprocity' of both agents for optimal child development, and in turn at the quality of the overall interaction to support cognition later in life.

Considering interactive learning specifically, the majority of work on children comes from studies on language acquisition: these robustly and consistently show that social interaction is a critical and constraining factor for successful language development (Kuhl, 2007). In a pioneering study, Kuhl et al. (2003) trained 9-month old American babies to distinguish Chinese Mandarin sounds in three different conditions: in interaction with a native speaker, or by exposure to either videos or sound recordings from the same native speaker. Despite equivalent exposure time and content of Chinese sounds, only the group who engaged in live-interaction with the teacher showed learning, and being exposed to videos or sound recordings was associated with no learning. While this study is not strictly looking at knowledge-based learning (e.g. new words/concepts), it provides strong evidence for the crucial role of interaction in children's learning over non-interactive learning methods.

More work on word acquisition during child-caregiver interaction has been conducted by Yu and Smith (Yu et al., 2017; Yu & Smith, 2012a, 2013, 2016). In their experimental paradigm, the infant and the caregiver engage in a series of free-play sessions during which they manipulate and name various objects (toys), while both wear head-movement sensors and eye-trackers. Crucially, in all their studies, the parent (teacher) is not aware that their infant's learning of the objects' names will be tested after the free-play session. This ensures that child-caregiver interactions are as natural as possible. By conducting a series of dyadic analyses, Yu and Smith (2012) showed that 18-month old infants were more likely to successfully learn objects' names if two things happened concurrently: 1) the infant (learner) held the object closer so that it was visually dominant within their visual field (over other competitor objects on the play table), and 2) the caregiver (teacher) named the object. Overall, these papers demonstrate the importance of social interaction in young children's word-learning.

The critical role of social interaction for optimal language development is relatively unsurprising, considering how heavily human language relies on the 'social brain' and vice versa (Blakemore, 2010; Kuhl, 2003, 2007). Also, in many developmental pathologies such as autism spectrum disorder, social cognition deficits and communicative disorders are co-occurring (Happé, 1995; Whitehouse et al., 2007). Given the highly-interconnected nature of social cognition and language processing (Binney & Ramsey, 2020), learning language within a social context would be expected to be beneficial.

Because of the strong relationship between communication and sociality, one may argue that the social-interaction advantage is limited to language development. However, the beneficial effect of social interaction during development is not limited to language. Evidence from a variety of studies show that social interaction supports learning more generally across different domains, including visuospatial categorization (Lauricella et al., 2011), procedural learning (Sauppé & Mutlu, 2014) and mathematical reasoning (Kostyrka-Allchorne et al., 2019a).

In their study, Kostyrka-Allchorne and colleagues found that in a large group of five-year-olds (n=215) the physical presence of a teacher (versus having  the teacher on screen) was associated with the highest learning, independent of whether children were observing the teacher playing with a shape or they were playing with it themselves. This study is a great example of why student-teacher interaction should be studied as interdependent: it is certainly unlikely that the mere presence of the teacher had something ‘magical’ about it so that the child learned more when the teacher was physically there. Equally, audience effects – defined as “a change in behaviour caused by being observed by another person” (Hamilton & Lind, 2016, pp. 160) – cannot explain the results, as students were explicitly observed in the on-screen condition too. Rather, there may well be something within the dynamic of learner-teacher interaction as it occurs face-to-face that positively impacted learning.

For example, Marsh, Ropar & Hamilton (2013) found that children showed more over-imitation – i.e. unnecessarily copying actions of others – when an adult was demonstrating goal-oriented actions, compared to when demonstration was presented through a recorded video. This suggests that social factors may directly give rise to different behaviours (e.g. overimitation) during learning with others, and this may increase with age as people think more about social norms (Clay, Over & Tennie, 2018, see below for a discussion of the mechanisms subserving interactive learning).

### 1.2.2 *Social interaction for adult learning*

Studies on interactive learning are predominantly focusing on children and young people as they are considered as the typical learners. Little is known about social learning in adults, and even less in interactive learning specifically. There is some evidence suggesting that, similar to what is found in children, social interaction acts as a catalyst for learning in adults too (Verga & Kotz, 2013). Again similarly to the literature on children, the majority of studies on interactive learning in adults has considered the domain of language, the focus being on second language acquisition (Jeong et al., 2010, 2021; Li & Jeong, 2020; Verga & Kotz, 2013).

In their study, Jeong and colleagues (2021) asked 36 Japanese adults to learn two sets of unknown Korean spoken words via either translation or videos depicting social situations. Words encoded from the social condition showed significantly higher accuracy rates and faster reaction times (RTs) than words encoded from the translation condition, and the social-video learning condition was also associated with higher activity in the right temporal parietal junction, right hippocampus, and motor areas, as measured with functional magnetic resonance imaging (fMRI). In their review, Verga and Kotz (2013) reported evidence for the importance of sociality in adult learning: specifically, learning a second language in interaction with another person significantly improves long-term retention of new vocabulary. Noteworthy, language is tightly linked to social cognition (Garrod & Pickering, 2004), and as such it is relatively unsurprising that language learning would benefit when occurring within social contexts.

Considering domains other than language, studies on adults have mainly looked at the impact of social interaction in comparison to online/virtual learning environments. Results are less conclusive: the majority of studies found no difference in learning outcomes between teaching live versus teaching through recorded videos (Brockfeld et al., 2018; Davis et al., 2008; Phillips, 2015; Schreiber et al., 2010; Solomon et al., 2004; Vaccani et al., 2016). In their intervention study, Brokfeld et al. (2018), divided 296 medical students into four groups, three of which received 41 four-hour lessons live, while the last group watched videos of the same lessons. The group assigned to the video condition changed daily, so that all students saw both live and video lectures. The effectiveness of the teaching method was evaluated by looking at students' performance on 301 multiple-choice questions of the medical exam. Similar approaches were adopted by the other studies cited here, and all found that learning performance did not differ across teaching methods. Despite no difference in objective performance, all these studies found that there were some differences in subjective evaluation, with the majority of students preferring live lessons. However, these studies did not control for exposure time: recorded material could be replayed multiple times while the live session was only live once. Studies that controlled for content and exposure time across conditions, found a significant improvement in the learning of medical students during social interactive lectures compared to recorded tutorials (John et al., 2016; Ramlogan et al., 2013).

Overall, studies of interactive learning in adults yield similar results to those in children. However, scholars are less unified on the notion that social factors matter in adult learning, possibly because adult learning – specifically of new concepts and information – is generally under-studied compared to children learning, and social factors in adulthood may be less critical than during development. Also, crucially, the studies reviewed above did not directly control the social factor during learning.

In the next section, I review neuroimaging studies to describe the neural signature of social interactive learning.

### 1.2.3 *The neural signature of interactive learning*

Hyperscanning has become increasingly popular over the last decade, as it has the advantage of measuring brain activity from more than one individual at the same time, meaning the social brain can be studied while people engage in social interaction rather than in isolation (Babiloni & Astolfi, 2014; Czeszumski et al., 2020; Dumas et al., 2011; Kelsen et al., 2020a).

In a five-person electroencephalography (EEG) hyperscanning study, Davidesco et al. (2019) simultaneously measured brain activity from four students and their teacher during a science class. They found that alpha-band (8-12Hz) brain-to-brain coherence, but not intra-brain alpha synchrony (i.e. within individuals), significantly predicted students' learning, as measured via performance in an immediate and a delayed test a week after the class. Moreover, moment-to-moment variation in alpha-band brain-to-brain coherence during the class specifically predicted what information was retained by the students a week later. Alpha frequency band is a well-established neural index of attention (Klimesch et al., 2007), which suggests that learning was better predicted by moments when student was attuned (or paying attention) to the teacher, and concurrently the teacher was attuned (or paying attention) to the students.

The same research group conducted another EEG hyperscanning study where brain activity was recorded from 12 high school seniors simultaneously over a semester (Dikker et al., 2017). Recording took place during students' regular biology class and was repeated over 11 sessions. Results showed that the degree to which brain activity was synchronized across students predicted student class engagement (quantified as student appreciation ratings of different teaching styles and student daily self-reported focus). In particular, they conducted a group-based neural coherence analysis to link student-to-group brain synchrony to different predictors. They found that student focus predicted student-to-group synchrony above and beyond teaching style, and also students who were more focused on a given day showed higher synchrony for that day.

Given the association between learner-to-group synchrony and class engagement (Dikker et al., 2017), and the link between engagement, attentional processes and learning (Nissen & Bullemer, 1987a), Dikker's group extended their work to ask whether learner-to-group or learner-to-teacher neural synchrony predicts learner's content retention (Bevilacqua et al., 2019). Using a similar real-world classroom scenario, biology class materials were presented in either videos or live lectures, and students completed a multiple-choice quiz after each class. Results showed that brain-to-brain synchrony was higher for video than for live lectures (as expected by greater similarity in low-level processes during watching of the same video content). However, for live lectures only, social closeness to the teacher was related to learner-teacher brain synchrony: in other words, when there was a contingent learner-teacher interaction, this was reflected in their brain activity. In addition, learning performance correlated with learner–teacher closeness, but not with learner-teacher brain synchrony.

Using functional near-infrared spectroscopy (fNIRS), Holper et al (2013) recorded prefrontal brain activity during the Socratic dialog simultaneously in seventeen teacher–student pairs. The Socratic dialog is a classical teaching model where the teacher encourages learning by interrogating the students via a set of structured questions. They found that learning - as measured by students' correct responses – was associated with higher correlation of student-teacher brain activity.

Similar findings were obtained by another group who also used fNIRS to measure brain activity from learner and teacher dyads during the acquisition of a music song (Pan et al., 2018). They found that brain activity in the bilateral Inferior Frontal Cortex showed learner-teacher synchronization. This was specifically associated with moments when the learner was observing the teacher and when learning was more interactive (measured in terms of turn-taking). Importantly, learner-teacher brain synchronization could predict student's performance on the learned song. The same research group conducted a further study to investigate the causal role of such synchronization in learning (Pan et al., 2020). They used transcranial alternating current stimulation (tACS) to induce (or disrupt) brain synchrony in different conditions and found that induced teacher-learner neural coupling facilitated motor coordination, which in turn was associated with enhanced learning of novel Chinese songs. This intriguing work hints at many further interesting questions, and it will be useful to see it replicated and extended.

Overall, these studies demonstrate that brain-to-brain synchrony can be measured during interactive learning and may correlate with learning performance either across sessions (Dikker et al., 2017) or even across individual events (Davidesco et al., 2019; Pan et al., 2018). However, the presence of a correlation does not necessarily reveal the causal mechanism behind the effect, and I consider possible cognitive processes in the next section.

## 1.3 Learning *from* and *with* others: Cognitive mechanisms of interactive learning

In the previous sections I have reviewed evidence showing that social interaction plays a key role in human learning across the life span and in a variety of cognitive domains, and also that it has a distinctive neural signature in the brain. It remains unclear, however, what social and cognitive mechanisms enhance learning in an interactive context. In this section, we consider some of the possible mechanisms which have been suggested and how these might be studied.

A number of cognitive mechanisms have been proposed to account for the advantage of interactive learning over non-interactive learning, including stimulus saliency (End & Gamer, 2017), social arousal (Berger, 2011), internal motivation (Evans & Boucher, 2015), sustained attention (Yu et al., 2017), audience effects (Hamilton & Lind, 2016), eye-contact and gaze (Ho et al., 2015), joint attention (Mundy & Newell, 2007), common ground (Bohn et al., 2019), attunement and shared intentionality (Sabbagh & Baldwin, 2001) and mutual predictions within inter-agents dynamics (Kingsbury et al., 2019). These can be distinguished on the basis of whether they describe effects within one individual alone (e.g. the learner) or whether they describe the learner-teacher relational dynamic (Brown & Brüne, 2012). Here, we discuss these systematically and evaluate them in relation to interactive learning.

### 1.3.1 *Individual-based mechanisms: learner and teacher-based approaches*

Individual-based mechanisms include stimulus saliency (i.e. how obvious or prominent a stimulus is in someone's environment), social arousal (i.e. a state of excitement or energy cost in relation to the physical intensity of a social situation), internal motivation (i.e. the individual willingness to engage in a situation or perform action), and sustained attention (i.e. a focused state maintained on a stimulus over an extended period of time). Stimulus saliency and social arousal have been proposed as possible explanations for the social learning advantage on the basis of the well-established effects that faces (and social stimuli more generally) are processed differently than other type of non-social stimuli (Bentin et al., 1981), and that we get more aroused in social than non-social situations (Cacioppo et al., 1996). In other words, according to these accounts, interactive learning is not ‘special’ because it is social per se, rather because social contexts share some features that make encoding of information somehow more memorable for future recalls (Jeong et al., 2021). In line with this, it has been found that distinct neural patterns of activation are associated with encoding and retrieving information learned in social contexts (Jeong et al., 2010, 2021).

In addition to external bottom-up influences, the internal motivation of the learner may be fundamental to direct sustained attention, which in turn is an essential pre-requisite of learning (Nissen & Bullemer, 1987b). There is no doubt that engaging with the learning material, by attending and processing the target information, is a strong predictor of how well we may be doing on a follow-up test. Yu et al (2017) demonstrated that this may be a very early mechanism that we engage in from a young age. They found that, in 9-month old infants, sustained attention predicted the learning of new vocabulary above and beyond joint attention between infant and their caregiver.

However, these factors seem to be telling only part of the story, and specifically the part concerning the learner. For example, Kostyrka-Allchorne and colleagues (2019) found that 5-year olds learned about atypical geometric shapes better when there was a teacher physically present in the room with them. The observed learning benefit may well reflect some degree of arousal given by the physical presence of the teacher. However, it cannot be excluded that the presence of the teacher improved learning via mechanisms of relational dynamics (Schertz et al., 2013). We simply cannot exclude either option, due to the way the study was designed and the fact that the focus of the analysis was limited to the learner.

In fact, the role played by the teacher is crucial in determining the learning outcomes, and yet it is often underrepresented in the learning literature. Social communicative signals (both verbal and non-verbal e.g. pointing, eye-gaze) are overtly employed by an expert (the teacher) to transfer information to a novice (the learner, Csibra & Gergely, 2009, 2011). Teacher's communication is functional to achieve successful teaching (and in turn someone else's learning), and as such is explicitly adjusted to maintain the learner's attention and assist information transfer. Teacher's communicative actions are therefore the other fundamental aspects to consider in the study of human learning.

The fact that teachers can adjust their verbal and non-verbal behaviour to assist the learner has been demonstrated in the case of both children and adult learners. For example, in their study, Brand et al. (2002) showed that caregivers deliberately modified both their language and their action when sharing information about novel object properties to infant compared to adults (who presumably were not novice to those objects). Similarly, Vigliocco et al. (2019) further showed that caregivers adapted their language and their actions when presenting unknown vs known objects to their 2-3 years old children. Similar modification of action with pedagogical intentions has been demonstrated in adults (McEllin, Knoblich and Sebanz, 2018). In three experiments, McEllin et al. recorded movements of participants playing simple xylophone melodies either alone, for a learner watching them, or together with another participant. They found that movement velocity was altered specifically in the condition when participants were playing to demonstrate a musical sequence to a novice, compared to when they were playing alone or with someone else who was expert in the melody.

This literature demonstrates the importance of considering the teacher as well as the learner when studying how humans learn in interaction. However, looking at one or the other may not be enough. Studies that only consider one side of the interaction may overlook the social dynamic unfolding during interpersonal communication, and in turn reach partial and/or inaccurate conclusions regarding the mechanisms of human interactive learning.

### 1.3.2 *For every learner there is a teacher, and vice-versa: interaction-based approaches*

Real-time social interaction involves rich and complex behavioural dynamics, with bi-directional responses and input between two or more people (Clark, 1996; Garrod & Pickering, 2004; Holler & Levinson, 2019; Yu & Smith, 2012). Such a multifaceted phenomenon is unlikely to rely on a single cognitive mechanism but rather a number of cognitive processes, which may be absent in a non-interactive situation. During interactive learning, learner-teacher dynamics may be characterised by joint attention (Schertz et al., 2013), common ground (Bohn et al., 2019), shared intentionality (Sabbagh & Baldwin, 2001), or all these processes together (Schmitz, 2014). These mechanisms of attunement between two or more conversational partners may allow information to be shared more effectively, and in turn be advantageous in those situations when we learn socially (Frith & Frith, 2012; Hu et al., 2015).

One approach to examining the rich and complex dynamics of interpersonal interaction is to argue that social interaction is more than just a context for social cognitive processes, but in fact replaces individual mechanisms (De Jaegher et al., 2010). In such an enactive model, the inter-personal relational dynamics becomes autonomous from the single individual parts making up the interaction. This implies that traditional single-person models have little relevance to the two-person interaction, and that researchers need to find a new type of dynamic model to understand interaction at a more abstract level.

However, this work argues for a more incremental approach, where social interaction is included as an additional element in the study of human cognition. As such, interpersonal interactions can be integrated into – and understood by building on – models of the solo brain. For example, we know that learning a new concept from a video will involve processes of perception, language and memory that allow the learner to integrate the new information into their existing knowledge structures.  Learning a concept in interaction is likely to engage the same processes plus additional cognitive systems (e.g., joint attention, common ground etc), where the moment-by-moment coordination of gaze and speech allows these additional processes to function smoothly.

There is evidence that the quality and quantity of social cues present in a given interaction substantially affects the communicative outcome of that interaction (Cartmill et al., 2013; Sauppé & Mutlu, 2014). Rich visual cues may enable stronger attunement by providing more information about the interaction partner's gaze and mental states (Kajopoulos et al., 2020; Marotta et al., 2012). Alksne (2016) looked at what features in teaching videos improved the quality of the lecture in a group of young adults: they found that speaking over the presentation and making eye-contact significantly improved student engagement, which in turn has been positively associated with learning outcomes (Bevilacqua et al., 2014).

The fact that we somehow use our bodies to achieve a better attunment with our intelocutor(s) during social communication has been recently well demonstrated by Fini and colleagues (2021). In their study, they asked adults to guess concrete and abstract concepts from some photos, while being in interaction with an avatar. The avatar moved following the kinematics of a real actor's arm previously recorded, from which human movement were implemented on the avatar. They found an association between sociality (as measured by motor imitation and motor synchrony between the participant and the avatar) and guessing of abstract concepts. They argued that greater motor imitation showed by the learner specifically during more difficult trials (abstract words) reflected a greater longing for help: participants would try to attune more to the avatar to receive more hints and support in the guessing task. This interpretation is in line with the argument that social attunement may be a way to support efficient information transfer across interlocutors (Shafto et al., 2012).

However, by looking only at the student, this study does not tell us much about whether the direction of such synchrony is unidirectional (from learner to teacher) or rather bi-directional. For example, Davidesco et al. (2019) found that while learner-to-learner brain synchrony was instantaneous, learner-to-teacher brain synchrony could best predict learning when adjusting for a temporal lag of ~200 milliseconds. Specifically, student brain activity would 'tune in' to brain activity of the teacher only after a short delay, suggesting a sequential, lagged transfer of information from teachers to students. This type of data shows that, to fully grasp the neural mechanisms of interactive learning, it may be insufficient to focus on one social agent alone, instead dyadic analysis may carry more interesting and comprehensive information about these complex dynamics.

### 1.3.3 *Synchronization as a signature of social learning*

A growing body of literature is emerging showing that a signature of interactive learning may be a bi-directional synchrony during teacher-learner interaction (see Section 2 for a review of the literature on this). When A interacts with B, both A and B would share some processing linked to the experience they are both part of, while the brain of A would process information about B and the brain of B would process information about A. By looking at individual brain systems as part of an interaction, we can start to understand the full temporal and behavioural dynamics that are reflected into individual brain activity (of interactive agents). These patterns of bi-directional coordination can be interpreted within the framework of the mutual-prediction hypothesis (Hamilton, 2020; Kingsbury et al., 2019; Nishimura et al., 2020). This claims that, when interacting with others, we engage in social prediction all the time in order to anticipate other people's actions and mental states (Holler & Levinson, 2019; Tamir & Thornton, 2018). Furthermore, when two people are both engaged in mutual prediction, their brain states will correlate and thus the signals recorded from their brains will correlate, giving rise to interbrain synchrony. Thus, predictive mechanisms present in individual brains can give rise to a consistent cross-brain signal that may predict learning (Bevilacqua et al., 2019; Davidesco et al., 2019; Dikker et al., 2017; Holper et al., 2013; Pan et al., 2018).

However, claiming that brain-to-brain coupling on its own can tell us something conclusive about the quality of the social interaction, and even further, on the learning mechanisms of teacher-student social exchange, is at best ambitious – if not misleading (see Novembre & Iannetti, 2021 and Hamilton, 2020 for a discussion on this). In conjunction with studying interpersonal brain synchrony, it is critical to understand the coordination of actions and how that relates to shared knowledge states (see Hasson et al., 2012 for a comprehensive framework of neural synchrony and its behavioural references). This may be particularly useful when learning from or teaching to someone else. In the case of interactive learning, the co-creation of knowledge and understanding is fundamental to the learning process: ideally, the teacher would want to share information, and the learner would want to tune in to their teacher to receive and process that information, while both would remain sensitive to feedback coming from their interlocutor to adjust their behaviour accordingly. It has been proposed that the extent to which people synchronise may be a proxy of ongoing exchanges during human social interaction (Balconi et al., 2017; Reinero et al., 2020): in other words, high brain-to-brain synchrony across social agents should reflect behavioural inter-personal dynamics. Possibly, the objective is that of reducing prediction errors and increasing affiliation and communicative benefits (Hoehl et al., 2020). Therefore, integrating behavioural data into hyperscanning studies is necessary to achieve a more comprehensive and meaningful knowledge of how humans learn from and with others.

In fact, being a form of social interaction, good pedagogy would be therefore characterised by continuous reciprocity: the teacher would monitor the audience's engagement and understanding, and use the audience feedback to adapt their performance as needed. Such mutual-prediction engages the brain in a constant probabilistic estimate of occurrence of external experiences based on expected outcome. These may be plausible mechanisms underlying inter-personal synchrony and shared neural representations typical of social situations (Brown & Brüne, 2012). Studies have shown that interpersonal synchrony manifests across multiple levels during social interaction, including motor coordination (Chang et al., 2020; Kruppa et al., 2020), action coordination and decision-making (Heggli et al., 2020), and verbal coordination (Jiang et al., 2020; M. Nguyen et al., 2019; von Zimmermann & Richardson, 2016). In addition, person-to-person synchrony has been reported even at the physiological (Gordon et al., 2020; Kragness & Cirelli, 2020) and neural level (Hasson et al., 2012b; Hasson & Frith, 2016; Hoehl et al., 2020; T. Nguyen et al., 2020).

## 1.4 Gap in the literature and rationale for the present work

The scope of this introduction was to look at the state-of-the-art in the neuroscience of human learning as it most naturally occurs, i.e. socially. I have presented evidence showing the crucial role that social interaction plays in human learning across the life-span. Social interactions feature as a catalyst for the human ability to acquire and retain new information (Csibra, 2006; Meltzoff et al., 2009). However, taking the cited literature together, a few issues emerge. I consider these below, and outline how this PhD work attempts to overcome them.

### 1.4.1 ***Learning *via* the social has been (wrongly) interchanged with learning *about* the social.***

The majority of work presented in the literature reviewed above has used the term 'social learning' without clearly distinguishing between situations where people learn *via social* interaction to situations where people learn *about the social* world. Generally, studies focused on the investigation of learning about the social world (e.g. facial expression, language), either via social or via non-social contexts. In fact, there are disproportionally more studies on interactive learning in the domain of language acquisition than any other domain (both in children and adults). A tight link exists between language and social interaction, given the social nature of communication

(Garrod & Pickering, 2004; Kuhl, 2007; Seyfarth & Cheney, 2014). The problem with studies that did not distinguish the medium through which we learn from the object of learning, is that it is difficult to disentangle the contribution of social interaction in building new long-term knowledge (e.g. via one or multiple cognitive mechanisms described in section 2), in contrast to mechanisms of general social cognition (Holtgraves & Kashima, 2008). A clear separation between learning via the social and about the social is therefore important to build accurate cognitive models, and in turn should guide experimental practice.

My PhD aims to understand how social interaction supports adult learning, and identify its distinct contribution. As such, we adopt a definition of social learning that is learning explicitly (i.e. the learner is aware that their aim is to learn new facts) via social interaction about non-social knowledge. To do so, I develop a task which employs social interaction as the medium through which learning occurs, while the object of learning (learning content) is carefully conceived to be non-social. Specifically, I design a series of fact-based descriptions about several obscure items (e.g. exotic animals, antiques, more details are reported in experimental chapter 2 and Appendix) which participants learn in different social contexts. Furthermore, in line with this idea, in chapter 3, I present a study which looks at the case of adults with Autistic Spectrum Condition (ASC). The blend (or lack of clear separation) between learning via the social and learning about the social is particularly evident in the literature on autism. In fact, based on decades of studies of how people with autism learn about social signals, it is commonly assumed that people with ASC also learn less well via social contexts and may prefer asocial learning (e.g. computer-based learning). This however has never been formally tested.

### 1.4.2 *Learning has been mainly studied within a specific context: in isolation, and in children.*

One of the main issues emerging from the review above is that the behavioural and cognitive mechanisms involved in learning are mainly studied in isolation: it remains largely unexplored how the experience of learning from others modulates the dynamic of people involved in the interaction, both within individuals and between individuals, as a coordinated system. Social agents will inevitably influence and be influenced by each other, and as such the inter-personal dynamics need to be taken into account to fully grasp the cognitive and neural mechanisms subserving social interactive learning.

It is the case, however, that learner-teacher social dynamics has been largely neglected by modern cognitive neuroscience research, especially in adulthood. Methods adopted to study human learning have often included single-user tasks, where participants were required to memorise things from cards/screens, in very repetitive and highly constrained experimental paradigms (e.g. Batial & Shmueli, 1997; Duff & Hulme, 2012). More recently, there has been a trend towards studying human learning in more dynamic social contexts (e.g. Jeong et al., 2021; Kostyrka-Allchorne et al., 2019). Despite the creditable effort to move away from the traditional reductionist approach (Putnam, 1973), this new line of research has only partially included the social aspect in the study of human learning. Namely, at best social context has been included in the study design and data collection, while the focus of data analysis has been almost exclusively either on the learner (Gilbert et al., 2001) or (less often) on the teacher (Battro, 2010), and only rarely on the interaction (Holper et al., 2013). This is also the case in many developmental studies that see the teacher (caregiver) as providing an input to the learner (the child) (e.g., Cartmill et al., 2013). The problem with this is that conclusions may be based on a partial view of what is happening during a real-world student-teacher interaction.

Understanding inter-personal coordination in social learning may be particularly important in conceptual learning (see point 1.4.1) because of the particular challenges required in acquiring new concepts. Unlike learning an arbitrary word list, a new concept must be integrated with other existing knowledge of many different types (e.g., Tremont et al. 2000). Interactive teaching may allow a learner to try out a new concept and explore how it relates to other concepts with immediate feedback, which is likely to provide richer and more robust learning. Furthermore, within the domain of interactive learning research, more direct comparisons of observational and interactive social learning are needed. This is essential to disentangle the contribution of specific factors associated with social contexts that benefit human learning.

In addition, very little is known about social learning specifically in adults. I have presented a large body of evidence accumulated over the last decades showing how social interaction is crucial for development. Nevertheless, social interaction remains a fundamental human need beyond childhood, with implications for adulthood wellbeing and mental health (e.g. Tomova et al., 2020). Despite this, there is no work on whether and how social interaction supports learning in adults.

**Figure 1.3 Experimental set-up for the hyperscanning study**

Illustration of two participants during the hyperscanning study on learning described in Chapter 5. Teacher-student exchange is investigated in high-ecological set-up where participants are free to engage in conversation on the learning content while behavioural and brain metrics are recorded. Picture published with participants' permission.

This work attempts to fill this gap, by bringing the investigation of social learning beyond the niche within which it has been mainly studied so far. Namely, here we examine social learning i) in interaction and ii) within the adult population (e.g. figure 1.3). To do so, studies reported in this thesis employ paradigms which directly compare social interactive conditions (e.g. live conversation) with social non-interactive conditions (e.g. observational learning) and recruit participants aged from early 20s to their mid-thirties. Interactive learning is considered in a variety of contexts, including online (chapter 2 and 3) and face-to-face (chapter 5), and the teacher-learner social exchange dynamic is placed at the core of the investigation.

### 1.4.3 ***Social interaction is complex and multimodal, and should be studied as such.***

The relation between social interaction and learning may be modulated by complex dynamics spanning across behaviour, physiology and the brain (García & Ibáñez, 2014; Redcay & Schilbach, 2019). Previous studies have mainly either examined only behaviour or the brain. However, complex dynamics are unlikely to be fully grasped by experiments that look solely at one modality (e.g. speech or neural activity). Instead, the study of interactive conceptual learning should be integrated with the study of interactive communication, to include coordination in a number of verbal and non-verbal behaviours (e.g. eye-gaze, Holler & Levinson, 2019b; Murgiano et al., 2021), as well as neural dynamics. We know very little about how coordination within and across these different channels can support conceptual learning.

This work attempts to integrate multimodal experimental designs to grasp the complexity of interactive learning and the mechanisms subserving it. To do so, we make use of video recordings, behavioural metrics (e.g. learning performance, but also enjoyment and anxiety score, as well as social signals such as eye-gaze and joint attention), and measures brain activity while the interaction unfolds. Specifically, we employ hyperscanning (i.e. measuring brain activity simultaneously from more than one brain) functional-Infrared Spectroscopy (fNIRS).

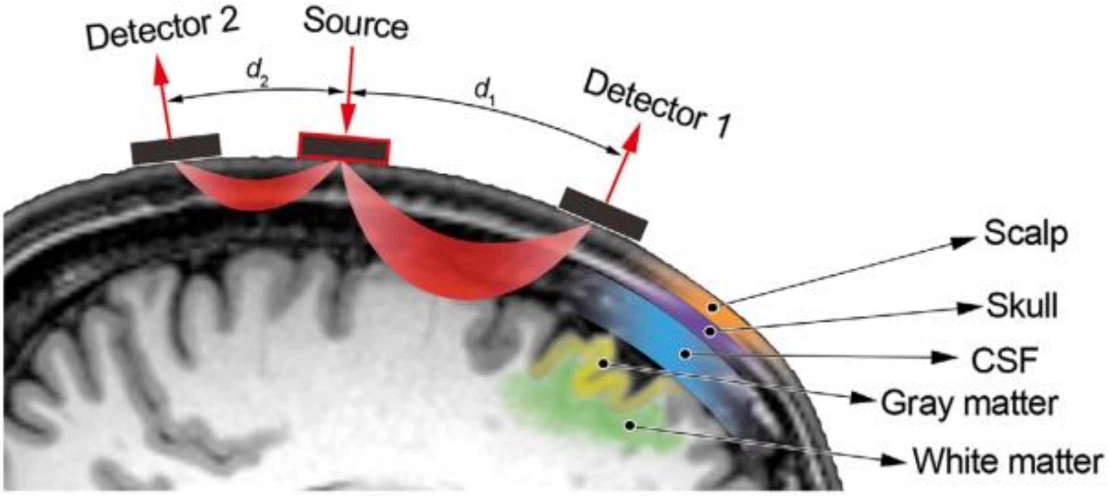


**Figure 1.4 Schematic of fNIRS principles**

Illustration of the path (shown in red) followed by the NIR photons from the light source to the detector travelling through the different tissues of the head. The penetration depth of the light is proportional to the source–detector distance (d1: deeper channel; d2: superficial channel). A channel is composed by the pair source–detector and is located at the midpoint between the source and the detector and at a depth of around the half of the source–detector separation. Standard source-detector distance to detect signal from brain cortex is 3cm. Picture adapted with permission from Pinti et al. 2020.

fNIRS is an optical, non-invasive neuroimaging technique that measures brain tissue concentration changes of oxygenated ($HbO_2$) and deoxygenated (HbR) haemoglobin following neuronal activation (BOLD response). This is achieved by shining NIR light (650–950 nm) into the head, which can travel through the scalp and reach the brain cortex (figure 1.4). Crucially, $HbO_2$ and HbR absorb the NIR light differently: $HbO_2$ absorption is higher for > 800 nm; on the contrary, HbR absorption coefficient is higher for < 800 nm. This difference in absorption depends on the colour of the blood that is more red for oxygenated blood (arterial blood) and more purple for venous blood. When a brain area is active and involved in the execution of a certain task, the brain's metabolic demand for oxygen and glucose increases, leading to an oversupply in regional cerebral blood flow (CBF) to meet the increased metabolic demand of the brain. This in turn leads to an increase in $HbO_2$ and a decrease in HbR concentrations in the area which is most active. These fluctuations in oxygenated and de-oxygenated blood cause changes in light attenuation that can be measured by fNIRS.

The great advantage of fNIRS is that it allows us to measure brain activity in naturalistic context, as it is wearable and relatively robust to movement (Czeszumski et al., 2020; Kelsen et al., 2020b; Pinti et al., 2018). Therefore, this technique enables the study of social interaction with a high level of ecological validity, to an extent that is not possible with other neuroimaging modalities such as fMRI. For example, it allows us to measure *brain-to-brain coherence* across real-time interacting people (see section 1.1.2, Figure 1.2 and Figure 1.3). In addition, being wearable it ensures that other metrics can be collected in combination with brain data, with relatively little interference across measurements (e.g. video-recordings, movement and physiological measures such as breathing can be collected alongside brain signals, see Figure 1.3). This makes this technique very attractive for multi-modal experimental work, including the experiments presented in chapter 4 and 5 of this thesis, and guarantees that a certain degree of real-world complexity can be incorporated in the experimental study of social interaction (Czeszumski et al., 2020; Kelsen et al., 2020).

## 1.5 Overview of experimental chapters

This thesis attempts to fill the existing gap in the literature by investigating the role of social interaction in adult human acquisition of new knowledge. First, it looks at social learning online (chapter 2 and 3), and then moves to a face-to-face context to integrate the study of neural mechanisms during social interaction (chapter 4) and how these relate to learning (chapter 5). Across all studies, this work prioritises the ecological validity of the design, allowing – within certain constrains (e.g. time limit) – a relatively free exchange between the teacher and the learner. Since the aim was to understand the cognitive and neural mechanisms of social learning in relation to how it usually unfolds in the real world, participants have not been deceived at any point and the aim of the study was always made explicit. In addition, to minimise any confound arising from reward processing, participants were compensated for their time in the form of monetary reward, but did not receive further performance-based reward.

In particular, chapter 2 investigates learning online as a function of interactivity (social contingency) and social signals (e.g. exposure to face). This is achieved by directly comparing learning during live video-calls versus pre-recorded videos over two experiments in a large sample of neurotypicals. In chapter 3, we ask whether the same pattern of results found in neurotypicals would be replicated in adults with ASC. Chapter 4 investigates whether social interaction (e.g. engaging in conversation) modulates the neural response for non-social processing (e.g. movie watching). Chapter 5 asks whether inter-personal neural dynamics (brain-to-brain coherence) can be used as a good metric to predict learning, along behavioural measures (eye-gaze and joint attention).

Please note that throughout the thesis the pronoun ‘we’ will be used to refer to the experimental team who contributed to the work, with the default assumption that the candidate Sara De Felice has been the leading researcher who carried out the experiment. When other researchers completed a significant part in any of the studies presented in this thesis (e.g. data collection), details about their contribution are reported at the beginning of each chapter. In addition, it is worth bringing to the attention of the reader that the great majority of the data reported in this thesis has been collected during the Covid-19 global pandemic, with severe delays and great challenges for the study of naturalistic social interaction. This has not only deferred the whole research work (e.g. applying for extra ethical procedures; almost doubling the testing session duration to allow for sanitation of the research facilities etc), but has also led to a series of adaptations of the original study plan, based on constrains imposed by the unprecedented situation (e.g. testing people within the same household for Study 3 and 4 to follow government regulations).

# 2. Chapter 2 - Learning Online in Neurotypical Adults

Data from this chapter has been published on 14th September 2021, full reference is:

De Felice, S., Vigliocco, G. & Hamilton, A. (2021) Social interaction is a catalyst for adult human learning in online contexts. *Current Biology*, *31*(21), 4853-4859. https://doi.org/10.1016/j.cub.2021.08.045

Design, hypotheses and analysis plan for Experiment 2 of this chapter was pre-registered on the 21st December 2020 on OSF (De Felice & Hamilton, 2020 10.17605/OSF.IO/NXS37).

## 2.1 Background

Social learning refers to any learning happening between two or more individuals. Observational learning (Bandura, 2019) involves acquisition of information through passive exposure to the material (e.g. learning from a pre-recorded video). In contrast, interaction-based learning (Shamay-Tsoory, 2021) requires mutual-feedback between student and teacher (e.g. learning in live conversations, Morgan et al., 2015). 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 (Kourtis et al., 2020) 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(Hertel et al., 2008). Alternatively, socially contingent teaching might boost learning, as seen in children (Kuhl, 2007) but not always in adults (Ramlogan et al., 2013; Solomon et al., 2004, see Introduction, Chapater 1 for an extensive litterature review on this).

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 (D. Zhang et al., 2006), multimedia characters (Kwok et al., 2016), recorded slides (Vaccani et al., 2016) or podcasts (Schreiber et al., 2010). 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 (Kourtis et al., 2020; Moore & Barresi, 2017) and/or distract learners (Kajopoulos et al., 2020). Alternatively, social cues such as eye-gaze (Marotta et al., 2012) and gestures of a teacher (Wakefield et al., 2018) could benefit learning by facilitating the coordination and 'attunement' between student and teacher (Mundy & Newell, 2007), via mechanisms of joint attention and social engagement (Hoehl et al., 2020a; Kasari et al., 2008; Kawai, 2011; Saito et al., 2010).

Over the 2020 Covid-19 pandemic, learning – as well as many other activities – has been moved online in many countries around the world. A whole generation of students, who were used to busy schools and universities, have now moved all their learning activities to a virtual environment. This has changed the patterns of social interaction during learning, but has also opened up several new options in terms of the format in which lectures are delivered: either live or recorded, with different degree of interactivity and/or social cues (e.g. face of the teacher being visible or not). However, it is not yet clear which of these formats represent the best way to help students to learn. Therefore, studying the role played by social interaction in learning in such contexts not only becomes timely relevant for educational purposes, but also makes it extremely interesting for our understanding of cognition and how human learning adapts to these new environments.

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 2.1). 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.

## 2.2 Experiment 1

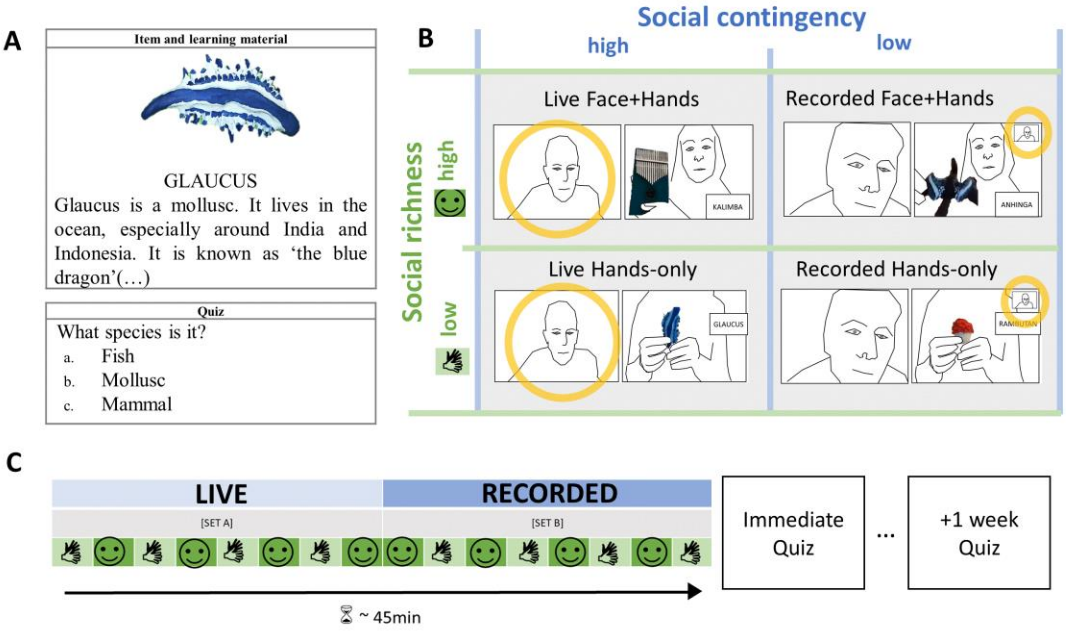
### 2.2.1 *Methods*

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.

#### *2.2.1.1 Participants*

The 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 take 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.

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 video-call quality (N=2, we only accepted subjects who reported 4 and above on a 1(poor)-5(excellent) video-call 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.



Glaucus is a mollusc. It lives in the ocean, especially around India and Indonesia. It is known as 'the blue dragon'(…)

What species is it?
a. Fish
b. Mollusc
c. Mammal

**Figure 2.1 Experimental paradigm**

**A.** Sample item. Participants learnt facts about unusual items in a 2 min structured conversation. See appendix (Table 1) for the full set of items. **B.** Design for Experiment 1. In a 2x2 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 time-line. In each 45 minute session, participants learnt 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 in Appendix and online at https://osf.io/tuz9n/.

### 2.2.1.2 Materials

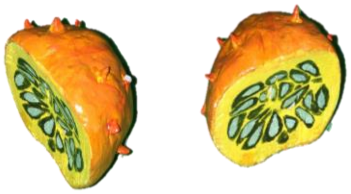
Two learning sets were created, each including eight items, two from each of the following four categories: animals, food, ancient objects and musical instruments. Item selection started from a pool used in Vigliocco et al., 2019. Final items were selected on the basis of an initial pilot (N = 15) run 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 handmade 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). 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). Full information and question sets for all items are reported in Table 2.1.

**Table 2.1 Full set of items for the learning task.**

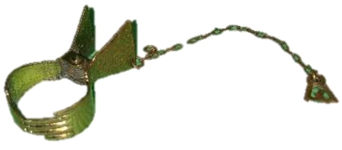
Learning material with picture and full description on the left column, multiple choice quiz per item on the right column. The same multiple-choice quiz was used for immediate and delay testing.

| **Item and learning material** | **Multiple-choice quiz** |
|---|---|
| 1. GLAUCUS<br><br>Glaucus is a mollusc. It lives in the ocean, especially around India and Indonesia. It is known as 'the blue | 1. What species is it?<br>a. Fish<br>b. Mollusc<br>c. Mammal<br>2. What is its habitat?<br>a. Ocean<br>b. Lake<br>c. Tropical forest<br>3. How is it also known as and why? |

| | |
|---|---|
| dragon’ because of its colour and ray-shape tentacles. It feeds on jellyfish. However the jellyfish is not digested completely: the most poisoning cells are stored in the tips of the Glaucus tongue (in-between teeth) and used as a self-defence tool. It has a gas-filled sac in its stomach which acts as a balloon allowing it to float on the surface of the water: they don't swim but are carried along by the winds and ocean currents. | a. "Blue dragon" because of its colour and shape<br>b. "Cold fire" because of the blue ray-shape fingers<br>c. "Sea snake" because of the long tail<br>4. What does it eat?<br>a. Fish<br>b. Leaves<br>c. Jellyfish<br>5. How can it float?<br>a. Thanks to its ray-shape fingers<br>b. It doesn't: it lives deep on the bottom of the ocean<br>c. Via a gas-filled sac in its stomach |
| <br>2. TARSIER<br>Tarsier is a mammal. Originally from south-east Asia, it lives in the jungle. Its name comes from its very long tarsal (ankle) bone. The elongated legs allow it to jump from tree to tree hunting prey (it can jump up to 40 meters). It is totally carnivorous and eats mainly insects. It has special eyes: these are bigger and heavier than its brain. Because its eyes are so huge, it cannot move them and has to rotate its head to look around. The head can rotate 180 degrees. Such big eyes allow for very good night vision. | 1. What species is it?<br>a. Bird<br>b. Amphibian<br>c. Mammal<br>2. What is its habitat?<br>a. Jungle<br>b. Mountain<br>c. Fresh water<br>3. Where does the name come from?<br>a. From its long fingers<br>b. From its long ankle bone<br>c. From the region it lives in<br>4. What does it eat?<br>a. Only carnivorous, mostly insects<br>b. Only herbivorous, mainly bananas<br>c. Omnivorous, mainly insects<br>5. What's special about its eyes?<br>a. They act like magnifying glasses to detect extremely small insects<br>b. They are bigger and heavier than its brain and can see in the dark<br>c. They are bright yellow to be able to see in the dark |

| | |
|---|---|
| <br>3. KIWANO<br>Kiwano is a fruit native of Sub-Saharan Africa and needs a warm climate to grow. It is also known as the 'horned melon' because of its melon-like colour and the fact that it presents some horns/spikes on its skin. All its parts, including seeds, flesh, and peel, are edible. Its flavour is a combination of banana, cucumber, and lime: so quite fresh and with a gelatine-like texture. It is a good source of vitamins and liquids, and represents one of the few sources of water in Africa deserts during the dry season. It grows for half of the year from January until July. | 1. Where is it originally from?<br>a. South-East Asia<br>b. Sub-Saharan Africa<br>c. Northern Africa<br>2. How is it also known and why?<br>a. Orange melon because of its colour<br>b. Spiked peach because of its skin<br>c. Horned melon because of its skin<br>3. What can/cannot be eaten?<br>a. It is all edible<br>b. The peel cannot be eaten<br>c. The seeds are poisonous<br>4. What does it taste like?<br>a. Strawberries, lemon, and melon<br>b. Banana, cucumber, and lime<br>c. Spinach, peas and lemon<br>5. In which season(s) does it grow?<br>a. Only during summer<br>b. From July to November<br>c. From January to July |
| <br>4. CHERIMOYA<br>Cherimoya is a fruit native of South America. Its name means “cold seeds“, because it grows at high altitude where the climate is cold. The whitish flesh is edible, however its seeds are poisonous especially if eaten in big quantities (these are usually used to make pesticides). The flesh tastes like a blend of banana, pineapple, peach, and strawberry. It is very nutrient and has high calories with lots of good vitamins. It ripens from autumn until spring. When ripe the skin is green, but it turns brown when goes rotten. | 1. Where is it originally from?<br>a. South America<br>b. North America<br>c. South-east Asia<br>2. What does its name mean?<br>a. "Spring taste"<br>b. "Rough skin"<br>c. "Cold seeds"<br>3. What can/cannot be eaten?<br>a. It is all edible<br>b. The seeds are poisonous<br>c. The peel cannot be eaten<br>4. What does it taste like?<br>a. Sweet potatos<br>b. Grapes, lemon, apple and pear<br>c. Banana, pineapple, peach and strawberry<br>5. In which season(s) does it grow?<br>a. Autumn to spring<br>b. Winter to summer<br>c. Summer only |

| | |
|---|---|
| <br>5. CAXIXI<br>Caxixi is originally from Africa. It belongs to the group of idiophones, as it is a percussion instrument played by shaking it: the small particles inside (made by little stones or sand) hit the walls of the instrument producing sound. Depending on whether you hit the walls or the bottom, it produces different sound: softer on the sides and sharper on the bottom. Its name resembles the sound it makes. It is still played in some communities in Africa and its music is believed to call for enchanted spirits and to ward off evil ones. | 1. Where is it originally from?<br>a. Indonesia<br>b. Africa<br>c. Mexico<br>2. What type of musical instrument is (classification)?<br>a. Idiophone<br>b. Membranophone<br>c. Aerophone<br>3. Where does the name come from?<br>a. From african language and means rattle snake<br>b. From the name of the seeds inside the basket<br>c. From the sound it makes<br>4. How is it played?<br>a. Shaking it<br>b. Hitting it<br>c. tying it around the wristle while one dances<br>5. What is its social valence (in what context is it usually played)?<br>a. Used at weddings to bring fortune to the couple<br>b. Used to sing children to sleep<br>c. To call for enchanted spirits and ward off evil ones |
| <br>6. KALIMBA<br>Kalimba is originally from Africa. It belongs to the group of idiophones, as it is a percussion instrument played by plucking the tines with the thumbs. Its name means "little music", because of the sound it makes which is quite delicate. When out of tune, can be | 1. Where is it originally from?<br>a. Australia<br>b. Central America<br>c. Africa<br>2. What type of musical instrument is (classification)?<br>a. Idiophone<br>b. Membranophone<br>c. Aerophone<br>3. What does the name mean?<br>a. Hand guitar because of how it is played<br>b. Tiny Keyboard because of its shape<br>c. Little music because of its sound<br>4. How is it played?<br>a. By plucking the tines with your thumbs<br>b. With a pick similar to a guitar |

| | |
|---|---|
| tuned by regulating the individual tines either a bit higher or a bit lower across the horizontal bar. It is usually associated with joyful events and played at religious ceremonies, weddings, and social gatherings to celebrate the sense of community. | c. By pressing on the tines like a piano key<br>5. What is its social valence (in what context is it usually played)?<br>a. Played at baby showers to symbolize new life<br>b. Played at funerals and wakes symbolizing transition<br>c. Played a religious ceremonies, weddings, and social gathering to celebrate sense of community |
| 7. PORTE-JUPE<br><br>Porte-jupe used to be a female accessory popular during the Victorian age in England. It was a tong used by women to lift up their long skirt. The small ring at the top would have held a cord, ribbon, or chain to suspend the tool just below the waist. Its function was that of making women more agile when they started to engage more in outdoor activities. For this reason, it was a fashion accessory associated with emancipation of women. Its name comes from french and literally means 'dress holder'. | 1. What is its function?<br>a. Used as a clip for long hair<br>b. Used to lift long skirts<br>c. Used to hold multiple items<br>2. When was this item particularly popular?<br>a. Middle ages in Italy<br>b. Early 1900s in China<br>c. Victorian age in England<br>3. Who was used by mainly?<br>a. Athletes with long hair<br>b. Children<br>c. Women<br>4. Where does its name mean?<br>a. "Dress holder"<br>b. "Hair holder"<br>c. "Elegant jewellery"<br>5. What was its social valence?<br>a. It reflected style and wealth<br>b. It was a sign of women's emancipation<br>c. It was associated with elegance among athletes |
| <br>8. SCOTCH HANDS<br>Scotch hands are kitchen tools used for making butter. They are made of wood | 1. What is its function?<br>a. Kneading dough<br>b. Mixing alcoholic beverages<br>c. Making butter<br>2. When was this item particularly popular?<br>a. Early 1900s across the UK<br>b. Late 1700s in Ireland<br>c. 1800s throughout Europe<br>3. Who was used by mainly?<br>a. People living in industrial cities |

| | |
|---|---|
| but can also be found in metal. The external side is smooth while inside there are some small ribs. They were popular in early 1900s across the UK. They were mainly used by women and chefs, especially in big farms where they had cows and were making diary products. Their name comes from their Scottish origin and their hand-like appearance. Usually one pad stays firm while the other is moved to shape or mix salt in the butter pat. | b. Women during wedding parties<br>c. Women and chefs<br>4. Where does its name come from?<br>a. From its Scottish origin and hand-like appearance<br>b. From its usage of mixing scotch and soda<br>c. From its inventor<br>5. In which contexts was it found?<br>a. In wealthy houses only<br>b. In big farms<br>c. In pubs and bars |
| <br>9. ANHINGA<br>Anhinga is a bird. It lives around fresh water, including rivers and lakes. It can be found in warm parts of America (e.g. Florida, Mexico). It is also known as the water turkey, because it has a big tail similar to the one of turkeys. During breeding season, male anhinga have a blue ring around their eyes to attract female attention. It feeds on fish that it catches with its long beak. Body and feet designed to swim efficiently. It has a gland underneath its tail: this produces an oil that makes its feathers waterproof. | 1. What species is it?<br>a. Bird<br>b. Mammal<br>c. Fish<br>2. What is its habitat?<br>a. Forest<br>b. Sea<br>c. Fresh water<br>3. How is it also known as and why?<br>a. Air duck because of its feet<br>b. Water turkey because of its tail<br>c. Water giraffe because of its long neck<br>4. What does it eat?<br>a. Algae<br>b. Fish<br>c. Worms and insects<br>5. What's its peculiarity?<br>a. It has special organs that allow to breathe underwater<br>b. It has a gland that produces an oil to waterproof its feather<br>c. Its saliva acts as a glue to catch little fish |
| 10. AXOLOTL<br> | 1. What species is it?<br>a. Fish<br>b. Amphibian<br>c. Mammal<br>2. What is its habitat?<br>a. Lakes<br>b. Ocean<br>c. Rivers<br>3. How is it also known as and why? |

| | |
|---|---|
| Axolotl is an amphibian which lives in the lake of Mexico. It is also known as the 'walking fish' for its way to 'walk' in water. It feeds on worms, insects and small fish, but can sometimes show cannibalism (eating another axolotl). It has a special power: it can regenerate limbs within days multiple times. For this reasons it is used a lot in staminal cell research against cancer. It reproduces asexually: male deposits the sperm into his habitat and then do a small dance to attract female so that she can find it and her eggs can be fertilised. | a. Swimming lion because of its ray shape head<br>b. Sea lizard because of its body shape<br>c. Walking fish because of the way it moves in water<br>4. What does it eat?<br>a. Algae<br>b. Worms, insects and small fish<br>c. Dead body of big fish<br>5. What's its special power?<br>a. Can regenerate limbs<br>b. Can see through obstacles (rocks, other animals etc)<br>c. Can blend in and become invisible |
| 11. RAMBUTAN<br><br>Rambutan is a fruit originally from southeast Asia. Because of the shape of the skin, which has thin spikes, its name means "hairy". It is all edible: it has only one big seed that can be cooked and eaten. It has a very sweet flavour which resembles that of grape, and for this reason is often used to make jams and jelly. It grows on ever-green trees, which fruit twice a year, in summer and winter. | 1. Where is it originally from?<br>a. Middle east<br>b. Southeast Asia<br>c. South America<br>2. What does the name mean?<br>a. Hairy<br>b. Spikes<br>c. Spider<br>3. What can/cannot be eaten?<br>a. It is completely edible a part from the central seed<br>b. It is completely edible including its seed<br>c. It is poisonous<br>4. What does it taste like?<br>a. Banana<br>b. Peach<br>c. Grape<br>5. In which season(s) does it grow?<br>a. From spirng to late summer<br>b. Only in summer<br>c. Twice per year in summer and in winter |
| 12. HULUSI<br><br>Hulusi is a musical instrument originally from China. It belongs to the group of | 1. Where is it originally from?<br>a. Australia<br>b. China<br>c. Japan<br>2. What type of musical instrument is (classification)?<br>a. Aerophone<br>b. Idiophone |

| | |
|---|---|
| aerophone, as it is played by blowing air through it like a flute. It has three bamboo pipes, one with finger holes while the others serve to make harmonic sound effects. Sometimes one of the two pipes is clogged up and it is only ornamental. Its name means "silky gourd": 'silky' refers to the delicate sound it makes, and 'gourd' refers to its round shape on top. It is usually played by minority groups in some regions of China, Vietnam and Thailand and has generally been associated with social struggles. | c. Membraphone<br>3. What does the name mean?<br>a. Sacred wind<br>b. Silky gourd<br>c. Long flute<br>4. How is it played?<br>a. Hitting it with a stick<br>b. Shaking it<br>c. Like a flute<br>5. What is its social valence (in what context is it usually played)?<br>a. It symbolizes minority groups<br>b. It represents national unity<br>c. It is played at big festivals |
| <br>13. AGOGO<br>Agogo is a musical instrument originally from Africa. Because of its shape, its name means ‘bells’, although the bell is in fact empty (there is no stem inside). It belongs to the group of idiophones, as it is a percussion instrument played by hitting the bells with the stick. It is played during religious ceremonies to give the right rhythm to the dance: it is believed that being in tune with the music will make it easier to be possessed by the spirits that will guide worshippers to a successful life. | 1. Where is it originally from?<br>a. Africa<br>b. China<br>c. Brazil<br>2. What type of musical instrument is (classification)?<br>a. Aerophone<br>b. Idiophone<br>c. Membranophone<br>3. What does the name mean?<br>a. Cones<br>b. Bells<br>c. Corn<br>4. How is it played?<br>a. Hitting it<br>b. Shaking it<br>c. Folding it<br>5. What is its social valence (in what context is it usually played)?<br>a. Played at Olympic games<br>b. Played during religious ceremonies<br>c. Played at private parties to dance |
| 14. STRIGIL<br> | 1. What is its function?<br>a. Scratch one’s back<br>b. Clean the body<br>c. Work the clay<br>2. When was this item particularly popular?<br>a. First half on 20th century |

| | |
|---|---|
| Strigil is an ancient object used during the Roman empire and ancient Greece to clean the body. It was mainly used by male athletes after their sport performance to scrape off the dirtiness from their skin before the bath. Because of its function, the name strigil literally means "scraper", from Latin. It could differ in the type of metal used and its design, depending on the social status of the owner and time period. It was often found in the tombs of important athletes to represent their social status in the afterlife. | b. Ancient Egyptian empire<br>c. Ancient Roman and Greek times<br>3. Who was used by mainly?<br>a. Women<br>b. Athletes<br>c. Doctors<br>4. What does its name mean?<br>a. Scraper<br>b. Knife<br>c. Curved metal<br>5. In which contexts was it found?<br>a. Tombs of athletes<br>b. Hospitals<br>c. House of nobles and aristocrats |
| <br>15. CHATLAINE<br>Chatlaine is a decorative belt carrying household items including scissors, keys, needles etc. It was very popular in the middle age across Europe, and mainly wore by women. Its name comes from French and literally means "Lady of the castle". It could be made of metal, from gold to bronze or iron depending on how wealthy the owner was. As holding keys to different lockers and doors around the house, it represented a symbol of authority for the woman who wore it as other people in the houselhold had to ask her for access. | 1. What is its function?<br>a. Necklace which carries household items<br>b. Belt which carries household items<br>c. Decorative wristlet<br>2. When was this item particularly popular?<br>a. Late 1800s in Russia<br>b. 20th century in North America<br>c. Middle ages around Europe<br>3. Who was used by mainly?<br>a. Woman<br>b. Children<br>c. Doctors<br>4. Where does its name mean?<br>a. House care<br>b. Lady of the castle<br>c. Lease of personal belongings<br>5. What was its social valence?<br>a. Slavery within the household<br>b. Political responsibilities<br>c. Women’s authority within the household |

#### *2.2.1.3 Procedure*

Procedure involved four main parts: invitation on Prolific, the video-call (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 minutes 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 video-call 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 whether they reviewed of any experimental item during the one week-delay period. This part lasted about 5-10 mins.

#### *2.2.1.4 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.

#### *2.2.1.5 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 S1. Statistical tests, p-values and Confidence Intervals are reported for all contrasts for both experiments in Table S2.

### 2.2.2 ***Results***

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 2.2A 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 ($F_{(1, 23)}=25.81$,

$p<.0001$, $\eta^2=.53$). There was also a main effect of *Social contingency* ($F_{(1,\ 23)}=33.34$, $p<.0001$, $\eta^2=.59$): 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* ($F_{(1,\ 23)}=1.28$, $p=.27$, $\eta^2=.05$). However, we found an interaction effect between *Social contingency* and *Social richness* ($F_{(1,\ 23)}=6.28$, $p=.017$, $\eta^2=.22$; 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 (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 ($t_{(23)}=2.15$, $p=.04$). 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.

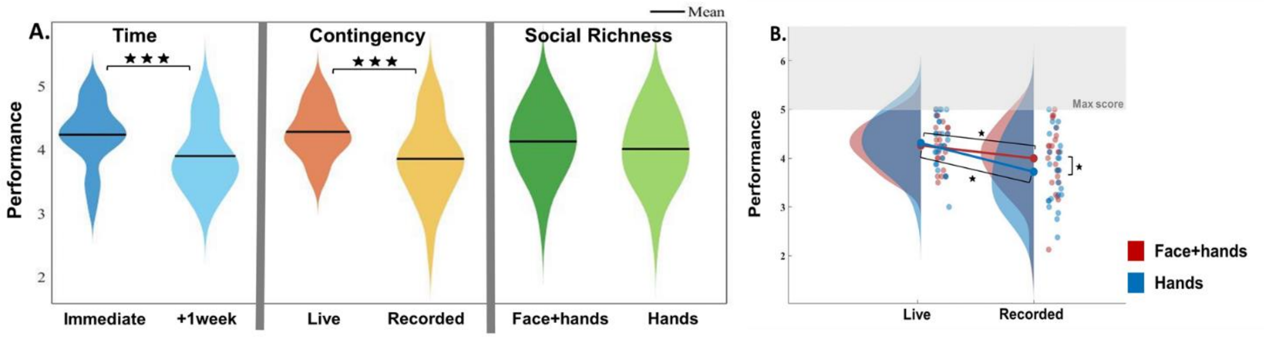


**Figure 2.2 Results for Experiment 1**

**A.** Main effects of Time, Contingency and View on learning performance. Significant effects of Time and Contingency were found, the symbol *** indicates $p<.0001$. **B.** Interaction effects. There was an interaction effect between Contingency and Richness $F(1, 23) = 6.28$, $p = .017$, η2=.22. The symbol * indicates significant difference as measured by paired t-test $p< .05$. Here, performance has been averaged across time (immediate and +1week test). The violin plots show the probability density function (Kernel density estimation), which can go beyond the smallest and largest data point. For clarity, we have indicated the max possible score on performance (x axis). Dots are showing individual scores.

## 2.3 Experiment 2

Experiment 2 was pre-registered on the 21st December 2020 on OSF (De Felice & Hamilton, 2020 10.17605/OSF.IO/NXS37).

### 2.3.1 *Methods*

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.

#### *2.3.1.1 Participants*

We used the software program G*Power to conduct a power analysis. From experiment 1, we used the minimum effect size of interest of $\eta^2 = .05$ (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.

Participants' recruitment, material, procedure, data pre-processing and data analysis for Experiment 2 was the same as described for Experiment 1.

### 2.3.2 *Results*

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* ($F_{(1, 26)}=30.68$, $p<.0001$, $\eta^2=.54$; Figure 3A). However, we did not find a main effect of *Social contingency* ($F_{(1, 26)}=1.67$, $p=.21$, $\eta^2=.06$) or of *Social richness* ($F_{(1, 26)}=.04$, $p=.84$, $\eta^2=.002$). Importantly, we replicated the interaction effect between *Social contingency* and *Social richness* ($F_{(1, 26)}=5.28$, $p=.03$, $\eta^2=.16$; Figure 3B).

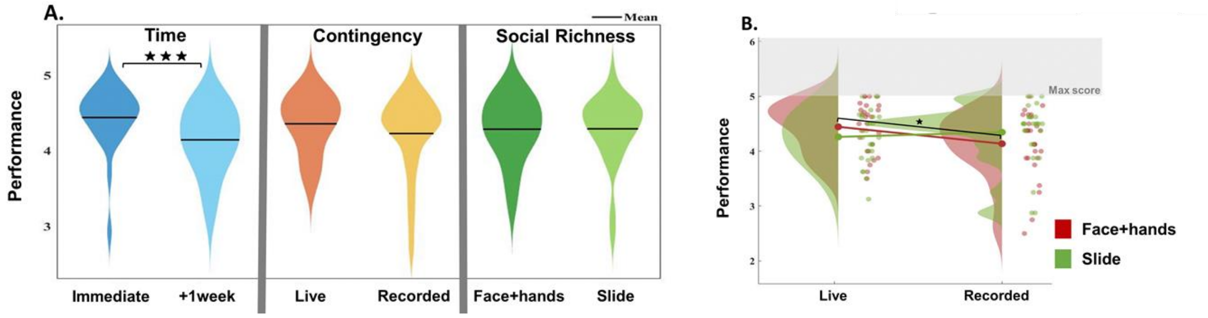


**Figure 2.3 Results for Experiment 2**

**A.** Main effects. A significant effect of Time was found, the symbol *** indicates $p<.0001$. **B.** Interaction effects. There was an interaction effect between Contingency and Richness $F(1, 26) = 5.28$, $p = .03$, $\eta2=.16$. The symbol * indicates significant difference as measured by paired t-test $p< .05$. Here, performance has been averaged across time (immediate and +1week test). The violin plots show the probability density function (Kernel density estimation), which can go beyond the smallest and largest data point. For clarity, we have indicated the max possible score on performance (x axis). Dots are showing individual scores.

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 2.2 and 2.3). 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 might lead to more learning than slides ($t_{(26)}= 1.77$, $p=.09$), while the opposite was observed in the recorded condition ($t_{(26)}=-1.87$, $p=.07$; see Table 2.3 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.

**Table 2.2 Sample size (N), Means and SDs for Experiment 1 and 2.**

Mean number of items recalled at test (max of 5) for each condition.

| **Experiment 1** | **N** | **Mean** | **SD** |
|---|---|---|---|
| Immediate – Live – Face+Hands | 24 | 4.46 | .42 |
| Immediate – Live – Hands | 24 | 4.52 | .45 |
| Immediate – Recorded – Face+Hands | 24 | 4.07 | .80 |
| Immediate – Recorded – Hands | 24 | 3.91 | .75 |
| +1week – Live – Face+Hands | 24 | 4.06 | .58 |
| +1week – Live – Hands | 24 | 4.10 | .69 |
| +1week – Recorded – Face+Hands | 24 | 3.92 | .66 |
| +1week – Recorded – Hands | 24 | 3.53 | .74 |
| **Experiment 2** | | | |
| Immediate – Live – Face+Hands | 27 | 4.56 | .43 |
| Immediate – Live – Slide | 27 | 4.45 | .55 |
| Immediate – Recorded – Face+Hands | 27 | 4.32 | .59 |
| Immediate – Recorded – Slide | 27 | 4.44 | .52 |
| +1week – Live – Face+Hands | 27 | 4.33 | .56 |
| +1week – Live – Slide | 27 | 4.07 | .63 |
| +1week – Recorded – Face+Hands | 27 | 3.95 | .70 |
| +1week – Recorded – Slide | 27 | 4.25 | .69 |

**Table 2.3 Statistics for all tests**

Results reported for Experiment 1 and 2 for all comparisons.

| Comparisons | F or T value | df | p-value | 95% CI | |
|---|---|---|---|---|---|
| | | | | Lower bound | Upper bound |
| ***Experiment 1 main effects*** | F | | | | |
| Time: Immediate vs +1week | 25.81 | 23 | .000 | .200 | .474 |
| Contingency: Live vs Recorded | 33.34 | 23 | .000 | .274 | .579 |
| Richness: Face+Hands vs Hands | 1.28 | 23 | .270 | -.094 | .321 |
| | | | | | |
| ***Experiment 1 simple effects*** | t | | | | |
| Live Face+Hands > Live Hands | -.448 | 23 | .66 | -.273 | .176 |
| Recorded Face+Hands > Recorded Hands | 2.15 | 23 | .04 | .011 | .540 |
| Live Face+Hands > Recorded Face+Hands | 2.99 | 23 | .007 | .081 | .448 |
| Live Hands > Recorded Hands | 5.61 | 23 | .001 | .372 | .805 |
| | | | | | |
| ***Experiment 2 main effects*** | F | | | | |
| Time: Immediate vs +1week | 30.68 | 26 | .000 | .186 | .405 |
| Contingency: Live vs Recorded | 1.67 | 26 | .208 | -.066 | .290 |
| Richness: Face+Hands vs Slide | -.039 | 26 | .844 | -.149 | .123 |
| | | | | | |
| ***Experiment 2 simple effects*** | t | | | | |
| Live Face+Hands > Live Slide | 1.77 | 26 | .088 | -.029 | .397 |
| Recorded Face+Hands > Recorded Slide | -1.87 | 26 | .073 | -.441 | .021 |
| Live Face+Hands > Recorded Face+Hands | 2.45 | 26 | .021 | .050 | .568 |
| Live Slide > Recorded Slide | -.72 | 26 | .476 | -.326 | .157 |

### 2.3.3 *Additional analysis*

As a supplement to our pre-registered ANOVA analysis, we used a Logistic Regression Model to evaluate what factors predicted recall on each individual question answered by participants. The model used the *glmer* function in the lme4 package in R (Bates et al., 2015). It 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 ($R2=.098 \pm .046$, $p<.01$). 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 ($R^2=.195 \pm .052$, $p<.0001$).

## 2.4 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 on learning differently depending on whether learning is interactive or observational (Figure 2.4): 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.

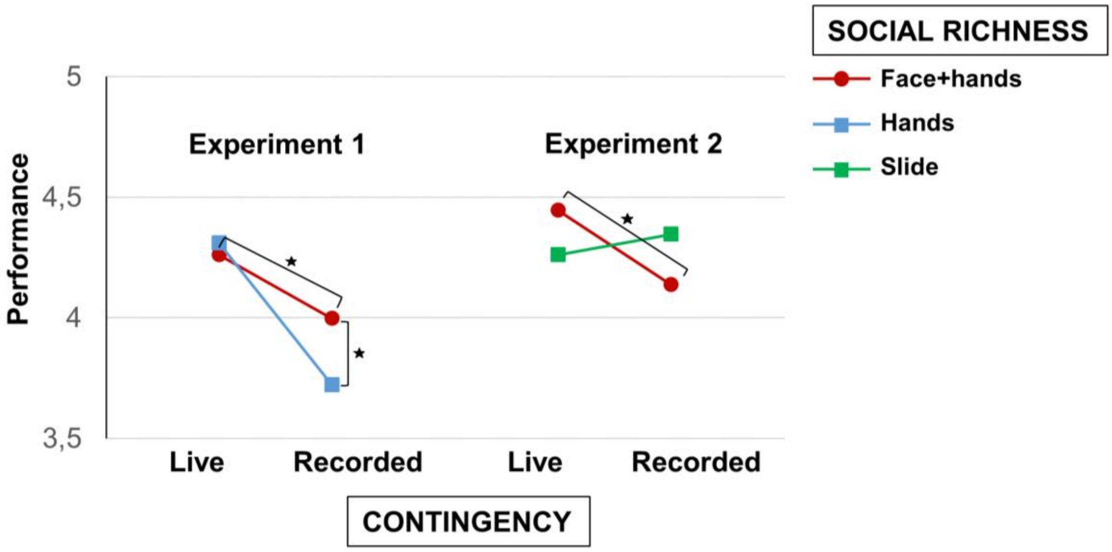


**Figure 2.4 Summary of results from experiment 1 and experiment 2**

Performance has been averaged across time (immediate and +1week test). For Experiment 1, we found a main effect of Contingency (F(1, 23)=33.34, p<.0001, η2=.59) and an interaction effect between Contingency and Social Richness (F(1, 23)=6.28, p=.017, η2=.22). Paired-sample t-tests revealed significant differences between Recorded-Face+hands and Recorded-Hands condition (t(23)=2.15, p=.04); Live-Face+hands and Recorded-Face+hands (t(23)=2.99, p=.007), and Live-Hands and Recorded-Hands (t(23)=5.61, p=.001). For Experiment 2, we found an interaction effect between Contingency and Social Richness (F(1, 26)=5.28, p=.03, η2=.16). Paired-sample t-tests revealed significant differences between Live-Face+hands and Recorded-Face+hands (t(26)=2.45, p=.02). We also observed some trends for Live-Face+hands vs Live-Slide (t(26)= 1.77, p=.09) and for Recorded-Face+hands vs Recorded-Slide (t(26)=-1.87, p=.07).

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 2.4): 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, Kuhl et al., 2003; Lauricella et al., 2011) and social contingency (Lauricella et al., 2011; Myers et al., 2017; O'Doherty et al., 2011; Roseberry et al., 2014) significantly improve learning in a variety of contexts (Calvert et al., 2007; Goldstein & Schwade, 2008; Troseth et al., 2006).

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 (John et al., 2016; Ramlogan et al., 2013; D. Zhang et al., 2006). 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. 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 (Dale et al., 2013; Dideriksen et al., 2020). While a systematic analysis of verbal and non-verbal behaviours observed during the sessions is beyond the scope of this study, 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 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 (Schertz et al., 2013), common ground (Bohn et al., 2019), shared intentionality (Sabbagh & Baldwin, 2001) or all these processes together in order to attune with the teacher(Schmitz, 2014). This attunement may allow information to be shared more effectively (Hu et al., 2015; Kuhl, 2007).  Rich visual cues may enable stronger attunement by providing more information about the interaction partner's gaze and mental states (Kajopoulos et al., 2020; Marotta et al., 2012).  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

(Kourtis et al., 2020) and social cues may become distracting (Friedman & Förster, 2010; Kirkorian, 2016; Phillips et al., 2007). 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), 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 (Hamilton & Lind, 2016), 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 (Rice & Redcay, 2016; Seuren et al., 2021).

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 behavioural 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-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 (De Jaegher et al., 2010). Future studies using virtual reality might be able to do so (X. Pan & Hamilton, 2018), 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, 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 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 (John et al., 2016; Ramlogan et al., 2013). 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). 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 (Guerin, 1986), which in turn could either significantly improve learning (Lytle et al., 2018), or making it harder (Skuballa et al., 2019). Given that our design only involves one student, we do not know how our results generalise 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 optimise learning in real-life educational contexts.

# 3. Chapter 3 – Learning online in adults with Autistic Spectrum Condition

Design, hypotheses and analysis plan for Experiment 2 of this chapter was pre-registered on 30th November 2022 on OSF (De Felice & Hamilton, 2022 https://archive.org/details/osf-registrations-5pga3-v1).

*Contribution notes*: MSc student Anna Hatilova collected the raw data for Experiment 1. Medical student Filip Trojan and BSc student Iris Tsui collected raw data for Experiment 2.

## 3.1 Background

In the previous chapter we have shown how social interaction can catalyse learning in a online context, with a strong learning advantage for material learned via a live video-call compared to pre-recorded videos. We identified *social contingency* (the bi-directional and time-dependent exchange between a sender and a receiver) as the critical predictor of such learning advantage. Our conclusions however are limited to neurotypical (NT) adults. In this chapter, we asked whether the same live-learning advantage would be replicated in a sample of adults with Autistic Spectrum Condition (ASC).

The question of whether people with ASC benefit from interactive learning is interesting for at least two main reasons. The first one is more practical and directly relates to pedagogy: ASC is often associated with poor outcomes at school. Despite this, studies on the experience of education (both online and offline) on ASC are scarce and limited to self-report questionnaires, and results are inconsistent (Adams et al., 2019; Odom et al., 2015). Experimentally testing the effect of interactive learning on ASC would therefore contribute to fill this gap and directly inform educational practices.

The second reason to motivate this study is more theoretical, with implications for cognitive models of both social learning and ASC. ASC is a heterogeneous group of neurodevelopmental conditions manifesting in infancy or early childhood (American Psychiatric Association, 2013). While the constellation and severity of symptoms can vary significantly across individuals, the main common feature of ASC is disruption of processes crucial in communication and social interaction, including implicit imitation (Bernier et al., 2007; Edwards, 2014), joint attention (Roos et al., 2008), pragmatic language use (Philofsky et al., 2007; Whitehouse et al., 2007) and affect sharing (Abell et al., 2000; Happé & Frith, 2014).

Furthermore, Vivanti and Rogers (2014) identified three aspects of social learning, which may be specifically impaired in ASC. Namely, they claim that social learning is characterised by being 1) implicit (without explicit instructions, e.g. via imitation), 2) intrinsically rewarding and 3) flexible (use social signals adaptively according to the communicative contexts, e.g. Senju et al., 2013). While the implicit aspect of Vivant & Rogers' definition of social learning may be less relevant here – as we are interested in declarative learning of non-social knowledge (see Introduction, Chapter 1) –, the rewarding and the flexible nature of social exchange may be part of the interactive-learning advantage (or the lack of it) over less interactive contexts. In fact, clinical and experimental observations found that people with ASC show disrupted processing on both social rewarding (Corbett et al., 2014) and flexible behaviour (Semrud-Clikeman et al., 2010).

The social motivation hypothesis argues that people with ASC engage less in social contexts as they do not find these as rewarding as NT adults (Chevallier et al., 2012; G. Dawson et al., 1998; Mundy, 1995; Mundy & Newell, 2007): in other words, according to this hypothesis, people with ASC are less drawn to attend and process social signals (and therefore benefit from them) as they do not enjoy it as much, with consequences for their cognitive development and ultimately resulting in cognitive deficits of the social sphere (Corbett et al., 2014). This idea is supported by studies showing that people with ASC find social stimuli less attractive than non-social ones, in contrast to what observed in NT people (Čeponiene et al., 2003; Dubey et al., 2015; Klin, 1991; Riby & Hancock, 2008). However, a meta-analysis of 13 fMRI studies found disrupted rewarding processes in people with ASC for both social and non-social signals (Clements et al., 2018), suggesting that there may be a *global* deficit in reward processing.

Other evidence suggests that people with ASC cannot flexibly adapt their attention to use social signals in social learning contexts (Spengler et al., 2010). This may be due to attentional processes and repetitive and restrictive interests, which may expand beyond the social world. Sasson et al. (2008) presented NT children and children with ASC with some complex scenes of both social and non-social stimuli. They found that on average, children with ASC looked at fewer images, fixated each image for longer and were much more detailed-oriented. Crucially, they did not report any difference between social and non-social stimuli. Similarly, a study on eye-gaze and word-learning found that children with ASC were able to follow the gaze of the instructor as well as NT children, but they use it in qualitatively different ways: while NT children were able to infer the social informativeness of gaze (resulting in higher semantic learning of the object-word), children with ASC used eye-gaze mainly as an associative aid to consolidate phonological feature of the word, and showed significantly less contingent gaze-to-object fixation (Norbury et al., 2010).

Taken together, this evidence seem to predict that people with ASC would fail to show an interactive-learning advantage (either because live interaction is less rewarding or because it may be harder to attend to compared to a less interactive situation). However, this conclusion is mainly drawn from evidence on *implicit* learning of *social* material (e.g. language, Fitch et al., 2010; Holtgraves & Kashima, 2008; Seyfarth & Cheney, 2014). It is not clear whether the same predictions can refer to *explicit, declarative* learning of *non-social* material *via* social exchange. The question of how people with ASC acquire new factual knowledge in social interaction remains largely unexplored empirically.

In fact, the object of investigation here is how *social interaction* supports learning of factual *non-social* knowledge, in contrast to other forms of social learning, namely motor learning, observational learning and imitation (Bandura, 2019, also see Introduction, Chapter 1). The latter forms of social learning have been extensively studied both in children and adults with ASC (M. Dawson et al., 2007; Pearson, 2004; Webb et al., 2017), and overall results are consistent in finding idiosyncratic processing associated with ASC. What remains unclear is whether acquisition of *non-social* knowledge benefits from *social interaction* equally in people with ASC as in NT (Chapter 2). To the best of our knowledge, there is no experimental work that has investigated this.

In this study, we compare learning performance during three different social learning conditions (one live and two recorded), where learning-content is always delivered online by a (human) teacher (similar to Study 1, Chapter 2). In all conditions, participants are explicitly instructed to learn facts about obscure items (e.g. exotic animals). In the live condition, the participant joins a live video-call where they can interact with the teacher. In the recorded-observant condition, the participant learns the material from a pre-recorded video of a previous session (observing a previously recorded student-teacher interaction). In the recorded-alone condition, the participant learns the material from a pre-recorded video of the teacher alone.

We expect to replicate results from Study 1 (Chapter 2), showing a live-learning advantage in NT. In line with our results from Study 1 (Chapter 2) that showed that social cues are beneficial for learning specifically during interactive learning, but not during observational leaning (also see Tylén et al., 2012), we also expect the recorded-observant condition to be associated with the lowest performance. With regards to ASC, according to the literature showing social cognition deficits in ASC, one could speculate that no advantage in the live condition would be observed in this group. However, this speculation would be based on previous studies which mainly looked at implicit learning of social content, in contrast to the present study, which focuses on declarative interactive learning of non-social knowledge. We therefore do not have strong predictions about ASC performance. Here we first present results from a pilot study (N=46), and then show findings from a larger pre-registered replication (N=82).

## 3.2 Experiment 1

### 3.2.1 *Methods*

#### *3.2.1.1 Design*

This study aims to investigate whether 1) participating in a live learning session improves learning online compared to recorded videos of either a previous interaction or of a teacher alone and ii) whether these conditions impact learning differently in adults with ASC compared to neurotypicals. To answer these questions, this study adopted a 2 (group) x3 (learning condition) x2 (time) repeated-measures design, with between- and within-subjects factors. The between-subjects factor is group (Autistic Spectrum Condition (ASC) vs neurotypical (NT)), the within-subjects factors are i) learning condition (live vs recorded of another social learning episode vs recorded of the teacher alone), and ii) time of recall (immediate vs delay quiz). Specifically, facts about 15 items were presented with two minutes per item (see materials below and Study 1, Chapter 2). Five items were assigned to each condition: 1) live condition: participant participated in a live video-call when they learned in interaction with the teacher; 2) recorded-observant condition: participant was shown a pre-recorded video of the teacher presenting the learning material to a student (confederate); 3) recorded alone condition: participant was shown a pre-recorded video of the teacher alone presenting the learning material (Figure 3.1). Learning score (outcome measure) for each participant was obtained from a multiple-choice quiz (see Materials and Study 1, Chapter 2).

Items assigned to each condition and trial order within each condition remained fixed for the whole duration of this experiment (see below for details on materials). Order of conditions were randomised across participant.

Live

Recorded-Observant

Recorded-Alone

**Figure 3.1 Schematics of the three experimental conditions**
Learning sessions are represented as appearing to participants. From left to right: Live condition, participant learn about 5 items as they interact with the teacher in a real-time video-call; Recorded-observant condition: participant learn about 5 items from a pre-recorded sessions with a confederate acting as a previous participant; Recorded-Alone condition: participant learns about 5 items from a pre-recorded session of the teacher alone. In each condition, participant learns about 5 different items. Items were assigned to each condition randomly and remained fixed within each experiment, and counterbalanced between experiments.

### *3.2.1.2 Materials*

A selection of 15 items were selected from Study 1 (Chapter 2), three from the exotic food category (*Rambutan, Kiwano, Cherimoya*), four from the antiques category (*Strigil, Porte-joupe, Scotch Hands, Chatelaine*), four from the animals category (*Tarsier, Axolotl, Glaucus, Anhinga*) and four from the musical instruments category (*Kalimba, Caxixi, Agogo, Hulusi*). Each condition presented a mixture of objects from these categories, which remained fixed for all participants for this experiment, as follow: 1) live condition [*Tarsier, Kalimba, Strigil, Axolotl, Rambutan*]; 2) recorded-observant condition [*Porte-joupe, Kiwano, Caxixi, Scotch Hands, Glaucus*]; 3) recorded-alone condition [*Agogo, Cherimoya, Anhinga, Chatelaine, Hulusi*]. Learning was tested via the same multiple-choice quiz used in Study 1 (Chapter 2). Full details of item information and multiple-choice quiz are reported in Table 1.

#### *3.2.1.3 Procedure & participants recruitment*

This study was approved by the UCL ethic committee. Participants were recruited via the online platform Prolific (www.prolific.co). The platform retains demographic details as well as information on any disabilities/diagnosis of users, as reported by the users at the time of account registration. Such anonymous information can be used by researchers to create adverts targeting a specific pool. Two separate adverts were published: one targeting neurotypical participants and one targeting people with Autistic Spectrum Disorder (ASC). As a further check, users who responded to our adverts were asked to confirm their diagnosis via a questionnaire on Gorilla Experiment Builder (www.gorilla.sc). To ensure that the experimenter was blind to participants' diagnosis, recruitment was done by a researcher who was not involved in data collection.

To be eligible, all participants had to i) be fluent in English (speaking English regularly for >5 years); ii) be aged 18-65; iii) give consent to having their camera and microphone on; and iv) give consent to being recorded for the whole duration of the experiment. Participants were paid at the hourly rate of £7.50 for a total of £15 over two hours. An additional £3 were offered for those who completed a 10 min quiz a week later.

Participants who responded to our advert were asked to complete four main parts: 1) background battery (independently online, on Gorilla Experiment Builder), 2) learning session (over a video-call), 3) online learning multiple-choice quiz immediately after the learning session (independently online on Gorilla Experiment Builder), and 4) repeat the quiz a week later.

#### *3.2.1.4 Background Battery*

Users who responded to the Prolific adverts were redirected to Gorilla Experiment Builder (www.gorilla.sc) where they received instructions on the study and gave consent for participation. They then completed the Background Battery tasks. This comprises of i) Spot-the-word test, a measure of verbal fluency (Baddeley et al., 1993), ii) matrix reasoning item bank (MaRs-IB), a measure of non-verbal reasoning (Chierchia et al., 2019) and iii) Animated Triangle test, a measure of mentalising (Abell et al., 2000; Livingston et al., 2021; White et al., 2011).

Upon completion of the Background battery task, an independent researcher sent the participant ID to the experimenter (teacher), who arranged a video-call with the participant (via Prolific chat), while remaining blind to their diagnosis.

#### *3.2.1.5 Video-call*

The experimenter greeted the participant and checked that audio and video worked properly. The participant was asked to open the zoom window in full-screen mode and choose the gallery view (i.e. everyone in the call is shown equal size, this ensured that view during live session was comparable to view during pre-recorded video watching). Participants were told that the aim of the study was to investigate how people learned online and whether this differed in people with ASC, and were asked not to disclose any personal information to the teacher, who was blind to their diagnosis. The experimenter then explained that the participant will learn some facts about exotic food, animals, antiques and rare musical instruments over three formats: in live interaction with the teacher (live condition), and through watching of pre-recorded videos showing either the teacher with a previous participant (recorded-observant condition) or the teacher alone (recorded-alone condition). They were instructed to memorize as much information as possible, as at the end of the video-call they will complete a multiple-choice quiz to test their learning. During the live condition, participants were told that they were free to ask questions and interact with the teacher. Before starting the learning sessions, participant’s pre-knowledge was tested. If any item was known, it was excluded from the analysis (but not from the learning session). Learning sessions started with either the live, recorded-alone or recorded-observant condition, in a counterbalanced and semi-randomised order. The whole call lasted approximately 40 minutes (i.e. 10 minutes per condition, with 2 min per item and five items in each condition, plus 10 minutes for instructions).

#### *3.2.1.6 Learning quiz*

Immediately after the learning session, participants were redirected to Prolific, where their IDs were included into a ‘white list’, so that a new advert was visible to them only. By replying to that advert, participants were redirected to Gorilla Experiment Builder (www.gorilla.sc), where they reported on the quality of the video call (audio and video), before completing the learning quiz. After the learning quiz, they also completed an ‘enjoyment questionnaire’ and inclusion of other in self-questionnaire (Aron et al., 1992). This part lasted approximately 20 minutes and was completed by the participant independently (note that the ‘immediateness’ of the quiz was ensured by the experimenter who terminated the video-call only after participant initiated the quiz part on Gorilla Experimenter Builder).

Exactly one week after the learning sessions, participants invited through Prolific to the final stage of the experiment, and directed to Gorilla Experimenter Builder to complete the same learning quiz. Additionally, participants filled in a history questionnaire, to check for potential revision of any of the items (e.g. search on Google). This part lasted approximately 10 minutes.

### 3.2.2 ***Results***

#### *3.2.2.1 Sample*

53 participants took part in the study (Table 2). Participants were excluded when reporting 3 or less on a 1 (poor) to 5 (excellent) video-call quality scale (N=3), and being visibly distracted during the video-call (N=1). Of the remaining 49 participants, 46 ($N_{ASC}$=20, male=11, female=4, non-binary=4; $N_{NT}$=26, male=11, female=15) completed the full experiment, including the one-week delay quiz (see 2.1 Design and 2.3 Procedure). We lost demographic data from one participant (ASC group) due to a technical fault.

Participant with ASC either received a diagnosis by a clinician (N=18) or were self-diagnosed (N=2). ASC and NT group did not differ on age ($\text{mean}_{ASC}$ (sd) = 27.79 (9.22), $\text{mean}_{NT}$ (sd) = 29.85 (9.90), $t_{(43)}$=.71, $p$=.48), verbal fluency (Spot the word test, $\text{mean}_{ASC}$ (sd) = 47.63 (6.71), $\text{mean}_{NT}$ (sd) = 44.73 (6.23), $t_{(43)}$=-1.48, $p$=.15) non-verbal reasoning (MaRs-IB, $\text{mean}_{ASC}$ (sd) = 64.48 (18.57), $\text{mean}_{NT}$ (sd) = 61.64 (16.98), $t_{(43)}$=-.52, $p$=.60) and mentalising test (Animated Triangle, $\text{mean}_{ASC}$ (sd) = 9 (2.54), $\text{mean}_{NT}$ (sd) = 9.33 (1.92), $t_{(43)}$=.38, $p$=.71). ASC scored significantly higher on AQ than NT ($\text{mean}_{ASC}$ (sd) = 33.37 (6.73), $\text{mean}_{NT}$ (sd) = 19.19 (7.29); $t_{(43)}$ = -6.73, $p$<.0001).

#### *3.2.2.2 Data pre-processing*

Single trials were excluded when: i) participants reported that they knew the item; ii) internet connection dropped during the single trial but was good for the rest of the experiment; iii) the experimenter reported incorrect information about the item; iv) the participant reported revising information about a given item before the delay quiz (excluded from delay performance only). Performance was calculated for each learning condition separately, as an average over the included trials (score= points collected on all trials / total points available on all included trials).

#### *3.2.2.3 Analysis of Variance*

An analysis of variance (ANOVA) was run to test the difference in learning performance between 2 (groups: ASC and NT) x 3 (learning conditions: live, recorded-alone and recorded-observant) x 2 (time of learning quiz: immediate and delay). Means and SD for all conditions are reported in Table 3. Results for main and interaction effects are reported in Table 4 and Figure 2.

*Main effects.* Findings show a main effect of time: unsurprisingly, people remembered more things straight after the learning session (mean (sd) = 4.2(.58) than a week later (Mean=3.8, *sd*=.73; $F_{(1,44)}$=56.16, $p<.001$, $\eta^2$=.56, large effect size; Bakeman, 2005). More interestingly, we found a main effect of learning condition ($F_{(2,43)}$=3.86, $p=.03$, $\eta^2$=.15, medium effect size; Bakeman, 2005). The pairwise comparison revealed a significant learning advantage associated with Live compared to Recorded-alone condition ($M_{\text{Live}}$ (sd) = 4.1 (.09); $M_{\text{Recorded-alone}}$ (sd) = 3.9 (.1); $t_{(44)}$ = .19 *p* = *.008*). No other significant difference between learning conditions were found. No main effect was found for group: in other words, people with ASC showed a NT-equivalent performance ($\text{mean}_{\text{ASC}}$ (sd) = 4.01 (.09); $\text{mean}_{\text{NT}}$ (sd) = 3.99 (.1); $t_{(44)}$=-.01 *p*=*.93*).

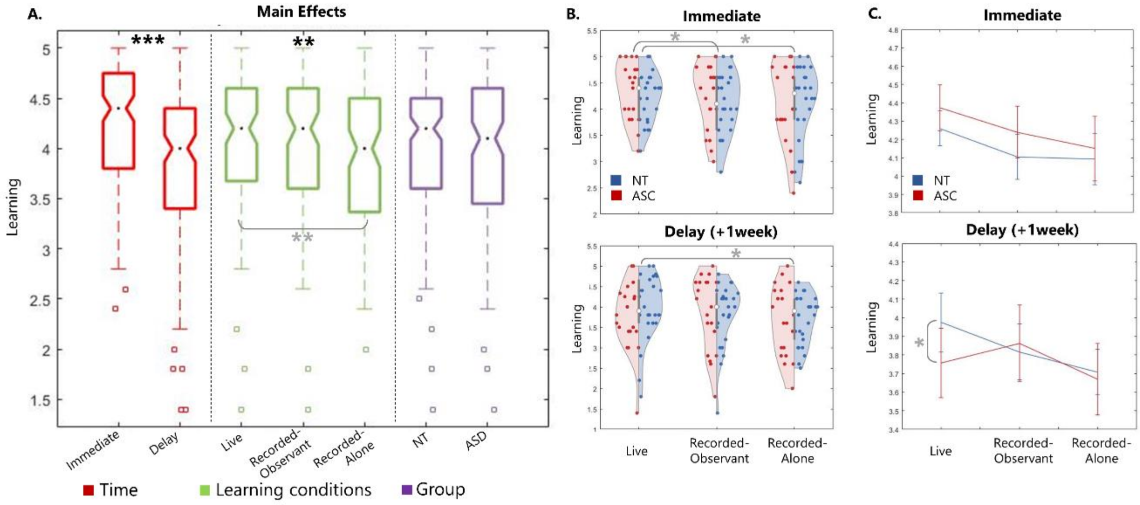


**Figure 3.2 Results for Experiment 1**

Results from dataset of Experiment 1. *p<.05; **p<.001;***p<.0001 **A.** Boxplots of the three main factors of interest: Time, Learning condition and Group. **B.** Violin plots of learning performance immediately after the learning session (top) and a week later (bottom), for the three learning conditions. Violins are split in half showing the distribution of NT (blue) and ASD (red) sample separately. **C.** Line plot for learning performance immediately after the learning session (top) and after one week (bottom), plotted separately for NT (blue) and ASD (red). Error bars represent the standard error of the mean.

*Interaction effects.* No significant interaction effects where found between the main factors of interest (group, learning conditions and time). However, visualization of the data (Fig. 1.C) revealed a trend specific to ASC group: while immediate recall showed a similar pattern across conditions between groups, delayed recall dropped specifically for items learned during the Live condition for ASC group. A 2(group) x2(time) was therefore run to test the hypothesis that delay performance was significantly more affected for ASC compared to NT specifically for things learned during Live condition. Results revealed a group*time interaction effect: $F_{(1,44)}$=4.88, *p*=*.03*, $\eta^2$=.1): for things learned during Live condition, a week later people with ASC recalled significantly less things compared to NT ($mean_{ASC}$ (sd) = 3.75 (.18); $mean_{NT}$ (sd) = 3.97 (.16)).

#### *3.2.2.4 Conclusions from Experiment 1 & Hypothesis for Experiment 2*

Experiment 1 found that for both NT and people with ASC, learning during Live session was associated with better recall both immediately after the session and one week later. In addition, it was found that the ASC group exhibited a decline in recall for items learned over Live interaction specifically, to a significantly greater extent than to that observed in the NT group.

Based on these results, a follow-up experiment was pre-registered (https://archive.org/details/osf-registrations-5pga3-v1) to confirm two main hypotheses:

1. Participants from both groups will learn more from live calls (Live condition) compared to pre-recorded video calls (Recorded-alone and Recorded-observant condition).
2. There will be an interaction between learning condition, group and time: while neurotypical adults will show a consistent advantage for interactive learning (Live condition) over time, the ASC group will show better immediate learning for material learnt in Live condition, and better long-term learning for materials learned from pre-recorded videos (Recorded-alone and Recorded-observant condition).

Experiment 2 consisted of two equal sub-experiments, which followed the same methods and procedure as experiment 1. Each sub-experiment was run by a different teacher.

**Table 3.1 Demographics for dataset from Experiment 1, 2 and combined**

Sample size (N), age, AQ score and performance on background battery tests for neurotypical (NT) and Autistic Spectrum Condition (ASC) group. Note that higher the AQ score, greater the autistic traits. 1. (Baddeley et al., 1993) 2. (Chierchia et al., 2019) 3. (Livingston et al., 2021) 4. (Baron-Cohen et al., 2001).

| **Experiment 1 (N=46)** | **Neurotypical (NT) N=26** | | **Autistic Spectrum Condition (ASC) N=20** | | **NT vs ASC** | |
|---|---|---|---|---|---|---|
| | **Mean** | **SD** | **Mean** | **SD** | **t-test** | ***p*** |
| Age | 29.85 | 9.90 | 27.79 | 9.22 | .71 | .48 |
| Verbal fluency (Spot the word)[1] | 44.73 | 6.23 | 47.63 | 6.71 | -1.49 | .14 |
| Non-verbal reasoning (Matrix reasoning item bank, MaRs-IB)[2] | 61.64 | 16.98 | 64.48 | 18.57 | -.52 | .60 |
| Mentalising (Animated Triangle)[3] | 9.62 | 1.87 | 8.74 | 2.35 | .88 | .18 |
| Autistic Quotient (AQ)[4] | 19.19 | 7.29 | 33.37 | 6.73 | -6.65 | <.0001 |
| **Experiment 2 (N=82)** | **Neurotypical (NT) N=42** | | **Autistic Spectrum Condition (ASC) N=40** | | **NT vs ASC** | |
| | **Mean** | **SD** | **Mean** | **SD** | **t-test** | ***p*** |
| Age | 27.50 | 4.91 | 27.63 | 5.13 | .13 | .89 |
| Verbal fluency (Spot the word)[1] | 44.54 | 8.46 | 45.73 | 9.83 | -.59 | .56 |
| Non-verbal reasoning (Matrix reasoning item bank, MaRs-IB)[2] | 61.37 | 17.26 | 63.08 | 18.68 | -.43 | .67 |
| Mentalising (Animated Triangle)[3] | 8.93 | 2.22 | 9.37 | 1.85 | -.44 | .33 |
| Autistic Quotient (AQ)[4] | 19.98 | 6.74 | 28.39 | 8.60 | -4.93 | <.0001 |
| **Combined (N=128)** | **Neurotypical (NT) N=68** | | **Autistic Spectrum Condition (ASC) N=60** | | **NT vs ASC** | |
| | **Mean** | **SD** | **Mean** | **SD** | **t-test** | ***p*** |
| Age | 28.49 | 7.27 | 27.58 | 6.62 | .73 | .46 |
| Verbal fluency (Spot the word)[1] | 44.61 | 7.63 | 46.33 | 8.95 | -1.17 | .24 |
| Non-verbal reasoning (Matrix reasoning item bank, MaRs-IB)[2] | 61.47 | 17.02 | 63.52 | 18.50 | -.65 | .51 |
| Mentalising (Animated Triangle)[3] | 9.19 | 2.10 | 9.17 | 2.03 | .08 | .94 |
| Autistic Quotient (AQ)[4] | 19.67 | 6.91 | 29.97 | 8.33 | -7.60 | <.0001 |

## 3.3 Experiment 2

### 3.3.1 ***Results***

#### *3.3.1.1 Sample*

86 participants took part in this study (Table 2), split equally between two researchers playing the role of the teacher. Each researcher completed the recruitment for the other, so that each teacher was blind to the diagnosis of the student during data collection. Participants were excluded when reporting 3 or less on a 1 (poor) to 5 (excellent) video-call quality scale (N=4). The final sample included 82 participants ($N_{ASC}$=41, male=17, female=20, non-binary=4; $N_{NT}$=41, male=12, female=27, non-binary=2).

Participants in the ASC group either received a diagnosed by a clinician (N=13) or were self-diagnosed (N=28; see section below and appendix for further analysis excluding the self-diagnosed participants). The ASC and NT group did not differ on age ($mean_{ASC}$ (sd) = 27,49 (5,13), $mean_{NT}$ (sd) = 27,63 (4,91), $t_{(80)}$=.13, $p$=.89), verbal fluency (Spot the word, $mean_{ASC}$ (sd) = 45.73 (9.83), $mean_{NT}$ (sd) = 44.54 (8.46), $t_{(80)}$=-.59, $p$=.56), non-verbal reasoning (MaRs-IB, $mean_{ASC}$ (sd) = 63.08 (18.68), $mean_{NT}$ (sd) = 61.37 (17.26), $t_{(80)}$=-.43, $p$=.67) and mentalising test (Animated Triangle, $mean_{ASC}$ (sd) = 9.37 (1.85), $mean_{NT}$ (sd) = 9 (2.20), $t_{(43)}$=-.81, $p$=.42). ASC scored significantly higher on AQ than NT ($mean_{ASC}$ (sd) = 28.39, (8.6), $mean_{NT}$ (sd) = 19.98 (6.74); $t_{(80)}$=-6.73, $p$<.0001).

**Table 3.2 Sample size (N), Means and SDs for all conditions for Experiment 1, 2 and combined**

Mean number of items recalled at test (max of 5) for each condition for neurotypical (NT) and Autistic Spectrum Condition (ASC) group for each Experiment and for combined datasets.

| **Experiment 1 – Recall scores** | | | |
|---|---|---|---|
| ***Neurotypical (NT)*** | **N** | **Mean** | **SD** |
| Immediate: Live | 26 | 4.26 | .10 |
| Immediate: Recorded-observant | 26 | 4.10 | .12 |
| Immediate: Recorded-alone | 26 | 4.09 | .15 |
| +1week: Live | 26 | 3.98 | .16 |
| +1week: Recorded-observant | 26 | 3.82 | .16 |
| +1week: Recorded-alone | 26 | 3.71 | .14 |

| ***Autistic Spectrum Condition (ASC)*** | N | Mean | SD |
|---|---|---|---|
| Immediate: Live | 20 | 4.37 | .12 |
| Immediate: Recorded-observant | 20 | 4.24 | .14 |
| Immediate: Recorded-alone | 20 | 4.15 | .17 |
| +1week: Live | 20 | 3.75 | .18 |
| +1week: Recorded-observant | 20 | 3.86 | .19 |
| +1week: Recorded-alone | 20 | 3.67 | .16 |
| **Experiment 2 – Recall scores** | | | |
| ***Neurotypical (NT)*** | **N** | **Mean** | **SD** |
| Immediate: Live | 42 | 4.33 | .08 |
| Immediate: Recorded-observant | 42 | 3.93 | .10 |
| Immediate: Recorded-alone | 42 | 4.15 | .10 |
| +1week: Live | 42 | 3.84 | .12 |
| +1week: Recorded-observant | 42 | 3.58 | .13 |
| +1week: Recorded-alone | 42 | 3.75 | .14 |
| ***Autistic Spectrum Condition (ASC)*** | **N** | **Mean** | **SD** |
| Immediate: Live | 42 | 4.46 | .08 |
| Immediate: Recorded-observant | 42 | 4.13 | .10 |
| Immediate: Recorded-alone | 42 | 4.33 | .10 |
| +1week: Live | 42 | 4.01 | .13 |
| +1week: Recorded-observant | 42 | 3.81 | .13 |
| +1week: Recorded-alone | 42 | 4 | .14 |
| **Combined – Recall scores** | | | |
| ***Neurotypical (NT)*** | **N** | **Mean** | **SD** |
| Immediate: Live | 68 | 4.30 | .06 |
| Immediate: Recorded-observant | 68 | 4 | .08 |
| Immediate: Recorded-alone | 68 | 4.13 | .08 |
| +1week: Live | 68 | 3.89 | .10 |
| +1week: Recorded-observant | 68 | 3.67 | .10 |
| +1week: Recorded-alone | 68 | 3.75 | .10 |
| ***Autistic Spectrum Condition (ASC)*** | **N** | **Mean** | **SD** |
| Immediate: Live | 60 | 4.43 | .07 |
| Immediate: Recorded-observant | 60 | 4.16 | .08 |
| Immediate: Recorded-alone | 60 | 4.27 | .09 |
| +1week: Live | 60 | 3.93 | .10 |
| +1week: Recorded-observant | 60 | 3.83 | .11 |
| +1week: Recorded-alone | 60 | 3.89 | .11 |

#### *3.3.1.2 Analysis of Variance*

An analysis of variance (ANOVA) was run to test the difference in learning performance between 2 (groups: ASC and NT) x 3 (learning conditions: live, recorded-alone and recorded-observant) x 2 (time of learning quiz: immediate and delay). Mean and SD for each conditions are reported in Table 3. Results for main and interaction effects are reported in Table 4 and Figure 3.

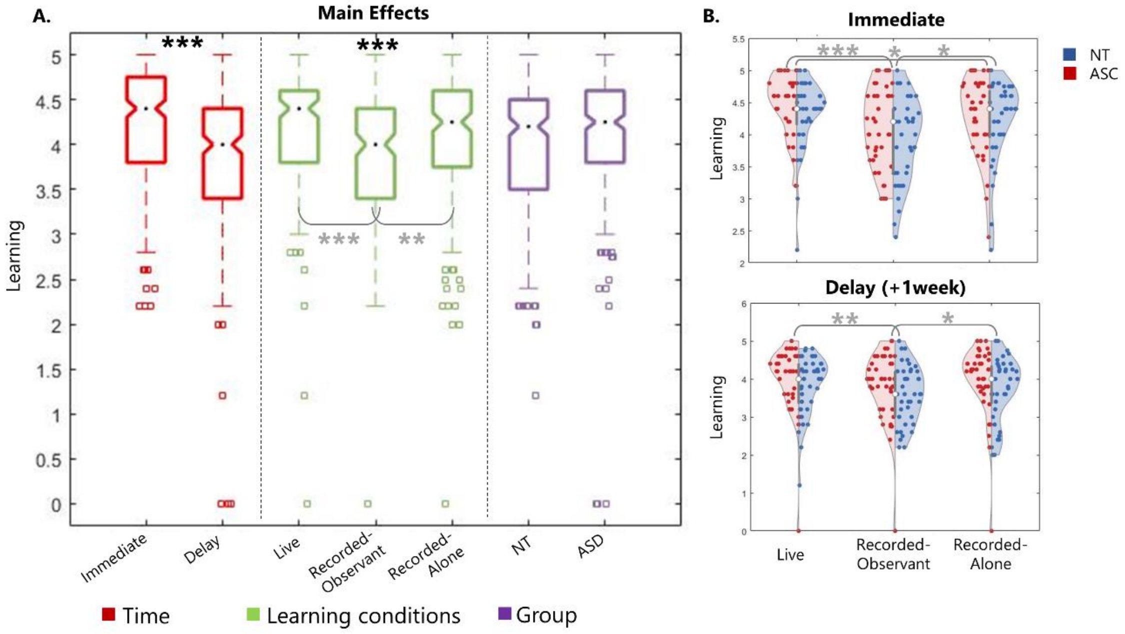


**Figure 3.3 Results for Experiment 2**

Results from dataset of Experiment 2. $*p<.05$; $**p<.001$;$***p<.0001$ **A.** Boxplots of the three main factors of interest: Time, Learning condition and Group. **B.** Violin plots of learning performance immediately after the learning session (top) and a week later (bottom), for the three learning conditions. Violins are split in half showing the distribution of NT (blue) and ASC (red) sample separately.

*Main effects.* Findings show a main effect of time: unsurprisingly, people remembered more things straight after the learning session (mean (sd) = 4.22 (.6)) than a week later (mean (sd) = 3.83 (.08); $F_{(1,80)}=38.56$, $p<.0001$, $\eta^2=.32$, large effect size; Bakeman, 2005). More interestingly, there was a main effect of learning condition ($F_{(2,80)}=13.53$, $p<.0001$, $\eta^2=.15$, large effect size; Bakeman, 2005). Pairwise comparisons revealed that Live interaction condition was the one associated with the highest learning, while Recorded-observant condition was associated with the worst learning: specifically, there was a significant learning advantage associated with Live compared to Recorded-observant condition ($M_{Live}$ (sd)= 4.16 (.06); $M_{Recorded-observant}$ (sd) = 3.86 (.07); $t_{(80)}=.29$, $p<.0001$), and an advantage approaching significance level compared to Recorded-alone condition ($M_{Recorded-alone}$ (sd) = 4.06 (.08); $t_{(80)}=.10$, $p=.08$). Recorded-observant condition was associated with significantly worse performance than Recorded-alone condition ($t_{(80)}$=-.19, *p=.001*). No main effect was found for group: in other words, people with ASC showed NT-equivalent performance ($M_{ASC}$ (sd) = 4.12 (.09); $M_{NT}$ (sd) = 3.93 (.09); $t_{(44)}$=.19, *p=.13*).

*Interaction effects*. No significant interaction effects where found between the main factors of interest (group, learning conditions and time). Data visualisation showed a similar pattern across groups and times, with Recorded-observant condition producing the worst learning performance in both groups.

To summarise, this experiment found support for hypothesis 1, showing that for both NT and people with ASC, learning during Live session was associated with better recall over time. In contrast, hypothesis 2 was not supported: ASC and NT group showed the same pattern of performance over time, with learning over Recorded-observant condition being associated with the worst performance for both groups.

**Table 3.3 Statistical results for all tests for Experiment 1, 2 and combined**

ANOVA results for experiment 1, experiment 2 and for the combined dataset. Factors of interest: Group (Neurotypical vs Autistic Spectrum Condition), Learning Condition (Live vs Recorded-observant vs Recorded-alone) and Time (Immediate vs Delay).

| **Comparisons** | **F or T value** | **df** | ***p*** | **Partial Eta Squared** | **95% CI** | |
|---|---|---|---|---|---|---|
| | | | | | **Lower bound** | **Upper bound** |
| **Experiment 1 (N=46)** | **F** | | | | | |
| Main effects | | | | | | |
| Group | .007 | 1 | .93 | .00 | -.39 | .36 |
| Learning condition | 4.06 | 2 | .016 | .09 | -- | -- |
| Time | 56.16 | 1 | <.0001 | .56 | .30 | .51 |
| Interaction effects | F | | | | | |
| Group*Time | 2.56 | 1 | .12 | .06 | -- | -- |
| Group*Learning condition | .62 | 2 | .54 | .01 | -- | -- |
| Learning condition*Time | .92 | 2 | .40 | .02 | -- | -- |
| Learning condition*Time*Group | 1.08 | 2 | .34 | .02 | -- | -- |
| Simple effects | t | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Live vs Recorded-observant | .08 | 45 | .14 | -- | -.03 | .2 |
| Live vs Recorded-alone | .19 | 45 | .008 | -- | .05 | .32 |
| Recorded-observant vs Recorded-alone | .10 | 45 | .16 | -- | -.04 | .24 |
| Live$_{imm}$ vs Recorded-observant$_{imm}$ | 2.51 | 45 | .016 | -- | .03 | .26 |
| Live$_{imm}$ vs Recorded-alone$_{imm}$ | 2.41 | 45 | .02 | -- | .03 | .35 |
| Recorded-observant$_{imm}$ vs Recorded-alone$_{imm}$ | .62 | 45 | .53 | -- | -.10 | .19 |
| Live$_{del}$ vs Recorded-observant$_{del}$ | .54 | 45 | .59 | -- | -.11 | .20 |
| Live$_{del}$ vs Recorded-alone$_{del}$ | 2.13 | 45 | .04 | -- | .01 | .37 |
| Recorded-observant$_{del}$ vs Recorded-alone$_{del}$ | 1.56 | 45 | .12 | -- | -.04 | .33 |
| | | | | | | |
| **Experiment 2 (N=82)** | **F** | | | | | |
| *Main effects* | | | | | | |
| Group | 2.24 | 1 | .14 | .03 | -.45 | .06 |
| Learning condition | 13.54 | 2 | <.0001 | .14 | -- | -- |
| Time | 38.56 | 1 | <.0001 | .33 | .26 | .51 |
| Interaction effects | F | | | | | |
| Group*Time | .16 | 1 | .69 | .002 | -- | -- |
| Group*Learning condition | .18 | 2 | .83 | .002 | -- | -- |
| Learning condition*Time | 2.12 | 2 | .12 | .03 | -- | -- |
| Learning condition*Time*Group | .03 | 2 | .97 | .00 | -- | -- |
| *Simple effects* | **t** | | | | | |
| Live vs Recorded-observant | .30 | 1 | <.0001 | -- | .18 | .41 |
| Live vs Recorded-alone | .10 | 1 | .09 | -- | -.02 | .22 |
| Recorded-observant vs Recorded-alone | -.19 | 1 | .001 | -- | -.31 | -.08 |
| Live$_{imm}$ vs Recorded-observant$_{imm}$ | 5.86 | 45 | >.0001 | -- | .24 | .49 |
| Live$_{imm}$ vs Recorded-alone$_{imm}$ | 2.32 | 45 | .02 | -- | .02 | .28 |
| Recorded-observant$_{imm}$ vs Recorded-alone$_{imm}$ | -3.21 | 45 | .002 | -- | -.34 | -.08 |
| Live$_{del}$ vs Recorded-observant$_{del}$ | 3.32 | 45 | .001 | -- | .09 | .36 |
| Live$_{del}$ vs Recorded-alone$_{del}$ | .74 | 45 | .46 | -- | -.08 | .18 |
| Recorded-observant$_{del}$ vs Recorded-alone$_{del}$ | -2.49 | 45 | .01 | -- | -.32 | -.04 |
| | | | | | | |
| **Combined (N=128)** | **F** | | | | | |
| *Main effects* | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Group | 1.54 | 1 | .22 | .01 | -.34 | .08 |
| Learning condition | 12.63 | 2 | <.0001 | .09 | -- | -- |
| Time | 77.32 | 1 | <.0001 | .38 | .30 | .49 |
| *Interaction effects* | **F** | | | | | |
| Group*Time | .11 | 1 | .75 | .001 | -- | -- |
| Group*Learning condition | .45 | 2 | .64 | .004 | -- | -- |
| Learning condition*Time | 2.61 | 2 | .07 | .02 | -- | -- |
| Learning condition*Time*Group | .5 | 2 | .6 | .004 | -- | -- |
| Simple effects | **t** | | | | | |
| Live vs Recorded-observant | .22 | 1 | <.0001 | -- | .14 | .31 |
| Live vs Recorded-alone | .13 | 1 | .004 | -- | .04 | .22 |
| Recorded-observant vs Recorded-alone | -09 | 1 | .05 | -- | .18 | .002 |
| Live$_{imm}$ vs Recorded-observant$_{imm}$ | 6.24 | 127 | <.0001 | -- | .19 | .37 |
| Live$_{imm}$ vs Recorded-alone$_{imm}$ | 3.28 | 127 | .001 | -- | .07 | .27 |
| Recorded-observant$_{imm}$ vs Recorded-alone$_{imm}$ | -2.34 | 127 | .02 | -- | -.22 | -.02 |
| Live$_{del}$ vs Recorded-observant$_{del}$ | 3.06 | 127 | .003 | -- | .06 | .27 |
| Live$_{del}$ vs Recorded-alone$_{del}$ | 1.86 | 127 | .06 | -- | -.01 | .20 |
| Recorded-observant$_{del}$ vs Recorded-alone$_{del}$ | -1.06 | 127 | .29 | -- | .-18 | .05 |

## 3.4 Combined analysis for Experiment 1 and 2

In the previous section we presented results from experiment 1 (N=46) and experiment 2 (N=82). Overall, in both experiments we found that learning in Live video-call was associated with the best performance for both groups. However, while experiment 1 showed a significantly greater decline in recall over time for ASC specifically for things learned in Live condition (as compared to NT), experiment 2 did not confirm this pattern. In this section I present a combined analysis to better understand the effect of social learning online in ASC and NT over time.

### 3.4.1 ***Results***

#### *3.4.1.1 Sample*

The combined dataset included a sample of 128 participants (Table 2, NT N= 67; ASC N= 61). For the ASC group, either received a diagnosed by a clinician (N=31) or were self-diagnosed (N=29). Note that all analyses were also run by excluding the self-diagnosed participants in the ASC group, and as results did not differ, here we report the full sample including the self-diagnosed participants (for results considering only participants who were clinically-diagnosed refer to Appendix, Table 2, 3 and 4). The ASC and NT group did not differ on age ($\text{mean}_{\text{ASC}}$ (sd) = 27.58 (6.62), $\text{mean}_{\text{NT}}$ (sd) = 28.49 (7.27), $t_{(125)}$=.73, *p*=.46), verbal fluency (Spot the word, $\text{mean}_{\text{ASC}}$ (sd)= 46.33, (8.95), $\text{mean}_{\text{NT}}$ (sd) = 44.61 (7.63), $t_{(125)}$=.64, *p*=.24) and non-verbal reasoning (MaRs-IB, $\text{mean}_{\text{ASC}}$ (sd) = 63.52 (18.50), $\text{mean}_{\text{NT}}$ = 61.47 (17.02); $t_{(80)}$=-.65, *p*=.51) and mentalising test (Animated Triangle, $\text{mean}_{\text{ASC}}$ (sd) = 9.27 (2.03), $\text{mean}_{\text{NT}}$ (sd) = 9.08 (2.13), $t_{(43)}$=-.49, *p*=.62). ASC group scored significantly higher on AQ than NT ($\text{mean}_{\text{ASC}}$ (sd) = 29.97 (8.33), $\text{mean}_{\text{NT}}$ (sd) = 19.67 (6.91), $t_{(125)}$=-7.61, *p*<.0001). We also confirmed that teacher was not a significant factor for learning performance ($F_{(2,126)}$=.55, *p=.58*), ensuring the dataset could be combined into one despite being collected by different experimenters.

#### *3.4.1.2 Analysis of Variance*

An analysis of variance (ANOVA) was run to test the difference in learning performance between 2 (groups: ASC and NT) x 3 (learning conditions: live, recorded-alone and recorded-observant) x 2 (time of learning quiz: immediate and delay). Means and SD for all conditions are reported in Table 3. Results for main and interaction effects are reported in Table 4 and Figure 4.

*Main effects.* Findings show a main effect of time: unsurprisingly, people remembered more things straight after the learning session (M (sd) =4.2 (0.6)) than a week later (M (sd) = 3.83 (0.08); $F_{(1,128)}$=77.31, *p<.0001*, $\eta^2$=.38, large effect size; Bakeman, 2005). The main effect of learning condition was confirmed ($F_{(2,128)}=12.63$, $p<.0001$, $\eta^2=.09$, medium effect size; Bakeman, 2005). Pairwise comparisons confirmed that Live interaction condition was the one associated with the highest learning, while Recorded-observant condition was associated with the worst learning: specifically, there was a significant learning advantage associated with Live compared to Recorded-observant condition ($M_{Live}$ (sd) = 4.31 (.05); $M_{Recorded\text{-}observant}$ (sd) = 3.92 (.06), $t_{(126)}=.22$, $p<.0001$), and compared to Recorded-alone condition ($M_{Recorded\text{-}alone}$ (sd)= 4.01 (.06), $t_{(126)}=.13$, $p=.004$). In addition, Recorded-observant condition was associated with the worst learning performance, showing also significant difference with Recorded-alone condition ($t_{(126)}=-.09$, $p=.05$). No main effect was found for group: in other words, people with ASC showed a NT-equivalent performance ($M_{ASC}$ (sd) = 4.09 (.08); $M_{NT}$ (sd) = 3.95 (07), $t_{(126)}=-.13$, $p=.22$).

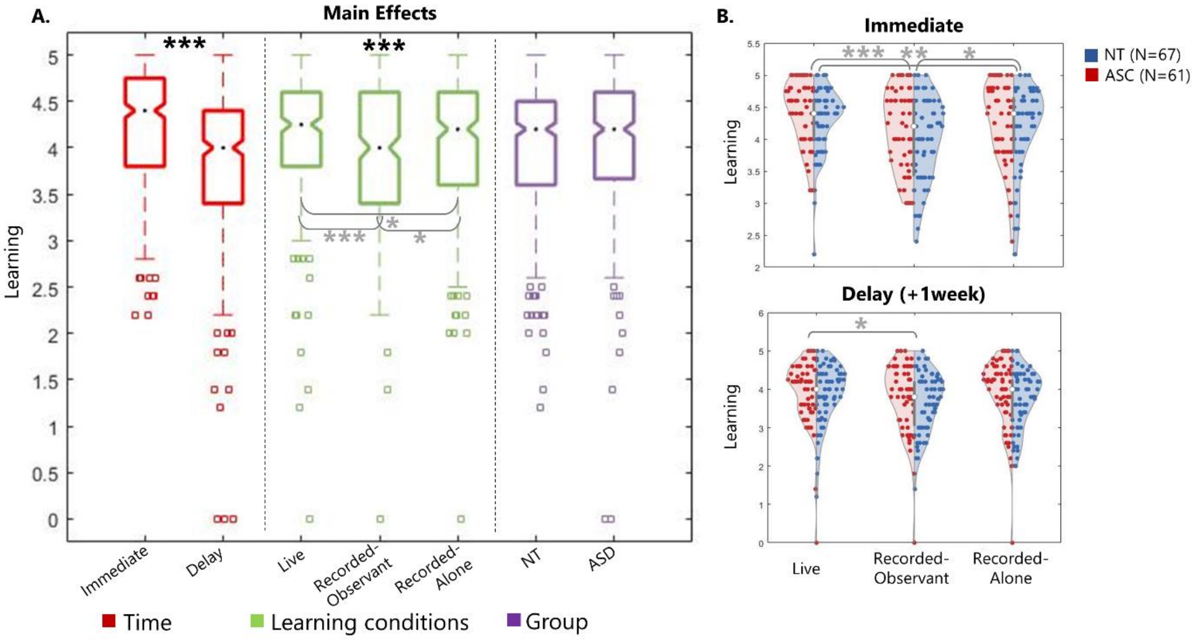


**Figure 3.4 Results for combined analysis (Experiment 1 and 2)**

**A.** Boxplots of the three main factors of interest across the two experiments: Time, Learning condition and Group. **B.** Violin plots of learning performance immediately after the learning session (top) and a week later (bottom), for the three learning conditions across the two experiments. Violins are split in half showing the distribution of NT (blue) and ASC (red) sample separately.

*Interaction effects*. No significant interaction effects where found between the main factors of interest (group, learning conditions and time). However, the learning condition * time interaction approached significance ($F_{(2,128)}$=2.61, *p=.07*). Post-hoc analysis revealed that all contrasts were significant (*p<.01*), but for delay recall, Live condition vs Recorded-alone condition only approached significance ($t_{(128)}$=1.86, p=.06), while Recorded-observant and Recorded-alone condition did not significantly differ ($t_{(128)}$=-1.06, *p=.29*).

#### *3.4.1.3 Exploratory Analyses*

In addition to the analysis of variance, we ran exploratory analyses to investigate the effect of other variables on social learning. Here we use autistic traits (AQ) as a continuous measure of autistic traits to minimise any confound arising from the fact that our ASC group included both self-diagnosed and clinically-diagnosed participants. We built two linear mixed-effects regression models to predict learning. Models were run in Matlab R2020b using the function *fitlme*. Full outcomes for both models are reported in table 5.

First, we built a model to predict learning performance from learning condition (Live vs Recorded-observant vs Recorded-alone) along with others variables including AQ, mentalising (Animated Triangle test), verbal fluency (Spot the word test) and non-verbal reasoning (MaRs-IB) measures, while controlling for variability coming from the teachers and individual participants:

Learning ~ Condition + AQ + Animated Triangle + Spot the word + MaRs-IB + (1 | Participant) + (1|Teacher)

Results confirmed 'Condition' as being a significant predictor of learning performance (beta = -.07, *p = .003*). In addition, we found that both verbal fluency (Spot the word test, beta = .02, *p = .002*) and non-verbal reasoning (MaRs-IB, beta = .01, *p = .0003*) were significant predictors of learning performance. This is not surprising: these measures have been linked to fluid cognition and intelligence (Baddeley et al., 1993;

Chierchia et al., 2019), which has been robustly associated with learning and academic performance more generally (Primi et al., 2010). With regards to measures of autistic traits (AQ) and mentalising (Animated Triangle), we did not find any significant effect, in line with the previous analysis of variance showing no difference between neurotypical and ASC group.

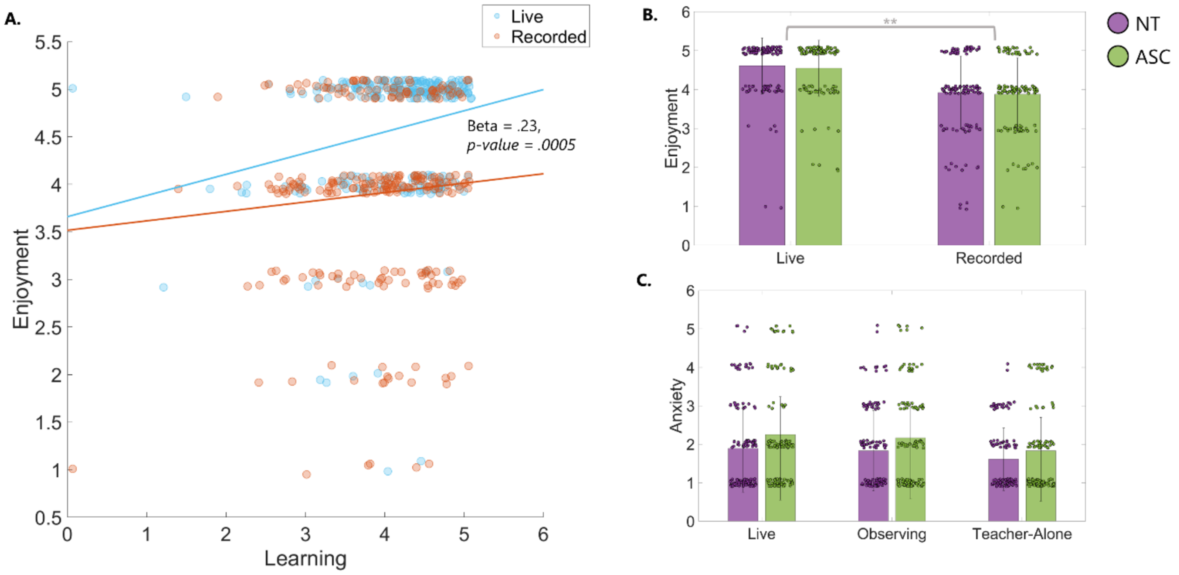


**Figure 3.5 Enjoyment and Anxiety level across conditions and groups**

**A.** Scatter plot of learning performance by enjoyment score for both groups, divided by condition (Live and Recorded, note: our questionnaire did not distinguish between the two recorded conditions). **B.** Box plots of enjoyment score divided by condition (Live and Recorded, note: our questionnaire did not distinguish between the two recorded conditions) and by group. **C.** Box plots of Anxiety score divided by condition and by group. All data plotted here refers to the combined sample (N=128; NT=67, ASC=61). ** p<.001

Second, in addition to the predictors included in Model 1, we included a measure of Enjoyment and a measure of Anxiety, as well as their interaction with learning condition.

Learning ~ Condition + AQ + Animated Triangle + Spot the word + MaRs-IB Enjoyment + Enjoyment*Condition + Anxiety + Anxiety*Condition + (1 | Participant) + (1|Teacher)

These measures were collected via participant self-report questionnaires after they completed the experiment: Enjoyment measure reflected a score from 1 to 5 for the question '*How much did you enjoy learning from the experimenter during [the video-call? / the pre-recorded video?]*' (1 = Not at all, 5 = Extremely much); Anxiety measure reflected a score from 1 to 5 for the question '*How much anxious / uncomfortable did you feel when you learned [live from the experimenter? / from the recorded video of another participant? / from the recorded video of the experimenter only?]*' (1 = not at all, 5 = extremely much).

Results show that Enjoyment was a significant predictor of learning for both groups (beta = .23, *p* = *.0005*), with people enjoying live interaction (M (sd) = 4.58 (0.71)) significantly more than learning from pre-recorded videos (M (sd) = 3.91 (0.95), $t_{(128)}$ = 7.52, *p<.00001*). We also found Enjoyment*Condition interaction effect approaching significance (beta = -.05, *p= .07*), with Enjoyment boosting Learning performance slightly more for the Live Condition than the Recorded one (Figure 3.5.A), although this is hard to interpret given that we only have Enjoyment scores for Recorded condition overall (i.e. our questionnaire did not make a distinction between Recorded-Observant and Recorded-Teacher-Alone condition). Interestingly, compared to Model 1, Condition was no longer a significant predictor of learning performance (beta = -.07, *p* = *.089*). Anxiety was not found to be a significant predictor of learning. Neither Enjoyment nor Anxiety scores differ between groups (Figure 3.5.B and 3.5.C).

**Table 3.4 Results from the linear mixed-effects models**

Outcome of the linear mixed-effects regression models. **Model 1:** Learning ~ Condition + AQ + Animated Triangle + Spot the word + MaRs-IB + (1 | Participant) + (1|Teacher); **Model 2:** Learning ~ Condition + AQ + Animated Triangle + Spot the word + MaRs-IB Enjoyment + Enjoyment*Condition + Anxiety + Anxiety*Condition + (1 | Participant) + (1|Teacher).

| | **Beta estimate** | | **SE** | | **df** | | ***p*** | | **Lower bound** | | **Upper bound** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| Predictors | | | | | | | | | | | | |
| Condition | -.07 | .24 | .02 | .14 | 750 | 746 | .003 | .089 | -.11 | -.04 | -.02 | .52 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Verbal fluency (Spot the word)[1] | .02 | .02 | .006 | .006 | 750 | 746 | .002 | .0007 | .01 | .01 | .03 | .03 |
| Non-verbal reasoning (Matrix reasoning item bank, MaRs-IB)[2] | .01 | .01 | .002 | .002 | 750 | 746 | .0003 | .001 | .004 | .003 | .02 | .01 |
| Mentalising (Animated Triangle)[3] | .002 | -.01 | .02 | .02 | 750 | 746 | .91 | .57 | -.04 | -.06 | .05 | .03 |
| Autistic Quotient (AQ)[4] | .001 | .003 | .005 | .14 | 750 | 746 | .78 | .53 | -.01 | -.01 | .01 | .01 |
| Enjoyment | -- | .23 | -- | .07 | -- | 746 | -- | .0005 | -- | .10 | -- | .37 |
| Anxiety | -- | .02 | -- | .04 | -- | 746 | -- | .63 | -- | -.07 | -- | .11 |
| Enjoyment*Condition | -- | -.05 | -- | .03 | -- | 746 | -- | .07 | -- | -.11 | -- | .005 |
| Anxiety*Condition | -- | -.02 | -- | .02 | -- | 746 | -- | .33 | -- | -.07 | -- | .02 |

## 3.5 Discussion

In the previous study (Chapter 2), we found evidence for social interaction boosting learning in neurotypical (NT) adults in online contexts. Here we asked whether the same interactive-learning advantage would replicate in adults with Autism Spectrum Condition (ASC). Previous literature on ASC has mainly focused on infancy and childhood, and studied social learning (e.g. imitation and observation, Bandura, 2019) of *social* knowledge (e.g. face perception: M. Dawson et al., 2007; Webb et al., 2017; or language: Norbury et al., 2010; Whitehouse et al., 2007). To the best of our knowledge, this is the first large empirical investigation to test interactive-learning of non-social factual knowledge in adults with ASC.

We tested learning over two blinded experiments (one pre-registered) in 128 adults, equally split in NT and ASC group, as they were presented with facts about obscure items online (over zoom) in three conditions (figure 1): in a live video-call with the teacher (interactive condition), by watching a recorded video of a previous student-teacher session (recorded-observant condition) and by watching a recorded video of the teacher alone (recorded-alone condition).

Results revealed two important findings: first, overall and across groups, learning in the live video-call was significantly greater than learning over pre-recorded videos, replicating results from study 1 (Chapter 2, De Felice et al., 2021). Second, and possibly more crucial for the scope of this study, we found that the interactive-learning advantage was present in participants with ASC too, in accordance with our pre-registered hypothesis 1: in fact, this group not only showed NT-equivalent performance overall, but benefitted from learning in live video-call over pre-recorded video as much as NT individuals did. With regards to our pre-registered hypothesis 2, the present data does not support it: we found no difference between groups in learning-advantage over time. We therefore discuss the results for the NT and ASD ignoring the factor of time and instead focusing on the other contrast of interest: the learning condition.

As discussed in previous chapters (Introduction, Chapter 1 and Study 1, Chapter 2), there are a number of mechanisms (e.g. attention, mutual-understanding) that may be supporting learning in a contingent learner-teacher interaction. Our investigation does not allow to disentangle these different mechanisms, nor to interpret NT and ASC group performance with reference to specific cognitive processes. Overall however, all the mechanisms identified as possible candidate to explain the interactive-learning benefit – including attention, social motivation and reward, back channelling, mentalising, arousal, monitoring and language – have been found to various degrees to be abnormal in ASC (Abell et al., 2000; Clements et al., 2018; Hamilton & Lind, 2016; Hill, 2004; Klin, 1991; Webb et al., 2017; White, 2013). This would predict poorer learning in ASC in social contexts. However, our results robustly contradict this prediction.

If the cognitive processes implicated in social interactive learning are also those typically disrupted in ASC, why do we still find that participants with ASC learned as well as NT in our study? We identified two possible explanations: i) ASC showed NT-equivalent performance, but at greater cost (compensatory hypothesis); ii) while ASC may struggle to learn *about* 'the social', they benefit from learning *via* 'the social' as much as NT do (about-the-social versus via-the-social hypothesis): in other words, evidence from previous studies come from experimental designs looking at how people with ASC learns *about* the social. In contrast, the present task disentangled *the means* through which learning occurs from *the object* of learning: this may have allowed people with ASC to benefit from the same interactive mechanisms supporting learning in NT individuals.

First, let's consider the possibility that similar learning between groups comes at greater cost for participants with ASC (compensatory hypothesis). In fact, both hyper- or hypo-arousal during social information processing has been associated with ASC (Yi et al., 2022). People with ASC also show abnormalities in executive function, sensory processing and emotional regulation (Fernandez-Prieto et al., 2021; Kilroy et al., 2019; Semrud-Clikeman et al., 2010). This may result in the social environment being cognitively demanding for autistic people, and would predict that they may show discomfort and/or less enjoyment during the task, despite overall NT-equivalent performance. Results however clearly reject this interpretation in several ways. Our conditions were *all* social, as the only variation was the level of contingency (live vs recorded) and the social richness of the scene (observing a learner-teacher interaction vs observing the teacher alone). Despite this, ASC group learned better specifically over contingent live sessions: it does not seem reasonable to hypothesise that any compensatory mechanisms would act differently for the live condition compared to the recorded ones. Furthermore, ASC group *enjoyed* the experiment as much as NT, and significantly more when learning in the interactive condition compared to the less-interactive ones (pre-recorded videos). Crucially, anxiety level during the task (as measured via self-report) was also similar between the two groups and across conditions.

Instead, we suggest that this data supports the *about-the-social vs via-the-social hypothesis*. The majority of the past literature has failed to disentangle these two mechanisms experimentally: previous experimental work looked at how people with ASC either learn *about* the social, or – e.g. in imitation studies – make *implicit* use of social signals to learn. In contrast, the present work disentangled *the means* through which *explicit* learning occurs from *the object* of learning: this may have allowed ASC to benefit from the same interactive mechanisms supporting learning in NT. While the present data does not allow us to identify the specific process, or set of processes, responsible for supporting learning in interactive contexts, it robustly showed how crucial it is to separate the medium through which we learn from the content of learning, and how contingent social interaction in online context can act as a catalyst across a variety of neuro-population, inclusive of ASC groups.

This work has the strength of studying a novel question in a large sample, with findings robustly replicated over two separate experiments (one pre-registered) and three different blinded experimenters performing the role of the teacher. Results are ground-breaking in that they demonstrate, for the first time, that adults with ASC not only *benefit* but also *enjoy* learning via live-interaction more than from pre-recorded videos. Some considerations however must be made in order to place these findings within the correct panorama. Participants with ASC who took part in this study were all high-functioning individuals, and conclusions may not extend to the entire autism population. This consideration is particularly relevant for our sample, which included a number of self-diagnosed participants in the ASC group. Moreover, recruiting participants over online platforms (e.g. Prolific) may attract people who are generally keen to participate in a social interactive experiments, possibly introducing a bias in the sample, while leaving out those people who are less likely to engage in social interaction (either because they struggle more or because they enjoy it less). Future work should investigate this question in a sample recruited via diverse sources (e.g. schools), to ensure a more diverse sample within the autistic community.

Our results predict that the interactive-learning advantage would also apply to in-person contexts. In fact, the present study – despite being online – resembled more the typical in-person situation: participants did not engage with any of the defining characteristics of a typical e-learning experience, e.g. pausing, repeat, forwarding etc. We however acknowledge that other aspects specific to in-person interaction, e.g. arousal due to physical proximity (Lougheed et al., 2016; McBride et al., 1965), may play a role in learner-teacher interaction, and online versus face-to-face contexts should be directly tested.

In conclusion, we showed that people with high-functioning ASC benefit and enjoy learning in interactive contexts as much as NT people do, and did not report being more anxious in any of our social interactive conditions. The present work has implications for classic cognitive models of social learning, arguing for a distinct separation of the context through which learning occurs from the content of learning. Designing experiments which explicitly separate these two factors is essential to better understand the underlying cognitive mechanisms supporting interactive social learning in both neurotypical as well as neuro-diverse populations. This in turn would help identify specific dysfunctions of social cognition, without making assumptions about a certain condition in relation to one factor (e.g. via-the-social) based solely on the other (e.g. about-the-social). Practical implications for pedagogy are obvious, including re-thinking about how we deal with education in ASC.

# 4. Chapter 4 – Inter-subject neural coherence as a function of social interaction

Data from this chapter has been published as a pre-print on 24th April 2023, full reference is:

De Felice, S., Hakim, U., Gunasekara, N., Pinti, P., Tachtsidis, I., & Hamilton, A. (2023, April 24). Social interaction increases brain synchrony during co-watching of novel movies. https://doi.org/10.31234/osf.io/ruhmc

*Contribution notes:* Uzair Hakim, Natalie Gunasekara and Paula Wicher helped to co-collect the hyperscanning fNIRS data.

## 4.1 Background

In Study 1 (Chapter 2) and Study 2 (Chapter 3) we demonstrated how social interaction boosts learning online in both neurotypical and people with Autistic Spectrum Condition (ASC). We discussed some of the behavioural mechanisms that may be engaged during contingent social exchange, including shared-attention, arousal and back channelling. In the next two chapters, we shift our attention to the neural underpinnings of social interaction (and how these relate to behaviour), by presenting a large hyperscanning dataset collected using functional-infrared spectroscopy (fNIRS). First, we study how engaging in face-to-face conversation synchronises neural activity (current chapter), before asking how brain synchronisation relates to learning in interaction (chapter 5). The study of neural response across multiple brains has grown over the last few decades, giving rise to an important debate about what neural dynamics can tell us about different cognitive mechanisms, within and between brains. We discuss the literature here by considering these different mechanisms, before presenting the rationale for our study.

First, attending to the same sensory stimuli has been found to elicit similar neural activity across brains. This has been referred to as neural alignment or inter-subject correlation (ISC, see Introduction, section 1.1.2). Using fMRI, Hasson et al. (2004) found that five different individuals showed similar neural response during free watching of a movie in occipital, parietal and temporal areas. Since between-brain synchronisation was found to extend beyond typical auditory and visual sensory-processing cortices to high-level association areas, the authors concluded that this reflected shared understanding of the movie narrative. Following this work, a number of studies confirmed this interpretation, showing neural alignment over areas involved in reasoning and abstract thinking, including pre-frontal and frontal regions (Jääskeläinen et al., 2008), and extending from visual stimuli to speech comprehension (Wilson et al., 2008), and from fMRI to EEG group-analysis (Poulsen et al., 2017).

Furthermore, similar neural representations have been found during interpretation and recall of shared events. For example, when people were given different interpretation of an ambiguous story, ISC was greater between individuals who were given the same interpretation (Yeshurun et al., 2017), and the same results replicated when participants were free to develop their own interpretation of a movie showing interacting abstract shapes (M. Nguyen et al., 2019). Likewise, in an fMRI study, Chen et al. (2017) found that neural patterns were more similar between people recalling the same event than between recall and perception of that event.

If ISC reflects shared understanding, how is it modulated by relational dynamics between people that tend to be psychologically closer and 'understand each other better', like partners, family and friends? Using fMRI, Parkinson et al. (2018) tried to answer this exact question. They collected information about the social network proximity between undergraduate students, and used it to predict similarity in neural response across participants during free watching of naturalistic movies. They found that neural responses during movie watching were particularly similar among friends, and ISC decreased as distance in real-life social network increased. Taken together, this data suggests that ISC during processing of external stimuli reflects mechanisms involved in attending to, understanding, interpreting and eventually remembering those stimuli similarly across individuals, in ways that are modulated by inter-personal dynamics. However, this work cannot answer questions about real-time interactive minds.

Going beyond single-brain scanning, brain-to-brain coherence or synchrony is obtained by recording brain activity from multiple people simultaneously, and therefore can give information about the real-time neural dynamics between interactive brains (see Introduction 1.1.2). Azhari and colleagues designed a series of hyperscanning studies where they measured brain-to-brain coherence in parent-child dyads (Azhari et al., 2019, 2020, 2021, 2023). In their paradigm, the child sits on their parent’s lap to co-watch a series of short cartoons, while brain activity of the parent and the child is recorded simultaneously via fNIRS. Findings showed that real father-child dyads exhibited greater coherence than pseudo dyads (i.e. shuffled parent-child pairs) in the medial left pre-frontal cortex (mPFC). Also, this was modulated by father’s age (Azhari et al., 2021), parenting stress (Azhari et al., 2019) and maternal anxiety (Azhari et al., 2023), with older, more stressed and more anxious parents resulting in less synchrony. Interestingly, co-parenting couples attending social salient signals (e.g. child laughing) also exhibited greater synchrony when being physically in the same room, compared to attending to the same stimuli at separate times, and significantly more than pseudo couples (Azhari et al., 2020). The small sample sizes of these studies however, together with the fact that they restrict to specific relationships (parent-child/co-parenting couples), call for more studies to extent their conclusions to broader social network dynamics.

Overall, results on neural alignment suggest that not only do different brains respond similarly to the same reality, but that such similarity is modulated by relationship closeness, possibly reflecting affinity in the way people perceive, experience and make sense of the world. In addition, hyperscanning studies reveal that social factors including co-presence and psychological factors affecting inter-personal dynamics (e.g. parental stress), modulate brain synchrony between people beyond simple stimulus-driven response. However, some questions still need to be addressed. All the studies discussed so far considered long-term social dynamics, i.e. relationship that built over several years (parent-child, romantic couples, and friends) and did not consider real-time (short-term) social interaction. It remains unclear whether face-to-face communication (e.g. having a conversation) modulates synchrony in brain activity between people as they co-experience the world around them.

The closest attempt to answer the question of whether social interaction modulates similarity for later brain response has been investigated in an fMRI study by Sievers et al. (pre-print). In their paradigm, participants' brains were first scanned during presentation of novel movie clips with ambiguous narratives. Then participants were assembled into small groups and asked to reach a consensus (via conversation) about each movie clip's narrative. Finally, participants received a second brain scanning, during which they were presented with the same clips as well as new ones from the same movies. Results revealed more ISC after conversation, and distinctive patterns of similarity in brain activity were observed within members of the same group, a finding interpreted by the authors as reflecting the group's unique discussion.

Sievers and colleagues' work innovatively demonstrated the effect of real-world conversation in modulating inter-subject *neural alignment* to later stimuli. However, by comparing neural response across solo brains, this work cannot inform our understanding of how people's brain synchronise as they *co*-experience the world, nor elucidate how a conversation changes inter-brain synchrony during shared experiences. In addition, the question remains as to whether the observed neural alignment by Sievers and colleagues specifically reflects shared common-ground over a given topic (i.e. building a consensus about ambiguous narratives) or could potentially arise from any social interaction episode. The contribution of social interaction alone is difficult to disentangle, as while increased ISC after conversation extended to novel clips, these were still part of the same movie, and thus related to the conversation's content.

The question we ask here is not whether similarity in brain responses reflects shared understanding of a given experience (as this has been convincingly demonstrated elsewhere). Rather we ask what the minimal social factor able to modulate brain synchrony between people is: in other words, is talking *about* the experience necessary to observe conversation-related increase in brain coherence, or is social interaction per se between two people (e.g. having a conversation *unrelated* to the experience) enough to further attune their brain responses for later events?

In this study we asked pairs to watch an episode of the BBC cartoon DipDap, as they sit next to each-other, while we measure their brain activity using fNIRS. Note that all pairs were familiar pairs (e.g. friends, flatmates or partners). This has been mainly dictated by the fact that data was collected during the covid-19 global pandemic, where UK government only allowed people within the same household to meet and interact face-to-face. After the first co-watching phase, participants engage in a conversation on unrelated topics (see Chapter 5). They then take part in a second co-watching phase, when they are presented with a new episode of the same cartoon. By comparing brain coherence computed from real-dyads to brain coherence computed from pseudo-dyads and separately for co-watching pre-conversation and co-watching post-conversation, we ask two questions: 1) is brain coherence during co-watching greater between real dyads (who sit next to each other and are familiar with each other) different from pseudo-dyads (i.e. baseline for stimulus-related activity)? And 2) what is the effect of an *unrelated* conversation on brain coherence during later co-watching?

## 4.2 Methods

### 4.2.1 *Participants*

62 volunteers took part in the study, paired in 31 dyads. Participants were recruited via online platforms including university participant databases and social media, as well as flyers placed at local libraries and cafes around university campus. 1 dyad was excluded from the final sample due to data recording failure, and 3 dyads did not pass the pre-processing data quality check (see nirs signal processing section). The final sample included 27 dyads (N=54, 34 females, 1 non-binary, age range = 19-37, age mean (sd) = 26.61 (4.76), years of education mean (sd) = 19.66 (2.99)). Demographic information are reported in Table 1. All participants gave written consent to participate in the study and were reminded of their right to withdraw at any point.

Data collection took place during a time of severe covid-19 restrictions, which made it necessary for participants and their partner to be from the same household in order for them to participate in a face to face experiments (with no mask on). In our study, there were 17 'friends' dyads, 10 'romantic' dyads and 3 'flatmate' dyads. The average relationship duration in years was 6.95 (sd = 4.45). Overall, on a scale from 1 (not at all) to 5 (very), they reported being close to their partner on average 4.43 (sd = .77), with no significant difference across dyad sub-groups.

### 4.2.2 *Materials and Procedure*

Participants set next to each other facing a screen (Figure 1*a*). Once the NIRS cap was placed and localised (see section 4.2.3 and 4.2.4), the experiment started. Participants first watched one episode of the BBC Dipdap animated series (Phase 1), then chatted about unrelated topics for about 20 minutes (Phase 2), before watching another (new) Dipdap episode (Phase 3). Each episode lasts 2 minutes and shows the adventures of Dipdap, an animated puppet who has to face a series of challenges created by an imaginary drawing line. The episodes are all non-verbal and can be watched in any order as they are all self-contained. They are particularly useful to engage the watcher's imagination, as one follows the drawing line creating new and surprising scenarios for the puppet. The two episodes, '*Balloon*' (no.7) and '*Headphones*' (no.38), were selected randomly from the full list of episodes (available at https://www.bbc.co.uk/iplayer/episodes/b00xgpj9/dipdap?page=1), and were presented randomly in a counter-balanced order over phase 1 (pre-conversation) and phase 3 (post-conversation). During the co-watching parts (phase 1 and 3), a separator was placed in between the two participants, ensuring they did not engage in any social communication *during* the presentation of the episode. During the social interaction part (Phase 2), participants were taking part in another experiment (Chapter 5) where they engaged in a structured conversation (conversation phase). None of the participants ever mentioned the Dipdap episode during the conversation phase.

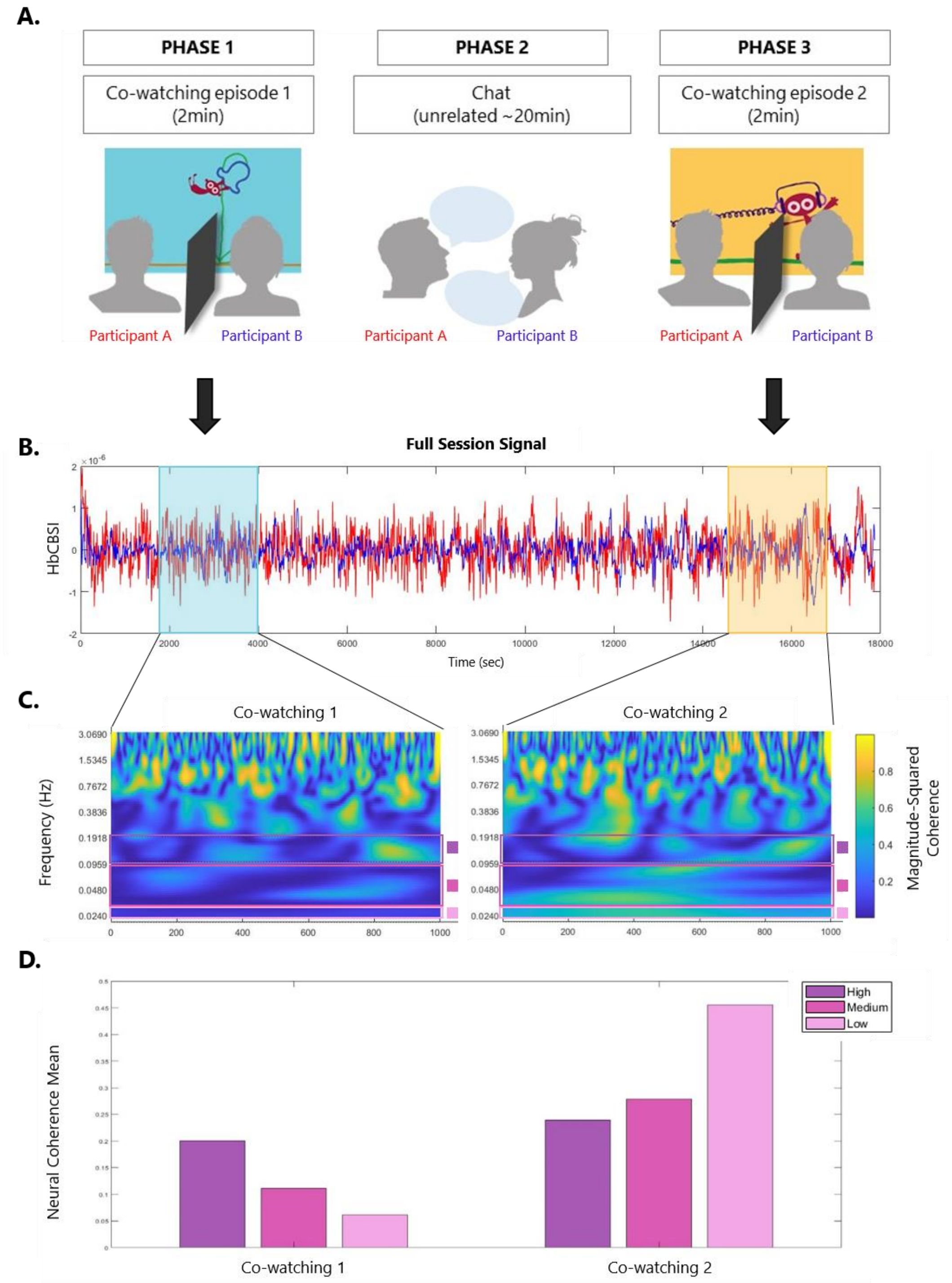


**Figure 4.1 . Example of data processing streamline for one dyad (and one channel/ROI).**

**A.** participants seat next to each other and watch an episode of the BBC series 'Dipdap' (Phase 1). After watching one episode, the two participants engage in a social interaction task, when they chat about unrelated topics (Phase 2). They than watch another – novel - episode of Dipdap (Phase 3). During co-watching (phase 1 and 3), a separator ensures that participants do not engage in any form of communication. The two Dipdap episodes were randomly allocated to phase 1 or 3 (counterbalanced across dyads). They are all non-verbal, self-contained, identical for duration and comparable in terms of audio/visual features. **B.** Full session Nirs Signal (HbCBSI) plotted for participant A (red) and participant B (blue). Nirs signal during each video co-watching is highlighted. **C.** Wavelet coherence spectrogram for video 1 and video 2. Bars show the frequency of interest used in analysis. **D.** Bars plot of the mean for the three frequencies of interest (High: 0.1-0.2 Hz, Medium: 0.03-0.1 Hz, Low: 0.02-0.03 Hz) for video 1 and video 2. Data plotted in B., C. and D. belongs to the same dyad.

### 4.2.3 *fNIRS signal acquisition*

Hemodynamic signals were acquired using a 56 optodes (28 sources and 28 detectors, split between two heads) continuous wave NIRS system (Shimadzu LABNIRS, Kyoto, Japan) with three wavelengths of light (780, 805 and 830 nm). Each participant in a dyad had the same distribution of 38 channels over both hemispheres (7 source and 7 detectors per hemisphere, Figure 4.2*A*), with a source-detector distance of 3cm. Before starting the recording, data quality was optimized by adjusting the detectors gain to maximize signals' intensities and improving the optical coupling between the optodes and the scalp (e.g. by moving the hair away from underneath the optodes). Data was collected at a sampling frequency of 8.33 Hz.

### 4.2.4 *fNIRS data pre-processing*

The full data processing pipeline is illustrated in Figure 3. Raw intensity data from all optodes of both participants was converted into a readable format using Homer toolbox in Matlab. Each converted raw file was then split into two, based on the channel configuration of each individual participant within one dyad. Individual channel inspection for inclusion was assessed following the pipeline in Pinti et al. 2019. Specifically, channels were excluded if no heart beat oscillation was visible in the frequency spectrogram, or if light saturation artefacts and/or large motion errors were present. After the data quality check and exclusion, on average each channel had 44 data points (out of 54 participants, min =31; max = 53).

Of the included channels, raw intensity signals at the 3 wavelenghts were pre-processed using the Homer2 toolbox. In particular, intensity data were converted into changes in optical density (function: hmrIntensity2OD). Optical densities were then corrected for motion artefacts using the wavelet-based method (function: hmrMotionCorrectWavelet , iqr=1.5) and band-pass filtered in the range [0.01 0.4] Hz (5th order Butterwort filter, function: hmrBandPassFilt). Changes in HbO2 and HbR were calculated using the modified Beer-Lambert law assuming a fixed DPF of [6 6 6] (function: hmrOD2Conc). HbO2 and HbR were then combined into the 'activation signal' through the CBSI approach (Burgess et al., 2022; Cui et al., 2011).

### 4.2.5 ***Channel to Regions of Interest (ROIs) allocation***

Figure 4.2*B* shows the location of each channel in standard space across my whole sample (one colour per channel). The variability between participants is visibly large. Such location variability across participant is not unique to this experiment, but rather a very common issue in studies using fNIRS (Zimeo Morais et al., 2018). This is usually not addressed in neuroscientific studies, and it is instead assumed that each probe falls in the same location across participants. However, figure 4.2*B* shows how this is not a safe assumption to make. In the section below, I outline some steps I took before data analysis to minimize the negative effects of the variability in probe locations in my data.

#### *4.2.5.1 From real to standard space*

Before starting fNIRS data acquisition, we collected 3D space coordinates of the location of each optode from all participants, using a Polhemus Electromagnetic Tracking system (https://polhemus.com). We then converted these coordinates from real space (specific to the individual) to Montreal Neurological Institute (MNI) space (where individual locations can be compared), using the NIRS SPM-12 toolbox (Tak et al. 2016). To make sure that the location of each optode was registered correctly, we performed a check on each one of them within each participant: to be classified as correctly registered, we tested the assumption that each optode should be located between 2.5 and 3.5 cm from any neighbour optode (given the 3cm distance on the cap configuration and taking into account errors in the measurements obtained through the 3D digitizer due to environmental electromagnetic interferences). When this assumption was met, the channel location was computed as the MNI coordinates of the middle point of two adjacent optodes (reflecting the original head configuration, figure 4.2*A*) for each participant. When an optode location was clearly off the standard grid (distance from neighbour optodes was either <2.5cm or >3.5cm), the MNI coordinates for that optode were discarded and the  location of any channel forming from the mis-located optode(s) was computed based on the mean MNI coordinate of well-located optodes. For participants where either more than 50% of optodes were mis-located (N=7) or Polhemus registration failed all together (N=2), the mean MNI coordinates were used to computed all channel locations.

At the end of this process, every channel in every participant has an MNI coordinate as shown in Figure 4.2*B*. This allowed to compare channel locations across my sample. Noticeably, there is still substantial individual variability in the locations of each channel.

#### *4.2.5.2 From channel to regions of interest*

Using the database neurosynth (https://neurosynth.org/), I identified functional ROIs that would be potentially engaged in my study and that were of interest for our hypotheses. Specifically, I identified xyz coordinates for left and right hemisphere for the following terms: working memory - planning - DLPFC [1091 studies]; speech production [107 studies]; speech comprehension [424 studies]; TPJ/theory of mind [181 studies]; visual cortex [488 studies]; parietal - memory retrieval/episodic memory/joint attention [324 studies]. For each term, the area with the highest activation was identified as the 'centre' of the ROI. When terms produced extensive cortical activation (e.g. language terms over left hemisphere), an extra 'centre' was selected. A total of 18 centres of interest where identified across the two hemispheres.

To check that these ROIs were in line with my headset configuration, and that I had enough data points for each of my ROIs, I plotted the xyz coordinates for each centre of interest along with the mean MNI for all channels. I then generated a spheres for each of my ROIs, having as centre the ROIs centre, and a radius of 2cm. From here, all ROIs that have less than 44 data points falling within the 2cm radius sphere were excluded. This threshold has been chosen so that our ROIs reflected the distribution of our dataset (44 was the average data points contributing to each channel after the quality check, see 'nirs data quality check and signal pre-processing', section 4.2.8).

After this process, 8 ROIs were considered for the final analysis, 4 for each hemisphere: Dorso-Lateral Pre-Frontal cortex (DLPF), ventral Pre-Motor cortex (vPM), Temporo-Parietal Junction (TPJ), and Superior-Parietal Lobe (SPL; Figure 4.2*C*). The MNI coordinates for all ROIs are reported in Table 4.1. For each participant, each channel was assigned to the closest ROI, and in turn it was checked that each ROI was receiving signal from the closest channel (e.g. if channel 27 was *the* clos*est* channel to SPL, but in turn it was clos*er* to TPJ, it was assigned to TPJ, and the next closest channel to SPL was instead assigned to SPL). All channels contributing to any ROI met the assumption that were not further distant than 2cm, and they each contributed to just one ROI.

At the end of this process, each ROI had 54 allocated channels from 54 participants, apart from the left SPL which had 41 (figure 4.2*D*).

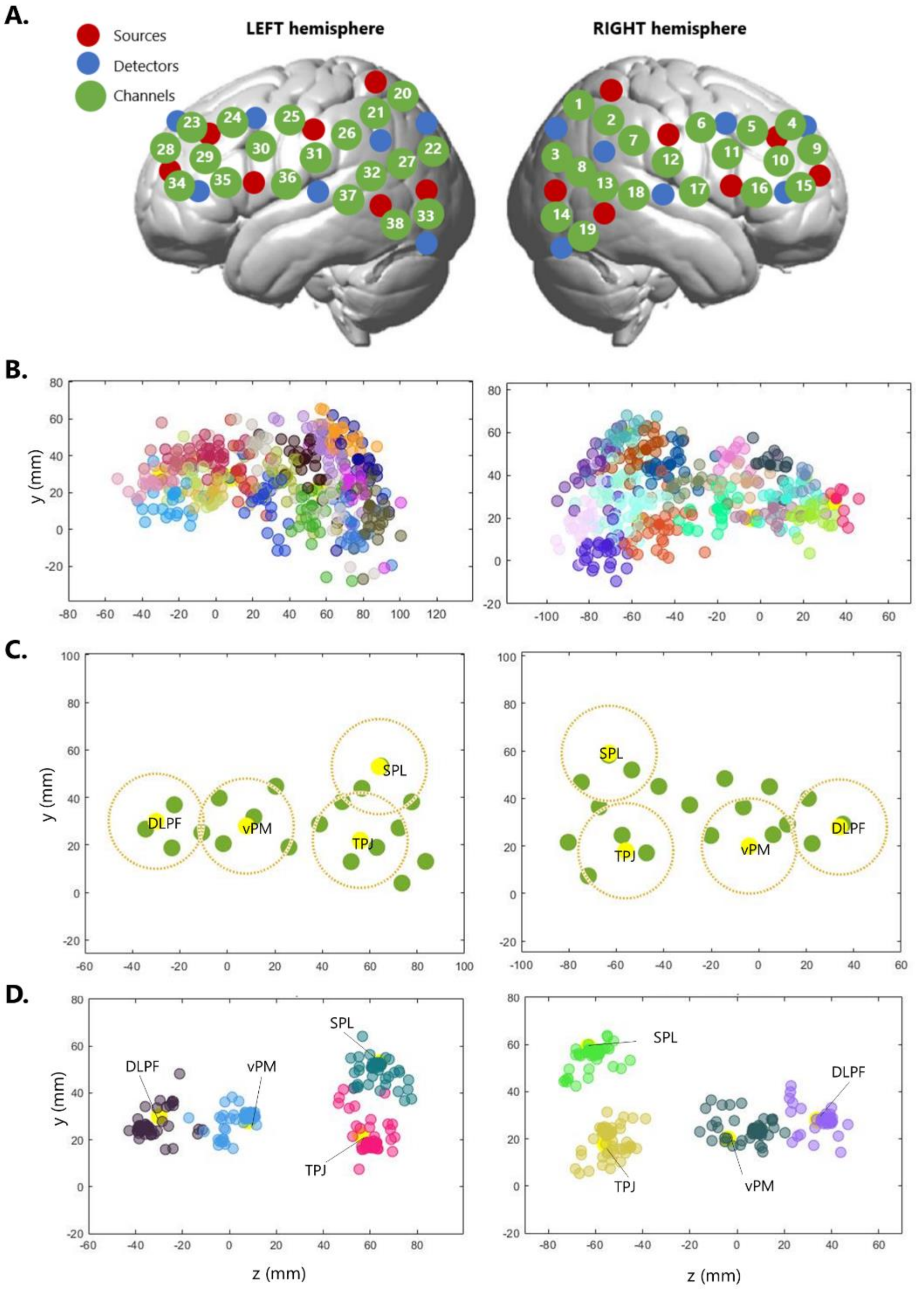


**Figure 4.2 . From headset probe locations to Region of Interest**

**A.** NIRS headset configuration. Optodes are divided by 7 sources and 7 detectors per hemisphere, spreading from parietal to frontal regions. This configuration forms 19 channels per hemisphere, for a total of 38 channels per participant. **B.** Channel localization in standard space. Channels (1-38) are plotted. Each dot is one channel per participant, each nominal channel is assigned to one colour. **C.** 8 functional ROIs are plotted in yellow, 4 in each hemisphere. Green dots are channels for one participant. For each ROI, the closest channel to the centre would be assigned and contribute with data. No more than one channel would contribute to each ROI per participant. To be assigned to an ROI, channels must be located within the area marked by the dark yellow dotted line around that ROI centre (radius

2cm). **D.** Channels plotted after being assigned to one of the 8 ROIs. Each colour represent one ROI. DLPF = Dorso-Lateral Pre-Frontal cortex; vPM = ventral Pre-Motor cortex; TPJ = Temporo-Parietal Junction; SPL = Superior Parietal Lobe.

**Table 4.1 MNI coordinates for the 8 ROIs included in this analysis**

MNI coordinates for the centre of each ROIs as taken from neurosynth database (https://neurosynth.org/). DLPF = Dorso-Lateral Pre-Frontal cortex; vPM = ventral Pre-Motor cortex; TPJ = Temporo-Parietal Junction; SPL = Superior Parietal Lobe; L = Left; R = Right.

| Region | Laterality | X | Y | Z |
|---|---|---|---|---|
| DLPF | R | 44 | 34 | 28 |
| DLPF | L | -46 | 30 | 30 |
| vPM | R | 64 | -4 | 20 |
| vPM | L | -58 | -8 | 28 |
| TPJ | R | 58 | -56 | 18 |
| TPJ | L | -54 | -56 | 22 |
| SPL | R | 37 | -63 | 59 |
| SPL | L | -40 | -64 | 53 |

### 4.2.6 ***Pseudo-dyads***

The aim of this study (and the next study, Chapter 5), was to investigate neural synchrony as a potential marker for social cognition. In order to distinguish the neural coupling arising from simply being exposed to the same sensory experiences (e.g. watching a video), from neural synchrony arising from social cognitive processes, pseudo-dyads were created. Pseudo-dyads were computed respecting the same experimental characteristics of real dyads, including video presentations and condition order. For example, if real dyad 2 (formed by participant blue 2 and red 2) and real dyad 25 (participant blue 25 and red 25) had both watched the *Balloon* Dipdap episode first, and the *Headphones* Dipdap episode last, and both would start the learning condition with separator ON (see experimental procedure in chapter X) they will form one real-dyad subgroup. Then, all possible combination of pseudo-dyads were computed within each real-dyads sub-group (e.g. pseudo-dyads 1: blue 2 and red 25, pseudo-dyad 2: blue 25 and red 2). These ensured that pseudo-dyads would be exactly the same as real-dyads in all aspects of the experimental procedure, a part from the main factor of interest, i.e. having participated in the experiment *together*. A total of 198 pseudo-dyads were created.

## 4.3 Data Analysis

### 4.3.1 *Wavelet Coherence Analysis*

We measured brain synchrony by running a wavelet coherence analysis using the MATLAB R2020b function *wcoherence* (Figure 4.1*C*). The main strength of this analysis over more simple correlation analyses is that it takes into account both the temporal and frequency characteristics of the two signals (Grinsted et al., 2004; Müller et al., 2004). Wavelet coherence was calculated for each ROI within both real and pseudo dyads, for each trial separately (video 1 and video 2). This gave the spectrogram for each dyad in the time-frequency space. Given the large range of frequencies, we selected three frequency bands of interest, namely high (0.1-0.2 Hz, i.e. 5-10 sec period), medium (0.03-0.1 Hz, i.e. 10-30 sec period), and low (0.02-0.03 Hz, i.e. 30-60 sec period; Figure 4.1*C* and 4.1*D*). This decision was informed by both a general agreement in the literature that different frequencies in hemodynamic rhythms are reflecting different cognitive processes (Cannon et al., 2014; Ward, 2003), and more specifically previous studies looking at brain coherence in social interaction contexts (Cui et al., 2011). Note that frequency components higher than 0.2Hz were disregarded altogether, as these would not reflect true brain signal as measured by fNIRS (the hemodynamic response has a frequency <0.2Hz), but rather physiological components such as respiration.

Once all real and pseudo-dyads had a coherence index for each ROI separately for the video 1 and video 2 trial, the mean in brain coherence across videos was also computed for each ROI. This gave us a general measure of brain coherence during video co-watching. In addition, in order to investigate whether social interaction was responsible for any *change* in brain coherence between participants, the brain coherence difference between video 2 and video 1 was also computed. Therefore, our final matrix had 27 dyads (real, or 198 pseudo) x 2 brain coherence measures (mean and change) x 3 frequency bands (high, medium and low) x 8 ROIs (DLPF left and right, vPM left and right, TPJ left and right, and SPL left and right).

### 4.3.2 ***Permutation testing***

In order to answer the question of whether i) being physically in the same room/being familiar with one another and ii) having a conversation would drive brain coherence, above and beyond what would be explained by simply processing the same stimulus (e.g. watching the same video), 10.000 permutations were computed between real and pseudo dyads (Figure 4.3). Permutation test has been proved to be a robust analysis tool to control for risk of type-1 error in multiple comparisons (Lage-castellanos et al., 2010; Pesarin, 2001).

The logic here is that real and pseudo dyads share the same features (they all watched the same videos, in the same order, and participated in the same experiments, in the same room), a part from the one factor of interest: pseudo dyads, in contrast to real dyads, did not experience those things *together,* and did not have a conversation *between each other*.

For the permutation testing, we computed two input statistics. The first input statistic was 'mean brain coherence V1+V2', calculated for each real dyad and each pseudo dyad. We then calculated a t-statistic for the difference between real and pseudo dyads. We permuted the labels on the data (real or pseudo) 10,000 times and calculated a distribution of t-statistics. Then we tested whether the true t-stat was different from the permuted t-stat. This helped us answer the questions of whether real dyads had greater brain coherence beyond what would be expected by simply being exposed to the same stimuli (pseudo dyads).

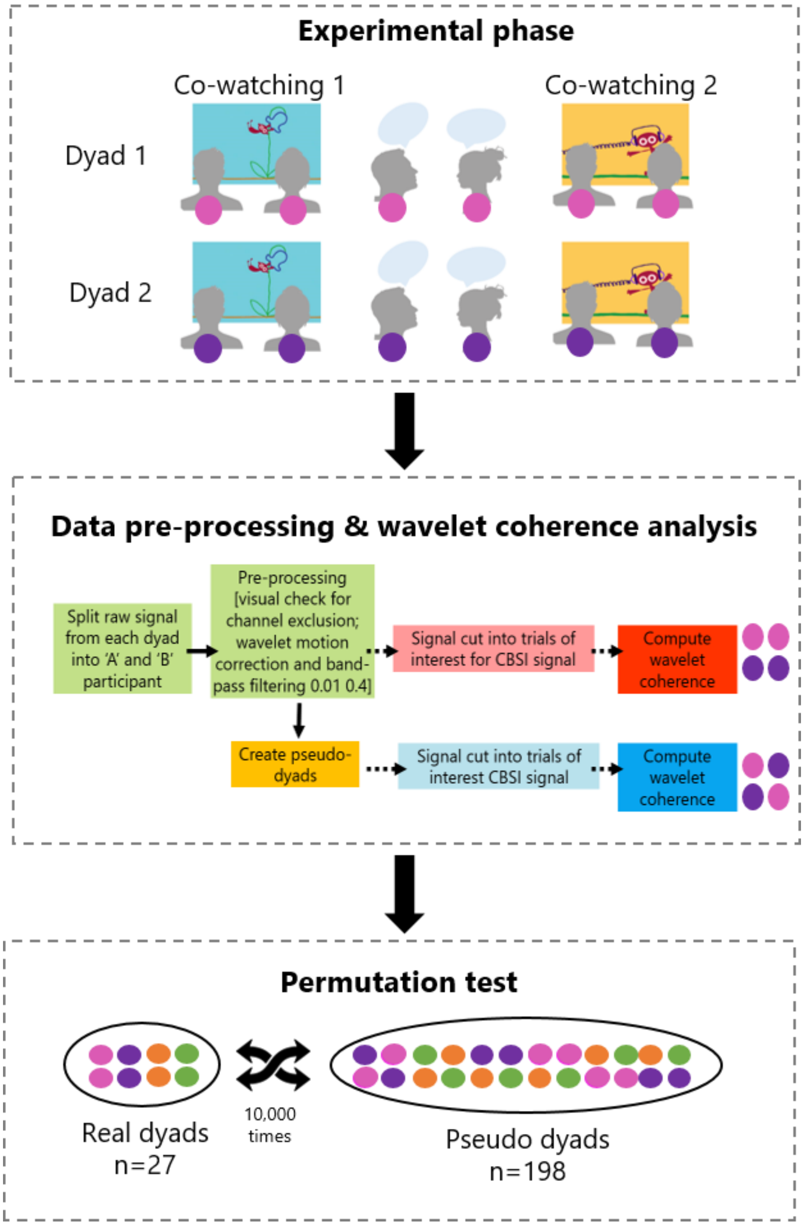


**Figure 4.3 Permutation analysis streamline.**

Illustration of data analysis pipeline to compute pseudo dyads and run permutation test. Pseudo dyads were computed on the basis of some pre-assigned characteristics to match real dyads on all experimental factors (e.g. trial order) but the social factor (e.g. being physically next to each other, see text). The same data pre-processing pipeline and wavelet coherence analysis has been performed for real and pseudo dyads. 10000 permutations were run between real dyads and pseudo dyads for the two variable of interest: mean brain coherence over co-watching 1 and co-watching 2, and brain coherence difference between co-watching 2 and co-watching 1 (see Hypothesis).

Our second input statistic was computed as the brain coherence change as a function of conversation (i.e. brain coherence co-watching phase 2 - brain coherence co-watching phase 1). Again, this was computed for real dyads and pseudo dyads, and the t-statistic was calculated for the difference between real and pseudo dyads. The true t-statistic was then compared to the t-statistic from 10,000 samples of permuted data. This helped us answer the question of whether social interaction would lead to change in brain coherence.

For each of the two measures of interest (mean and change), 10,000 permutations were repeated separately for the 8 ROIs and for the three frequency bands, for a total of 240.000 permutations per measure.

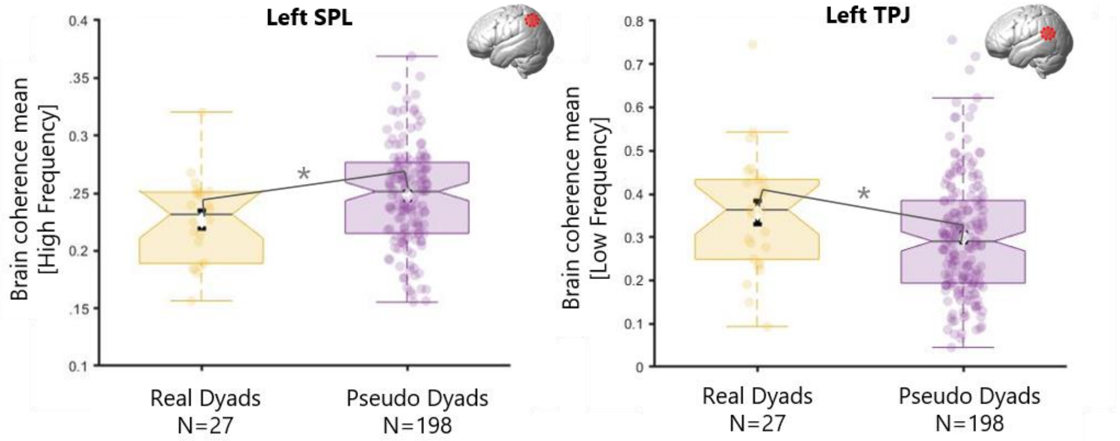
## 4.4 Results

Full results are reported in Table 4.2 and 4.3 for all ROI and all frequency bands. Main findings are presented in Figure 4.4. A reminder that our frequency bands include high band (0.1-0.2 Hz, i.e. 5-10 sec period), medium band (0.03-0.1 Hz, i.e. 10-30 sec period), and low band (0.02-0.03 Hz, i.e. 30-60 sec period). Results reported here come from 10,000 permutations test (Lage-castellanos et al., 2010; Pesarin, 2001).

### 4.4.1 *Does brain coherence change between real and pseudo dyads?*

Brain coherence mean during co-watching of both videos was significantly different between real and pseudo dyads over left SPL for the high frequency band and left TPJ for the low frequency band (Figure 4.4*A*). Interestingly, these effects were in opposite directions: we found that over left parietal ROI, brain coherence was smaller in real dyads compared to pseudo dyads (high frequency band, t = -.027, *p* = *.04*), and over left TPJ, brain coherence was greater in real dyads compared to pseudo dyads (low frequency band, t = .058; *p* = *.04*). In other words, being physically in the same room with someone familiar during co-watching (real dyads) was associated with *less* brain coherence for the high frequency band over the left SPL and *more* brain coherence for the low frequency band over left TPJ, compared to what would be expected on average for processing the movie (pseudo dyads). No difference was found in brain coherence (in any frequency bands) for all other ROIs between real and pseudo dyads.

**A. Does brain coherence change overall between real and pseudo dyads?**



**B. Does brain coherence change as a function of recent social interaction?**

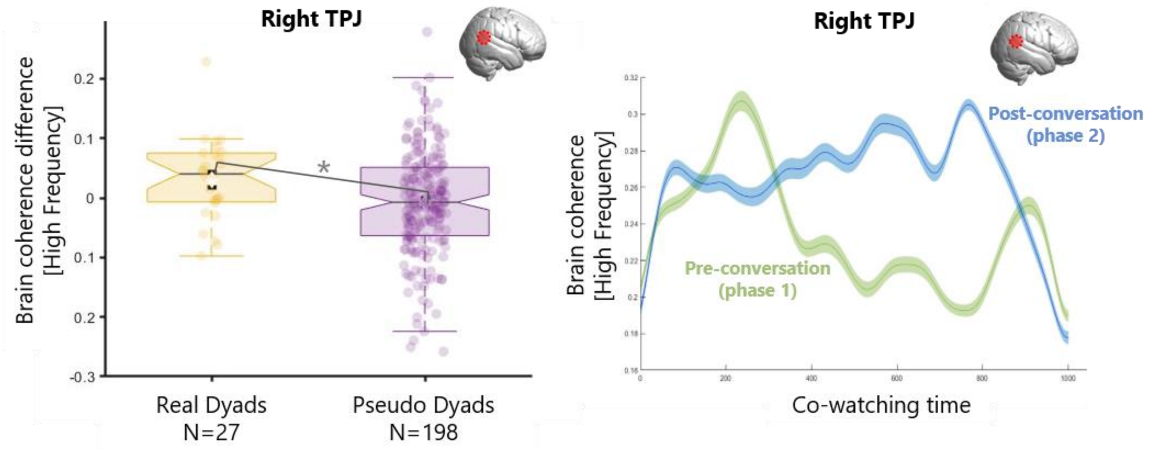


**Figure 4.4 Results**

**A.** Boxplots showing the distribution for real dyads (yellow) and for pseudo dyads (purple) of the brain coherence mean across the two co-watching phases. Right panel: there was significantly less brain coherence (high frequency 0.1-0.2 Hz) in real vs pseudo dyads over left SPL. Left panel: there was significantly more brain coherence (low frequency 0.02-0.03 Hz) in real vs pseudo dyads over left TPJ. **B.** Plots of brain coherence difference between co-watching 2 (post-conversation) and co-watching 1 (pre-conversation). Left panel: boxplots of the distribution of brain coherence difference across co-watching phases for real (yellow) and pseudo (purple) dyads. After a conversation, there was significantly more coherence in real vs pseudo dyads over right TPJ. Right panel: brain coherence for real dyads (sample mean) during co-watching pre-conversation (phase 1) and co-watching post-conversation (phase 2) over session duration. *p<.05. brain coherence = Inter-Subject Correlation, SPL = Superior Parietal Lobe, TPJ = Temporo-Parietal Junction

**Table 4.2 Results for brain coherence over both co-watching phases (average)**

Results from 10.000 permutations test for brain coherence between real and pseudo dyads. Permutation statistic used was the average brain coherence across both videos [(co-watching phase 1 + co-watching phase 2) / 2].

| Region | Real dyads mean (sd) | Pseudo dyads mean (sd) | Observed diff. | p-value | Effect size | Confidence Interval | |
|---|---|---|---|---|---|---|---|
| High Frequency Band (0.1-0.2 Hz) | | | | | | | |
| DLPF right | .25 (.04) | .25 (.06) | .00 | .90 | -.02 | -.42 | .38 |
| DLPF left | .26 (.05) | .25 (.05) | .00 | .98 | .07 | -.33 | .47 |
| vPM right | .23 (.05) | .24 (.04) | -.01 | .35 | -.27 | -.67 | .13 |
| vPM left | .26 (.04) | .25 (.05) | .01 | .23 | .14 | -.26 | .55 |
| TPJ right | .25 (.05) | .25 (.04) | .00 | .95 | -.01 | -.41 | .39 |
| TPJ left | .24 (.04) | .25 (.05) | -.01 | .41 | -.27 | -.67 | .13 |
| SPL right | .25 (.06) | .25 (.06) | .00 | .82 | -.05 | -.45 | .35 |
| **SPL left** | **.23 (.04)** | **.25 (.07)** | **-.02** | **.04** | **-.28** | **-.68** | **.12** |
| Medium Frequency Band (0.03-0.1 Hz) | | | | | | | |
| DLPF right | .24 (.07) | .23 (.07) | .00 | .93 | .05 | -.35 | .45 |
| DLPF left | .21 (.05) | .24 (.06) | -.02 | .07 | -.45 | -.85 | -.04 |
| vPM right | .24 (.06) | .24 (.06) | .00 | .93 | .08 | -.32 | .49 |
| vPM left | .24 (.06) | .23 (.05) | .01 | .43 | .11 | -.29 | .51 |
| TPJ right | .23 (.05) | .24 (.06) | .00 | .72 | -.07 | -.47 | .33 |
| TPJ left | .25 (.07) | .24 (.06) | .02 | .19 | .19 | -.21 | .59 |
| SPL right | .23 (.05) | .25 (.08) | .00 | .85 | -.22 | -.62 | .18 |
| SPL left | .22 (.07) | .21 (.07) | .00 | .80 | .17 | -.24 | .57 |
| Low Frequency Band (0.02-0.03 Hz) | | | | | | | |
| DLPF right | .38 (.17) | .37 (.20) | .07 | .07 | .09 | -.31 | .49 |
| DLPF left | .28 (.15) | .31 (.14) | -.02 | .57 | -.15 | -.55 | .25 |
| vPM right | .34 (.17) | .34 (.15) | .00 | .90 | .01 | -.39 | .41 |
| vPM left | .35 (.14) | .33 (.14) | .04 | .18 | .15 | -.25 | .55 |
| TPJ right | .33 (.12) | .30 (.15) | .03 | .30 | .21 | -.19 | .61 |
| **TPJ left** | **.35 (.14)** | **.31 (.15)** | **.06** | **.05** | **.29** | **-.11** | **.70** |
| SPL right | .35 (.18) | .33 (.17) | .04 | .22 | .16 | -.24 | .56 |
| SPL left | .34 (.15) | .30 (.17) | .05 | .17 | .20 | -.20 | .60 |

### 4.4.2 *Does brain coherence change as a function of recent social interaction?*

The difference in brain coherence between co-watching phases (coherence video 2 – coherence video 1) was significantly higher in real dyads compared to pseudo dyads over right TPJ for the high frequency band (t = .04, *p = .03;* Figure 4.4*B*). In other words, brain coherence for the high frequency band over the right TPJ during subsequent co-watching of a novel video was higher between two interlocutors (real dyads), than between two people who did *not* have a chat *with each other* (pseudo-dyads).

**Table 4.3 Results for brain coherence difference after conversation**

Results from 10.000 permutations test for brain coherence difference across co-watching phases between real and pseudo dyads. Permutation statistic used was the difference brain coherence across videos (co-watching phase 2 - co-watching phase 1).

| **Region** | **Real dyads mean (sd)** | **Pseudo dyads mean (sd)** | **Observed diff.** | **p** | **Effect size** | **Confidence Interval** | |
|---|---|---|---|---|---|---|---|
| High Frequency Band (0.1-0.2 Hz) | | | | | | | |
| DLPF right | -.017 (.08) | -.005 (.07) | -.02 | .44 | -.17 | -.57 | .23 |
| DLPF left | -.004 (.08) | -.003 (.08) | .00 | .96 | -.01 | -.41 | .39 |
| vPM right | .016 (.06) | .001 (.09) | .02 | .37 | .18 | -.22 | .59 |
| vPM left | -.018 (.07) | 0 (.08) | -.02 | .30 | -.21 | -.62 | .19 |
| **TPJ right** | **.030 (.07)** | **-.008 (.09)** | **.04** | **.03** | **.44** | **.03** | **.84** |
| TPJ left | .006 (.09) | -.002 (.08) | .01 | .66 | .09 | -.31 | .50 |
| SPL right | .007 (.08) | -.004 (.07) | .01 | .50 | .14 | -.26 | .54 |
| SPL left | .019 (.07) | .01 (.08) | .01 | .68 | .11 | -.29 | .52 |
| Medium Frequency Band (0.03-0.1 Hz) | | | | | | | |
| DLPF right | -.003 (.09) | -.007 (.10) | .00 | .86 | .03 | -.37 | .43 |
| DLPF left | -.016 (.11) | .005 (.10) | -.02 | .31 | -.20 | -.60 | .20 |
| vPM right | -.004 (.13) | .001 (.11) | -.01 | .82 | -.05 | -.45 | .35 |
| vPM left | -.004 (.12) | -.007 (.11) | .00 | .91 | .02 | -.38 | .42 |
| TPJ right | .031 (.13) | -.01 (.11) | .04 | .08 | .37 | -.03 | .77 |
| TPJ left | .023 (.10) | -.003 (.10) | .03 | .24 | .25 | -.15 | .65 |
| SPL right | -.005 (.10) | -.006 (.10) | .00 | .97 | .01 | -.39 | .41 |
| SPL left | .034 (.09) | .009 (.09) | .03 | .26 | .27 | -.13 | .67 |
| Low Frequency Band (0.02-0.03 Hz) | | | | | | | |
| DLPF right | -.056 (.25) | .002 (.24) | -.08 | .26 | -.24 | -.64 | .17 |
| DLPF left | -.028 (.29) | -.003 (.31) | -.03 | .69 | -.08 | -.48 | .32 |
| vPM right | .015 (.32) | .009 (.30) | .01 | .93 | .02 | -.38 | .42 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| vPM left | .048 (.26) | .002 (.28) | .05 | .42 | .17 | -.23 | .57 |
| TPJ right | .054 (.30) | .043 (.27) | .01 | .85 | .04 | -.36 | .44 |
| TPJ left | .112 (.29) | .006 (.26) | .11 | .07 | .40 | 0 | .80 |
| SPL right | -.052 (.27) | .034 (.26) | -.10 | .11 | -.33 | -.73 | .07 |
| SPL left | .088 (.26) | .041 (.23) | .05 | .44 | .20 | -.20 | .60 |

## 4.5 Discussion

In this study, we asked whether social factors (e.g. co-presence and face-to-face interaction) could specifically contribute to brain synchrony above what would be expected by processing the same stimulus. We measured neural response during movie co-watching in pairs before and after they engaged in a conversation. We report two main findings: first, over the left hemisphere, real pairs showed *increased* brain synchrony over temporo-parietal junction (TPJ) and *reduced* brain synchrony over Superior Parietal Lobe (SPL), compared to pseudo pairs (who had never seen each other and watched the same movie at different times). Second, real pairs who engaged in conversation showed *increased* synchrony over right TPJ during subsequent novel movie co-watching, significantly more than what was observed in pseudo pairs. We discuss each of these findings in turn.

First, when comparing brain-to-brain synchrony between real pairs and pseudo pairs across both co-watching phases (average over two movies), we found that real pairs showed increased synchrony over left TPJ and reduced synchrony over left SPL. Importantly, real pairs differ from pseudo pairs for two main features: they are familiar with their partner *and* they *co*-experienced the movie watching in time and space. While the contribution from these two factors is difficult to disentangle here, we can make some speculations on how these may have modulated synchrony in this study.

Increased neural synchrony over left TPJ between real dyads during movie watching is consistent with previous studies showing that similarity in neural response to narratives was greater between people who were also closer within their social network

(Parkinson et al., 2018). Our real pairs were all familiar with their partner and shared the same household, as data was collected during the covid-19 pandemic when face-to-face social interaction was only allowed within restricted social networks. Therefore, contrasting our real pairs with pseudo pairs *de facto* resulted in contrasting familiar versus unfamiliar pairs. Although the exact mechanisms remains unclear, greater brain coherence between familiar pairs (real dyads) may reflect emotional attunement (Nummenmaa et al., 2012), shared psychological perspectives (Lahnakoski et al., 2014) and social closeness (Wolf & Tomasello, 2020) typical of intimate relationships. However, in contrast to previous studies on neural alignment, this study measured brain activity from each pair *simultaneously*. It may be that co-experiencing movie-watching would additionally modulate brain synchrony in real pairs (beyond familiarity), in ways that are not possible when watching the same movie alone. Here, we therefore refer to *stimulus-driven brain coherence* to distinguish cognitive processing during co-experiences (like in this study), from neural alignment of cognitive processing happening solo (like in Parkinson et al., 2008). Sharing a physical environment activates processes of self-location and vestibular regulation with reference to the external world (Ionta et al., 2011): physical proximity may therefore engage a series of computations that may align the brains of people immersed in the same spatial-temporal context (also see Hamilton, 2020), in ways that do not occur in alone experiences. Mechanisms of familiarity and co-presence are unlikely to be mutually exclusive and possibly modulate brain-to-brain dynamics in tandem. Future studies should disentangle the effect of physical proximity from familiarity in aligning brain activities during sensory processing, by contrasting familiar and unfamiliar pairs (or groups) and directly comparing solo experiences with shared experiences.

Decreased neural synchrony over parietal regions in real compared to pseudo pairs also seems consistent with the interpretation that co-experiences activate qualitatively different neural response compared to experiencing the same thing alone. Previous work has identified the SPL as part of a global functional system termed as the 'intrinsic system' (Golland et al., 2007; Konishi et al., 2015; Miura & Noguchi, 2022; Yeshurun et al., 2021). Using fMRI, Golland et al. (2007) found that superior and posterior parietal regions showed a low level of within-subject correlation during multiple presentations of the same video, suggesting that this region processes information in a way that is relatively de-coupled from external sensory inputs. In line with this interpretation, a large body of evidence points to SPL as the hub for a neurobiological model of internal cognitive processes, including body self-consciousness and body-ownership (see Blanke, 2012 for a review). It is possible that the observed reduction in brain coherence during movie watching may be due to real pairs engaging in mechanisms designated to maintain a first-person perspective over shared experience (Benedek et al., 2016), therefore showing de-coupling over SPL from the external contextual inputs, including their partner.

Our second main finding was an increased level of brain coherence over right TPJ after face-to-face social interaction (difference in brain coherence between co-watching phase 2 and co-watching phase 1). In other words, *after* people engaged in conversation, their brain response to a novel movie watching was *more similar* compared to what was observed before the conversation, and increased to a significantly greater degree in real pairs compared to pseudo pairs. We refer to this effect as *socially-driven brain coherence*. There are two important points to highlight here: first, both real dyads and pseudo dyads engaged in conversation, but crucially only real dyads conversed *with one another*, while pseudo dyads conversed with someone else (i.e. their real partner rather the one forming the pseudo pair). Therefore, the observed effect cannot be explained by simply engaging in any social exchange, but specifically arise from interaction *with one another*. The second important element to consider is that conversation did not touch upon the content of the movie at any point, and the movie presented after the conversation was a novel one. This means that the observed increase in neural coherence cannot be interpreted as reflecting explicit consensus over a specific instance (e.g. one particular movie), but rather suggests that social interaction may support the development of general common ground and shared-understanding for future events, in ways that are not attached to a specific paradigm or context.

These results are consistent with previous studies showing conversation-related neural similarity between people (Sievers et al., pre-print). Importantly, they go beyond existing literature by demonstrating that social interaction distinctly contributes to increase synchrony over right TPJ between people co-experiencing later events. The right TPJ has been found to be associated with shared-understanding of external reality (M. Nguyen et al., 2019; Salazar et al., 2021; Yeshurun et al., 2017), as well as being heavily involved in social processes including mentalising (Molenberghs et al., 2016) and face-to-face conversation (Cañigueral, Zhang, et al., 2021a). The non-verbal Dipdap episodes used in this study are likely to elicit internal narratives of what is about to happen next, as a puppet is challenged by an imaginary line creating new unpredictable adventures. Here we show that neural response in the right TPJ becomes more similar between people who just engaged in social interaction, even to later unpredictable events. Future studies should further test this hypothesis by comparing brain coherence with explicit individual reports of their interpretation of the new event.

While we cannot be sure about the exact mechanisms behind the observed increase in brain coherence after conversation, one can suggest some speculations. It may be that the common ground and shared-understanding built over the conversation extended beyond it to immediately later events. Studies on mimicry suggests that during conversation people tend to naturally mirror their interlocutors body posture, speech rate, and even word-choice, with the ultimate goal of alignment of high-level mental representation (Garrod & Pickering, 2004, 2009). This effect has been shown to last also after the conversation event (e.g. Richardson et al., 2007). This may have been particularly true in the case of a pedagogical interaction (like the one here, see Chapter 5) where common understanding is crucial, as the teacher and the learner have the common goal of transferring and receiving information efficiently. If this interpretation is correct, synchrony would be observed not only at the neural level but also for other physiological signals (e.g. breathing, Konvalinka et al., 2023; McFarland, 2001), and eye-movements (Richardson et al., 2007; Richardson & Dale, 2005), and would also be stronger in dyads where such signals coupled more *during* the conversation. Future studies should test this hypothesis.

Our results also suggest that brain coherence is affected by social processes differently across different frequency bands. We hesitate to make strong conclusions about the specific underlying cognitive mechanisms subserving these different frequencies, as these are difficult to interpret. Previous fNIRS studies have mainly looked at one frequency band only (e.g. Cui et al., 2012; Lu & Hao, 2019), and more direct investigations are needed to test different frequency components in relation to specific cognitive processes. Also, fNIRS has a relatively slow temporal resolution and cannot measure changes happening faster than the hemodynamic response (~5 seconds). Future work should combine different neuroimaging modalities to investigate a broader range of frequency components. However, the present findings suggest that different frequency bands may reflect difference in neuronal rhythms, possibly mirroring the complexity of spatio-temporal dynamics in social interaction, in line with previous work on both social and non-social processing (Cannon et al., 2014; Ward, 2003).

Using fNIRS in this study has allowed us to investigate real-world interaction in ways that other neuroimaging modalities would have not made possible (e.g. fMRI, see Introduction, Chapter 1). Specifically, it provided a way to study *real-time brain synchrony* to investigate questions which have only been considered in terms of neural alignment so far. However, the downside of this includes relatively poor spatial resolution. By re-allocating channels to specific ROIs based on their MNI coordinates (see methods), we have tried to minimize this limitation. Exact comparisons across studies in terms of neuro-anatomical regions and associated functional processes, especially when data comes from tools with high spatial precision like fMRI, should however still be inferred with caution. Future studies should combine the use of multiple techniques to integrate strengths from different neuroimaging modalities, as well as making use of other technologies to include behavioural and physiological data to disentangle the contribution from different factors in driving brain coherence.

In conclusion, in this study we showed how social interaction can distinctly affect brain response across people in real-time, for later processing of non-social signals (movie watching). We demonstrated how co-experiencing a simple activity like watching a movie can both couple and de-couple brain regions. This possibly reflects mechanisms of internal and external processing, namely how we experience the world within ourselves and with others. Furthermore, we were able to specifically isolate the role of social interaction and show how interacting *with* someone in particular synchronises brain signals for later events. These results are in line with the two previous chapters in highlighting the role that social interaction has in shaping our experiences of non-social processing (e.g. learning online in chapter 2 and 3, and movie-watching in this chapter), and have implications for our understanding of social dynamics and how we share experiences and align interpretation with our friends and family in the real world.

# 5. Chapter 5 – Neural synchrony as a marker of learning in interaction

*Contribution notes*: Uzair Hakim, Natalie Gunasekara and Paula Wicher helped to co-collect the hyperscanning fNIRS data. MSc student Daniel Tompkins manually scored the free-recall learning test. Medical student Aliakber Dewji and Research Assistant Kamilla Bobyreva manually video-coded participants' eye-gaze during experimental sessions.

## 5.1 Background

In the previous experimental chapters, we demonstrated that social interaction boosts learning in online contexts, both in neurotypicals (chapter 2) and adults with ASC (chapter 3). We then showed that social interaction modulates brain-to-brain coherence during presentation of naturalistic stimulus (watching a movie, Chapter 4). Given that social interaction supports learning and also modulates brain synchrony between interactive people, can therefore brain synchrony used to predict learning? Here, we build on the findings from previous chapters and go further by asking whether brain-to-brain coherence during teacher-learner interaction can be a biomarker of successful learning.

A growing body of literature is emerging investigating the success of teacher-learner interaction (as a function of learning performance) in terms of brain synchrony (e.g. Bevilacqua et al., 2019; Davidesco et al., 2023; Dikker et al., 2017; Y. Pan et al., 2020, 2021, 2023; also see Introduction Section 1.2.3 for a comprehensive review of the literature on this). *Brain-to-brain synchrony* (or *coherence*) reflects the degree of real-time co-variance across brains, as measured simultaneously from multiple people as they interact (note that the term 'interaction' here refers to a range of situations along a continuum, from full-body and speech interaction to just co-presence, e.g. chapter 4). The study of brain-synchrony has been increasing thanks to the more widely available use of hyperscanning methods, where multiple people can engage in interaction and share experiences while their neural response are monitored simultaneously (Babiloni & Astolfi, 2014b; Czeszumski et al., 2020).

Studies agree in showing that teacher-learner brain coherence is linked to learning performance (e.g. Davidesco et al., 2023; Y. Pan et al., 2023). However, the direction of the relationship between brain synchrony and learning is not always consistent across contextual situations and/or different brain regions: in other words, more synchrony is not always associated with more learning (and vice-versa), but the relationship may be context-specific and reflect other mediating mechanisms. For example, Bevilacqua et al. (2019) showed that brain-to-brain synchrony was greater during video class compared to live lecture, but that specifically in live lectures, brain synchrony could predict teacher-student closeness, which in turn predicted learning. Most importantly, brain synchrony alone could not account for how well students retained information after class.

Similarly, Davidesco et al. (2019) found that specifically EEG alpha-band brain-to-brain coherence predicted learning, but not other frequency bands. Alpha frequency band is a well-established neural index of attention (Klimesch et al., 2007), which suggests that learning may have been better predicted by attentional dynamics between student and teacher, and brain synchrony simply reflected this cognitive process, rather than being the explanatory mechanisms per se. This hypothesis is consistent with another study by the same group, showing that the degree to which brain activity was synchronized across students reflected student class engagement (Dikker et al., 2017). Other mechanisms such as turn-taking (Y. Pan et al., 2018) and motor coordination (Y. Pan, Dikker, et al., 2020) have also been suggested to explain the non-linear relationship observed between brain synchrony and learning.

Although publications on the topic have proliferated over the last few years, it remains unclear what are the cognitive and neural mechanisms driving this effect: in other words, why does social interaction supports learning, and how? Existing literature suggests that brain-to-brain synchrony on its own cannot explain conclusively why certain interaction are 'better' or 'more successful' than others, nor can it account for the full complexity of teacher-student social exchange. In conjunction with studying brain-to-brain synchrony, it is therefore crucial to study the coordination of (joint) actions and how that relates to shared knowledge states (see Novembre & Iannetti, 2021 and Hamilton, 2020 for a discussion on this). Related to this, a useful theoretical standpoint is the *embodied mutual prediction framework* (Hamilton, 2020; Kingsbury et al., 2019), which explicitly acknowledge that interactive brains exist *within* interactive bodies. In other words, any synchrony between multiple brains results from auditory, visual and motor signals exchanged between social agents, and as such it should be studied and interpreted in conjunction with behavioural data.

The mutual prediction framework makes the claim that synchrony between brains in a social interaction arises because the two people involved in the interaction each engage both brain systems for performing their own actions and also brain systems for perceiving/predicting the actions of their partner. In cases where both partners are acting in a predictable fashion, there is a good match between A's brain activity for performing own action and B's brain activity for predicting A's action, because both of these brain states are related to the same action that A is doing (e.g. talking, gazing, etc). Similarly, there will be a good match between B's brain activity for performing her own action and A's brain activity for predicting B's action because both are related to the action B is doing (e.g. listening, gaze following etc). Thus, this sensorimotor coordination between A and B gives rise to coherent brain activity between A and B. Within this framework, high coherence can potentially be taken as an indicator that two participants are doing well at prediction and that their motor and cognitive processes are well matched, while low coherence might indicate unpredictable actions, a failure of prediction or a divergence of attention between the two people.

Tracking behaviour is important to understand social cognition and inter-personal dynamics particularly in the case of teacher-learning interaction. In fact, in interactive learning, the co-creation of knowledge and shared understanding is functional to the learning process, as the teacher share information, and the learner tunes in to their teacher to receive and process that information, while both adjust their behaviour as they receive feedback from their interlocutor. If brain-to-brain synchronisation is a proxy of behavioural and psychological dynamics, and particularly reflects the mutual prediction, increasing affiliation and communicative benefits across different minds (Hamilton, 2020; Hoehl et al., 2020b), then it would makes sense to study brain-synchrony while also manipulating and studying behaviour. Integrating behavioural data into hyperscanning studies means that a more comprehensive and meaningful understanding of how humans learn from and with others can be achieved.

One aspect of behaviour that may be particularly useful to consider in the context of social learning is eye-gaze (Emery, 2000). Gaze has been found to be an important social signal during communication (Richardson et al., 2007; Richardson & Dale, 2005; Wohltjen & Wheatley, 2021), and is considered one of the strongest factor driving brain synchrony (Dumas et al., 2011; Kelsen et al., 2020; Noah et al., 2020; Saito et al., 2010). In the context of social learning, eye-gaze has been shown to be an important predictor of children's learning (Elsabbagh et al., 2012; Ho et al., 2015; Lanthier et al., 2021; Yu et al., 2017) and adult class engagement (Alksne, 2016). Also, it has been demonstrated that inter-subject gaze behaviour is a proxy of joint attention, i.e. when two or more people attend to the same thing simultaneously (Kourtis et al., 2020; Wohltjen & Wheatley, 2021), considered an essential pre-requisite for learning and social communication more generally (Mundy & Newell, 2007). In chapter 2, we showed that seeing the full face of the teacher (including their eyes) improved learning in interaction. However, what remains unclear is the relationship between mutual-gaze, the emergence of joint attention episodes, brain synchrony and learning in face to face teacher-learner interaction (see Figure 5.1 for definition of gaze behaviours in interaction).

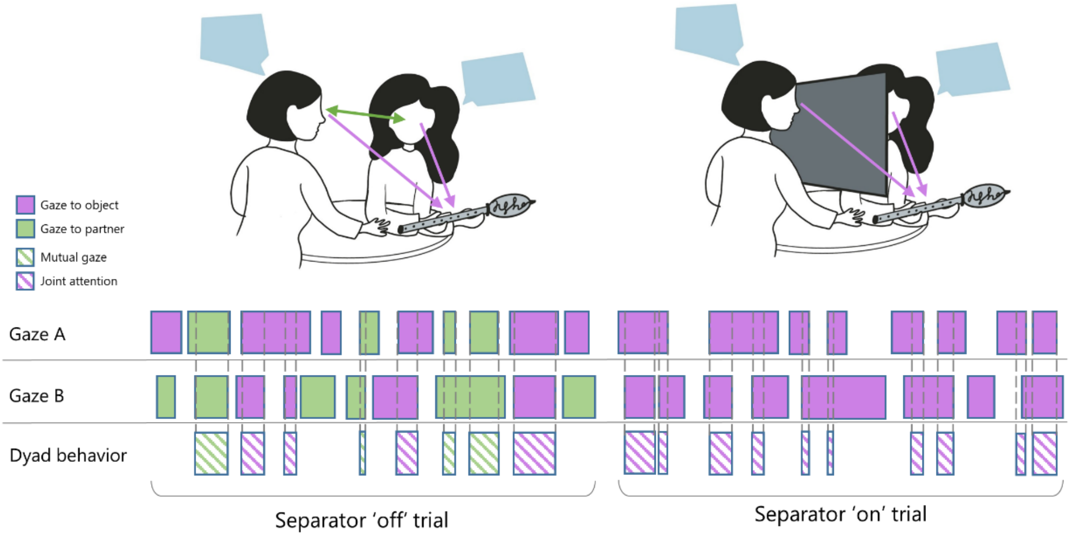


**Figure 5.1 Eye-gaze behaviour definitions and experimental-condition manipulation**

The separator 'off' condition (left) resembles a typical naturalistic interaction where both mutual gaze and joint attention episodes occur. The separator 'on' condition creates a disrupted-view where partners cannot see each other's faces, where mutual-gaze is not possible and only joint attention episode can occur. The lower panel illustrates how we coded gaze in interactions. Coders first recorded the gaze of A and B to the object / partner / neither and then an algorithm categorised each time point as mutual gaze (both people look to partner at the same time) or joint attention (both people look to the object at the same time).

In this study, we use fNIRS hyperscanning to measure brain activity simultaneously from two people, as they interact in a learner-teacher dynamic. Before coming to the lab and independently in their own time, participants learned some facts about obscure items (e.g. exotic animals). In the lab, they share those facts with their partner. Each participant alternates being a teacher and being a learner. In addition to the brain data, we also video-recorded the interaction and coded participants eye-gaze throughout the session. At the end of the interaction, participants' learning was tested via free-recall and multiple choice quiz. The multiple choice quiz was also repeated a week later to include a measure long-term learning. Thus, the primary data we will analyse in this study comprises (1) interpersonal brain coherence in specific regions of interest across both people in the dyad for each trial (2) the amount of time spent in joint attention or mutual gaze on each trial (3) average learning score for the item learnt in each trial. We will examine how both the brain and behavioural factors predict the learning outcomes for the dyad.

The complexity and novelty of this design refrains us to make strong predictions about our results. Informed by the mutual-prediction framework and acknowledging that interactive brains exist only within interactive bodies, we plan to build a model which includes brain synchrony along with behavioural data, namely eye-gaze behaviour, to investigate the dynamic between social signals and brain synchrony in predicting learning during face-to-face teacher-learner interaction. We plan to build a large model where contribution from predictors and their interactions will be evaluated via a stepwise iterative method.

Previous work on some of the aspects investigated here can nevertheless hint at some hypotheses. We regards to brain signals, we hypothesise that teacher-learner synchrony would predict learning, especially in areas involved in brain networks implicated in social cognition and communication, including dlPFC and TPJ, in line with previous studies on face-to-face communication and information sharing (Cañigueral, Zhang, et al., 2021a; Fronda & Balconi, 2020; Jiang et al., 2015). However, we do not have strong predictions on the direction of this relationship: it may be that greater learning would be associated with more synchrony in certain areas while de-coupling in other areas, possibly reflecting different behavioural dynamic unfolding over time between the teacher and the learner. In line with studies showing the importance of joint attention in effective communication and information sharing (e.g. Kourtis et al., 2020; Richardson et al., 2007), we hypothesise that joint attention would positively predict learning. With regards to mutual-gaze behaviour, the literature is mixed in reporting a positive association between mutual-gaze and learning (e.g. Schneider & Pea, 2013), and also showing that in naturalistic conversation mutual-gaze is avoided when speaking, possibly to minimise cognitive load and prioritise speech planning and production (Cañigueral, Ward, et al., 2021). In fact, previous studies on learning in naturalistic interaction did not specifically study the effect of joint attention and mutual-gaze in the same experiment, and it is not clear how these interact to modulate both brain synchrony and learning.

To better isolate different eye-gaze behaviours, we include a 'separator' condition, where learner-teacher interaction is 'disrupted' by a physical barrier (see Figure 5.1). Importantly, note that even during the separator 'on' condition (obstructed view), participants can engage in joint attention by attending to the object simultaneously and communicating via gestures and speech. Therefore, in line with the previously stated hypothesis that joint attention would be positively associated with learning, and acknowledging that even in the separator 'on' condition teacher-learner exchange is highly interactive, we do not expect difference in learning performance between conditions. However, by blocking the interlocutor view, the separator 'on' condition allows us to isolate joint attention and mutual-gaze episodes and to look at these different social signals separately, including their impact on brain synchrony. We predict that joint attention and mutual gaze would modulate brain coherence differently, possibly engaging different brain areas, but we do not have strong hypotheses on the neural topography or directionality of these effects.

## 5.2 Methods

This experiment corresponds to 'phase 2' of Study 3 (Chapter 4, i.e. the conversation stage in-between the two co-watching phases). Therefore, information about participants and nirs signal acquisition as well as some steps of the data analysis pipeline (including nirs data quality check and signal pre-processing, channel to regions of interest (ROIs) allocation and wavelet transform coherence analysis) applies also to this dataset. Also, the learning material used for this study is the same used in Study 1 (Chapter 2). Therefore, here we report enough details for the reader to understand the work done in this study, while we refer to previous chapters for more details on specific sections.

### 5.2.1 *Participants*

Participants were the same who participated in Study 3 (Chapter 4). The final sample included 27 dyads (N=54, 34 females, 1 non-binary, age range = 19-37, age mean (sd) = 26.61 (4.76), years of education mean (sd) = 19.66 (2.99)). All participants gave written consent to participate in the study and were reminded of their right to withdraw at any point.

### 5.2.2 *Material and Procedure*

A schematic of experimental procedure is illustrated in figure 5.2.A. This Study was composed of three main parts: 1) the pre-learning phase, 2) the experimental-session phase and 3) the learning-test phase. Participants were instructed to learn facts about obscure items in their own time before the experimental session (pre-learning phase), and told that they will teach those facts to their partner (experimental-session phase), whose learning will in turn be tested (learning-test phase). During the experimental-session phase, for certain trials a separator obstructed the view of the two participants.

Learning material was the same used in Study 1 (Chapter 2). 16 items formed two learning sets (set 'red' and set 'blue', 8 items in each set, 2 item from each of the four categories 'animals', 'antiques', 'exotic food' and 'musical instruments'). In each dyad, one participant was randomly assigned to the 'red' role (and received the 'red' set) and the other was assigned to the 'blue' role (and received the 'blue' set). The learning set assigned to each participant formed the material they had to *teach to* their partner during the experimental-session phase, while the other set formed the material they had to *learn* as students (and on which they were then being tested): in other words, participant 'blue' would pre-learn the set 'blue' (pre-learning phase), then teach the set 'blue' to their partner and learn the set 'red' from their partner (experimental-session phase), and be tested on the set 'red' (learning-test phase).

To ensure the learning material was truly novel to participants, a few days before the experimental session the researcher scheduled a short call and asked each participant to give a 'yes/no' answer to the question 'have you ever heard of e.g. *Axolotl*?'. If any of the items were known to just one participant within the dyad, then the learning set containing the known item was assigned to that participant (i.e. this ensured that any pre-knowledge would contribute to the teacher performance, but not the student performance). If both participants within the dyad knew about the item, than that item was excluded post-hoc from the analysis. At the end of the screening call, participants were each sent a unique link via email, containing their learning set (pre-learning phase).

The *pre-learning phase* was conducted online (in Gorilla Experiment Builder) and lasted on average 33 (sd = 2.7) minutes. This included a mixture of text and pictures about 8 items as well as MCQs with feedback to facilitate memorisation of the facts. Participants were instructed to complete the pre-learning phase in their own time and away from their partner, and were informed that while they could repeat this phase as many time as they wanted, to participate to the experimental session they had to achieve a 100% learning score (as measured in Gorilla Experimenter Builder).

The *experimental-session phase* was conducted in the lab and lasted approximately 1.5 hour. Participants set next to each other at approximately 30 degrees angle (Figure 5.1). Next to each participant and at reaching distance, there was a box containing the 8 items learned during the pre-learning phase. In addition, a cardboard separator was placed next to participant 'red'. Before starting the experiment, polhemus data was collected and the nirs cap was mounted on participants' heads by two researchers (see Chapter 4). During the experiment, a monitor in front of them displayed the name of the item to discuss in that trial and whether to place or remove the separator in between them. Participants were instructed to read the name of the item on the monitor and, if that item belonged to their learning set, pick it from their box and share their knowledge about it. Each trial lasted 1.5 min and a sound signalled the start and the end of that trial. Items’ order were presented randomly alternating from the ‘red’ set and the ‘blue’ set, so that each participant was never playing the same role (student/teacher) twice consecutively. The separator was placed (or removed) every four trials.

Participants were told to share information about a given item as well as possible as teachers, and to learn as much as possible as students, as their knowledge would have been tested at the end of the session. Beyond these instructions, they were welcomed to use the time in each trial autonomously and freely (e.g. ask questions, repeat information, interrupt their partner, play with the items’ model etc). Apart from the item model, no prompts nor scripts were given to participants during this phase.

In the *learning-test phase*, learning was measured immediately after the experimental-session phase via a free-recall questionnaire (‘*can you write down all you can remember about* e.g. *Axolotl*?’). Once finished, participant were reimbursed for their times and reminded to complete a multiple-choice quiz by the end of the day and in a week time. The multiple-choice quiz link was sent to participants at the end of the experimental session and again 7 days after. This was run in Gorilla Experiment Builder and was the same used in Study 1 (Chapter 2) and Study 2 (Chapter 3).

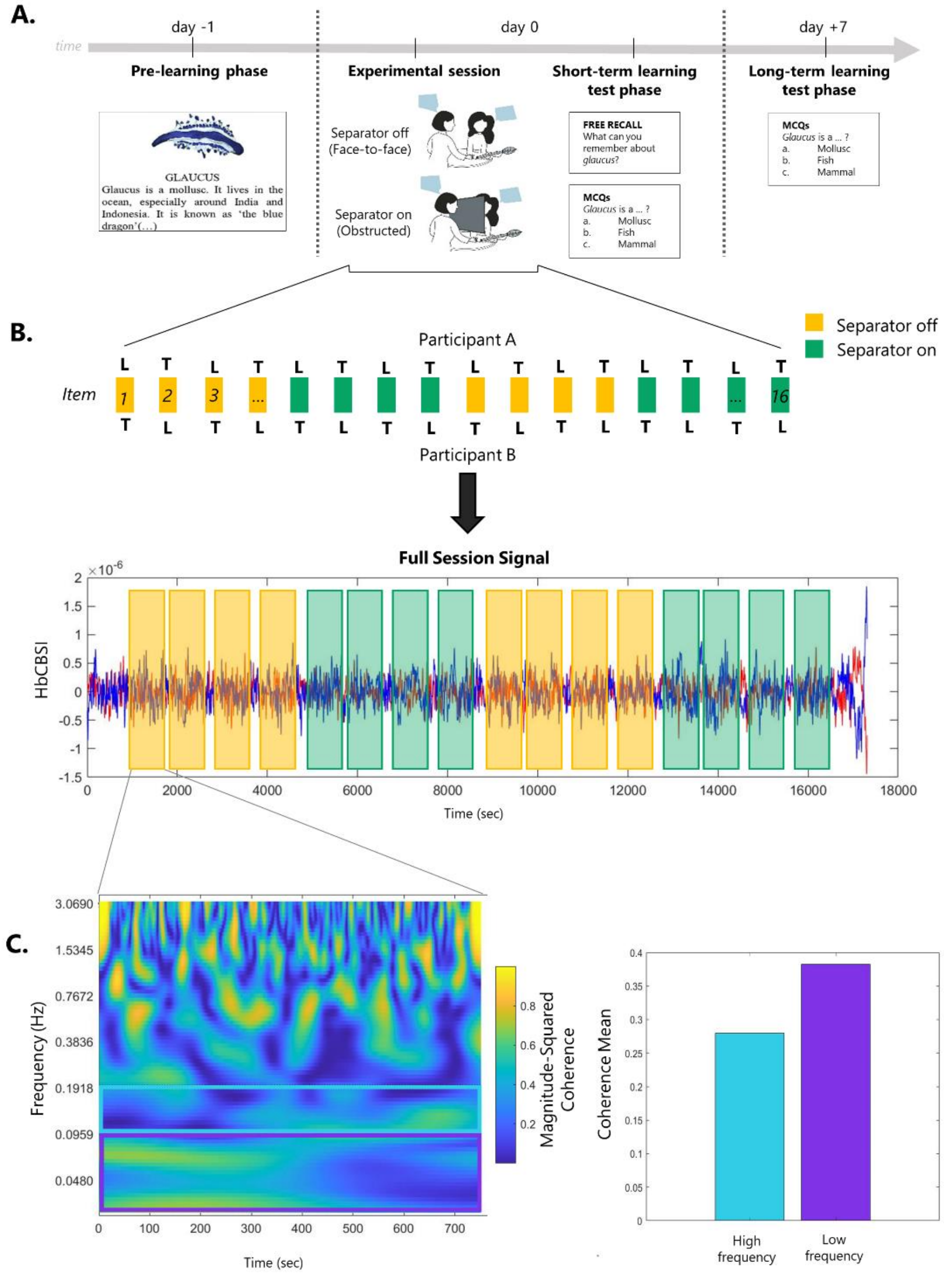


**Figure 5.2 Data collection and processing streamline for one dyad**

**A.** Experimental design **B.** Full session Nirs Signal (HbCBSI) plotted for participant A (red) and participant B (blue). Yellow (face-to-face condition) and green (obstructed condition) bars represent separate experimental trials. **C.** *Right*: Wavelet coherence spectrogram for one trial. Bars show the two frequency bands of interest. *Left*: Bars plot of the mean for the two frequency bands of interest (High: 0.1-0.2 Hz, Low: 0.03-0.1 Hz).

### 5.2.3 *Nirs signal acquisition*

Please refer to Chapter 4 Section 4.2.3 for details on this section.

### 5.2.4 *Channel to Regions of Interest (ROIs) allocation*

We analysed data in 8 ROIs comprising the dlPFC, vPM, TPJ and SPL in each hemisphere. Figure 4 shows channel configuration and ROIs. Please refer to Chapter 4 Section 4.2.5 for details on how we assigned channels to ROIs.

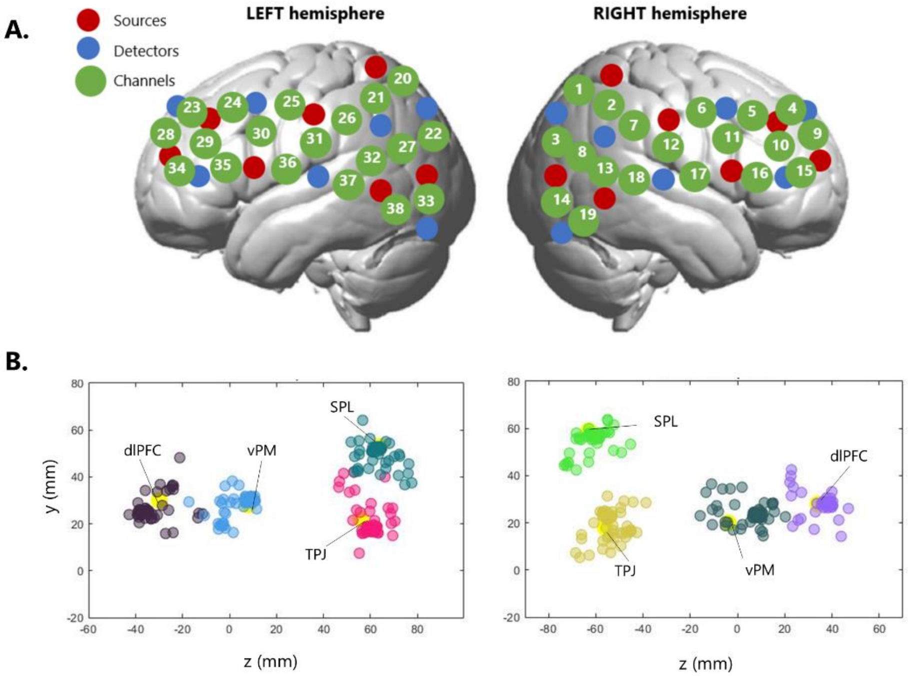


**Figure 5.3 Headset probe locations and Region of Interest**

**A.** NIRS headset configuration. Optodes are divided by 7 sources and 7 detectors per hemisphere, spreading from parietal to frontal regions. This configuration forms 19 channels per hemisphere, for a total of 38 channels per participant. **B.** Channels plotted after being assigned to one of the 8 ROIs. Each colour represent one ROI. DLPF: Dorso-Lateral Pre-Frontal cortex; vPMC: ventral Pre-Motor cortex; TPJ: Temporo-Parietal Junction. See Chapter 4 and Figure 4.2 for details on how we got from A to B.

### 5.2.5 *Nirs data quality check and signal pre-processing*

The full data processing pipeline is illustrated in Figure 2 and reported in details in Chapter 4 Section 4.2.4. After the data quality check and channel exclusion, on average each channel had 44 data points (out of 54 participants, min = 31; max = 53).

### 5.2.6 ***Data Analysis***

#### *5.2.6.1 Wavelet Coherence Analysis*

To measure neural synchrony, we calculated the *coherence* between the signals of the two brains within each dyad. This was obtained by running wavelet coherence analysis separately for each dyad during single trials (Figure 2.C), over the 8 ROIs, using the MATLAB R2020b function *wcoherence*. We selected two frequency bands of interest, namely high (0.1-0.2 Hz, i.e. 5-10 sec period) and low (0.03-0.1 Hz, i.e. 10-30 sec period). Compared to chapter 4, here the time series where too short to select a third frequency bands (trial duration in this Study = 90 sec vs co-watching phase duration in Study 3 = 120sec). This decision was informed by both a general agreement in the literature that different frequencies in neuronal rhythms are reflecting different cognitive processes (Cannon et al., 2014; Ward, 2003), and more specifically previous fNIRS studies looking at brain coherence in social interaction contexts (Cui et al., 2011).

#### *5.2.6.2 Learning performance scoring*

Across the whole sample, 16 trials were excluded based on participants' pre-knowledge screening (see Section 5.2.2). Learning performance on the included trials was scored from MCQs (immediate and delay) and free recall questionnaire. For MCQs performance, a score of 0 (incorrect) or 1 (correct) was assigned to each question (five per item), and a global score was then computed for each item (max score 5). With regards to free recall, an MSc postgraduate was trained to score the content recalled ('student performance'). 1 point was given for every fact correctly recalled (each item had a total of 15 facts). If facts were reported only partially, these were scored as 0.5. When facts were not recalled entirely, or something else was reported altogether, this was scored as 0.

In some instances, students misremembered the item's name while still remembering facts about that item: e.g. reporting facts about *Anhinga* while referring to it as *Axolotl*. Awarding 0 to these cases would have unfairly assessed learning performance: the student recalled considerable information and only failed to remember the target name for those facts. To ensure a fair scoring, we added an additional scoring section, referred to as 'Correct name-description association'. Students scored 1 on this section if they reported facts about an object and referred to it with the correct name (e.g. reporting facts about *Anhinga* and also referring to it as *Anhinga*). In cases where participants recalled information about an object but under the wrong name, they were awarded 0 for 'Correct Name-Description Association', but marked normally for the other fifteen facts.

In addition, the audio-recordings of all dyads were scored for how many facts were provided by the teacher in each trial ('teacher performance'). This allowed us to account for variability in learning caused by variation in teaching quality which arose because our procedure prioritised ecological validity and encouraged free-flowing interaction between participants. Having a measure of 'teacher performance' alongside the 'student performance' allowed us to compute a rigorous measure of learning (student performance/teacher performance), which reflected what the student was truly taught during a naturalistic teacher-student interaction. In some cases, facts were recalled when these were not discussed verbally, but were observed visually e.g. the colour of a fruit. In these cases, facts scored a point for both student performance and teacher performance.

Overall, learning was characterised by a single score calculated as the average of the scores in the free recall test, the immediate multiple choice test and the delayed multiple choice test. Scores on each of these three tests were expressed as a % of the maximum available points given the constraints listed above, so that the total learning score for each person and each item gives the best estimate of how much the participant has learnt about that item.

#### *5.2.6.3 Eye-gaze behaviour scoring*

Video recordings of each session were manually scored for all dyads included in the wavelet transform coherence analysis (N=27, see section 4.2.4 in Chapter 4). Due to a technical fault, we lost video recording for one dyad. This gives us a total sample of 26 dyads for which we extracted eye-gaze behavioural measures. Each video was coded separately for participant A and participant B in each dyad, specifically for gaze to object and gaze to partner (see Figure 5.1). From this coding, we computed two measures reflecting the sum of instances (all in seconds) when a particular behaviour occurred: *joint attention* was computed as the time when both participants looked at the learning-object simultaneously; *mutual-gaze*, was computed as the time when participants looked at each other (only available for separator 'off' trials). In addition, we computed an extra measure for *sustained attention* separately for participant A and B as the time each one of them looked at the learning-object (independently on their partner's behavior). We used this to control for time participants attended to the object which was not related to the interaction.

#### *5.2.6.4 Linear mixed-effects regression models*

To investigate what factors could best predict learning in a face-to-face naturalistic teacher-learner interaction, we built a model to predict learning that included all the behavioural and brain signals we collected. We then looked at data coming from the separator 'on' condition and separator 'off' condition separately. This allowed us to include mutual-gaze among the predictors of the model for separator 'off' data. We used a stepwise removal iterative method to exclude terms which did not significantly explain the variance in learning. The predicted factor in all our models was average learning (see section 5.2.6.2 for details on how this was computed). Mixed-linear model analyses were run using the function *fitlme* and *step* in MatlabR2023a.

We first describe the models in more details below, and then present results from each of this analysis in the Result section.

*'Big exploratory model'*: predictors included coherence values for each ROI in the two frequency bands (8 ROIs x 2 frequency bands, 16 terms), joint attention (seconds, 1 term) and the separator condition (0/1, 1 term). The model also included all possible interaction terms: namely, separator x joint attention (1 term), separator x coherence in all ROIs (16 terms) and joint attention x coherence in all ROIs (16 terms). This gave us a model with 51 predicting terms. To control for variability coming from other terms, we included random effects of dyad sustained attention (i.e. average of time spent by both teacher and learner looking the object, not necessarily simultaneously), learner sustained attention, learning-items, teacher performance (how much the teacher reported) and dyad. A schematic of the model's formula is reported below:

> *Learning ~ separator + brain-coherence*$_i$ *+ joint attention + separator*brain-coherence*$_i$ *+ separator*joint-attention + joint-attention*brain-coherence*$_i$ *+ (1 | learner sustained-attention) + (1 | dyad sustained-attention) + (1 | item) + (1 | teacher performance) + (1 | dyad)*

where subscript *i* indicates that 8 different brain coherence values were included for the 8 ROIs.

*Separator 'on' and separator 'off' models*: to better investigate the relationship between eye-gaze behaviour and brain coherence in predicting learning, we considered data coming from the separator 'on' and the separator 'off' condition separately. For separator 'on' condition, we run the same model as the 'all-inclusive model', without the 'separator' term:

> *Learning in separator 'on' ~ brain-coherence*$_i$ *+ joint attention + joint-attention*brain-coherence*$_i$ *+ (1 | learner sustained-attention) + (1 | dyad sustained-attention) + (1 | item) + (1 | teacher performance) + (1 | dyad)*

For separator 'off' condition, we added mutual-gaze as an independent predictor and in interaction with brain coherence values:

> *Learning in separator 'off' ~ brain-coherence*$_i$ *+ joint attention + mutual-gaze + joint-attention*brain-coherence*$_i$ *+ mutual-gaze*brain-coherence*$_i$

*+ (1 | learner sustained-attention) + (1 | dyad sustained-attention) + (1 | item) + (1 | teacher performance) + (1 | dyad)*

## 5.3 Results

Descriptive statistics for all measures are reported in Table 5.1. Brain synchrony data is also plotted for each ROIs in Figure 5.4 (high-frequency band) and 5.5 (low frequency band). Overall, learning distribution was very similar across experimental conditions ($t_{(368)}$=-.74, *p=.45*), as shown in Figure 5.4.

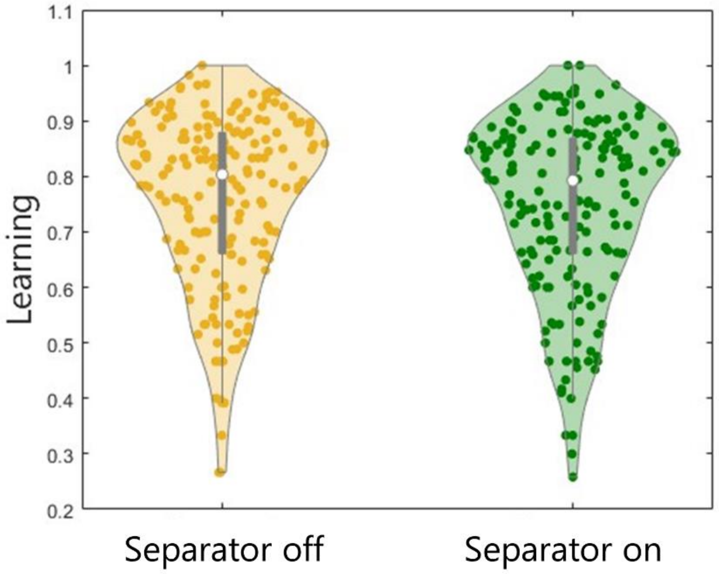


**Figure 5.4 Distribution of learning performance (%) for the two experimental conditions**

Each dot in the scatterplot represents learning performance for one item (trail).

We first report outcomes from the big exploratory linear-mixed effect model, and then present results separately for separator 'on' and separator 'off' condition. Given the high number of predictors in our models, we set our alpha level for significance at 0.01.

**Table 5.1 Descriptive statistics for all experimental variables**

MCQ: multiple-choice question; dlPFC: dorsolateral pre-frontal cortex; vPMC: ventral pre-motor cortex; TPJ: temporo-parietal junction; SPL: superior parietal lobe; wtc: wavelet transform coherence; JA: joint attention.

| | ***Overall*** | | ***Separator on*** | | ***Separator off*** | |
|---|---|---|---|---|---|---|
| **Variable** | Mean | SD | Mean | SD | Mean | SD |
| Learning immediate free-recall (%) | .56 | .22 | .54 | .23 | .57 | .22 |

|  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|
| Learning immediate MCQ (%) | .87 | .18 | .88 | .17 | .86 | .17 |
| Learning +1week MCQ (%) | .83 | .20 | .82 | .21 | .84 | .20 |
| Learning average (%) | .75 | .16 | .75 | .16 | .76 | .15 |
| MG (sec) | -- | -- | -- | -- | 8.75 | 8.23 |
| JA (sec) | 58.22 | 21.25 | 68.12 | 21.22 | 48.69 | 16.41 |
| Sustained Attention learner (sec) | 68.14 | 26.67 | 75.65 | 27.08 | 60.72 | 24.16 |
| Sustained Attention dyad (sec) | 65.34 | 25.14 | 72.40 | 25.69 | 58.38 | 22.59 |
| dlPFC right high-freq (wtc) | .27 | .07 | .27 | .07 | .26 | .08 |
| dlPFC left high-freq (wtc) | .27 | .07 | .26 | .06 | .27 | .07 |
| vPMC right high-freq (wtc) | .26 | .07 | .26 | .07 | .26 | .07 |
| vPMC left high-freq (wtc) | .28 | .07 | .28 | .07 | .28 | .07 |
| TPJ right high-freq (wtc) | .28 | .07 | .28 | .07 | .27 | .07 |
| TPJ left high-freq (wtc) | .27 | .06 | .26 | .06 | .28 | .07 |
| SPL right high-freq (wtc) | .27 | .07 | .27 | .08 | .27 | .06 |
| SPL left high-freq (wtc) | .27 | .07 | .27 | .07 | .26 | .06 |
| dlPFC right low-freq (wtc) | .30 | .10 | .30 | .10 | .29 | .10 |
| dlPFC left low-freq (wtc) | .31 | .10 | .32 | .11 | .30 | .10 |
| vPMC right low-freq (wtc) | .29 | .10 | .29 | .19 | .29 | .09 |
| vPMC left low-freq (wtc) | .31 | .09 | .31 | .09 | .31 | .09 |
| TPJ right low-freq (wtc) | .30 | .10 | .30 | .09 | .30 | .10 |
| TPJ left low-freq (wtc) | .29 | .10 | .29 | .10 | .29 | .10 |
| SPL right low-freq (wtc) | .30 | .10 | .30 | .11 | .30 | .10 |
| SPL left low-freq (wtc) | .29 | .09 | .29 | .10 | .28 | .09 |

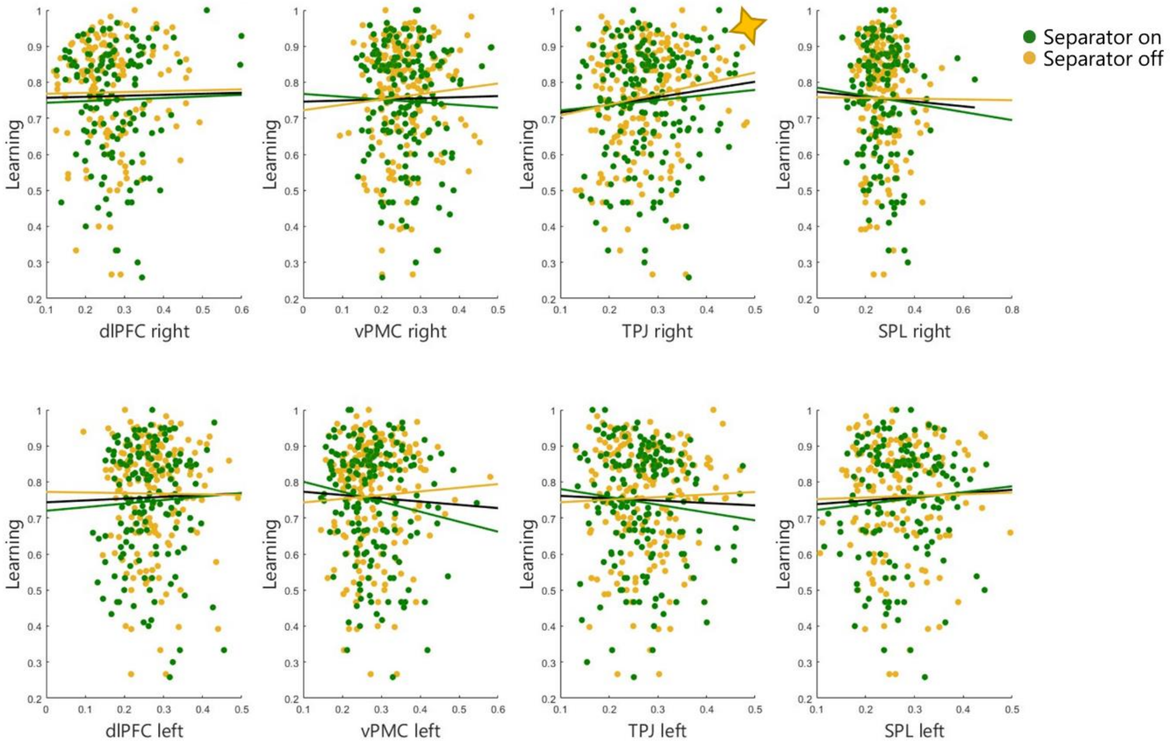


**Figure 5.5 High Frequency Band – brain synchrony (x) for each ROI and learning (y)**

Significant main effects are indicated with a yellow star. dlPFC: dorsolateral pre-frontal cortex; vPMC: ventral pre-motor cortex; TPJ: temporo-parietal junction; SPL: superior parietal lobe.

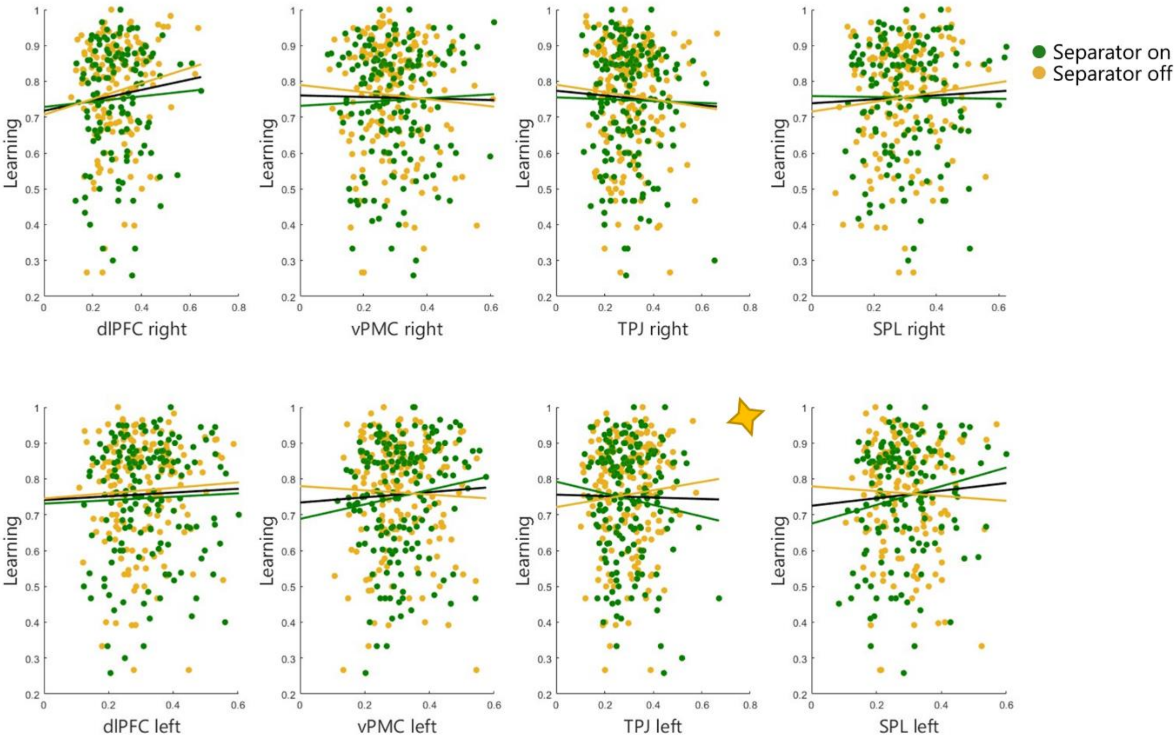


**Figure 5.6 Low Frequency Band – brain synchrony (x) for each ROI and learning (y)**

Significant main effects are indicated with a yellow star. dlPFC: dorsolateral pre-frontal cortex; vPMC: ventral pre-motor cortex; TPJ: temporo-parietal junction; SPL: superior parietal lobe.

*Big exploratory model.* As a reminder, the formula for this model is reported below:

> *Learning ~ separator + brain-coherence*$_i$ *+ joint attention + separator*brain-coherence*$_i$ *+ separator*joint-attention + joint-attention*brain-coherence*$_i$ *+ (1 | learner sustained-attention) + (1 | dyad sustained-attention) + (1 | item) + (1 | teacher performance) + (1 | dyad)*

Complete results for the big exploratory model are reported in Table 5.2. We found a main effect of brain synchrony in right TPJ for high frequency band (beta=1.16, *p-value<.001*) and in left TPJ for low frequency band (beta=.8, *p-value<.001*). We also found an interaction effect between right TPJ for high frequency band and joint attention (beta=-.02, *p-value=.01*), between left dlPFC for high frequency band and joint attention (beta=.02, *p-value<.001*) and between left vPMC for high frequency band and separator condition.

**Table 5.2 Big exploratory model**

dlPFC: dorsolateral pre-frontal cortex; vPMC: ventral pre-motor cortex; TPJ: temporo-parietal junction; SPL: superior parietal lobe; JA: joint attention.

| **Predictor Term** | **Estimate** | **SE** | **tStat** | **DF** | ***p*** | **Lower** | **Upper** |
|---|---|---|---|---|---|---|---|
| (Intercept) | 0.46 | 0.30 | 1.52 | 189 | *0.13* | -0.14 | 1.05 |
| separator | 0.43 | 0.23 | 1.89 | 189 | *0.06* | -0.02 | 0.87 |
| JA | 0.00 | 0.01 | -0.10 | 189 | *0.92* | -0.01 | 0.01 |
| dlPFC right high-freq | -0.47 | 0.39 | -1.22 | 189 | *0.22* | -1.24 | 0.29 |
| dlPFC left high-freq | -0.96 | 0.41 | -2.34 | 189 | *0.02* | -1.77 | -0.15 |
| vPMC right high-freq | 0.21 | 0.44 | 0.47 | 189 | *0.64* | -0.66 | 1.08 |
| vPMC left high-freq | -0.45 | 0.31 | -1.45 | 189 | *0.15* | -1.06 | 0.16 |
| **TPJ right high-freq** | **1.16** | **0.35** | **3.37** | **189** | ***<0.001*** | **0.48** | **1.85** |
| TPJ left high-freq | -0.18 | 0.31 | -0.59 | 189 | *0.56* | -0.80 | 0.43 |
| SPL right high-freq | -0.25 | 0.43 | -0.57 | 189 | *0.57* | -1.10 | 0.61 |
| SPL left high-freq | 0.62 | 0.37 | 1.69 | 189 | *0.09* | -0.11 | 1.34 |
| dlPFC right low-freq | 0.00 | 0.31 | 0.00 | 189 | *1.00* | -0.62 | 0.62 |
| dlPFC left low-freq | -0.44 | 0.24 | -1.82 | 189 | *0.07* | -0.92 | 0.04 |
| vPMC right low-freq | 0.28 | 0.28 | 1.03 | 189 | *0.31* | -0.26 | 0.83 |
| vPMC left low-freq | 0.13 | 0.24 | 0.54 | 189 | *0.59* | -0.35 | 0.60 |
| TPJ right low-freq | 0.19 | 0.22 | 0.85 | 189 | *0.40* | -0.25 | 0.62 |
| **TPJ left low-freq** | **0.80** | **0.23** | **3.46** | **189** | ***<0.001*** | **0.35** | **1.26** |
| SPL right low-freq | 0.30 | 0.24 | 1.25 | 189 | *0.21* | -0.17 | 0.77 |
| SPL left low-freq | -0.05 | 0.25 | -0.20 | 189 | *0.84* | -0.54 | 0.44 |
| separator:dlPFC right high-freq | 0.06 | 0.26 | 0.22 | 189 | *0.82* | -0.46 | 0.57 |
| separator:dlPFC left high-freq | -0.46 | 0.29 | -1.57 | 189 | *0.12* | -1.04 | 0.12 |
| separator:vPMC right high-freq | -0.49 | 0.30 | -1.62 | 189 | *0.11* | -1.09 | 0.11 |
| **separator:vPMC left high-freq** | **-0.92** | **0.25** | **-3.72** | **189** | ***<0.001*** | **-1.41** | **-0.43** |
| separator:TPJ right high-freq | -0.05 | 0.27 | -0.18 | 189 | *0.85* | -0.58 | 0.49 |
| separator:TPJ left high-freq | -0.28 | 0.24 | -1.15 | 189 | *0.25* | -0.76 | 0.20 |
| separator:SPL right high-freq | -0.18 | 0.32 | -0.58 | 189 | *0.56* | -0.81 | 0.44 |
| separator:SPL left high-freq | -0.03 | 0.27 | -0.10 | 189 | *0.92* | -0.56 | 0.51 |
| separator:dlPFC right low-freq | 0.02 | 0.23 | 0.09 | 189 | *0.93* | -0.43 | 0.47 |
| separator:dlPFC left low-freq | 0.06 | 0.19 | 0.32 | 189 | *0.75* | -0.31 | 0.43 |
| separator:vPMC right low-freq | -0.03 | 0.20 | -0.16 | 189 | *0.88* | -0.43 | 0.37 |
| separator:vPMC left low-freq | 0.26 | 0.20 | 1.31 | 189 | *0.19* | -0.13 | 0.66 |
| separator:TPJ right low-freq | 0.43 | 0.19 | 2.21 | 189 | *0.03* | 0.05 | 0.82 |
| separator:TPJ left low-freq | 0.12 | 0.19 | 0.67 | 189 | *0.51* | -0.25 | 0.49 |
| separator:SPL right low-freq | -0.43 | 0.19 | -2.19 | 189 | *0.03* | -0.81 | -0.04 |
| separator:SPL left low-freq | 0.08 | 0.19 | 0.41 | 189 | *0.68* | -0.30 | 0.45 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| separator:JA | 0.00 | 0.00 | 0.13 | 189 | *0.90* | 0.00 | 0.00 |
| dlPFC right high-freq:JA | 0.01 | 0.01 | 1.22 | 189 | *0.22* | -0.01 | 0.02 |
| **dlPFC left high-freq:JA** | **0.02** | **0.01** | **2.86** | **189** | ***<0.001*** | **0.01** | **0.04** |
| vPMC right high-freq:JA | 0.00 | 0.01 | 0.09 | 189 | *0.93* | -0.02 | 0.02 |
| vPMC left high-freq:JA | 0.01 | 0.01 | 2.17 | 189 | *0.03* | 0.00 | 0.02 |
| **TPJ right high-freq:JA** | **-0.02** | **0.01** | **-2.54** | **189** | ***0.01*** | **-0.03** | **0.00** |
| TPJ left high-freq:JA | 0.00 | 0.01 | -0.14 | 189 | *0.89* | -0.01 | 0.01 |
| PAR_r_high:JA | 0.01 | 0.01 | 0.60 | 189 | *0.55* | -0.01 | 0.02 |
| PAR_l_high:JA | -0.01 | 0.01 | -1.24 | 189 | *0.22* | -0.02 | 0.00 |
| DLPF_r_low:JA | 0.00 | 0.01 | 0.53 | 189 | *0.60* | -0.01 | 0.01 |
| DLPF_l_low:JA | 0.00 | 0.00 | 1.14 | 189 | *0.26* | 0.00 | 0.01 |
| vPMC right low-freq:JA | 0.00 | 0.00 | -0.48 | 189 | *0.63* | -0.01 | 0.01 |
| vPMC left low-freq:JA | 0.00 | 0.00 | -1.04 | 189 | *0.30* | -0.01 | 0.00 |
| TPJ right low-freq:JA | -0.01 | 0.00 | -1.52 | 189 | *0.13* | -0.02 | 0.00 |
| **TPJ left low-freq:JA** | **-0.01** | **0.00** | **-2.56** | **189** | ***0.01*** | **-0.02** | **0.00** |
| SPL right low-freq:JA | 0.00 | 0.00 | 0.42 | 189 | *0.67* | -0.01 | 0.01 |
| SPL left low-freq:JA | 0.00 | 0.00 | -0.17 | 189 | *0.86* | -0.01 | 0.01 |

*Separator 'on' model.* As a reminder, the formula for this model is reported below:

> *Learning in separator 'on' ~ brain-coherence*$_i$ *+ joint attention + joint-attention*brain-coherence*$_i$ *+ (1 | learner sustained-attention) + (1 | dyad sustained-attention) + (1 | item) + (1 | teacher performance) + (1 | dyad)*

Results for separator 'on' model are reported in Table 5.3. The final model after the stepwise iteration analysis is reported in Table 5.4. The stepwise analysis showed that variability in learning was significantly predicted by joint attention (beta=.01, *p-value<.001*), brain synchrony in left TPJ for low frequency band (beta=1.49, *p-value<.001*) and the interaction between the two (joint attention * left TPJ for low frequency band, beta=-.02, *p-value<.001*). In addition, we found a main effect of brain synchrony in vPMC for high frequency band (beta=-.54, *p-value<.001*).

**Table 5.3 Separator 'on' model**

dlPFC: dorsolateral pre-frontal cortex; vPMC: ventral pre-motor cortex; TPJ: temporo-parietal junction; SPL: superior parietal lobe; JA: joint attention.

| **Predictor Term** | **Estimate** | **SE** | **tStat** | **DF** | ***p*** | **Lower** | **Upper** |
|---|---|---|---|---|---|---|---|
| (Intercept) | -0.14 | 0.70 | -0.19 | 85 | *0.85* | -1.53 | 1.26 |
| dlPFC right high-freq | 0.76 | 0.73 | 1.03 | 85 | *0.31* | -0.70 | 2.22 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| dlPFC left high-freq | -1.87 | 0.94 | -2.00 | 85 | *0.05* | -3.73 | -0.01 |
| vPMC right high-freq | -0.87 | 1.13 | -0.77 | 85 | *0.45* | -3.12 | 1.38 |
| vPMC left high-freq | -1.64 | 0.69 | -2.38 | 85 | *0.02* | -3.00 | -0.27 |
| TPJ right high-freq | 1.15 | 0.74 | 1.56 | 85 | *0.12* | -0.32 | 2.63 |
| TPJ left high-freq | 0.06 | 0.61 | 0.09 | 85 | *0.93* | -1.16 | 1.27 |
| SPL right high-freq | 0.10 | 1.02 | 0.10 | 85 | *0.92* | -1.92 | 2.13 |
| SPL left high-freq | 0.06 | 0.83 | 0.07 | 85 | *0.95* | -1.59 | 1.70 |
| dlPFC right low-freq | -0.52 | 0.80 | -0.65 | 85 | *0.52* | -2.11 | 1.07 |
| dlPFC left low-freq | 0.07 | 0.52 | 0.13 | 85 | *0.90* | -0.96 | 1.09 |
| vPMC right low-freq | 1.14 | 0.94 | 1.21 | 85 | *0.23* | -0.74 | 3.01 |
| vPMC left low-freq | 0.64 | 0.57 | 1.12 | 85 | *0.26* | -0.50 | 1.79 |
| TPJ right low-freq | 0.95 | 0.56 | 1.68 | 85 | *0.10* | -0.17 | 2.07 |
| **TPJ left low-freq** | **1.93** | **0.48** | **3.99** | **85** | ***<0.001*** | **0.97** | **2.90** |
| SPL right low-freq | 0.37 | 0.56 | 0.66 | 85 | *0.51* | -0.74 | 1.48 |
| SPL left low-freq | -0.27 | 0.55 | -0.50 | 85 | *0.62* | -1.36 | 0.81 |
| JA | 0.01 | 0.01 | 1.32 | 85 | *0.19* | -0.01 | 0.03 |
| dlPFC right high-freq:JA | -0.01 | 0.01 | -0.98 | 85 | *0.33* | -0.03 | 0.01 |
| dlPFC left high-freq:JA | 0.03 | 0.01 | 2.28 | 85 | *0.03* | 0.00 | 0.05 |
| vPMC right high-freq:JA | 0.01 | 0.02 | 0.51 | 85 | *0.61* | -0.02 | 0.04 |
| vPMC left high-freq:JA | 0.02 | 0.01 | 1.61 | 85 | *0.11* | 0.00 | 0.03 |
| TPJ right high-freq:JA | -0.02 | 0.01 | -1.60 | 85 | *0.11* | -0.04 | 0.00 |
| TPJ left high-freq:JA | -0.01 | 0.01 | -0.89 | 85 | *0.38* | -0.03 | 0.01 |
| SPL right high-freq:JA | 0.00 | 0.01 | -0.17 | 85 | *0.87* | -0.03 | 0.02 |
| SPL left high-freq:JA | 0.00 | 0.01 | -0.12 | 85 | *0.90* | -0.02 | 0.02 |
| dlPFC right low-freq:JA | 0.01 | 0.01 | 0.78 | 85 | *0.44* | -0.01 | 0.03 |
| dlPFC left low-freq:JA | 0.00 | 0.01 | 0.05 | 85 | *0.96* | -0.01 | 0.01 |
| vPMC right low-freq:JA | -0.02 | 0.01 | -1.28 | 85 | *0.20* | -0.04 | 0.01 |
| vPMC left low-freq:JA | 0.00 | 0.01 | -0.50 | 85 | *0.62* | -0.02 | 0.01 |
| TPJ right low-freq:JA | -0.01 | 0.01 | -1.19 | 85 | *0.24* | -0.02 | 0.01 |
| **TPJ left low-freq:JA** | **-0.03** | **0.01** | **-3.96** | **85** | ***<0.001*** | **-0.04** | **-0.01** |
| SPL right low-freq:JA | -0.01 | 0.01 | -0.66 | 85 | *0.51* | -0.02 | 0.01 |
| SPL left low-freq:JA | 0.01 | 0.01 | 1.05 | 85 | *0.30* | -0.01 | 0.02 |

**Table 5.4 Separator ‘on’ model after the stepwise iterative analysis**

dlPFC: dorsolateral pre-frontal cortex; vPMC: ventral pre-motor cortex; TPJ: temporo-parietal junction; SPL: superior parietal lobe; JA: joint attention.

| **Predictor Term** | **Estimate** | **SE** | **tStat** | ***p*** |
|---|---|---|---|---|
| (Intercept) | -0.34 | 0.31 | -1.08 | *0.28* |
| **JA** | **0.01** | **0.00** | **3.70** | ***<0.001*** |
| **vPMC left high-freq** | **-0.54** | **0.18** | **-3.02** | ***<0.001*** |
| **TPJ left high-freq** | **-0.47** | **0.18** | **-2.57** | ***0.01*** |
| vPMC right low-freq | 1.60 | 0.78 | 2.06 | *0.04* |

| | | | | |
|---|---|---|---|---|
| **vPMC left low-freq** | **0,36** | **0,14** | **2,48** | ***0,01*** |
| **TPJ right low-freq** | **0,34** | **0,13** | **2,60** | ***0,01*** |
| **TPJ left low-freq** | **1,49** | **0,44** | **3,38** | ***<0,001*** |
| SPL left low-freq | 0,31 | 0,15 | 2,14 | *0,03* |
| vPMC right low-freq:JA | -0,02 | 0,01 | -2,08 | *0,04* |
| **TPJ left low-freq:JA** | **-0,02** | **0,01** | **-3,38** | ***<0,001*** |

*Separator 'off' model.* As a reminder, the formula for this model is reported below:

> *Learning in separator 'off' ~ brain-coherence*$_i$ *+ joint attention + mutual-gaze + joint-attention*brain-coherence*$_i$ *+ mutual-gaze*brain-coherence*$_i$ *+ (1 | learner sustained-attention) + (1 | dyad sustained-attention) + (1 | item) + (1 | teacher performance) + (1 | dyad)*

Results for separator 'off' model are reported in Table 5.5. The final model after the stepwise iteration analysis is reported in Table 5.6. The stepwise analysis showed that variability in learning was significantly predicted by joint attention (beta=-.02, *p=.01*) and mutual gaze (beta=-.08, *p<.001*). In addition, we found a main effect of brain synchrony in left dlPFC for low frequency band (beta=-2.89, *p=.01*). Left dlPFC for low frequency band also showed a significant interaction effect with both joint attention (beta=.04, *p=.01*) and mutual gaze (beta=.10, *p<.001*).

**Table 5.5 Separator 'off' model**

dlPFC: dorsolateral pre-frontal cortex; vPMC: ventral pre-motor cortex; TPJ: temporo-parietal junction; SPL: superior parietal lobe.

| **Name** | **Estimate** | **SE** | **tStat** | **DF** | ***p*** | **Lower** | **Upper** |
|---|---|---|---|---|---|---|---|
| (Intercept) | 2.88 | 1.59 | 1.81 | 69.00 | *0.07* | -0.30 | 6.05 |
| JA | -0.03 | 0.02 | -1.37 | 69.00 | *0.17* | -0.08 | 0.01 |
| MG | -0.10 | 0.05 | -1.88 | 69.00 | *0.06* | -0.20 | 0.01 |
| dlPFC right high-freq | -1.82 | 1.67 | -1.09 | 69.00 | *0.28* | -5.15 | 1.50 |
| dlPFC left high-freq | -3.57 | 1.59 | -2.25 | 69.00 | *0.03* | -6.74 | -0.41 |
| vPMC right high-freq | 0.00 | 1.86 | 0.00 | 69.00 | *1.00* | -3.72 | 3.71 |
| vPMC left high-freq | -1.15 | 1.43 | -0.81 | 69.00 | *0.42* | -4.00 | 1.70 |
| TPJ right high-freq | 0.65 | 1.72 | 0.38 | 69.00 | *0.71* | -2.79 | 4.09 |
| TPJ left high-freq | 1.42 | 1.66 | 0.86 | 69.00 | *0.39* | -1.88 | 4.72 |
| SPL right high-freq | -1.59 | 1.72 | -0.93 | 69.00 | *0.36* | -5.01 | 1.84 |
| SPL left high-freq | -1.80 | 1.94 | -0.93 | 69.00 | *0.36* | -5.67 | 2.06 |
| dlPFC right low-freq | -0.60 | 1.16 | -0.51 | 69.00 | *0.61* | -2.91 | 1.72 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| dlPFC left low-freq | -1.62 | 1.29 | -1.26 | 69.00 | *0.21* | -4.19 | 0.94 |
| vPMC right low-freq | 1.19 | 0.86 | 1.38 | 69.00 | *0.17* | -0.53 | 2.92 |
| vPMC left low-freq | -0.06 | 1.09 | -0.06 | 69.00 | *0.96* | -2.24 | 2.12 |
| TPJ right low-freq | 2.01 | 1.02 | 1.97 | 69.00 | *0.05* | -0.03 | 4.05 |
| TPJ left low-freq | -0.93 | 1.09 | -0.85 | 69.00 | *0.40* | -3.10 | 1.24 |
| SPL right low-freq | -1.31 | 1.21 | -1.08 | 69.00 | *0.29* | -3.73 | 1.11 |
| SPL left low-freq | 0.83 | 1.58 | 0.52 | 69.00 | *0.60* | -2.32 | 3.97 |
| dlPFC right high-freq:JA | 0.03 | 0.03 | 1.06 | 69.00 | *0.29* | -0.02 | 0.08 |
| dlPFC left high-freq:JA | 0.06 | 0.03 | 2.35 | 69.00 | *0.02* | 0.01 | 0.11 |
| vPMC right high-freq:JA | 0.00 | 0.03 | 0.13 | 69.00 | *0.90* | -0.05 | 0.06 |
| vPMC left high-freq:JA | 0.02 | 0.02 | 1.11 | 69.00 | *0.27* | -0.02 | 0.07 |
| TPJ right high-freq:JA | -0.02 | 0.03 | -0.63 | 69.00 | *0.53* | -0.07 | 0.04 |
| TPJ left high-freq:JA | -0.01 | 0.03 | -0.58 | 69.00 | *0.56* | -0.07 | 0.04 |
| SPL right high-freq:JA | 0.02 | 0.03 | 0.77 | 69.00 | *0.44* | -0.03 | 0.08 |
| SPL left high-freq:JA | 0.03 | 0.03 | 1.01 | 69.00 | *0.32* | -0.03 | 0.08 |
| dlPFC right low-freq:JA | 0.01 | 0.02 | 0.78 | 69.00 | *0.44* | -0.02 | 0.05 |
| dlPFC left low-freq:JA | 0.02 | 0.02 | 1.00 | 69.00 | *0.32* | -0.02 | 0.06 |
| vPMC right low-freq:JA | -0.02 | 0.01 | -1.47 | 69.00 | *0.15* | -0.04 | 0.01 |
| vPMC left low-freq:JA | -0.01 | 0.02 | -0.47 | 69.00 | *0.64* | -0.04 | 0.03 |
| TPJ right low-freq:JA | -0.03 | 0.02 | -1.84 | 69.00 | *0.07* | -0.06 | 0.00 |
| TPJ left low-freq:JA | 0.01 | 0.02 | 0.52 | 69.00 | *0.61* | -0.03 | 0.04 |
| SPL right low-freq:JA | 0.03 | 0.02 | 1.28 | 69.00 | *0.20* | -0.01 | 0.06 |
| SPL left low-freq:JA | -0.01 | 0.02 | -0.64 | 69.00 | *0.53* | -0.06 | 0.03 |
| dlPFC right high-freq:MG | 0.03 | 0.06 | 0.57 | 69.00 | *0.57* | -0.09 | 0.15 |
| dlPFC left high-freq:MG | 0.07 | 0.05 | 1.55 | 69.00 | *0.13* | -0.02 | 0.17 |
| vPMC right high-freq:MG | -0.01 | 0.06 | -0.12 | 69.00 | *0.90* | -0.12 | 0.11 |
| vPMC left high-freq:MG | 0.02 | 0.05 | 0.43 | 69.00 | *0.67* | -0.07 | 0.11 |
| TPJ right high-freq:MG | 0.04 | 0.06 | 0.69 | 69.00 | *0.50* | -0.08 | 0.16 |
| TPJ left high-freq:MG | -0.07 | 0.05 | -1.38 | 69.00 | *0.17* | -0.18 | 0.03 |
| SPL right high-freq:MG | 0.08 | 0.05 | 1.43 | 69.00 | *0.16* | -0.03 | 0.19 |
| SPL left high-freq:MG | 0.07 | 0.07 | 1.02 | 69.00 | *0.31* | -0.06 | 0.20 |
| dlPFC right low-freq:MG | 0.03 | 0.05 | 0.71 | 69.00 | *0.48* | -0.06 | 0.12 |
| dlPFC left low-freq:MG | 0.07 | 0.04 | 1.67 | 69.00 | *0.10* | -0.01 | 0.16 |
| vPMC right low-freq:MG | -0.04 | 0.03 | -1.27 | 69.00 | *0.21* | -0.11 | 0.02 |
| vPMC left low-freq:MG | 0.02 | 0.03 | 0.56 | 69.00 | *0.58* | -0.05 | 0.09 |
| TPJ right low-freq:MG | -0.07 | 0.04 | -1.93 | 69.00 | *0.06* | -0.14 | 0.00 |
| TPJ left low-freq:MG | 0.07 | 0.04 | 1.95 | 69.00 | *0.05* | 0.00 | 0.14 |
| SPL right low-freq:MG | 0.04 | 0.04 | 1.14 | 69.00 | *0.26* | -0.03 | 0.12 |
| SPL left low-freq:MG | -0.01 | 0.05 | -0.28 | 69.00 | *0.78* | -0.12 | 0.09 |

**Table 5.6 Separator ‘off’ model after the stepwise iterative analysis**

dlPFC: dorsolateral pre-frontal cortex; TPJ: temporo-parietal junction; SPL: superior parietal lobe; JA: joint attention; MG: mutual gaze.

| **Predictor Term** | **Estimate** | **SE** | **tStat** | ***p*** |
|---|---|---|---|---|
| (Intercept) | 2.18 | 0.44 | 4.99 | *0.00* |
| **JA** | **-0.02** | **0.01** | **-2.59** | ***0.01*** |
| **MG** | **-0.08** | **0.02** | **-4.13** | **<*0.001*** |
| SPL right high-freq | -0.60 | 0.36 | -1.66 | *0.10* |
| **dlPFC left low-freq** | **-2.89** | **1.03** | **-2.80** | ***0.01*** |
| TPJ left low-freq | -0.12 | 0.21 | -0.57 | *0.57* |
| SPL right low-freq | -1.80 | 0.95 | -1.90 | *0.06* |
| **dlPFC left low-freq:JA** | **0.04** | **0.02** | **2.51** | ***0.01*** |
| SPL right low-freq:JA | 0.03 | 0.02 | 1.72 | *0.09* |
| SPL right high-freq:MG | 0.07 | 0.03 | 1.98 | *0.05* |
| **dlPFC left low-freq:MG** | **0.10** | **0.03** | **3.08** | **<*0.001*** |
| TPJ left low-freq:MG | 0.04 | 0.02 | 2.11 | *0.04* |
| SPL right low-freq:MG | 0.06 | 0.03 | 2.23 | *0.03* |

### 5.3.1 *Is learning predicted by what *we* looked at OR by what *I* looked at?*

Consistently across different models, including the big exploratory model and the stepwise iterative analysis run separately for the two experimental conditions, joint attention emerged as a significant predictor of learning, with greater joint attention resulting in better learning. For the separator ‘off’ condition, also mutual gaze was found to significantly contribute to learning, with people generally learning more when engaging *less* in mutual-gaze. Note that both these measures reflect interactive behaviour, that is eye-gaze dynamic between learner and teacher. In other words, we observed that learning was predicted by inter-personal eye-gaze dynamic, even when controlling for individual eye-gaze pattern (sustained attention). To confirm that these effects reflect a truly interactive dynamic and are not simply driven by how long the learner spent fixating the object (regardless of what their partner was doing), we run a further much simpler model where we tested the isolated contribution of eye-gaze behavioural measures to learning, separately for separator ‘on’ and separator ‘off’ condition:

- *Learning in separator ‘on’ ~ joint attention + learner sustained-attention + (1 | item) + (1 | teacher performance) + (1 | dyad)*
- *Learning in separator ‘off’ ~ joint attention + mutual-gaze + learner sustained-attention + (1 | item) + (1 | teacher performance) + (1 | dyad)*

Results from these models are reported in table 5.7. Eye-gaze behaviour and learning data is also plotted in Figure 5.7 separately for each condition. For separator 'on' condition, the model showed that none of the eye-gaze behavioural measures included on its own could account for variability in learning. For separator 'off' condition, we found that joint attention (beta=.003, *p*=*.04*) was the only significant predictor of learning, above mutual gaze and learner sustained attention. Importantly, these results demonstrate that sustained attention cannot account for variability in learning, and that better learning was predicted specifically by time when learner and teacher attended to the object simultaneously.

**Table 5.7 Eye-gaze behaviour models**

Separator 'on'

| | **Estimate** | **SE** | **tStat** | **DF** | ***p*** | **Lower** | **Upper** |
|---|---|---|---|---|---|---|---|
| (Intercept) | 0.80 | 0.06 | 13.20 | 118 | *<0.001* | 0.68 | 0.92 |
| Joint attention | 0.00 | 0.00 | 0.33 | 118 | *0.74* | 0.00 | 0.00 |
| Sustained attention (learner) | 0.00 | 0.00 | -1.35 | 118 | *0.18* | 0.00 | 0.00 |

Separator 'off'

| | **Estimate** | **SE** | **tStat** | **DF** | ***p*** | **Lower** | **Upper** |
|---|---|---|---|---|---|---|---|
| (Intercept) | 0.625 | 0.090 | 6.948 | 118 | *<0.001* | 0.447 | 0.803 |
| **Joint attention** | **0.003** | **0.001** | **2.004** | **118** | ***0.04*** | **0.00** | **0.005** |
| Mutual gaze | 0.002 | 0.003 | 0.917 | 118 | *0.36* | -0.003 | 0.007 |
| Sustained attention (learner) | 0.000 | 0.001 | -0.329 | 118 | *0.74* | -0.002 | 0.002 |

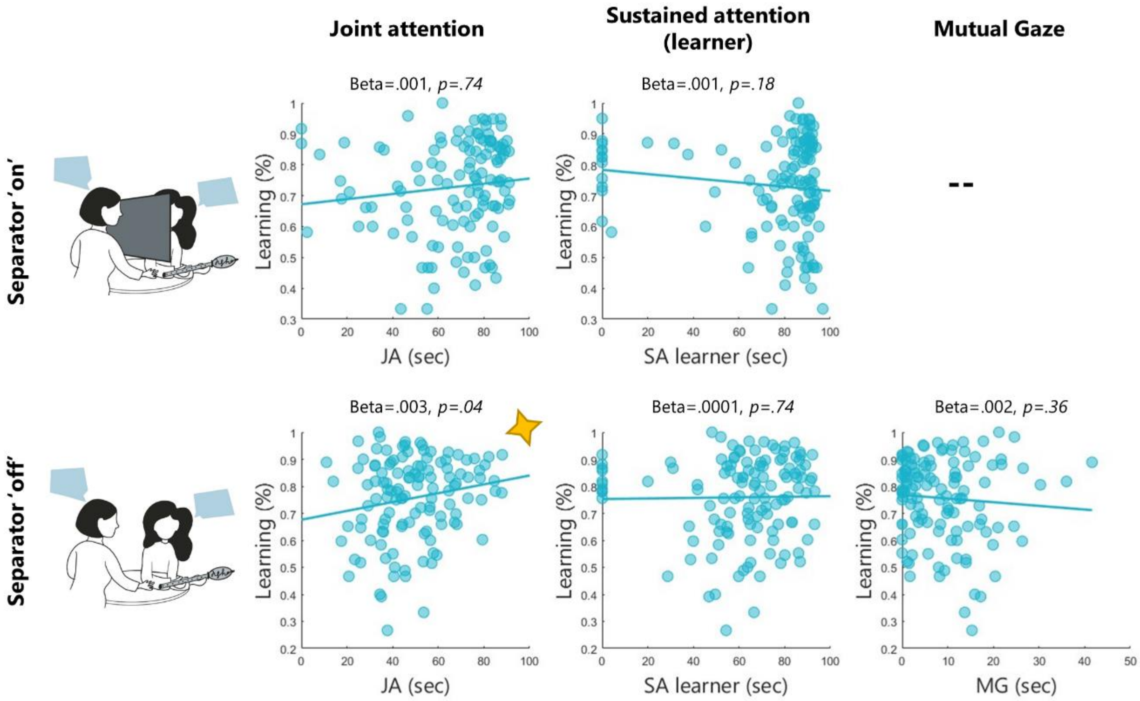


**Figure 5.7 Scatterplots of eye-gaze behaviour and learning**

Note that beta and *p* values reported here comes from the eye-gaze behavioural model only. Significant main effects are indicated with a yellow star. For statistics on these measures and their contribution in predicting learning when considered along brain synchrony measures refer to text and Table 5.4 and 5.6 and Figure 5.8. JA: joint attention; SA: sustained attention; MG: mutual gaze.

*Results summary*. Overall, brain synchrony emerged as a predictor of learning in the left hemisphere over dlPFC and TPJ. Interestingly, these same areas also showed interaction effects with eye-gaze behaviours, namely joint attention and mutual gaze. In left dlPFC, we found that during separator 'on' trials, only joint attention predicted learning (Fig5.8A). However, during separator 'off' trials when participants could see each other, coherence of left dLPFC predicted learning both alone and in interaction with both joint attention and mutual gaze (Fig 5.8B and C). While increased dlPFC coherence always related positively to learning, the effects of joint attention and mutual gaze went in opposite directions.

In left TPJ, we also found interactions of gaze brain coherence to predict learning. Specifically, left TPJ showed a positive interaction with joint attention during trials when learner and teacher could *not* see each other (separator 'on' condition, Figure

5.8D). Although no other interactions survived our alpha threshold for significance of .01, it is interesting to observe that the trends between brain synchrony and eye-gaze behaviour in predicting learning remain consistent: it seems that the best learning is associated with greater joint attention and less mutual gaze, as well as more brain synchrony in left dlPFC and left TPJ.

Finally, coherence in left vPMC, a regions heavily engaged in speech planning and production, showed a significant negative effect in predicting learning specifically during separator ‘on’ condition (see Table 5.4 and Figure 5.5): in other words, better learning was associated with *less* synchrony over this area.

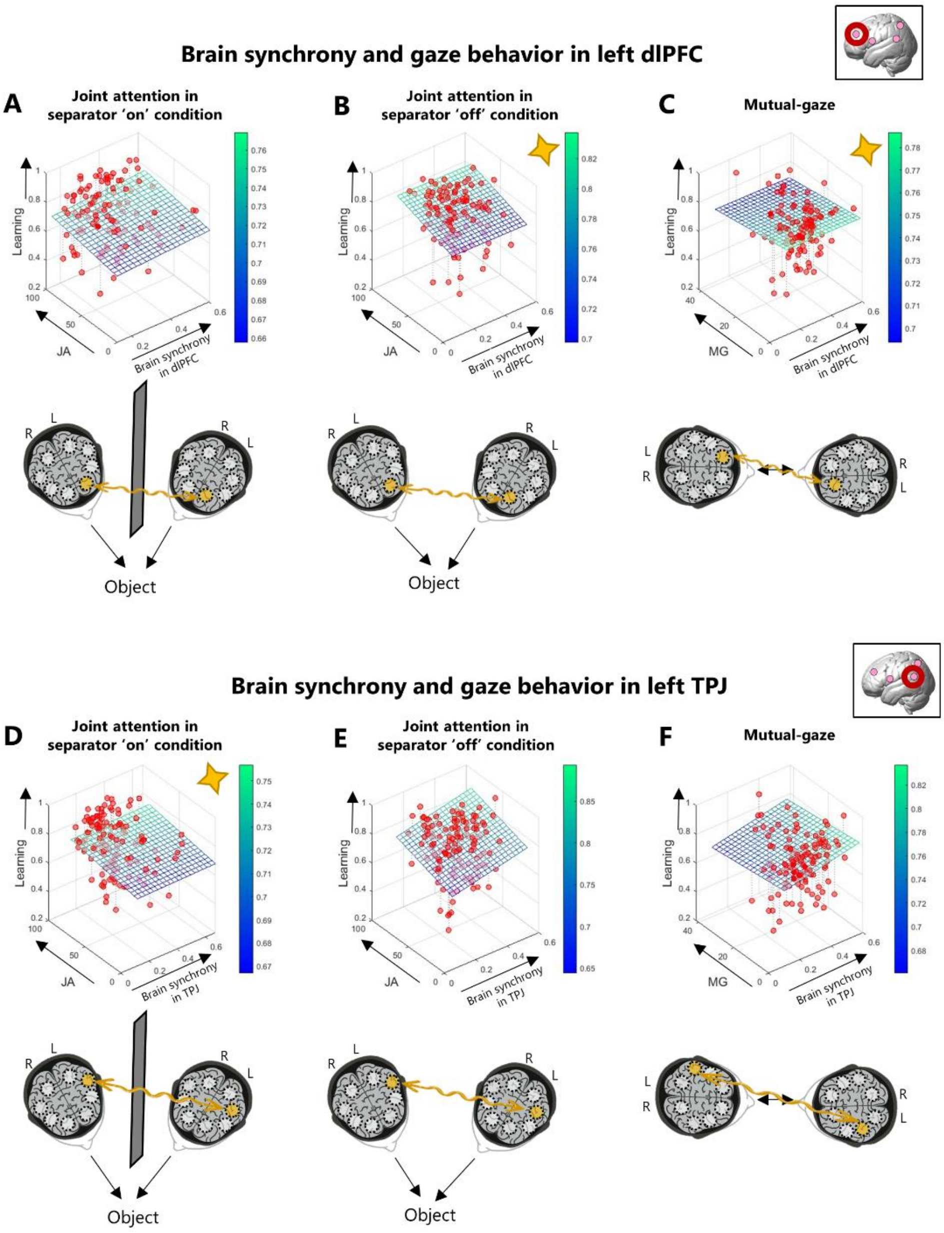


**Figure 5.8 3d plots of the two-way relationship between brain synchrony and eye-gaze behaviour over left dlPFC and TPJ**

Significant main effects are indicated with a yellow star. Below each plot is a schematic of the behavioural condition of the plotted data. L: left; R: right; JA: joint attention; MG: mutual gaze; dlPFC: dorsolateral pre-frontal cortex; TPJ: temporo-parietal junction.

## 5.4 Discussion

The aim of this study was to identify markers of successful interactive learning. To do so, we adopted a multi-modal approach to study of naturalistic teacher-learner interaction, in line with the idea that interactive brains exist within interactive bodies

(Hamilton, 2020). Using hyperscanning fNIRS, we measured brain activity from learner and teacher simultaneously as they exchanged information about obscure objects (e.g. exotic animals), while also coding their eye-gaze behaviour (see Figure 5.1). Specifically, we computed brain synchrony in 26 learner-teacher pairs over dorsolateral pre-frontal cortex (dlPFC), ventral pre-motor cortex (vPMC), temporo-parietal junction (TPJ) and superior parietal lobe (SPL) bilaterally, as well as measuring the time each pair spent in joint attention and mutual gaze behaviour, and asked whether and how these signals predicted learning.

Our findings show that learner-teacher brain synchrony is a good marker of learning specifically over left dlPFC and TPJ. Crucially, the relationship between brain synchrony and learning was *not* linear, i.e. more synchrony did not always resulted in better learning, but was modulated by eye-gaze behaviour, in particular joint attention and mutual-gaze. This was particularly evident as we experimentally manipulated the learner's and the teacher's view by placing a separator, i.e. a physical barrier, that in some trials obstructed their view, preventing them to engage in mutual-gaze behaviour, but not in joint attention (see Fig. 5.1 and 5.2). We discuss these results by considering learning performance in relation to joint attention and mutual-gaze separately.

First, we found that greater joint attention was associated with better learning, and showed a positive interaction with brain synchrony over both left dlPFC and TPJ. Joint attention has been shown to be a hallmark of good conversations (Garrod & Pickering, 2004; Richardson et al., 2007) and successful learning (Schertz et al., 2013; Striano et al., 2006). Also, previous literature has placed dlPFC and TPJ as important hubs within the brain network engaged in social cognition and specifically in face-to-face communication (Burgess et al., 2022; Jiang et al., 2015; Suda et al., 2010). Our results are also consistent with previous work in confirming the involvement of these areas in eye-gaze social dynamics (Jiang et al., 2017; Noah et al., 2020). However, when considering the two-way relationship between brain synchrony and joint attention in predicting learning, we suggest that the underlying cognitive mechanisms engaging left dlPFC and left TPJ may be specific to each of these two areas.

dlPFC has been shown to be particularly important in our ability to remain 'on-task' (Turnbull et al., 2019). Using fMRI, Turnbull and colleagues measured whole-brain neural activity as participants engaged in working-memory task varying in difficultly (a classic n-back paradigm where an always increasing number of digits has to be retrieved backwards). During the task, participants were randomly asked whether their thoughts were focused on the task or they were daydreaming. Results showed that activity in left dlPFC was correlated with being on-task specifically when task demands were higher, and off-task thoughts when demands were lower. In our study, participants were explicitly instructed to *teach and learn* as many information as possible, a task that involve high cognitive resources. Crucially, we found that higher synchrony over this area and the consequent benefit to learning was specifically associated with instances of joint attention, where teacher and learner where 'on task' together at the same time. Therefore, left dlPFC may have played a crucial role in supporting both teacher's task of sharing information clearly and accurately, and learner's task of *simultaneously* attending to that information in order to memorise as many facts as possible.

While synchrony over left dlPFC may support learning as reflecting teacher and learner *jointly attending to the task*, previous work suggests that synchrony over TPJ may instead reflect processing of inter-personal dynamics, or in other words teacher and learner *attending to each other*. This may contribute to co-creating a smoother communication which in turn increases learning. TPJ has been shown to be implicated in theory of mind and shared understanding (Doricchi et al., 2022; Jiang et al., 2015; Zheng et al., 2018), as well as being important in predicting other's action (Kayhan et al., 2022) and monitoring verbal exchange and turn-taking in contingent communication (Cañigueral, Zhang, et al., 2021a; Liu et al., 2019). More relevant for the present study, TPJ has been reported among the core areas to show brain synchrony during learner-teacher interaction in a number of studies on learning (Y. Pan, Novembre, et al., 2020; Zheng et al., 2018, 2020). Here, we show that synchrony over this area is directly modulated by joint attention in supporting effective communication, which consequently resulted in better learning.

In contrast to what we observed for joint attention, our second main finding shows that mutual-gaze was associated with both less brain synchrony and less learning over dlPFC. The fact that mutual gaze was associated with less brain synchrony over dlPFC is novel, yet relatively unsurprising. People frequently avert their eyes from interlocutors during natural conversation. Previous work has also confirmed this trend experimentally, showing that direct eye-contact disrupts our ability to produce speech (Kajimura & Nomura, 2016), therefore suggesting that smooth conversations may be associated with less mutual-gaze. In fact, it has been recently shown that eye-contact may be specifically employed to 'disrupt' synchrony (measured in terms of pupil size) in naturalistic interaction, by providing a signal to facilitate independent contributions to conversation (Wohltjen & Wheatley, 2021). Therefore, our results are consistent with previous work in showing that mutual-gaze is a social signal that lead people to 'de-synchronize' rather than 'synchronize' to their partner. Notably, overall we observe relatively little mutual-gaze behaviour in our data: precisely, on an average trial, mutual gaze was observed for 10% of the time, while joint attention was observed for 65% of the time. This is consistent with the interpretation that participants may have used this social signal to only briefly disengage from the shared task and allow independent internal processing, before 're-connecting' into shared information processing. In line with this interpretation, we do not find a significant reduction in learning as a function of mutual gaze (see Figure 5.7 and 5.8).

Lastly, we found that during separator 'on' condition, i.e. when learner and teacher could not see each other's face, better learning was predicted by *less* synchrony over vPMC. vPMC is typically engaged in speech planning and verbal production (Tremblay & Small, 2011). We identified two possible explanations for this result, which may be equally possible and not necessarily mutually exclusive. The first one develops from the mutual prediction framework: under this framework, decoupling is linked to learning because in a teacher-learner exchange, information to be learnt is novel and therefore – by definition – unpredictable. In other words, vPMC might fail to predict the exact words that teacher will say during conversation, given that the teacher's job is to share *unknown* information about a *novel* object. This would therefore results in low coherence and good learning. Note that this interpretation would assume that the same mechanism occurs during both separator 'on' (obstructed view) and separator 'off' (face-to-face view) condition, while our results specifically detect an effect during obstructed view only. It is possible that no effect is detected in face-to-face view (separator 'off' condition), because the additional visual signals may have helped good (speech) prediction. In fact, it has been recently demonstrated that multi-modal (non-linguistic) cues, including gestures and mouth movements, supports language comprehension (Y. Zhang et al., 2021). Our separator 'on' condition specifically isolates the contribution of linguistic (verbal) information from multi-modal cues, possibly resulting in less brain coherence associated with better learning.

Alternatively, it may be that greater de-coupling leads to better learning as it reflects internal mechanism of self-monitoring. Recently, vPMC has been included in models of speech perception and in particular is believed to monitor one's own speech (Ozker et al., 2022; Scott et al., 2020). Our task is heavily verbal: de-coupling from the partner may have reflected the ability of the speaker to 'tune in' with their own speech to integrate auditory feedback and adapt communication accordingly. Presumably, this would be particularly relevant in the case of the teacher, who would be the one predominantly speaking and sharing information about the object to learn. During this process, self-monitoring would possibly have a two-fold effect: in terms of brain coherence, it would result in more de-coupling as the teacher would focus on their own performance more than focusing on the interaction (e.g. what *they themselves* have just said, what *they themselves* are about to say next etc), resulting in a divergence in brain activity patterns. In addition, in terms of learning, greater self-monitoring could make for a better teacher, which would translate in better learning. This interpretation is consistent with the fact that the association between greater de-coupling and better learning was only found in our separator 'on' condition, i.e. when visual feedback was absent and the view between the teacher and the learner was obstructed. It may be that relying on auditory feedback from one's own speech may have been particular important during trials where it was not possible to rely on feedback from the partner (e.g. nodding to communicate understanding).

We acknowledge that both of these explanations rely on a number of assumptions and that many of these conclusions remain speculations. Future work is needed to disentangle the specific contribution of each brain networks in coupling and de-coupling and how these inter-brain dynamics impact learning. Overall however, our data points towards the conclusion that the relationship between learner-teacher brain-to-brain synchrony and learning is *non*-linear: different brain regions at different levels of the sensorimotor hierarchy supports learning by either coupling or de-coupling from their analogous region in the partner's brain. This is likely to depend on which signals those regions are responsible for computing and how accurately they are at predicting a partner's actions, beliefs and/or speech during the task. This could therefore mean that good learning would result, for example, from low coherence in areas engaged in predicting the precise word sequence in speech (vPMC), and from high coherence in areas engaged in predicting the overall arc of the conversation and the turn-taking structure (dlPFC and TPJ), in a complex non-linear system, that synergistically work to achieve an effective representation of the items under discussion.

### *Future Directions*

These results are novel in that they investigate social learning by directly linking brain and behavioural dynamics in naturalistic interaction. However, the focus is limited to eye-gaze behaviour. Future studies should go further and integrate other behavioural signals, especially verbal behaviour, to better disentangle the contribution of these different mechanisms to knowledge acquisition (Cartmill et al., 2013; Ponari et al., 2018). In addition, this work studied inter-personal dynamics by looking at the dyadic interaction, but did not investigate further the directionality of such dynamics. In the future, more sophisticated analysis tools (e.g. Granger causality) should be employed to make more precise inferences on mechanisms regulating mutual-prediction and information flowing from one person (e.g. teacher) to another (e.g. learner). While this study of dyadic interaction provides a good framework to build model of naturalistic conversation, the study of interactive learning should be extended to group dynamics, in order to make conclusions that can apply to educational setting more similar to the real-world (e.g. classroom).

Previous work has identified social networks as an important modulator of inter-personal brain dynamics (Parkinson et al., 2018; Sievers et al., pre-print). It remains unclear whether the relationship existing between teacher and learner affects the mechanisms studied here, e.g. by modulating the inter-personal brain and behavioural pattern and, in turn, learning outcomes. Due to restrictions imposed by the 2020 global pandemic, in this study, participants in each pair were recruited from the same household. Although we do not anticipate different results in dyads of strangers, as the cognitive mechanisms identified as potential contributors in supporting interactive learning should still hold, this will need to be experimentally tested. In line with this, it would be particularly interesting to study these questions in adolescence, a sensitive period for brain development, particularly for social cognition, and a time when relational social network are notably important for optimal development, mental health and wellbeing (Becht et al., 2021; Fuhrmann et al., 2015).

In conclusion, we found that during trials with more joint attention, inter-brain synchrony over left dlPFC and left TPJ increases and this is linked to better learning. This may reflect how participant's coordinating attending to the object drives both brain activity and learning.  In contrast, during trials with more mutual gaze, we observed a pattern in the opposite direction, where learner-teacher were less synchronized over these brain regions.  This may reflect mutual gaze crowding-out joint attention (because the two states cannot occur simultaneously) or it could reflect moments of internal *un*shared integration of information. In addition, specifically when partners could not see each other, learning was higher as their brain synchrony over vPMC was lower. This possibly reflects unsuccessful prediction of novel verbal information about to-be-learnt items; or alternatively it may reflect greater self-monitoring, as the teacher relies more on audio-feedback from their own speech when their view of the learner is obstructed, to achieve an effective contingent communication. Overall, we interpreted these results in line with previous work which identify dlPFC, TPJ and vPMC as the hubs for social brain networks and involved in shared-understanding, turn-taking, monitoring and mutual-prediction during naturalistic verbal communication. Importantly, these findings go beyond previous literature by showing the complex interplay between inter-personal brain synchrony and eye-gaze dynamics and how this supports learning in social contexts.

# 6. General Discussion

The goal of this thesis has been that of studying what factors of social interaction support learning of conceptual knowledge. On the premise that learning most naturally occurs socially, it investigated different aspects of naturalistic social exchange to identify the optimal condition for learning. Crucially, it distinguishes learning *via* the social from learning *about* the social, and asks the question of how *social interaction* supports declarative processing of *non*-social material. In doing so, we conducted four studies which prioritised ecological validity: all experiments involved relatively unconstrained teacher-learner interaction, and learning material resembled documentary-like content. The first two experimental chapters showed a series of studies on how adults learn in online contexts, while the third and fourth experimental chapter focused on face-to-face interaction and used functional Near-Infrared Spectroscopy (fNIRS) hyperscanning and wavelet transform coherence (WTC) analysis to measure brain synchrony in naturalistic interactions. We first summarise results from each chapter. We then discuss these in a broader context, outlining some general conclusions this work has reached as well as recognising its limitations. We finally outline some outstanding questions for future research in this area.

## 6.1 Summary of experimental chapters

Study 1 presents two online experiments, where social contingency (i.e. being part of a live interaction vs observing a pre-recorded one) and social cues (i.e. teacher's webcam on vs off vs showing a slide only) were manipulated. We asked participants to learn obscure items (e.g. exotic animals) either in live interaction with the teacher or from recorded videos. Results showed that learning in live interaction was associated with the best performance, and live social interaction with a full view of the teacher provided the optimal setting for learning, while seeing a slide had greater benefit during recorded sessions specifically. We suggest that while overall social interaction supports learning over less interactive contexts, social signals may specifically benefit learning during contingent interaction, as these serve the function of facilitating the feed-forward and feed-back exchange between interlocutors. In contrast, when the learner is observing a social interaction which they are not *part of*, social signals lose their functional role, while increasing the cognitive demand of the task, resulting in less learning. This interpretation is consistent with studies showing that being an observer of a social interaction is more cognitively demanding than actively engaging in that interaction (Kourtis et al., 2020), when social cues may become distracting (Friedman & Förster, 2010; Kirkorian, 2016; Phillips et al., 2007).

Study 2 replicates the live-learning advantage found in study 1, across two experiments and a large sample of adults with and without Autistic Spectrum Condition (ASC). Specifically, we asked participants to learned about obscure items (same as in Study 1) over zoom in three conditions: i) in live video-call with the teacher, ii) from a recorded video of a previous student-teacher interaction and iii) from a recorded video of the teacher alone. Results across both experiments and with three different teachers robustly confirmed that people learned better in live-video calls compared to both recorded conditions. Importantly, we found no difference in learning pattern between the NT and the ASC group. Mixed-linear effects model also revealed that enjoyment was the strongest predictor of learning. Both groups enjoyed learning significantly more in live interaction compared to learning from pre-recorded sessions, and reported similar levels of anxiety in all conditions. These findings advocate for a distinction between learning *about* the social versus learning *via* the social: while the former has been suggested to be atypical in ASC (Happé & Frith, 2014; Mundy & Newell, 2007), social interaction represents an effective medium through which learning can be boosted in both neurotypical and autistic people. Models of ASC suggesting dysfunctional social processes across social cognition should be revisited to consider social interaction not just as a puzzle to de-code, but rather a medium through which we experience the world around us (De Jaegher et al., 2010; Pfeiffer et al., 2014), which may support cognitive functioning across neuro-diverse populations.

Studies 3 and 4 aimed at identifying the effect of face-to-face social interaction on inter-subject neural synchrony and learning. Before looking at the complex interplay between social interaction, brain-to-brain dynamics and learning (Study 4), in study 3 we first asked whether and how co-presence and conversation change people's neural experience of the world. We invited pairs of familiar adults to co-watch two different episodes of the BBC DipDap series while we measured brain activity simultaneously from the two participants in each pair over frontal, temporal and parietal regions bilaterally. Compared to shuffled pseudo pairs (who had never seen each other and watched the same movie at different times), real pairs showed increased brain synchrony over left Temporo-Parietal Junction (TPJ) and reduced brain synchrony over left Superior Parietal Lobe (SPL). We interpreted these findings as reflecting complementary mechanisms by which people showed on one hand, in-synch brain activity in areas engaged in attunement and shared common-ground (i.e. TPJ, Cañigueral, Zhang, et al., 2021a; Jiang et al., 2015), and on the other hand de-coupled brain activity in areas recruited by internal cognitive processes (i.e. SPL, Golland et al., 2007), including body self-consciousness and body-ownership (Blanke, 2012). In addition, we found that after a conversation on topics unrelated to the cartoon episodes, real pairs showed increased synchrony over right TPJ during subsequent novel movie co-watching, to a significantly greater extent than what observed in pseudo pairs. The right TPJ has been found to be associated with shared-understanding of external reality (M. Nguyen et al., 2019; Salazar et al., 2021; Yeshurun et al., 2017). It may be that conversation with one another elicited more similar neural response between real partners by facilitating alignment of bodies and minds, whose effect extended to later co-watching (Garrod & Pickering, 2009; Richardson et al., 2007).

Study 4 asks whether teacher-student brain synchrony can be a marker of learning and, if so, how it is modulated by social signals such as eye-gaze behaviour. We asked participants to learn facts about a selection of obscure items from our previous studies (study 1 and 2) on their own, before coming to the lab and teach those facts to their partner. As they engaged with each other, we measured their brain activity via fNIRS hyperscanning and video recorded their interaction. Findings reveal that learning was best predicted by a complex dynamic between neural responses and behavioural metrics, namely mutual gaze and joint attention. In particular, during instances of joint attention, brain synchrony over left dlPFC and left TPJ increases and this resulted in better learning, possibly reflecting participants attending to the task (Turnbull et al., 2019) and to the partner (Doricchi et al., 2022; Stanley & Adolphs, 2013), respectively. In contrast, during mutual gaze, we observed that learner-teacher were *less* synchronized over these brain regions, potentially marking moments of *un*shared consolidation of information (Wohltjen & Wheatley, 2021). This may have been important to ensure a successful integration between internal 'solo' processes and external interactive processes, and in turn to achieve good learning. In addition, specifically when partners could not see each other, learning was higher when their brain synchrony was lower in left vPMC, an area involved in speech production, comprehension and also recently showed to process audio-feedback from one's own speech (Ozker et al., 2022). It may be that specifically for verbal information, de-coupling was linked to better learning as – by definition – learning material is unknown (and therefore unpredictable). While the difficulty in prediction may have been somehow attenuated by other multi-modal cues available during face-to-face view (e.g. gestures, mouth movement etc., Y. Zhang et al., 2021), this may have been particularly hard during the obstructed view condition when verbal information was the only modality available. This study showed how brain synchrony during learner-teacher interaction can be a good marker to predict learning outcomes, but can only be successfully interpreted with reference to behavioural metrics, in line with the mutual-prediction framework placing interactive brains within interactive bodies (Hamilton, 2020).

## 6.2 What can we conclude about the role of social interaction in learning? Strength and limitations of the present work

Taken together, the present experimental work led to findings which advance the field at the intersection between social neuroscience and educational neuroscience with three main discoveries. First, social interaction supports *adults* learning of *non-social* knowledge. Second, social contingency – including mechanisms engaged during contingent interaction (e.g. joint attention) – is the strongest factor predicting the social learning advantage, while social cues may *mediate* and *facilitate* social-learning advantage in interactive contexts, but are *not* essential. Third, while our results are in line with previous work showing that social interaction mediates brain-synchrony, we showed that behavioural measures (e.g. joint attention) are more informative predictors of the quality of the social interaction (e.g. learning success in a learner-teacher interaction) than brain measures (including inter-brains measures), especially when the brain is studied on its own without reference to behaviour. We discuss each of these points in turn.

Our first and strongest conclusion is that social interaction supports *adults* learning of *non-social* knowledge. We went beyond existing literature in showing that this is the case i) not only for learning of social knowledge, as it has been well demonstrated by previous work (e.g. see Verga & Kotz, 2013), and ii) not only in developmental sensitive periods such as childhood (e.g. Kostyrka-Allchorne et al., 2019; Roseberry et al., 2009), but most specifically it extends to learning of non-social factual knowledge in adults. We demonstrated that over 4 experiments and almost 200 participants (chapter 2 and chapter 3), where we directly compared interactive learning to non-interactive learning. Importantly, we showed that social interaction may support learning over mechanisms that are intact even when the networks supporting social cognition may not entirely be: by replicating the interactive-learning advantage in a large group of autistic adults (chapter 3) – a population typically associated with social cognition atypicalities (Happé & Frith, 2014) – we highlighted how learning *via* social interaction may engage mechanisms that are dissociable from mechanisms engaged during learning *about* the social world.

We cannot be sure that mechanisms responsible for the interactive-learning advantage in autistic adults are the exact same mechanisms responsible for the effect in neurotypicals. However, the conclusion that social interaction supports learning via mechanisms that are independent from processing of social signals, at least partially, is supported by two main findings: first, we observed a *within*-group advantage (and no between-group difference), and second, this was correlated with greater enjoyment and equivalent level of anxiety compared to less-interactive contexts in both groups. It should be noted that all our conditions where somehow social (all presented at the very least the recorded voice of the teacher, i.e. a social agent, through which the material was delivered). Future studies should directly compare learning via social contexts to learning via fully non-social medium (e.g. reading solo).

Nevertheless, the fact that all our conditions (across all our studies) were 'social' to a certain extent, allowed us to reach our second core conclusion: our effect can be more certainly attributed to the temporal *contingency* typical of real-world social interactions (Konvalinka et al., 2023; Redcay & Schilbach, 2019; Stanley & Adolphs, 2013), rather than the richness in social cues (e.g. voices, faces etc) that usually characterise social contexts. Our findings directly support this conclusion: in study 1 (chapter 2), social cues only interacted with social contingency, but did not show a main effect on learning; in study 2 (chapter 3), although learning from recorded videos was slightly better for videos with less social content (recorded teacher-alone condition > recorded teacher-student interaction), again this effect was marginal and only present for delay performance, and significantly less strong than the live-advantage effect; in study 4, we did not find any difference in learning performance when we compared full-face interaction to obstructed-view interaction.

Furthermore, some indirect evidence supporting the notion that social contingency characterises naturalistic social interaction (and its consequent benefit) above and beyond social cues per se, also comes from study 3: here we compared inter-personal neural dynamics in response to novel stimuli between real dyads and pseudo dyads. Note that all dyads (real and pseudo) were exposed to the same social environment, while only real dyads engaged in a real (contingent) interaction. We showed that inter-brain synchrony were specifically modulated by *real interaction with each other*, rather than by simple exposure to non-contingent social signals (pseudo dyads).

In line with our conclusion that social contingency plays a primary role in supporting mechanisms of naturalistic social interaction, we found that contingent behaviour can predict learning in interactive contexts better than neural measures, and may specifically *mediate* inter-subject neural dynamics (chapter 4). This leads to our third core conclusion: behavioural dynamics during naturalistic social interaction explain variability in learning above and beyond what neural measures on their own can do. This is in line with the recent paradigm shift towards studying *brains within bodies,* rather than brains isolated from the perceptual and sensory contexts of real-world experiences (De Jaegher et al., 2010; Hamilton, 2020). Our hyperscanning study on learning (chapter 4) provides strong evidence for this argument: we found that joint attention and mutual gaze were the strongest predictors of learning performance, and in turn modulated learner-teacher neural synchrony consistently over different brain regions. Specifically, we found that joint attention was associated with increased neural synchrony over areas typically part of the social brain network (TPJ and dlPFC, e.g. Kelsen et al., 2020) and facilitated learning, while mutual gaze was associated with less brain synchrony over the same regions, consistent with models of successful integration between internal and external interactive processes (Kajimura & Nomura, 2016; Wohltjen & Wheatley, 2021).

This work advances our understanding of naturalistic social learning. It also shows that – despite being challenging and resulting in complex multi-modal datasets – it is possible to study interactive brains within interactive bodies during real-world social dynamics, without compromising the rigour necessary in scientific experimental investigation. However, future studies are needed to resolve some of the limitations of the present work: for example, a more fine-grained study of eye-gaze dynamics (e.g. using eye-tracking which are more precise than post-hoc video coding) should confirm our conclusions with regards to the role of joint attention and mutual gaze in modulating brain synchrony and supporting learning. Also, due to the 2020 global pandemic, our face-to-face hyperscanning studies (chapter 4 and 5) included familiar pairs. Future work should directly compare how different inter-personal relational dynamics affect the social learning mechanisms. In addition, future work should include a greater range of behavioural metrics, in order to fully define the profile of a successful social interaction and find biomarkers for best learning. This would also call for more sophisticated analysis tools including cross brain general linear models, Principal component analysis and Granger causality.

Zooming out to incorporate the complexity while at the same time zooming in to increase the precision of each of the signals composing real-world social interaction is crucial to fully understand how people experience the world within themselves and with others. In the next section, we set out some outstanding research questions for future work in social neuroscience and educational neuroscience.

## 6.3 Future directions and outstanding questions

The aim of this thesis was to study adult learning as it most naturally occurs, that is socially, and to understand what aspects of naturalistic social interaction best support acquisition of new knowledge. As it contributes to advancing our understanding of real-world social learning, this thesis also opens up future pressing questions in the field of both social neuroscience and educational neuroscience.

The study of interactive learning cannot answer questions on how individual cognitive mechanisms work per se, unless research considers the individual agents alone and as part of an integrated social dynamic where they learn from (and/or with) one another. Taking a second-person neuroscience approach (Hamilton, 2020; Redcay & Schilbach, 2019) across all stages of the experimental work is particularly important as we are moving away from studying learning in isolation to study learning in social contexts: we must study interactive minds as they are found in the real world in the context of rich interactions, to fully understand interactive learning dynamics as they unfold (Shamay-Tsoory, 2021). We believe that this approach can give us a comprehensive understanding of what factors influence learning and its underlying cognitive mechanisms and neural markers.

It is hard to separate individual components because live interaction cannot be easily deconstructed. Future work should aim to develop paradigms that are able to disentangle different contributions while also refrain from 'de-constructing' the social interaction phenomenon. In other words, the challenge would be that of ensuring that ecological validity co-exists with experimental rigour (Vigliocco et al., pre-print). Investigations that would disentangle the different potential mechanisms engaged in contingent social interaction (e.g. arousal, attention, reciprocity, active attunement etc) may be able to do so by using virtual reality, which allows for experimental manipulation of different aspects of interaction and therefore to test which of these are most important to learning (Pan & Hamilton, 2018). In addition, using robots may be a valuable tool to isolate social mechanisms specifically responsible for efficient information transfer in interactive contexts (Jung et al., 2013; Tanaka et al., 2007; Westlund et al., 2015), and may allow us to answer questions about how human-human interaction differs from human-robot interaction.

In addition, it is not clear how social interaction benefits learning longitudinally throughout our lives. In particular, the majority of work has been conducted in children. This thesis contributed to fill the gap by looking at adulthood. However, adolescence remains a largely unexplored age-group, about which little is known with regards to the intersection between social interaction and learning (Blakemore, 2010). This would be important as adolescence is a sensitive period for the development of the social brain (Becht et al., 2021; Fuhrmann et al., 2015), as well as an age where peer-relationship and social networks are notoriously important. Also, very little is known about how we learn in interaction in ways that do not involve verbal communication, for example during large in-person classes where, although there is co-presence, there is no or very little active exchange. Studies addressing these questions are needed in order to assess to what extent the benefit of social interaction in learning is global and important throughout our life, or is content-dependent and age-sensitive.

## 6.4 Final remarks

In conclusion, this thesis investigated the role of social interaction in knowledge-based learning. It prioritised ecological validity and multi-modal paradigms to study learner-teacher interaction as it most naturally occurs. It included investigation of social learning over both online and face-to-face contexts, and looked at both behavioural measures and inter-personal neural dynamics. Findings revealed the strong beneficial role of social interaction in human acquisition of new knowledge, which extends to long-term learning and to neuro-diverse populations such as people with autistic spectrum condition (ASC). In addition, it provides novel insights into the complex dynamic between behaviour and neural signals in predicting learning, advocating for a study of interactive brains within interactive bodies, and pointing at joint attention as one of the core mechanisms through which we share information effectively to consolidate knowledge. This may be fundamental for humans as they do not exist in a vacuum, rather in a reality that it is shared, and therefore should also be interpreted, understood and co-constructed with others.

# References

Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, *15*(1), 1–16. https://doi.org/10.1016/S0885-2014(00)00014-9

Adams, D., Simpson, K., Davies, L., Campbell, C., & Macdonald, L. (2019). Online learning for university students on the autism spectrum: A systematic review and questionnaire study. *Australasian Journal of Educational Technology*, *35*(6), 111–131. https://doi.org/10.14742/ajet.5483

Al-Rukban, M. O. (2006). Guidelines for the construction of multiple choice questions tests. *Journal of Family & Community Medicine*, *13*(3), 125–133. http://www.ncbi.nlm.nih.gov/pubmed/23012132%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3410060

Alksne, L. (2016). How To Produce Video Lectures To Engage Students and Deliver the Maximum Amount of Information. *SOCIETY. INTEGRATION. EDUCATION. Proceedings of the International Scientific Conference*, *2*(May 2016), 503. https://doi.org/10.17770/sie2016vol2.1424

Azhari, A., Bizzego, A., Esposito, G., & Bizzego, A. (2021). Father-child dyads exhibit unique inter-subject synchronization during co-viewing of animation video stimuli. *Social Neuroscience*, *16*(5), 522–533. https://doi.org/10.1080/17470919.2021.1970016

Azhari, A., Gabrieli, G., Bizzego, A., & Bornstein, M. H. (2023). Probing the association between maternal anxious attachment style and mother-child brain-to- brain coupling during passive co-viewing of visual stimuli. *Attachment & Human Development*, *25*(1), 19–34. https://doi.org/10.1080/14616734.2020.1840790

Azhari, A., Leck, W. Q., Gabrieli, G., Bizzego, A., Rigo, P., Setoh, P., & Bornstein, M. H.

(2019). Parenting Stress Undermines Synchrony: A Hyperscanning Study. *Scientific Reports*, *July*, 1–9. https://doi.org/10.1038/s41598-019-47810-4

Azhari, A., Lim, M., Bizzego, A., Gabrieli, G., Bornstein, M. H., & Esposito, G. (2020). Physical presence of spouse enhances brain-to-brain synchrony in co-parenting couples. *Scientific Reports*, 1–11. https://doi.org/10.1038/s41598-020-63596-2

Babiloni, F., & Astolfi, L. (2014a). Social neuroscience and hyperscanning techniques: Past, present and future. *Neuroscience and Biobehavioral Reviews*, *44*, 76–93. https://doi.org/10.1016/j.neubiorev.2012.07.006

Babiloni, F., & Astolfi, L. (2014b). Social neuroscience and hyperscanning techniques: Past, present and future. *Neuroscience and Biobehavioral Reviews*, *44*, 76–93. https://doi.org/10.1016/j.neubiorev.2012.07.006

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot-the-Word test: A robust estimate of verbal intelligence based on lexical decision. *British Journal of Clinical Psychology*, *32*(1), 55–65. https://doi.org/10.1111/j.2044-8260.1993.tb01027.x

Bandura, A. (2019). Social learning: Theory. In *Encyclopedia of Animal Behavior* (pp. 380–386). https://doi.org/10.1016/B978-0-12-813251-7.00057-2

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Becht, A. I., Wierenga, L. M., Mills, K. L., Meuwese, R., van Duijvenvoorde, A., Blakemore, S. J., Güroğlu, B., & Crone, E. A. (2021). Beyond the average brain: individual differences in social brain development are associated with friendship quality. *Social Cognitive and Affective Neuroscience*, *16*(3), 292–301. https://doi.org/10.1093/scan/nsaa166

Benedek, M., Jauk, E., Beaty, R. E., Fink, A., Koschutnig, K., & Neubauer, A. C. (2016).

Brain mechanisms associated with internally directed attention and self-generated thought. *Nature Publishing Group*, *October 2015*, 1–8. https://doi.org/10.1038/srep22959

Bernier, R., Dawson, G., Webb, S., & Murias, M. (2007). EEG mu rhythm and imitation impairments in individuals with autism spectrum disorder. *Brain and Cognition*, *64*(3), 228–237. https://doi.org/10.1016/j.bandc.2007.03.004

Bevilacqua, D., Davidesco, I., Wan, L., Chaloner, K., Rowland, J., Ding, M., Poeppel, D., & Dikker, S. (2019). Brain-to-Brain Synchrony and Learning Outcomes Vary by Student – Teacher Dynamics : Evidence from a Real-world Classroom Electroencephalography Study. *Journal of Cognitive Neuroscience*, 401–411. https://doi.org/10.1162/jocn

Blakemore, S. J. (2010). The Developing Social Brain: Implications for Education. *Neuron*, *65*(6), 744–747. https://doi.org/10.1016/j.neuron.2010.03.004

Blanke, O. (2012). *Multisensory brain mechanisms of bodily self-consciousness*. *13*(August), 556–571. https://doi.org/10.1038/nrn3292

Blind. (2020). *"How engaged are you in most work meetings?"* https://www.teamblind.com/blog/index.php/2020/07/30/blind-polls-zoom-fatigue/

Bohn, M., Tessler, M. H., & Frank, M. (2019). *Integrating Common Ground and Informativeness in Pragmatic Word Learning*. https://doi.org/10.31234/osf.io/cbx46

Burgess, P. W., Crum, J., Pinti, P., Aichelburg, C., Oliver, D., Lind, F., Power, S., Swingler, E., Hakim, U., Merla, A., Gilbert, S., Tachtsidis, I., & Hamilton, A. (2022). Prefrontal cortical activation associated with prospective memory while walking around a real-world street environment. *NeuroImage*, *258*(March), 119392.

https://doi.org/10.1016/j.neuroimage.2022.119392

Calvert, S. L., Strong, B. L., Jacobs, E. L., & Conger, E. E. (2007). Interaction and participation for Young Hispanic and Caucasian Girls' and Boys' learning of media content. *Media Psychology*, *9*(2), 431–445. https://doi.org/10.1080/15213260701291379

Cañigueral, R., Ward, J. A., & Hamilton, A. F. d. C. (2021). Effects of being watched on eye gaze and facial displays of typical and autistic individuals during conversation. *Autism*, *25*(1), 210–226. https://doi.org/10.1177/1362361320951691

Cañigueral, R., Zhang, X., Noah, J. A., Tachtsidis, I., Hamilton, A. F. d. C., & Hirsch, J. (2021a). Facial and neural mechanisms during interactive disclosure of biographical information. *NeuroImage*, *226*(November 2020). https://doi.org/10.1016/j.neuroimage.2020.117572

Cañigueral, R., Zhang, X., Noah, J. A., Tachtsidis, I., Hamilton, A. F. D. C., & Hirsch, J. (2021b). *Facial and neural mechanisms during interactive disclosure of biographical information*. *226*(November 2020). https://doi.org/10.1016/j.neuroimage.2020.117572

Cannon, J., Mccarthy, M. M., Lee, S., Lee, J., Börgers, C., Whittington, M. A., & Kopell, N. (2014). Neurosystems: Brain rhythms and cognitive processing. *European Journal of Neuroscience*, *39*(5), 705–719. https://doi.org/10.1111/ejn.12453

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(28), 11278–11283. https://doi.org/10.1073/pnas.1309518110

Čeponiene, R., Lepistö, T., Shestakova, A., Vanhala, R., Alku, P., Näätänen, R., & Yaguchi, K. (2003). Speech-sound-selective auditory impairment in children with autism:

They can perceive but do not attend. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(9), 5567–5572. https://doi.org/10.1073/pnas.0835631100

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*(1), 115–125. https://doi.org/10.1038/nn.4450

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The Social Motivation Theory. *Trends in Cognitive Sciences*, *16*(4), 231–239. https://doi.org/10.1016/j.tics.2012.02.007.The

Chierchia, G., Fuhrmann, D., Knoll, L. J., Pi-Sunyer, B. P., Sakhardande, A. L., & Blakemore, S. J. (2019). The matrix reasoning item bank (MaRs-IB): Novel, open-access abstract reasoning items for adolescents and adults. *Royal Society Open Science*, *6*(10). https://doi.org/10.1098/rsos.190232

Clements, C. C., Zoltowski, A. R., Yankowitz, L. D., Yerys, B. E., Schultz, R. T., & Herrington, J. D. (2018). Evaluation of the social motivation hypothesis of autism a systematic review and meta-analysis. *JAMA Psychiatry*, *75*(8), 797–808. https://doi.org/10.1001/jamapsychiatry.2018.1100

Corbett, B. A., Swain, D. M., Newsom, C., Wang, L., Song, Y., & Edgerton, D. (2014). Biobehavioral profiles of arousal and social motivation in autism spectrum disorders. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *55*(8), 924–934. https://doi.org/10.1111/jcpp.12184

Cui, X., Bryant, D. M., & Reiss, A. L. (2011). NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. *NeuroImage*, *59*(3), 2430–2437. https://doi.org/10.1016/j.neuroimage.2011.09.003

Czeszumski, A., Eustergerling, S., Lang, A., Menrath, D., Gerstenberger, M., Schuberth, S., Schreiber, F., Rendon, Z. Z., & König, P. (2020). Hyperscanning: A Valid Method to Study Neural Inter-brain Underpinnings of Social Interaction. *Frontiers in Human Neuroscience*, *14*(February), 1–17. https://doi.org/10.3389/fnhum.2020.00039

Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2013). The Self-Organization of human interaction. In *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 59). Elsevier. https://doi.org/10.1016/B978-0-12-407187-2.00002-2

Davidesco, I., Laurent, E., Valk, H., West, T., Milne, C., Poeppel, D., & Dikker, S. (2023). *The Temporal Dynamics of Brain-to-Brain Synchrony Between Students and Teachers Predict Learning Outcomes*. https://doi.org/10.1177/09567976231163872

Davis, J., Crabb, S., Rogers, E., Zamora, J., & Khan, K. (2008). Computer-based teaching is as good as face to face lecture-based teaching of evidence based medicine: A randomized controlled trial. *Medical Teacher*, *30*(3), 302–307. https://doi.org/10.1080/01421590701784349

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *Journal of Autism and Developmental Disorders*, *28*(6), 479–485. https://doi.org/10.1023/A:1026043926488

Dawson, M., Mottron, L., & Gernsbacher, M. A. (2007). Learning in autism. *Learning and Memory: A Comprehensive Reference*, 759–772. https://doi.org/10.1016/B978-012370509-9.00152-2

De Felice, S., Vigliocco, G., & Hamilton, A. F. de C. (2021). Social interaction is a catalyst for adult human learning in online contexts. *Current Biology*, *31*(21), 4853-4859.e3. https://doi.org/10.1016/j.cub.2021.08.045

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14*(10), 441–447. https://doi.org/10.1016/j.tics.2010.06.009

Dideriksen, C., Christiansen, M. H., Kristian, T., Dingemanse, M., & Fusaroli, R. (2020). *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.

Dieffenbach, M. C., Gillespie, G. S. R., Burns, S. M., McCulloh, I. A., Ames, D. L., Dagher, M. M., Falk, E. B., & Lieberman, M. D. (2020). A Synchrony-Based Classification Approach for Predicting Attitudes Using fNIRS. *Social Cognitive and Affective Neuroscience*, *June 2020*, 117–128. https://doi.org/10.1093/scan/nsaa115

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., Rowland, J., Michalareas, G., Van Bavel, J. J., Ding, M., & Poeppel, D. (2017). Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Current Biology*, *27*(9), 1375–1380. https://doi.org/10.1016/j.cub.2017.04.002

Doricchi, F., Lasaponara, S., Pazzaglia, M., & Silvetti, M. (2022). Left and right temporal-parietal junctions ( TPJs ) as " match / mismatch " hedonic machines : A unifying account of TPJ function. *Physics of Life Reviews*, *42*, 56–92. https://doi.org/10.1016/j.plrev.2022.07.001

Dubey, I., Ropar, D., & De C Hamilton, A. F. (2015). Measuring the value of social engagement in adults with and without autism. *Molecular Autism*, *6*(1), 1–9. https://doi.org/10.1186/s13229-015-0031-2

Dumas, G., Lachat, F., Martinerie, J., Nadel, J., & George, N. (2011). From social behaviour to brain synchronization: Review and perspectives in hyperscanning. *Irbm*, *32*(1), 48–53. https://doi.org/10.1016/j.irbm.2011.01.002

Edwards, L. A. (2014). A meta-analysis of imitation abilities in individuals with autism spectrum disorders. *Autism Research*, *7*(3), 363–380. https://doi.org/10.1002/aur.1379

Elsabbagh, M., Mercure, E., Hudry, K., Chandler, S., Pasco, G., Charman, T., Pickles, A., Baron-Cohen, S., Bolton, P., & Johnson, M. H. (2012). Infant neural sensitivity to dynamic eye gaze is associated with later emerging autism. *Current Biology*, *22*(4), 338–342. https://doi.org/10.1016/j.cub.2011.12.056

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. Neuroscience and Biobehavioral Reviews. *Neuroscience and Biobehavioral Reviews*, *24*, 581–604. www.elsevier.com/locate/neubiorev%0Ahttps://www.sciencedirect.com/science/article/pii/S0149763400000257

Fernandez-Prieto, M., Moreira, C., Cruz, S., Campos, V., Martínez-Regueiro, R., Taboada, M., Carracedo, A., & Sampaio, A. (2021). Executive Functioning: A Mediator Between Sensory Processing and Behaviour in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *51*(6), 2091–2103. https://doi.org/10.1007/s10803-020-04648-4

Fitch, W. T., Huber, L., & Bugnyar, T. (2010). Social Cognition and the Evolution of Language: Constructing Cognitive Phylogenies. *Neuron*, *65*(6), 795–814. https://doi.org/10.1016/j.neuron.2010.03.011

Friedman, R. S., & Förster, J. (2010). *Implicit Affective Cues and Attentional Tuning : An Integrative Review*. *136*(5), 875–893. https://doi.org/10.1037/a0020495

Fronda, G., & Balconi, M. (2020). The effect of interbrain synchronization in gesture observation: A fNIRS study. *Brain and Behavior*, *10*(7), 1–13. https://doi.org/10.1002/brb3.1663

Fuhrmann, D., Knoll, L. J., & Blakemore, S. J. (2015). Adolescence as a Sensitive Period of Brain Development. *Trends in Cognitive Sciences*, *19*(10), 558–566. https://doi.org/10.1016/j.tics.2015.07.008

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8–11. https://doi.org/10.1016/j.tics.2003.10.016

Garrod, S., & Pickering, M. J. (2009). Joint Action, Interactive Alignment, and Dialog. *Topics in Cognitive Science*, *1*(2), 292–304. https://doi.org/10.1111/j.1756-8765.2009.01020.x

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523. https://doi.org/10.1111/j.1467-9280.2008.02117.x

Golland, Y., Bentin, S., Gelbard, H., Benjamini, Y., Heller, R., Nir, Y., Hasson, U., & Malach, R. (2007). Extrinsic and intrinsic systems in the posterior cortex of the human brain revealed during natural sensory stimulation. *Cerebral Cortex*, *17*(4), 766–777. https://doi.org/10.1093/cercor/bhk030

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models Published by : The Econometric Society Stable URL : https://www.jstor.org/stable/1912791 to Econometrica. *Econometrica*, *37*(3), 424–438.

Grinsted, A., Moore, J. C., & Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, *11*(5/6), 561–566. https://doi.org/10.5194/npg-11-561-2004

Guerin, B. (1986). Mere presence effects in humans: A review. *Journal of Experimental Social Psychology*, *22*(1), 38–77. https://doi.org/10.1016/0022-1031(86)90040-5

Hamilton, A. F. de C. (2020). Hyperscanning: Beyond the Hype. *Neuron*, *2020*, 8–11.

https://doi.org/10.1016/j.neuron.2020.11.008

Hamilton, A. F. de C., & Lind, F. (2016). Audience effects: what can they tell us about social neuroscience, theory of mind and autism? *Culture and Brain*, *4*(2), 159–177. https://doi.org/10.1007/s40167-016-0044-5

Happé, F., & Frith, U. (2014). Annual research review: Towards a developmental neuroscience of atypical social cognition. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *55*(6), 553–577. https://doi.org/10.1111/jcpp.12162

Hasson, U., & Frith, C. D. (2016). Mirroring and beyond: Coupled dynamics as a generalized framework for modelling social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693). https://doi.org/10.1098/rstb.2015.0366

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject Synchroinization of Cortical Activity During Natural Vision. *Science*, *303*(MARCH), 1634–1641.

Hertel, P. T., Brozovich, F., Joormann, J., & Gotlib, I. H. (2008). Biases in Interpretation and Memory in Generalized Social Phobia. *Journal of Abnormal Psychology*, *117*(2), 278–288. https://doi.org/10.1037/0021-843X.117.2.278

Hill, E. L. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences*, *8*(1), 26–32. https://doi.org/10.1016/j.tics.2003.11.003

Hirsch, J., Noah, J. A., Zhang, X., Dravida, S., & Ono, Y. (2018). A cross-brain neural mechanism for human-to-human verbal communication. *Social Cognitive and Affective Neuroscience*, *13*(9), 907–920. https://doi.org/10.1093/scan/nsy070

Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and listening with the eyes: Gaze signaling during dyadic interactions. *PLoS ONE*, *10*(8), 1–18.

https://doi.org/10.1371/journal.pone.0136905

Hoehl, S., Fairhurst, M., & Schirmer, A. (2020a). Interactional synchrony: signals, mechanisms and benefits. *Social Cognitive and Affective Neuroscience*, *September 2019*, 5–18. https://doi.org/10.1093/scan/nsaa024

Hoehl, S., Fairhurst, M., & Schirmer, A. (2020b). Interactional synchrony: signals, mechanisms and benefits. *Social Cognitive and Affective Neuroscience*, *September 2019*, 1–14. https://doi.org/10.1093/scan/nsaa024

Holtgraves, T. M., & Kashima, Y. (2008). Language, meaning, and social cognition. *Personality and Social Psychology Review*, *12*(1), 73–94. https://doi.org/10.1177/1088868307309605

Hu, J., Qi, S., Becker, B., Luo, L., Gao, S., Gong, Q., Hurlemann, R., & Kendrick, K. M. (2015). Oxytocin selectively facilitates learning with social feedback and increases activity and functional connectivity in emotional memory and reward processing regions. *Human Brain Mapping*, *36*(6), 2132–2146. https://doi.org/10.1002/hbm.22760

Ionta, S., Heydrich, L., Lenggenhager, B., Mouthon, M., Fornari, E., & Chapuis, D. (2011). Multisensory Mechanisms in Temporo-Parietal Cortex Support Self-Location and First-Person Perspective. *Neuron*, *70*(2), 363–374. https://doi.org/10.1016/j.neuron.2011.03.009

Jääskeläinen, I. P., Koskentalo, K., Balk, M. H., Autti, T., Kauramäki, J., Pomren, C., & Sams, M. (2008). Inter-Subject Synchronization of Prefrontal Cortex Hemodynamic Activity During Natural Viewing. *The Open Neuroimaging Journal*, *2*(1), 14–19. https://doi.org/10.2174/1874440000802010014

Jiang, J., Borowiak, K., Tudge, L., Otto, C., & von Kriegstein, K. (2017). Neural mechanisms of eye contact when listening to another person talking. *Social*

*Cognitive and Affective Neuroscience*, *12*(2), 319–328. https://doi.org/10.1093/scan/nsw127

Jiang, J., Chen, C., Dai, B., Shi, G., Ding, G., Liu, L., Lu, C., & Fiske, S. T. (2015). Leader emergence through interpersonal neural synchronization. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(14), 4274–4279. https://doi.org/10.1073/pnas.1422930112

John, J. R., Priya, H., Hemasundar, A., & Rajendran, G. (2016). *A Collation of Two Forms of Demonstration : In-person Versus Video Demon- stration which Gains the Upper Hand ? 4*(4), 23–29.

Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the Basic Empathy Scale. *Journal of Adolescence*, *29*(4), 589–611. https://doi.org/10.1016/j.adolescence.2005.08.010

Jung, M. F., Lee, J. J., Depalma, N., Adalgeirsson, S. O., Hinds, P. J., & Breazeal, C. (2013). Engaging robots: Easing complex human-robot teamwork using backchanneling. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, *April 2015*, 1555–1566. https://doi.org/10.1145/2441776.2441954

Kajimura, S., & Nomura, M. (2016). When we cannot speak: Eye contact disrupts resources available to cognitive control processes during verb generation. *Cognition*, *157*, 352–357. https://doi.org/10.1016/j.cognition.2016.10.002

Kajopoulos, J., Cheng, G., Kise, K., Müller, H. J., & Wykowska, A. (2020). Focusing on the face or getting distracted by social signals? The effect of distracting gestures on attentional focus in natural interaction. *Psychological Research*, *85*(2), 491–502. https://doi.org/10.1007/s00426-020-01383-4

Kasari, C., Paparella, T., Freeman, S., & Jahromi, L. B. (2008). Language Outcome in Autism: Randomized Comparison of Joint Attention and Play Interventions.

*Journal of Consulting and Clinical Psychology*, *76*(1), 125–137. https://doi.org/10.1037/0022-006X.76.1.125

Kawai, N. (2011). Attentional shift by eye gaze requires joint attention: Eye gaze cues are unique to shift attention. *Japanese Psychological Research*, *53*(3), 292–301. https://doi.org/10.1111/j.1468-5884.2011.00470.x

Kayhan, E., Nguyen, T., Matthes, D., Langeloh, M., Michel, C., Jiang, J., & Hoehl, S. (2022). Interpersonal neural synchrony when predicting others' actions during a game of rock-paper-scissors. *Scientific Reports*, *12*(1), 1–11. https://doi.org/10.1038/s41598-022-16956-z

Kelsen, B. A., Sumich, A., Kasabov, N., Liang, S. H. Y., & Wang, G. Y. (2020). What has social neuroscience learned from hyperscanning studies of spoken communication? A systematic review. *Neuroscience and Biobehavioral Reviews*, *October*. https://doi.org/10.1016/j.neubiorev.2020.09.008

Kilroy, E., Aziz-Zadeh, L., & Cermak, S. (2019). Ayres theories of autism and sensory integration revisited: What contemporary neuroscience has to say. *Brain Sciences*, *9*(3). https://doi.org/10.3390/brainsci9030068

Kingsbury, L., Huang, S., Wang, J., Gu, K., Golshani, P., Wu, Y. E., & Hong, W. (2019). Correlated Neural Activity and Encoding of Behavior across Brains of Socially Interacting Animals. *Cell*, *178*(2), 429-446.e16. https://doi.org/10.1016/j.cell.2019.05.022

Kirkorian, H. L. (2016). *Hollins Digital Commons Toddlers ' Word Learning from Contingent and Non-Contingent Video on Touchscreens*. 405–413.

Klin, A. (1991). Young autistic children's listening preferences in regard to speech: A possible characterization of the symptom of social withdrawal. *Journal of Autism and Developmental Disorders*, *21*(1), 29–42.

http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L21129343

Konishi, M., Mclaren, D. G., Engen, H., & Smallwood, J. (2015). *Shaped by the Past : The Default Mode Network Supports Cognition that Is Independent of Immediate Perceptual Input*. 1–18. https://doi.org/10.1371/journal.pone.0132209

Konvalinka, I., Sebanz, N., & Knoblich, G. (2023). The role of reciprocity in dynamic interpersonal coordination of physiological rhythms. *Cognition*, *230*(August 2022), 105307. https://doi.org/10.1016/j.cognition.2022.105307

Kostyrka-Allchorne, K., Holland, A., Cooper, N. R., Ahamed, W., Marrow, R. K., & Simpson, A. (2019). What helps children learn difficult tasks: A teacher's presence may be worth more than a screen. *Trends in Neuroscience and Education*, *17*(May), 100114. https://doi.org/10.1016/j.tine.2019.100114

Kourtis, D., Jacob, P., Sebanz, N., Sperber, D., & Knoblich, G. (2020). Making sense of human interaction benefits from communicative cues. *Scientific Reports*, *10*(1), 1–8. https://doi.org/10.1038/s41598-020-75283-3

Kuhl, P. K. (2007). Is speech learning "gated" by the social brain? *Developmental Science*, *10*(1), 110–120. https://doi.org/10.1111/j.1467-7687.2007.00572.x

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(15), 9096–9101. https://doi.org/10.1073/pnas.1532872100

Kwok, K., Ghrear, S., Li, V., Haddock, T., Coleman, P., & Birch, S. A. J. (2016). Children can learn new facts equally well from interactive media versus face to face instruction. *Frontiers in Psychology*, *7*(OCT), 1–10. https://doi.org/10.3389/fpsyg.2016.01603

Lage-castellanos, A., Martínez-montes, E., Hernández-cabrera, J. A., & Galán, L. (2010). *False discovery rate and permutation test : An evaluation in ERP data analysis*. *August 2008*. https://doi.org/10.1002/sim.3784

Lahnakoski, J. M., Glerean, E., Jääskeläinen, I. P., Hyönä, J., Hari, R., Sams, M., & Nummenmaa, L. (2014). Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage*, *100*, 316–324. https://doi.org/10.1016/j.neuroimage.2014.06.022

Lanthier, S. N., Zhu, M. J. H., Byun, C. S. J., Jarick, M., & Kingstone, A. (2021). The costs and benefits to memory when observing and experiencing live eye contact. *Visual Cognition*, *0*(0), 1–15. https://doi.org/10.1080/13506285.2021.1926381

Lauricella, A. R., Gola, A. A. H., & Calvert, S. L. (2011). Toddlers' learning from socially meaningful video characters. *Media Psychology*, *14*(2), 216–232. https://doi.org/10.1080/15213269.2011.573465

Liebowitz, M. R. (1987). *Liebowitz-social-anxiety-scale - therapist administered*.

Liu, W., Branigan, H. P., Zheng, L., Long, Y., Bai, X., Li, K., Zhao, H., Zhou, S., Pickering, M. J., & Lu, C. (2019). Shared neural representations of syntax during online dyadic communication. *NeuroImage*, *198*(January), 63–72. https://doi.org/10.1016/j.neuroimage.2019.05.035

Livingston, L. A., Shah, P., White, S. J., & Happé, F. (2021). Further developing the Frith–Happé animations: A quicker, more objective, and web-based test of theory of mind for autistic and neurotypical adults. *Autism Research*, *14*(9), 1905–1912. https://doi.org/10.1002/aur.2575

Lomoriello, A. S., Sessa, P., Doro, M., & Konvalinka, I. (2022). Shared Attention Amplifies the Neural Processing of Emotional Faces. *Journal of Cognitive Neuroscience*, *34*(6), 917–932. https://doi.org/10.1162/jocn_a_01841

Lu, K., & Hao, N. (2019). When do we fall in neural synchrony with others? *Social Cognitive and Affective Neuroscience*, *14*(3), 253–261. https://doi.org/10.1093/scan/nsz012

Lytle, S. R., Garcia-Sierra, A., & Kuhl, P. K. (2018). Two are better than one: Infant language learning from video improves in the presence of peers. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(40), 9859–9866. https://doi.org/10.1073/pnas.1611621115

Madsen, J., & Parra, L. C. (2022). Cognitive processing of a common stimulus synchronizes brains, hearts, and eyes. *PNAS Nexus*, *1*(1), 1–14. https://doi.org/10.1093/pnasnexus/pgac020

Marotta, A., Lupiáñez, J., Martella, D., & Casagrande, M. (2012). Eye gaze versus arrows as spatial cues: Two qualitatively different modes of attentional selection. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 326–335. https://doi.org/10.1037/a0023959

McFarland, D. (2001). Respiratory Markers of Conversational Interaction. *Journal of Speech, Language, and Hearing Research*, *February*, 141–156. https://doi.org/10.4324/9780203891346-20

Miura, N., & Noguchi, S. (2022). The presence of adjacent others facilitates interpersonal neural synchronization in the left prefrontal cortex during a simple addition task. *Scientific Reports*, 1–9. https://doi.org/10.1038/s41598-022-16936-3

Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience and Biobehavioral Reviews*, *65*, 276–291. https://doi.org/10.1016/j.neubiorev.2016.03.020

Moore, C., & Barresi, J. (2017). The role of second-person information in the development of social understanding. *Frontiers in Psychology*, *8*(SEP), 1–15. https://doi.org/10.3389/fpsyg.2017.01667

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., Whiten, A., & Laland, K. N. (2015). Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nature Communications*, *6*, 4–11. https://doi.org/10.1038/ncomms7029

Müller, K., Lohmann, G., Neumann, J., Grigutsch, M., Mildner, T., & Von Cramon, D. Y. (2004). Investigating the wavelet coherence phase of the BOLD signal. *Journal of Magnetic Resonance Imaging*, *20*(1), 145–152. https://doi.org/10.1002/jmri.20064

Mundy, P. (1995). Joint attention and social-emotional approach behavior in children with autism. *Development and Psychopathology*, *7*(1), 63–82. https://doi.org/10.1017/S0954579400006349

Mundy, P., & Newell, L. (2007). Attention, joint attention & social cognition. *Social Cognition*, *16*(January 2008), 269–274. https://doi.org/10.1111/j.1467-8721.2007.00518.x.Attention

Myers, L. J., LeWitt, R. B., Gallo, R. E., & Maselli, N. M. (2017). Baby FaceTime: can toddlers learn from online video chat? *Developmental Science*, *20*(4). https://doi.org/10.1111/desc.12430

Nguyen, M., Vanderwal, T., & Hasson, U. (2019). Shared understanding of narratives is correlated with shared neural responses. *NeuroImage*, *184*(September 2018), 161–170. https://doi.org/10.1016/j.neuroimage.2018.09.010

Nguyen, T., Schleihauf, H., Kayhan, E., Matthes, D., Vrtička, P., & Hoehl, S. (2020). Neural synchrony in mother–child conversation: Exploring the role of conversation patterns. *Social Cognitive and Affective Neuroscience*, *June 2020*, 93–102. https://doi.org/10.1093/scan/nsaa079

Noah, J. A., Zhang, X., Dravida, S., Ono, Y., Naples, A., McPartland, J. C., & Hirsch, J. (2020). Real-Time Eye-to-Eye Contact Is Associated With Cross-Brain Neural Coupling in Angular Gyrus. *Frontiers in Human Neuroscience*, *14*(February), 1–10. https://doi.org/10.3389/fnhum.2020.00019

Norbury, C. F., Griffiths, H., & Nation, K. (2010). Sound before meaning: Word learning in autistic disorders. *Neuropsychologia*, *48*(14), 4012–4019. https://doi.org/10.1016/j.neuropsychologia.2010.10.015

Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I. P., Hari, R., & Sams, M. (2012). *Emotions promote social interaction by synchronizing brain activity across individuals*. 1–6. https://doi.org/10.1073/pnas.1206095109

O'Doherty, K., Troseth, G. L., Shimpi, P. M., Goldenberg, E., Akhtar, N., & Saylor, M. M. (2011). Third-Party Social Interaction and Word Learning From Video. *Child Development*, *82*(3), 902–915. https://doi.org/10.1111/j.1467-8624.2011.01579.x

Odom, S. L., Thompson, J. L., Hedges, S., Boyd, B. A., Dykstra, J. R., Duda, M. A., Szidon, K. L., Smith, L. E., & Bord, A. (2015). Technology-Aided Interventions and Instruction for Adolescents with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *45*(12), 3805–3819. https://doi.org/10.1007/s10803-014-2320-6

Ozker, M., Doyle, W., Devinsky, O., & Flinker, A. (2022). A cortical network processes auditory error signals during human speech production to maintain fluency. *PLoS Biology*, *20*(2), 1–21. https://doi.org/10.1371/journal.pbio.3001493

Pan, X., & Hamilton, A. F. d. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape.

*British Journal of Psychology*, *109*(3), 395–417. https://doi.org/10.1111/bjop.12290

Pan, Y., Cheng, X., & Hu, Y. (2023). Three heads are better than one: cooperative learning brains wire together when a consensus is reached. *Cerebral Cortex (New York, N.Y.: 1991)*, *33*(4), 1155–1169. https://doi.org/10.1093/cercor/bhac127

Pan, Y., Dikker, S., Goldstein, P., Zhu, Y., Yang, C., & Hu, Y. (2020). Instructor-learner brain coupling discriminates between instructional approaches and predicts learning. *NeuroImage*, *211*(February). https://doi.org/10.1016/j.neuroimage.2020.116657

Pan, Y., Guyon, C., Borragán, G., Hu, Y., & Peigneux, P. (2021). Interpersonal brain synchronization with instructor compensates for learner's sleep deprivation in interactive learning. *Biochemical Pharmacology*, *191*(May 2020). https://doi.org/10.1016/j.bcp.2020.114111

Pan, Y., Novembre, G., Song, B., Li, X., & Hu, Y. (2018). Interpersonal synchronization of inferior frontal cortices tracks social interactive learning of a song. *NeuroImage*, *183*(July), 280–290. https://doi.org/10.1016/j.neuroimage.2018.08.005

Pan, Y., Novembre, G., Song, B., Zhu, Y., & Hu, Y. (2020). Dual brain stimulation enhances interpersonal learning through spontaneous movement synchrony. *Social Cognitive and Affective Neuroscience*, *16*(1–2), 210–221. https://doi.org/10.1093/scan/nsaa080

Parkinson, C., Kleinbaum, A. M., & Wheatley, T. (2018). Similar neural responses predict friendship. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-017-02722-7

Pearson, J. (2004). *Autism and learning disability*. 50–60. https://doi.org/10.1177/1362361304042718

Pesarin, F. (2001). *Multivariate Permutation Tests: With Applications in Biostatistics*.

John Wiley & Sons, Ltd., New York.

Pfeiffer, U. J., Schilbach, L., Timmermans, B., Kuzmanovic, B., Georgescu, A. L., Bente, G., & Vogeley, K. (2014). NeuroImage Why we interact : On the functional role of the striatum in the subjective experience of social interaction. *NeuroImage*, *101*, 124–137. https://doi.org/10.1016/j.neuroimage.2014.06.061

Phillips, L. H., Tunstall, Æ. M., & Channon, Æ. S. (2007). *Exploring the Role of Working Memory in Dynamic Social Cue Decoding Using Dual Task Methodology*. 137–152. https://doi.org/10.1007/s10919-007-0026-6

Philofsky, A., Fidler, D. J., & Hepburn, S. (2007). Pragmatic language profiles of school-age children with autism spectrum disorders and Williams Syndrome. *American Journal of Speech-Language Pathology*, *16*(4), 368–380. https://doi.org/10.1044/1058-0360(2007/040)

Ponari, M., Norbury, C. F., Rotaru, A., Lenci, A., & Vigliocco, G. (2018). Learning abstract words and concepts: Insights from developmental language disorder. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1752). https://doi.org/10.1098/rstb.2017.0140

Poulsen, A. T., Kamronn, S., Dmochowski, J., Parra, L. C., & Hansen, L. K. (2017). *EEG in the classroom : Synchronised neural recordings during video presentation*. 1–9. https://doi.org/10.1038/srep43916

Primi, R., Ferrão, M. E., & Almeida, L. S. (2010). Fluid intelligence as a predictor of learning: A longitudinal multilevel approach applied to math. *Learning and Individual Differences*, *20*(5), 446–451. https://doi.org/10.1016/j.lindif.2010.05.001

Ramlogan, S., Raman, V., & Sweet, J. (2013). *A comparison of two forms of teaching instruction: video vs. live lecture for education in clinical periodontology* (pp. 31–38).

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, *20*(8), 495–505. https://doi.org/10.1038/s41583-019-0179-4

Riby, D. M., & Hancock, P. J. B. (2008). Viewing it differently: Social scene perception in Williams syndrome and Autism. *Neuropsychologia*, *46*(11), 2855–2860. https://doi.org/10.1016/j.neuropsychologia.2008.05.003

Rice, K., & Redcay, E. (2016). Interaction matters: A perceived social partner alters the neural processing of human speech. *NeuroImage*, *129*, 480–488. https://doi.org/10.1016/j.neuroimage.2015.11.041

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, *29*(6), 1045–1060. https://doi.org/10.1207/s15516709cog0000_29

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The Art of Conversation Is Coordination. *Psychological Science*, *18*(5), 407–413. https://doi.org/10.1111/j.1467-9280.2007.01914.x

Roos, E. M., McDuffie, A. S., Weismer, S. E., & Gernsbacher, M. A. (2008). A comparison of contexts for assessing joint attention in toddlers on the autism spectrum. *Autism*, *12*(3), 275–291. https://doi.org/10.1177/1362361307089521

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype Me! Socially Contingent Interactions Help Toddlers Learn Language. *Child Development*, *85*(3), 956–970. https://doi.org/10.1111/cdev.12166

Roseberry, S., Hirsh-Pasek, K., Parish-Morris, J., & Golinkoff, R. M. (2009). Live action: Can young children learn verbs from video? *Child Development*, *80*(5), 1360–1375. https://doi.org/10.1111/j.1467-8624.2009.01338.x

Sabbagh, M. A., & Baldwin, D. A. (2001). Learning Words from Knowledgeable versus Ignorant Speakers: Links between Preschoolers' Theory of Mind and Semantic Development. *Child Development*, *72*(4), 1054–1070. https://doi.org/10.1111/1467-8624.00334

Saito, D. N., Tanabe, H. C., Izuma, K., Hayashi, M. J., Morito, Y., Komeda, H., Uchiyama, H., Kosaka, H., Okazawa, H., Fujibayashi, Y., & Sadato, N. (2010). Stay tuned: Inter-individual neural synchronization during mutual gaze and joint attention. *Frontiers in Integrative Neuroscience*, *4*(NOVEMBER 2010), 1–12. https://doi.org/10.3389/fnint.2010.00127

Salazar, M., Shaw, D. J., Gajdo, M., Mare, R., Czekóová, K., Mikl, M., & Brázdil, M. (2021). *You took the words right out of my mouth: Dual-fMRI reveals intra- and inter-personal neural processes supporting verbal interaction*. *228*(December 2020). https://doi.org/10.1016/j.neuroimage.2020.117697

Sasson, N. J., Turner-Brown, L. M., Holtzclaw, T. N., Lam, K. S. L., & Bodfish, J. W. (2008). Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays. *Autism Research*, *1*(1), 31–42. https://doi.org/10.1002/aur.4

Schertz, H. H., Odom, S. L., Baggett, K. M., & Sideris, J. H. (2013). Effects of Joint Attention Mediated Learning for toddlers with autism spectrum disorders: An initial randomized controlled study. *Early Childhood Research Quarterly*, *28*(2), 249–258. https://doi.org/10.1016/j.ecresq.2012.06.006

Schmitz, M. (2014). Joint attention and understanding others. *Synthesis Philosophica*, *29*(2), 235–251.

Schneider, B., & Pea, R. (2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning*, *8*(4), 375–397.

https://doi.org/10.1007/s11412-013-9181-4

Schreiber, B. E., Junaid, F., & Fabiana, G. (2010). Live lecture versus video podcast in undergraduate medical education. *Medical Education*, *10*(68), 1–6. https://bmcmededuc.biomedcentral.com/track/pdf/10.1186/1472-6920-10-68

Scott, T. L., Haenchen, L., Daliri, A., Chartove, J., Guenther, F. H., & Perrachione, T. K. (2020). Noninvasive neurostimulation of left ventral motor cortex enhances sensorimotor adaptation in speech production. *Brain and Language*, *209*(July), 104840. https://doi.org/10.1016/j.bandl.2020.104840

Semrud-Clikeman, M., Walkowiak, J., Wilkinson, A., & Butcher, B. (2010). Executive functioning in children with asperger syndrome, ADHD-combined type, ADHD-predominately inattentive type, and controls. *Journal of Autism and Developmental Disorders*, *40*(8), 1017–1027. https://doi.org/10.1007/s10803-010-0951-9

Senju, A., Tucker, L., Pasco, G., Hudry, K., Elsabbagh, M., Charman, T., & Johnson, M. H. (2013). The importance of the eyes: Communication skills in infants of blind parents. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1760). https://doi.org/10.1098/rspb.2013.0436

Seuren, L. M., Wherton, J., Greenhalgh, T., & Shaw, S. E. (2021). Whose turn is it anyway? Latency and the organization of turn-taking in video-mediated interaction. *Journal of Pragmatics*, *172*, 63–78. https://doi.org/10.1016/j.pragma.2020.11.005

Seyfarth, R. M., & Cheney, D. L. (2014). The evolution of language from social cognition. *Current Opinion in Neurobiology*, *28*, 5–9. https://doi.org/10.1016/j.conb.2014.04.003

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

Sievers, B., Welker, C., Hasson, U., Kleinbaum, A. M., & Wheatley, T. (n.d.). *How consensus-building conversation changes our minds and aligns our brains*.

Skuballa, I. T., Xu, K. M., & Jarodzka, H. (2019). The impact of Co-actors on cognitive load: When the mere presence of others makes learning more difficult. *Computers in Human Behavior*, *101*(July), 30–41. https://doi.org/10.1016/j.chb.2019.06.016

Solomon, D. J., Ferenchick, G. S., Laird-Fick, H. S., & Kavanaugh, K. (2004). A randomized trial comparing digital and live lecture formats ISRCTN40455708. *BMC Medical Education*, *4*, 1–6. https://doi.org/10.1186/1472-6920-4-27

Spengler, S., Bird, G., & Brass, M. (2010). Hyperimitation of actions is related to reduced understanding of others' minds in autism spectrum conditions. *Biological Psychiatry*, *68*(12), 1148–1155. https://doi.org/10.1016/j.biopsych.2010.09.017

Stanley, D. A., & Adolphs, R. (2013). Toward a neural basis for social behavior. *Neuron*, *80*(3), 816–826. https://doi.org/10.1016/j.neuron.2013.10.038

Striano, T., Chen, X., Cleveland, A., & Bradshaw, S. (2006). Joint attention social cues influence infant learning. *European Journal of Developmental Psychology*, *3*(3), 289–299. https://doi.org/10.1080/17405620600879779

Suda, M., Takei, Y., Aoyama, Y., Narita, K., Sato, T., Fukuda, M., & Mikuni, M. (2010). Frontopolar activation during face-to-face conversation: An in situ study using near-infrared spectroscopy. *Neuropsychologia*, *48*(2), 441–447. https://doi.org/10.1016/j.neuropsychologia.2009.09.036

Tanaka, F., Cicourel, A., & Movellan, J. R. (2007). Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 17954–17958. https://doi.org/10.1073/pnas.0707769104

Tremblay, P., & Small, S. L. (2011). On the context-dependent nature of the contribution of the ventral premotor cortex to speech perception. *NeuroImage*, *57*(4), 1561–1571. https://doi.org/10.1016/j.neuroimage.2011.05.067

Troseth, G. L., Saylor, M. M., & Archer, A. H. (2006). Young children's use of video as a source of socially relevant information. *Child Development*, *77*(3), 786–799. https://doi.org/10.1111/j.1467-8624.2006.00903.x

Turnbull, A., Wang, H. T., Murphy, C., Ho, N. S. P., Wang, X., Sormaz, M., Karapanagiotidis, T., Leech, R. M., Bernhardt, B., Margulies, D. S., Vatansever, D., Jefferies, E., & Smallwood, J. (2019). Left dorsolateral prefrontal cortex supports context-dependent prioritisation of off-task thought. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-11764-y

Tylén, K., Allen, M., Hunter, B. K., & Roepstorff, A. (2012). Interaction versus observation: Distinctive modes of social cognition in human brain and behavior? a combined fMRI and eye-tracking study. *Frontiers in Human Neuroscience*, *6*(NOVEMBER 2012), 1–11. https://doi.org/10.3389/fnhum.2012.00331

Vaccani, J. P., Javidnia, H., & Humphrey-Murto, S. (2016). The effectiveness of webcast compared to live lectures as a teaching tool in medical school. *Medical Teacher*, *38*(1), 59–63. https://doi.org/10.3109/0142159X.2014.970990

Verga, L., & Kotz, S. A. (2013). How relevant is social interaction in second language learning? *Frontiers in Human Neuroscience*, *7*(SEP), 1–7. https://doi.org/10.3389/fnhum.2013.00550

Vigliocco, G., Conventino, L., Felice, S. De, & Gregorians, L. (n.d.). *Ecological Brain : Reframing the Study of Human Behaviour and Cognition*.

Vigliocco, G., Motamedi, Y., Murgiano, M., Wonnacott, E., Marshall, C., Milán-Maillo, I., & Perniss, P. (2019). Onomatopoeia, gestures, actions and words: how do caregivers use multimodal cues in their communication to children? *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, *41*, 1171–1177. https://doi.org/10.31234/osf.io/v263k

Vivanti, G., & Rogers, S. J. (2014). Autism and the mirror neuron system: Insights from learning and teaching. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1644). https://doi.org/10.1098/rstb.2013.0184

Wakefield, E., Novack, M. A., Congdon, E. L., Franconeri, S., & Goldin-Meadow, S. (2018). Gesture helps learners learn, but not merely by guiding their visual attention. *Developmental Science*, *21*(6), 1–12. https://doi.org/10.1111/desc.12664

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, *7*(12), 553–559. https://doi.org/10.1016/j.tics.2003.10.012

Webb, S. J., Neuhaus, E., & Faja, S. (2017). Face Perception and Learning in Autism Spectrum Disorders. *Quarterly Journal of Experimental Psychology*, *70*(5), 970–986. https://doi.org/10.1080/17470218.2016.1151059.Face

Westlund, J., Dickens, L., Jeong, S., Harris, P., Desteno, D., & Breazeal, C. (2015). A Comparison of Children Learning New Words from Robots, Tablets, & People. *New Friends*, 26–27. http://newfriends2015.org/Proceedings/Contents/Papers/New_Friends_2015_submission_18.pdf

White, S. J. (2013). The triple i hypothesis: Taking another('s) perspective on executive dysfunction in autism. *Journal of Autism and Developmental Disorders*, *43*(1), 114–121. https://doi.org/10.1007/s10803-012-1550-8

White, S. J., Coniston, D., Rogers, R., & Frith, U. (2011). Developing the Frith-Happé animations: A quick and objective test of Theory of Mind for adults with autism. *Autism Research*, *4*(2), 149–154. https://doi.org/10.1002/aur.174

Whitehouse, A. J. O., Barry, J. G., & Bishop, D. J. M. (2007). The broader language phenotype of autism: A comparison with specific language impairment. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *48*(8), 822–830. https://doi.org/10.1111/j.1469-7610.2007.01765.x

Wilson, S. M., Molnar-Szakacs, I., & Iacoboni, M. (2008). *Beyond Superior Temporal Cortex : Intersubject Correlations in Narrative Speech Comprehension*. *January*. https://doi.org/10.1093/cercor/bhm049

Wohltjen, S., & Wheatley, T. (2021). Eye contact marks the rise and fall of shared attention in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(37). https://doi.org/10.1073/pnas.2106645118

Wolf, W., & Tomasello, M. (2020). Watching a video together creates social closeness between children and adults. *Journal of Experimental Child Psychology*, *189*, 104712. https://doi.org/10.1016/j.jecp.2019.104712

Yeshurun, Y., Nguyen, M., & Hasson, U. (2021). The default mode network: where the idiosyncratic self meets the shared social world. *Nature Reviews Neuroscience*, *22*(March). https://doi.org/10.1038/s41583-020-00420-w

Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C. J., & Hasson, U. (2017). *Same Story , Different Story : The Neural Representation of Interpretive Frameworks*. https://doi.org/10.1177/0956797616682029

Yi, L., Wang, Q., Song, C., & Han, Z. R. (2022). Hypo- or hyperarousal? The mechanisms underlying social information processing in autism. *Child Development Perspectives*, *16*(4), 215–222. https://doi.org/10.1111/cdep.12466

Yu, C., Suanda, S. H., & Smith, L. B. (2017). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science*, *22*(1), 1–12. https://doi.org/10.1111/desc.12735

Zhang, D., Zhou, L., Briggs, R. O., & Nunamaker, J. F. (2006). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information and Management*, *43*(1), 15–27. https://doi.org/10.1016/j.im.2005.01.004

Zhang, Y., Frassinelli, D., Tuomainen, J., Skipper, J. I., & Vigliocco, G. (2021). More than words: Word predictability, prosody, gesture and mouth movements in natural language comprehension. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1955). https://doi.org/10.1098/rspb.2021.0500

Zheng, L., Chen, C., Liu, W., Long, Y., Zhao, H., Bai, X., Zhang, Z., Han, Z., Liu, L., Guo, T., Chen, B., Ding, G., & Lu, C. (2018). Enhancement of teaching outcome through neural prediction of the students' knowledge state. *Human Brain Mapping*, *39*(7), 3046–3057. https://doi.org/10.1002/hbm.24059

Zheng, L., Liu, W., Long, Y., Zhai, Y., Zhao, H., Bai, X., Zhou, S., Li, K., Zhang, H., Liu, L., Guo, T., Ding, G., & Lu, C. (2020). Affiliative bonding between teachers and students through interpersonal synchronisation in brain activity. *Social Cognitive and Affective Neuroscience*, *15*(1), 97–109. https://doi.org/10.1093/scan/nsaa016

Zimeo Morais, G. A., Balardin, J. B., & Sato, J. R. (2018). FNIRS Optodes' Location Decider (fOLD): A toolbox for probe arrangement guided by brain regions-of-interest. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-21716-z