‘I interact therefore I am’
Human becoming in and through social interaction

Dissertation

zum Erwerb des Doctor of Philosophy (Ph.D.)

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

vorgelegt von

Dimitrios Bolis

aus Thessaloniki

2020

Supervisor: Prof. Dr. Leonhard Schilbach

Second expert: Prof. Dr. Nikolaos Koutsouleris

Third expert: Prof. Dr. med. Alkomiet Hasan

Dean: Prof. Dr. med. dent. Reinhard Hickel

Date of oral defense: March 6th, 2020

“One drop plus one drop makes a bigger drop, not two.”

Andrei Tarkovsky

Abstract

That psychological sciences suffer from a profound crisis is probably not extremely controversial. Yet, arguably, the recently debated replication failure is nothing but a symptom of deeply rooted dichotomies and ontological commitments lying at its core. Undeniably, essential aspects of the human condition are typically studied in isolation via applying static categories, while philosophical considerations and human practice are largely neglected. In this context and drawing inspiration from real-life experience through a Vygotskian lens, this thesis attempts to motivate a systematic shift of focus from *being* to *becoming*; in fact *becoming-with*. More concretely, leaning on the dialectical method, cultural-historical theory and recent developments of social computational neuroscience, (i.) this thesis presents the *dialectical attunement* account which argues that a multiscale analysis of social interaction might allow us to scientifically reconsider the self, beyond the individual, where it really emerges, unfolds and manifests itself — in social relations. In this light, (ii.) it puts forward the *dialectical misattunement hypothesis*, which views autism and broadly psychopathology as a dynamic interpersonal mismatch, rather than a (disordered) function of single brains. Critically, (iii.) it operationalizes these hypotheses by establishing a novel empirical framework, namely *two-person psychophysiology*, which measures and analyzes the multiscale dynamics of social interaction. Deploying this framework, this thesis empirically demonstrates that (iv.) real-time dynamics of social interaction do matter in both collective and individual dimensions – even beyond awareness– lending support to second-person and enactivist proposals. With regard to psychopathology, this thesis demonstrates that (v.) it is primarily the mismatch of autistic traits –not traits *per se*– which predicts core aspects of interpersonal attunement in real-life social relations, offering a first empirical validation of the dialectical misattunement hypothesis. Taken together, this thesis tries to break free from dichotomies such as internalism/externalism or healthy/patient, in experiential, theoretical, methodological and empirical regards. Such a dialectical and empirical approach to human becoming in and through social interaction encourages a social change pertinent to various fields of human research and practice, ranging from psychiatry and pedagogy to ethics and artificial intelligence.

Keywords: *dialectical attunement, dialectical misattunement, two-person psychophysiology, second-person neuroscience, Bayesian intersubjectivity, social interaction, social relations, self, culture, autism, dialectics, enactivism, computational psychiatry, predictive processing, active inference, free energy principle, Vygotsky, Bayes.*

Acknowledgments

This thesis could not be accomplished without the contribution of several people. First of all, I would like to express my deep gratitude and admiration to my academic supervisor, Leonhard Schilbach, for the scientific inspiration, continuous support, kind feedback and limitless trust as well as freedom of action, throughout the duration of this endeavor. In fact, I could not imagine of a more productive collaboration. Additionally, I am thankful to the members of my Thesis Advisory Committee, Nikolaos Koutsouleris and Alkomiet Hasan, as our meetings and their always generous comments ensured the smooth progression of this work.

I also really appreciate my interactions with a number of experts in the field, including but not limited to Bahador Bahrami and Ali Mahmoodi, who shared with me the original material of their seminal social interaction studies, as well as Cristina Becchio, Guillaume Dumas, Karl Friston, Ivana Konvalinka, Anna Ciaunica and Manolis Dafermos for the insightful and inspiring discussions and support. Crucially, parts of the work accomplished in collaboration with the Japanese National Institute for Physiological Sciences (NIPS) and the University of Cambridge and specifically with the great support from Takahiko Koike, Norihiro Sadato, Victoria Leong and Stanimira Georgieva. Furthermore, I am thankful to my colleagues in Munich and especially my officemates Juha Lahnakoski, Marie Luise Brandi, Tore Erdmann, Lara Henco, Hella Parpart, Laura Albantakis, Imme Christina Zillekens and Raoul Haaf for the daily and encouraging communication, not only scientifically but also personally.

Additionally, the multifaceted help of a number of BSs, MD and MSc students, I had the luck to supervise and work with, have been crucial for accomplishing the great amount of empirical experiments, which involved the recruitment of approx. 400 participants over the last 4 years,

that is Alana Darcher, Beril Nisa Can, Bhagyashree Padalkar, Carolin Nafziger, Daniela Seidel, Dorontina Ismajli, Ebru Ecem Tavacioglu, Elena Wang, Emre Yavuz, Jeanette Tamm, Julia Folz, Kamile Giedraityte, Lioba Enk, Magdalena Seethaler, Milena Aleksic, Tannaz Mostafid and Uta Schneider. I am also deeply grateful to all the participants who kindly volunteered to participate in our studies and oftentimes provided their invaluable feedback. Furthermore, the help of the MPI technician Karl-Heinz Honsberg has been instrumental in solving a multitude of technical problems I faced while designing and putting together the electronics of the two-person psychophysiology platform.

Last but not least, I owe this endeavor to the few persons around me and deeply within, who have been unceasingly providing me not only with the courage and –often practical– support for continuing, but the literal meaning of life, which according to Albert Camus has been nothing more or less than ‘whatever prevents you from killing yourself’.

Contents

Preface

Through others we become ourselves, Lev Vygotsky proclaimed almost a century ago. This short and seemingly simple passage does not only serve as a point of departure for this thesis, but has been, along with its philosophical underpinnings, incessantly shaping my view on the world. It was in 2006 when I stumbled upon a collection of Vygotsky's articles on a dialectical perspective to developmental psychology and pedagogy. The timing could not have been better.

At that point in time, getting increasingly uninterested in my formal electrical and computer engineering studies at the Aristotle University of Thessaloniki, I have been taking a break for working on a side project for a competition calling *'Imagine a world where technology enables a better education for all'*. Together with three friends and an engineering professor, we had started developing an enhanced educational environment for autistic children. The aim of our project was to facilitate social inclusion and educational access for autistic children via *tuning* technology to social needs. On a conceptual level, we primarily drew inspiration from the Monotropism hypothesis (Murray *et al*., 2005), which had been published just the year before. What was important for us about this hypothesis was that it could serve as an opportunity to break away from an over-medicalized paradigm, yet in a scientific way. In this light, at least part of the autism spectrum condition could be considered not as a disorder *per se*, but as a difference in attention allocation. This reading aligned nicely with my personal intuitions. When Microsoft asked each of us to describe our project in a single sentence for a promo video, I chose *"we see autism not as an illness, but as a way of being"*.

Growing up on stories from my mother's daily work as an educator for autistic children and paying regular visits to her kindergarten, I had already begun strongly doubting whether autism

was a disorder to ‘cure’. Coming back from kindergarten, my mother one day said, ‘today is one of my happiest days at work, I achieved teaching a boy hugging’. This child was not able to speak, but this ‘detail’ completely changed the communication with the educators and parents. Indeed, several autistic children in the kindergarten are not able to speak or they use language idiosyncratically, i.e., in an unusual but still meaningful way. In other words, these children do desire to communicate but their way is not *attuned* to the way of others. I was especially surprised with the effectiveness of communicating gestures when verbal channels failed. Communication is not about *retuning* a child after all; it is about cultivating suitable *interpersonal attunement.*

‘I am so excited; I found a way of making a child get up from his little corner at last’, my mother happily said another time. This child always preferred sitting silent at a corner of the room. It was when my mother accidentally played an oriental song over the speakers that the boy started dancing. This served as a trigger for the educators approaching and teaching him various other skills in the following period. His parents came from the Middle East. I still suspect there might have been a kind of cultural *attunement* between the child and the music. Indeed, on several occasions I observed a child in mood swings, getting instantly calmer, upon listening to an educator singing. I am tempted to speculate here that a kind of *interpersonal attunement* may have actually facilitated *intrapersonal attunement*, a kind of self-regulation.

On another occasion, my mother came back home deeply touched. A mother of a child who had been at the kindergarten years ago had visited her, saying ‘you literally saved my child’s life’. The child could not talk, but my mother had taught him, when in pain, to touch the suffering body part in a particular way. At that day, the child, unable to breathe due to a piece of food stuck in his throat, approached his mother making the taught movement. Indeed, seemingly small

learning steps towards establishing and enhancing interpersonal communication can prove critical in the long run.

I, once, asked the educators what was their most effective approach to teaching. Without any hesitation, they responded the first and most crucial part is the development of an interpersonal relation of trust and safety between the child and the educator. This phase might even take several months, but skipping it, any other pedagogical effort becomes meaningless. With that, I stop putting down relevant incidents. Reflecting back, what I really gained was an experiential understanding of autism and other developmental conditions as a *misattunement between* the child and the social world — not a mere disorder of the individual.

Seriously adhering to such a perspective means treating autism by focusing on re-establishing an *attunement* between the child and the (social) world and not by exclusively *retuning* the child according to an incidental current societal normal. This was exactly what we tried to do with our competition project back then in 2006-2008, aiming at dynamically *tuning* the educational environment to the needs of each child and not the other way round. To this end, we developed a self-regulated digital system, which, taking into account the historicity of each child's reactions, levels of arousal and personal preferences, adjusted itself in *real-time* to the needs and specific interests of the child, aiming at the same time at expanding their repertoire. With regard to real-life social interactions, the platform deployed social robotics, as a mediator to the interaction between two children for alleviating initial social pressure, and a digital forum for facilitating the communication between parents and educators.

The project being awarded with various prizes, received funding promises by educational officials for further development and use in Greek public schools. In the meanwhile, I completed

my engineering studies and continued with researching assistive technology in the lab of artificial intelligence and information analysis of the same university with the hope of enriching our autism project. After almost a year there, I started realizing that the Greek economic crisis would be cancelling any relevant plan in the following years. I knew it was about time to abandon my excitement — for the time being.

I moved to Switzerland to pursue an MSc on biomedical engineering in ETH Zurich. There, I came across and immediately got enthusiastic with the work of Karl Friston and colleagues on the 'free energy principle' (e.g., Friston, 2013). Formally bringing together thermodynamics, information theory, biological and human sciences, the free energy principle appeared as a powerful toolbox for a principled understanding of a great variety of phenomena; ranging from organismic life, developmental and evolutionary processes to human brain function, perception, learning and action. Indeed, while studying human learning of perceptual regularities and decision-making, I tried to link my Bayesian modeling and experimental work to autism research, but this was not feasible organizationally, as at that time the translational neuromodeling unit, where I worked for my master thesis, was at its very first days. Yet, studying the free energy principle decisively shaped the way I think of not only the brain, but life altogether. In fact, this was my second intellectual shock, after my contact with dialectics and Vygotsky.

Yet, before finishing my MSc thesis, I have already started feeling increasingly uncomfortable with the early articulations of such Bayesian accounts. In brief, through a Bayesian lens one can view the brain as an organ which tracks perceptual regularities, via a combination of already gained experience and newly sensed information. Importantly, the more confident one is about the validity of previous experience, the less these beliefs are updated based on currently

incoming information. While admiring the potential of such a powerful computational framework to formally model a multitude of real-life situations, I found that an exclusive focus on mechanisms within the individual organism sharply contradicted my understanding of the human condition as inherently social.

In the meanwhile, my readings on social philosophy have been progressing. In my first semester's philosophical class, I named my little project, *"Consciousness as a social product"*, perhaps to a slight disappointment of my instructors, as the officially given title of the class has been *"The neurobiology of consciousness".* The contradiction I felt between the mechanistic accounts of brain function, I used in my research work, and the theories of social philosophy, I read in parallel, was huge. Then it was when I decided to pursue a PhD with the dual aim of resolving such a personal intellectual contradiction in connection to my long-lasting desire to work on practically improving the quality of life for autistic persons and other socially excluded groups of people.

While preparing my PhD proposal in 2014, I accepted an offer by the Neural Control of Movement lab to work on researching autism in Ireland in a joint project between ETH Zurich and Trinity College Dublin. For the needs of this project, I worked on conducting neuroimaging experiments, as well as long face-to-face interviews with autistic persons and their families. This experience and especially the direct and close contact with autistic families over an extended period of time proved to be decisive. This was my second experiential shock with regards not only to autism this time, but the broader sociopolitical dimensions of social exclusion.

Interviews with autistic families further strengthened my intuition about autism, as a *difference* rather than a *disorder per se*. Sadly, almost every autistic person had a story of social exclusion

to report. Intriguingly, most of those stories were centered on the school life, while later in life social interaction appeared often –at least up to a degree– smoother. But what is it improved after school? After finishing school people usually have more freedom of choice with regard to their own daily activities, as well as people to interact with. I can recall that most of the young autistic persons reported as favorite hobby playing video games within autistic social groups. They regularly mentioned that they felt more comfortable when interacting with other autistic persons in everyday life. It has started getting clear to me that a 'problematic' social interaction was not a merely individual issue of the autistic person, but rather an interpersonal one.

Nevertheless, it was not only interpersonal difficulties autistic persons and their families discussed as prominent in these interviews. For instance, a lot of caregivers reported that their children, especially at an early age, could not tolerate unexpected noise or intense light. Such kinds of stimuli would usually make the autistic child extremely anxious, who would typically cover her ears or eyes until such stimulation stops. Another prominent difficulty discussed by the families was relevant to repetitive patterns of behavior, including so called obsessive and compulsive tendencies. For example, an autistic person would not have his lunch if the different types of food on the plate were not clearly separated from each other. Another person would not touch the handle of any door in the family house. Actually, this person got his fingers injured multiple times due to his habit to open or close the door exclusively by touching the very upper part of it. Another person would not touch anything in his room which had been brought (and thus touched) by another member of the family. Another child when in his room would rock back and forth for hours. But do such kinds of difficulties really lack a social dimension?

Of course not. A regularity I noticed was that most of those cases referred to family life within the house. Why was that? I came to realize that as these kinds of behavior appear oftentimes

irritating to (*neurotypical*) others, initially autistic persons were taught to avoid expressing them all together. However, the discomfort due to such forced behavior suppression was so strong that clinical practice seemingly followed a compromising approach in many cases: the autistic person was still taught to avoid performing such atypical patterns of behavior in public, but they were allowed to do so in the private space. This ensured some –albeit partial– relief to the autistic person. While this was a step toward somewhat taking into account the autistic voice, it was still far from a genuinely balanced approach. What if neurotypical patterns of social expectations were informed in parallel — through anti-stigma school campaigns, to name but an example?

Indeed, several were the complaints by autistic persons about others –especially school authorities– for considering them as troublemakers. Of course, it is not rare for young students to appear radical to their older teachers, yet I had the feeling that autistic people felt more often labeled as such. Indeed, autistic persons are oftentimes more focused on actual facts than others' impressions. This might increase the possibilities of experiencing interpersonal conflict in everyday life on an individual level; however, on a societal level it might actually constitute an invaluable alternative perspective to social life. Therefore, it would be perhaps more fruitful – and not only for the autistic person– if we aimed at informing the neurotypical perspective, instead of thoughtlessly penalizing neurodivergent thinking and behavior. Yet, we should not mistake a *difference* in thinking and interacting with others as a *lack* of interest to do so.

I vividly remember a mother telling me about her young autistic son, who once said 'I wish my sister had a tail'. The mother puzzled asked why. The son replied that in such a case he could be confident about when his baby sister was feeling happy because, as dogs do, she would be moving her tail. The mother was rather disappointed as she had spent much time teaching her son about typical emotional responses –indeed especially 'smiling'– through repetitive

presentation of smiling photos. When she asked him whether he recalls all this training, the son naturally replied, ‘of course I remember; you are happy when you show your teeth, but my sister has no teeth yet’. It was not a lack of social motivation, but rather a different learning style that caused the confusion. Indeed, as we will discuss later, such kind of static, passive and out-of-context training of social skills might not be the most optimal.

Perhaps most importantly, those interviews made clear to me that autism could not be addressed via mere neurobiological approaches targeting the individual, however mathematically powerful and methodologically neat they might appear. Considering a sociopolitical dimension is equally critical, to say the least. One mother of two autistic children, whose husband had recently died, kept telling me she wanted nothing more than providing her kids with the necessary skills and conditions for them to survive independently after her death. I had then started realizing that often the primary anxiety associated with autism stems from not the difference itself, but rather an *interpersonal misattunement* in a neurotypical world, as well as an inability of our society to provide *social inclusion for all*.

Those experiences resonated perfectly with my parallel readings on dialectics, primarily the work of Lev Vygotsky, who did not address disabilities, such as bodily challenges, as the core element to be ‘fixed’, but rather primarily focused on alleviating processes of social exclusion. He claimed that, what we really need to focus on, is re-establishing the channels of communication with the *other* and the *society*. This is exactly where my perspective to autism and other conditions draws inspiration from.

I had been already putting down my PhD proposal, starting with discussing prevalent cognitive approaches to autism and how a Bayesian perspective could facilitate an inter-theoretical

dialogue and a potential synthesis, when discussed about phenomenological perspectives to psychotherapy with colleagues in Dublin. Getting familiar with phenomenology and existentialism highly impacted the way I think. It was then when I was captivated by related enactivist accounts, perhaps not quite surprisingly, as they –albeit discretely– heavily draw on dialectics. In brief, enactivism claims that cognition and meaning arise in interaction with the environment and others — not via passive information processing. Work from Humberto Maturana, Francisco Varela, Ezequiel Di Paolo, Hanne de Jaegher, Shaun Gallagher and Tom Froese (e.g., De Jaegher *et al.*, 2010; De Jaegher & Di Paolo, 2007; Froese & Di Paolo, 2011), among others, further strengthened my –intuitive at the moment– critical view on traditionally individualistic theories of autism.

This is when I came across Leonhard Schilbach and his work with colleagues on the paradigm of 'second-person neuroscience', drawing on both philosophical considerations and methodological developments of modern neuroscience. This account, while emphasizing the role of real-time social interaction in social cognition, suggests that different modes of thinking about others are not necessarily exclusive to each other. For instance, humans might make sense of others in various ways, such as through phenomenological, inferential and interactional processes. Yet, the first two groups of processes come up more frequently in neuropsychological research than interactional ones. But is that inherent to human cognition, or perhaps a reflection of a self-fulfilling paradigm which –in a methodological vicious circle– probes the mechanisms which it anticipates? Indeed, second-person neuroscience postulates that cognitive processes during *real-time* social interactions might be fundamentally different than the ones arising in social observational situations (Schilbach *et al.*, 2013). With regard to psychopathology, it further suggests that transdiagnostically observed social impairments are more likely or may only

manifest in *real-time* social interactions, whereas observational scenarios might be less problematic and thus less informative clinically (Schilbach, 2016).

The second-person perspective highly resonated with both my philosophical and methodological considerations. Thus, I enthusiastically decided to start working on a PhD with Leonhard Schilbach and his team, the Independent Max Planck Research Group for Social Neuroscience and the Outpatient and Day Clinic for Disorders of Social Interaction at the Max Planck Institute of Psychiatry within the International Max Planck Research School for Translational Psychiatry and the Medical School of the LMU in Munich.

We start this thesis with briefly discussing the ongoing crisis of the field, arguing that explicitly considering real-life practice and philosophy will be critical. As Vygotsky proclaimed (1917-1934/1997), *"practice and philosophy [the stone which the builders rejected] are becoming the head stone of the corner".* Subsequently, before presenting the technical material of the theoretical, methodological and empirical work of this thesis, we go through the core conceptions in an intuitive way and outline the novelty — here I would like to draw the opportunity and make clear that, from this point on, I am dropping the use of the singular first-person 'I' in favor of the plural 'we', as a minimum acknowledgment that this, as any human endeavor, has been a deeply collective product in so many aspects. After all, *through others we become ourselves.*

Dimitris Bolis
Munich
December 2019

1. Introduction — *'Through others we become ourselves'*

"Practice and philosophy
[the stone which the builders rejected]
are becoming the head stone of the corner".

Lev Vygotsky

1.1. Science is not a scientist

Scientists have been skeptical about psychology already from its cradle. The American psychologist William James (1892) appeared rather cynical when discussing the field of psychology, "a string of raw facts; a little gossip and wrangle about opinions; a little classification and generalization on the mere descriptive level …. This is no science, it is only the hope of a science" (James, 1892/2001, p. 335; as cited in Dafermos, 2014). The description by the Russian psychologist Nikolai Lange was also rather apocalyptic, "The psychologist of these days is like Priam sitting among the ruins of Troy" (as cited in Yaroshevsky, 1989, p. 171; cf. Dafermos, 2014). The clinical side has not been much freer from such skepticism. The psychiatrist van Praag, at the end of last century (1992), was still warning "today's classification of the major psychiatric disorders is as confusing as it used to be some 30 years ago. All things considered, the present situation is worse. Then, the psychiatrists were at least aware that diagnostic chaos reigned and many of them had not high opinion of diagnosis, anyhow. Now the chaos is codified and thus much more hidden" (as cited in Ghaemi, 2018). Another 30 years have passed and the situation does not seem to have been drastically improved.

In fact, after almost a century since Vygotsky wrote his book on *"The Historical Meaning of the Crisis in Psychology: A Methodological Investigation"* (1927/1997), the broad field still debates about another major (facet of the) crisis. Indeed, in an online survey by Nature, more than 1,500 scientists were asked about their impression with regard to reproducibility in empirical science (cf. Baker, 2016). More than half agreed that there's a significant crisis, while 90% reported that there is at least a slight crisis. Only 3% responded negatively to the question. Additionally, over 70% reported that they failed to reproduce another's experiments. The field of psychology has been central to such debates. A special issue in 2012 focused exactly on this, *"Special Section on*

Replicability in Psychological Science: A Crisis of Confidence?" (Pashler & Wagenmakers, 2012), echoing Elms when claiming "whether [psychologists] are experiencing an identity crisis, a paradigmatic crisis, or a crisis of confidence, most seem agreed that a crisis is at hand" (Elms, 1975, p. 968; as cited in Dafermos, 2014).

According to Thomas Kuhn (1970), crises in science are periodic phenomena caused by a mismatch between the dominant scientific paradigm and empirical evidence, a kind of prediction error. Therefore, a crisis could be seen not merely as a major complication, but rather as an opportunity for a scientific revolution, so that predictions of scientific paradigms are updated. The simplified sketch, we drew so far, indeed, resembles current computational theories of 'optimal' learning agents, as if science was an 'ideal' scientist trying to make sense of the world. However, *science is not a scientist*[1]. Or if it was, it would not be an idealized abstract form of a scientist, but rather an everyday human being driven by material necessities within a concrete stage of the historical development and sociocultural context.

Indeed, a multitude of factors –oftentimes contradictory and seemingly irrelevant– contribute to the progress and occasionally regress of science. For instance, the above-mentioned Nature survey showed that scientists do believe that not only methodological, but also socioeconomic processes are at play: "Many top-rated factors relate to intense competition and time pressure" (Baker, 2016). The survey concluded that better training, targeted incentives and robust experimental designs are important factors for overcoming the crisis. A major part of the discussion about the replicability crisis has, indeed, been devoted to –either intentional or

[1] to paraphrase a Bruineberg and colleagues' article title, "The anticipating brain is not a scientist" (2018).

unintentional– mistakes made by individual scientists. Here, we endorse Dafermos (2014), who pointed out that "the crisis in science could not be explained as a result of personal mistakes of its founders or "good" or "bad" intentions of its practitioners".

Of course, the sophistication of experimental means, proper scientific training and integrity are important, yet it is the –even implicit– adoption of a theory that essentially defines the generation and interpretation of empirical observations (cf. **Section 5.2** for a discussion relevant to the topic of this thesis). Put simply, no answer is valid when asking the wrong question. This idea is not new. Already Vygotsky had argued that a fact always contains an implicit theoretical concept, *"everything described as a fact is already a theory"* (Vygotsky, 1917-1934/1997, p. 249), anticipating Kuhn's proposal about the theory-dependence of observation, which challenged the positivist perspective to science as an accumulation of facts (Dafermos, 2014). Having said that, we should note that there are crucial differences between Kuhn's proposal of *'paradigm replacement'*, which emphasized non-accumulation and discontinuity in the history of science and Vygotsky's *dialectical perspective* which viewed a set of evolutionary (cf. quantitative) and revolutionary (cf. qualitative) moments.

Yet, the empiricism's illusion of objectivity still hinders a genuine consideration of the philosophical underpinnings of the various theories, leading scientists to uncritically adopt their implicit ontological and epistemological assumptions — away from the concreteness of the real-life phenomena (Dafermos, 2014). In a nutshell, *"the principle and philosophy of practice is –once again– the stone which the builders rejected and which became the head stone of the corner. Here we have the whole meaning of the crisis"* (Vygotsky, 1917-1934/1997).

1.2. Core conceptions at a glance

In this light, this thesis aims at experientially (**Chapter 1**), theoretically (**Chapter 2**), methodologically (**Chapter 3**) and empirically (**Chapter 4**) addressing, as well as integratively discussing (**Chapter 5**) its two core problems: (i.) *tracing the social origins of human becoming across scales* and (ii.) *reconsidering psychopathology as a dynamic interpersonal mismatch, rather than a mere disorder of the individual* (**1st** and **2nd part** of this **Section** and **Chapters 2-5**). It concludes (**Chapter 6**) with a question: *Does social interaction matter after all?*

1.2.1. Dialectical attunement

Leaning on dialectics and Bayesian accounts of cognition and action, we start our analyses with theoretically arguing that a fine-grained analysis of social interaction across scales might allow us to reconsider the self beyond the individual, where it really unfolds and manifests itself, in social relations (cf. **Section 2.1**; Bolis & Schilbach, 2018b). To this end, we construe *human becoming* as the dynamic interplay between (social) *internalization* and (collective) *externalization* in and through culturally mediated social interaction. On one hand, internalization is taken as the co-construction of bodily hierarchical models of the (social) world and the organism. As Mead insightfully put it (1912) “Inner consciousness is socially organized by the importation of the social organization of the outer world”. On the other hand, externalization is taken as the collective –tool mediated– transformation of the world. In a nutshell, interpersonal statistical regularities (collective level) are thought of as shaping multiscale hierarchical bodily structures (individual level) and *vice versa*. Below, we unpack these core ideas.

Our notion of internalization is largely based on predictive processing conceptualizations (cf. Friston, 2013; see also Clark, 2013). Predictive processing has been defined as a hierarchical bidirectional process through which an organism adjusts itself in order to 'optimally' predict environmental and bodily regularities. With regard to brain function, on one hand, predictions are continuously generated and propagated from higher levels of the neural hierarchy to lower ones in an attempt to explain away prediction errors, i.e., the discrepancy between incoming information and generated predictions. On the other hand, prediction errors are propagated from lower levels of the hierarchy to higher ones in order to reconfigure the organism. The ultimate goal of such a process is to minimize prediction error as precisely as possible, through processes such as perception and learning. Importantly, higher levels of the hierarchy deal with higher levels of abstraction (e.g., higher scales of time and space). Such hierarchical structures should be considered as collectively shaped. First, we dynamically embody others in and through social interaction, shaping each other's hierarchical structure (**Fig. 2.2**; cf. Bolis & Schilbach, 2018b), and, second, such structures might even be socially extended into interbodily configurations (cf. Ramstead *et al.,* 2018). Structure and culture of social groups are two plausible facets of such configurations. In brief, organisms are continuously trying to optimize their expectations across scales in order to survive. Notably, expectations in this framework span a broad range of regulated processes, from conscious beliefs to interoceptive states.

Yet, organisms such as humans are not passive observers of reality, who merely try to adapt to it. On the contrary, humans continuously interact with their world (including their own body), adjusting it according to their prior expectations. For instance, when the perceived partial pressure of carbon dioxide exceeds certain bounds, the respiratory system is responsible for keeping it within expected levels, ensuring bodily order maintenance and thus survival. To give

another example, the body temperature tends to fall behind expected values (~37 °C) in case a person stays idle in the cold for a long time. Starting trembling, putting on a fire or going into a heated building typically reverses this tendency and helps keep the body temperature within well-defined bounds. Such processes of actively controlling the body and the environment, for minimizing prediction error through activity, have been called active inference (cf. Friston, 2013; see also Clark, 2013). However, such processes should not be thought of as exclusively lying within the individual. For example, architecture and technology can be viewed as a collective effort for reducing overall uncertainty via transforming the environment according to bodily and interpersonal expectations. In brief, humans actively co-construct and co-regulate –in interaction with other organisms– their ecosocial niches, so that they increase survival chances of not only the individual, but also the social group and the species as a whole.

In a nutshell, we operationalize our conceptualizations about human becoming leaning on theories of predictive processing and active inference, attempting to unveil their inextricable interrelations. To this end, we situate these accounts in the sociocultural realm through a dialectical prism. Dialectics is an evolving school of thought, which emphasizes becoming over being, as well as wholeness, interconnectedness and historicity of phenomena over artificial, isolated and decontextualized dichotomies. Additionally, dialectics emphasize inherent contradictions, which dynamically lead to crises and resolutions via qualitative leaps (cf. Wong, 2006). Put simply, “dialectics as a way of thinking strives to grasp the essential relations of a developing thing and reveals its historical origin and the perspectives of its change” (Dafermos, 2018). To make these points more concrete we now briefly examine a series of illustrative examples across various scales.

For instance, in a relatively short time scale, such as the scale from seconds to minutes of sensorimotor loops, let us imagine a person manipulating a tool (cf. Leontyev, 1975/1983). In her effort to actively modify the environment, she –at the same time– modifies herself. In order to be successful, in even holding the tool against gravity, she needs to literally embody the physical dynamics of the interacting system of the extended body and the environment. Or let us imagine a skilled couple dancing tango in perfect coordination. In this case, both bodies implicitly anticipate each other's moves and act accordingly. These interbodily anticipatory and enactive processes can be thought of as inextricably linked, largely unfolding beyond the time scale of conscious reflection. Indeed, anybody, experienced in dancing, may have realized that taking the time to reflect on the sequential exchange of explicit predictions and reactions in real-time ends up oftentimes more detrimental than beneficial in the overall dancing duet. Notably, such sensorimotor patterns, if enacted persistently over time, are shaped within larger time scales. For instance, mastering a skill, such as learning how to bike, takes months to years to reach relative stability.

In an intermediate scale, such as several years, let us think of an ontogenetic example inspired by Vygotsky (1930–1935/1978). A baby, trying to regulate her hunger and thus interoception balance, tries to reach a fruit. However, she is unable to do so, despite trying with her whole body, even extending the index finger toward the fruit. Her action repertoire is still not sufficiently developed. A caregiver, observing the situation and predicting what the baby is trying for, easily grasps the fruit and gives it to her. After multiple repetitions, the extended index finger comes to symbolize *pointing* to a desired object in order to draw attention and seek for help. This was a simple but illustrative example of how interpersonal regularities can eventually be transformed and internalized within individuals during real-time social

interactions, taking up a symbolic form. Stretching this example to an intergenerational scale, we can imagine how certain cultural conventions and norms, such as the symbolic function of pointing, might have emerged in and through social interaction.

In the scale of decades and centuries that cultural transmission and technology production play out, let us consider the modern electrical network. Switching on a light by the mere press of a button would seem like magic to a human a few centuries ago. Even today, literally no person by herself would be able to recreate the whole electrical system neither practically, nor even conceptually, if left alone. Yet, how come it seems so simple and effortless to switch on the light that we often do it almost automatically? It is because synergies of people have been interacting with one another and the environment, across multiple time and space scales. These interactions involved inextricably linked multiscale processes of (social) predictions and (collective) action.

Considering possible interactions of these processes at the scale of phylogeny might allow us to examine the potential social origins and interrelation of certain aspects of human anatomy and human-specific skills. A typical example can be found in the human eye morphology (e.g., ratio of exposed sclera in the eye outline) and the enhanced gaze-based interaction in humans (cf. Dobson, 2012; Kobayashi & Kohshima, 2001; Powell *et al.*, 2010). These sorts of multiscale and inextricably linked processes of *human becoming while attuning* to one another and the environment, is what we describe as *dialectical attunement* (cf. **Section 2.1**).

1.2.2. Dialectical misattunement

Subsequently, extending our focus on psychopathology, we put forward the *dialectical misattunement hypothesis*, which views psychiatric conditions, such as autism spectrum conditions (ASC), not as (disordered) function within single brains, but rather as a dynamic

interpersonal mismatch (cf. **Section 2.2**; Bolis *et al*., 2016; 2017). Here, *misattunement* across persons is thought of as a series of disturbances of the dynamic and reciprocal unfolding of an interaction. Such misattunement results in potentially increasingly divergent prediction and interaction styles and *vice versa*. Prediction and interaction styles are defined as a set of prior expectations and reaction patterns a person dynamically develops in interaction with the world and others, as discussed above. A dynamic *interpersonal misattunement* should be similarly expected to unfold across various scales.

For instance, in the scale of minutes, let us consider the case of a dialogue between two persons. One might express an opinion or act in a way the other considers unexpected or even insulting. In turn, the second person might reply with a defensive phrase or action. This slight initial misattunement can potentially spiral out to a major argument in a feedback loop manner, while other factors such as emotional engagement and internalized cultural conventions might further reinforce such a vicious circle. An example of such an interacting dyad may consist of an autistic child who, when stressed, tends to react with repetitive movements and a neurotypical one who has –via exclusively interacting with neurotypical peers– formed a rather strictly tuned set of expectations of what a typical conversational reaction might be.

Furthermore, let us imagine a human relationship, on the scale of months or years. Short scale interpersonal disturbances might lead to a cumulative *misattunement*, which can oftentimes go beyond the conscious will of the interacting parts. In other words, slight day to day misunderstandings –if not timely resolved– can potentially lead to situations difficult to overcome. Eventually, the communicative gap can potentially reach a level which is impossible to bridge.

Now imagine an (autistic) child who along their whole development, in the scale of decades, repeatedly experiences such kind of interpersonal *misattunement* situations. In such a case, a persistent social exclusion might actually exert a higher impact on the development of this person, than an initial atypicality in the generation of expectations and reactions. This is likely to prevent them from naturally developing the knowledge and skills a typical person develops, almost effortlessly, in and through the daily social interactions within a particular culture.

Indeed, with regard to larger groups of persons, this kind of *misattunement* could even take on a cultural form, spanning a scale of several generations. For instance, culturally cultivated beliefs in a given society about a specific group of people might highly impact the effectiveness of interaction between *in*- and *out*-group persons. From a Bayesian perspective, one can imagine social stigma and stereotypes as a strict set of prior beliefs. Although certain rigid prior beliefs can potentially serve as useful heuristics for quick and effective decisions, in many cases they do get detrimental to human communication, separating social groups. This kind of groups could be autistic versus neurotypical persons, but could include any dichotomy; ranging from groups with different medical diagnosis status to groups with a different ethnic or cultural background, such as immigrants in a receiving country versus the local population. Further crucial factors to the instigation and maintenance of such vicious cycles might include certain financial interests and power dynamics.

Crucially, an initial (medical) condition can lead to a cascade of other ‘comorbid’ conditions, such as depression and anxiety, not through an actual biological causal link, as often taken for granted, but through the interplay of an actual condition and social expectations in a given sociocultural context. Let us consider as an illustrative example a person, who, diagnosed as HIV positive, develops depression the years following their diagnosis. In this case, the psychiatric

condition of depression should be examined more in relation to an interpersonally aversive environment due to social stigma, than the actual medical condition.

Finally, one could speculate about such kind of interpersonal misattunement processes in evolutionary scales. As we saw above, interpersonal misattunement is possible to develop and be reinforced across various time scales, especially when additional factors serve as stabilizing factors. If such processes prove robust enough to persist across multiple generations, resulting in a consistent segregation of groups of people, one could imagine an impact even at the level of (epi-)genetics.

So far, we examined scenarios of *misattunement between persons.* However, such interpersonal segregations and disturbances might even take the form of an *environmental misattunement* for certain groups of people through a highly selective cultural and technological artefact production. Let us contemplate a highly hypothetical scenario. In a world where human's height typically exceeds the 3m, a person of an average height in our world (1.60 - 1.70m) would face multiple difficulties in everyday life; even sitting on a chair in a restaurant would constitute a real challenge. Now imagine how an autistic person might feel in a neurotypically designed cultural and technological world. At least part of their anxiety could be alleviated by reconsidering such kind of design plans. This kind of inextricably linked and multiscale processes of dynamically disturbed interaction between persons –mediated by the (cultural) environment– is what we call *dialectical misattunement* (cf. **Section 2.2**). This thesis primarily focuses on the case of autism, but analogous rationale can be applied to various other groups.

Having considered situations of *typical and atypical attunement*, between persons, but also between them and the environment, we now turn into situations where attunement fails after

having been more or less typical up to a point in life. An extended involuntary solitary confinement constitutes such a case. Typically, isolated individuals report multiple perceptual hallucinations — frequently of social nature (Grassian, 2006). Adopting a dialectical misattunement perspective, we argue that this kind of hallucinations might be a way to reduce prediction error due to a discrepancy between strong bodily expectations to (socially) interact and unexpectedly unfolding reality. Indeed, "Heidegger (1962) provides an analysis of human existence in which being-with (Mitsein) or being-with-others is part of the very structure of human existence, shaping the way that we are in the world. […] In effect, one doesn't come to have a social constitution by way of interacting with others; one is "hard-wired" to be other-oriented, and this is an existential characteristic that makes human existence what it is" (Gallagher, 2014). Even worse, such situations are largely self-enhancing, as a person who experiences extended isolation can potentially gradually develop 'social atrophy', which in turn can contribute to further isolation and so on. We could consider various relevant examples, such as homeless persons, institutionalized or otherwise socially excluded.

Taken together, the theoretical investigation of interpersonal misattunement and how it could be potentially ameliorated and actually even prevented could be relevant, not only to psychiatry, but various other fields of research and societal practice.

1.2.3. Two-person psychophysiology

Yet, however informative relevant conceptual work might be, an effort to truly go beyond the individual will remain incomplete until put to the test empirically and in real-life practice. To this end, this thesis puts forward *two-person psychophysiology* which first embeds empirical studies within the context of social interactions in order to consider the *collective* as a core unit of

analysis and second synthesizes well-established empirical practices, ranging from observational and questionnaire-based approaches to multimodal neurobehavioral recordings in order to study social phenomena across scales.

Indeed, crucial aspects of *interpersonal attunement and misattunement* might not be always graspable by so called spectatorial paradigms, which primarily trigger and monitor either *third-person* inferential or *first-person* phenomenological processes. In contrast —but also complementarily— to such accounts, *second-person* accounts emphasize the role of the real-time and reciprocal dynamics of social interactions in making sense of others: "These accounts – sometimes contrastively described as the 'second-person' approach to other minds– ask whether social cognition from an observer's point of view is really the most pervasive way of knowing other minds and suggest that social cognition may be fundamentally different when we are actively engaged with others in ongoing social interaction, i.e., when we engage in social cognition from an interactor's point of view" (Schilbach, 2016; see also Schilbach *et al.*, 2013; Schneider *et al.*, 2013). In fact, it has been suggested that social interaction in and of itself might even constitute –rather than merely contextualize or enable– social cognition (cf. participatory sense making; De Jaegher & Di Paolo, 2007; 2010). What is crucial in this debate is to make sure the empirical paradigm is not *a priori* selective with regard to certain aspects of human experience and psychopathology.

With regard to psychopathology, we introduce an integrative research line that might allow for escaping mere contrasts of traditional dichotomies such as *'patient versus neurotypical'*, moving toward examining both *intra*- and *inter*personal processes — both within and across social groups. To make this point more clear, let us examine the case of autism. Neuropsychiatric research has been traditionally based on an individualistic paradigm that typically dichotomizes

participant samples into categories according to diagnostic labels, such as *autistic* and *neurotypical* groups of persons, before studying a relevant phenomenon. These diagnostic groups are then usually tested within a specific study task and subsequently the (averaged) behavioral or neural parameters are contrasted across groups. The calculated differences (e.g. lower scores on a given task or hypoactivation of certain brain areas) are usually taken as defining attributes of a given condition, such as autism. However, this is not necessarily the case.

What has been traditionally regarded as a difficulty or even impairment of the autistic person can be reviewed as a dynamic and cumulative interpersonal misattunement across time scales. For instance, it might not be precise to reduce the empathy difficulties, autistic persons typically face, to a trait of the autistic person *per se* (cf. the double empathy problem; Milton, 2012). In this light, the ontology of a psychiatric condition could (and should) be reconsidered as truly dynamic, multiscale, relational and interactional.

As a matter of fact, this thesis introduces a novel way of facilitating a shift away from an exclusive study and diagnosis of the individual, to considering types of interacting dyads and groups: i.e., autistic, neurotypical and mixed dyads/groups, or more broadly dyads/groups of varying similarity with regard to certain aspects (e.g. autistic traits), expecting smoother interactions within the more homogenous groups or dyads (cf. **Section 2.2**; Bolis *et al.*, 2017). In case these hypotheses prove true, the definition of conditions such as autism becomes relevant to the other and generally the social context.

In a nutshell, two-person psychophysiology comprises a novel research framework that synthesizes and extends experimental and observational approaches, ranging from strictly structured and free viewing tasks to real-time social interaction and real-life aspects (cf.,

Chapter 3). This allows for a formal and multiscale interpretation of multimodal and high-resolution datasets, while preserving ecological validity. In what follows, after discussing the theoretical underpinnings in **Chapter 2**, we describe how the framework of two-person psychophysiology is operationalized across time scales in studying, first, *real-time* social interactions (**Sections 3.1; 4.1; 5.1**) and, second, *real-life* social relations (**Sections 3.2; 4.2; 5.2**).

1.3. Novelty

Before presenting the technical material in depth, we below list the cardinal contributions of this thesis, reflecting its novelty.

1. **Dialectical attunement**

   This thesis introduces the concept of *dialectical attunement*, which brings together Bayesian and dialectical perspectives for the integrative study of human becoming, emphasizing the constitutive role of social interactions. This yields a potentially unified account of the multiscale dynamics of culture and the self (**Section 2.1**; Bolis & Schilbach, 2018b).

2. **A two-person psychophysiology framework**

   In order to operationalize and validate the hypotheses about the fundamental role of interactions, this thesis complements second-person neuroscience approaches via establishing a novel empirical framework, namely *two-person psychophysiology*. This helps move beyond the individual as the unit of analysis in empirical research, by virtue of measuring and analyzing the multiscale dynamics of social interactions (**Chapter 3;** Bolis & Schilbach, 2018a).

3. **Empirical validation of second-person neuroscience**

   This thesis provides an empirical validation of second-person approaches, which claim that real-time social interaction is fundamentally different than situations of social observation (**Section 4.1)**.

4. **Dialectical misattunement**

   This thesis introduces the *dialectical misattunement hypothesis*, which reviews autism (and other conditions) as a dynamic and cumulative interpersonal mismatch, rather than a mere disorder of the individual. This yields a potentially unified account of the multiscale dynamics of psychopathology and social exclusion (**Section 2.2**; Bolis *et al*., 2016; 2017).

5. **A relational paradigm for autism research**

   In doing so, this thesis articulates a novel research line for the research of autism (and other conditions) beyond the individual, via the study of social interactions and relations of not merely (i.) *neurotypical* and (ii.) *mixed*, but also crucially (iii.) *autistic dyads and groups*, explicitly predicting smoother interaction within homogeneous dyads and groups – above and beyond the degree of individual autistic and other traits. Ultimately, this research line points toward studying psychiatric conditions, transdiagnostically, as interpersonal distance in a multidimensional feature *space*, as opposed to a currently defined *spectrum*. (**Chapter 3;** Bolis *et al*., 2016; 2017).

6. **Empirical validation of the dialectical misattunement hypothesis**

   This thesis provides an empirical validation of the dialectical misattunement hypothesis, via demonstrating that it the mismatch of autistic traits –not traits *per se*– that primarily predicts core aspects of real-life interpersonal attunement (**Section 4.2)**.

2. Theoretical work — *'When Vygotsky met Bayes'*

"The experimenters in the natural sciences imagine that they free themselves from philosophy when they ignore it, but they turn out to be slaves of the worst philosophy, which consists of a medley of fragmentary and unsystematic views."

"Everything described as a fact is already a theory."

Lev Vygotsky

This chapter presents the theoretical part of this thesis, consisting of two published papers, which introduce the dialectical attunement (**Section 2.1**) and the dialectical misattunement (**Section 2.2**) accounts respectively:

1. *'I interact therefore I am': The self as a historical product of dialectical attunement*
2. *Beyond autism: Introducing the dialectical misattunement hypothesis and a Bayesian account of intersubjectivity.*

2.1. Dialectical attunement

Original publication details (submitted 2018; accepted 2018)

Bolis, D., & Schilbach, L. (2018). 'I interact therefore I am': The self as a historical product of dialectical attunement. Topoi, 1-14.

Contribution of authors

DB did the initial conceptualization work, planned the manuscript and organized its argumentative structure. DB wrote the initial draft of the manuscript. LS revised the manuscript, provided critical feedback and overall supervision. LS provided funding.

'I interact therefore I am': The self as a historical product of dialectical attunement

Dimitris Bolis[a,b,2], Leonhard Schilbach[a-d]

Affiliations

[a] Independent Max Planck Research Group for Social Neuroscience, Max Planck Institute of Psychiatry, Kraepelinstrasse 2-10, 80804 Munich-Schwabing, Germany

[b] International Max Planck Research School for Translational Psychiatry (IMPRS-TP), Munich, Germany

[c] Department of Psychiatry, Ludwig Maximiliam Universität, Munich, Germany

[d] Outpatient and Day Clinic for Disorders of Social Interaction, Max Planck Institute of Psychiatry, Kraepelinstrasse 2-10, 80804 Munich-Schwabing, Germany

Keywords

Social interaction · Intersubjectivity · Internalization · Externalization · Predictive processing · Active inference

[2] Corresponding author: Dimitris Bolis. E-mail: dimitris_bolis@psych.mpg.de

Abstract

In this article, moving from being to becoming, we construe the 'self' as a dynamic process rather than as a static entity. To this end we draw on dialectics and Bayesian accounts of cognition. The former allows us to holistically consider the 'self' as the interplay between internalization and externalization and the latter to operationalize our suggestion formally. Internalization is considered here as the co-construction of bodily hierarchical models of the (social) world and the organism, while externalization is taken as the collective transformation of the world. We do not consider these processes as sequentially linked, but rather as a dialectic between the collective and the individual. This leads us to the suggestion of the self as a historical product of dialectical attunement across multiple time scales, from species evolution and culture to individual development and everyday learning. Subsequently, we describe concrete means for empirically testing our proposal in the form of two-person psychophysiology and multi-level analyses of intersubjectivity. Taken together, we suggest that a fine-grained analysis of social interaction might allow us to reconsider the 'self' beyond the static individual, i.e., how it emerges and manifests itself in social relations. Such an approach, we believe, could be relevant in multiple fields, from ethics and psychiatry to pedagogy and artificial intelligence.

Through others, we become ourselves.
Lev Vygotsky (1896–1936)

I see nothing other than becoming.
Heraclitus (ca. 535–475 BC)
in Nietzsche's the birth of tragedy

2.1.1. What is the'Self'?

Questioning the Question

In modern societies people tend to consider the idea of the 'self' as self-evident. Certain civilizations have even suggested that one's ultimate goal in life is to 'know thyself' (e.g., the ancient Greeks referred to it in multiple instances as 'γνῶθι σεαυτόν'). However, the question of

the ‘self’ did not exist from the beginning of the history of culture and human thought, but it arose at a certain level of historical development as a result of deep societal transformations. At different stages of historical development, this question has been addressed in different ways. For instance, Plato (429–347 BC), and before him Homer (ca. Eighth century BC), imagined the ‘self’ as an immaterial spiritual substance (i.e., the psyche or the soul). More specifically, Plato contrasted the eternal form with the ephemeral body, which he thought of as an imperfect copy of the former (Kraut, 2017). In fact, we later meet dualistic views on the ‘self’ in various religious traditions, as well as in notable thinkers, such as Plotinus (ca. 204–270) and Descartes (1596–1650). Descartes, who famously declared “I think, therefore I am” (or “I doubt, therefore I think, therefore I am”, as paraphrased by Antoine Léonard Thomas), considered mind and body as two distinct entities, which could yet influence each other.

Nowadays, mainstream science has moved away from an idealistic and dualistic view of the ‘self’. Already, Aristotle had argued that the soul could not be separated from the body (cf. Sihvola, 2008). Yet, religion, offering the concept of immortality as a solution to the problem of death has played a pivotal role in hindering this transition (Barresi & Martin, 2011), ignoring alternative solutions such as the one put forward by Epicurus (341–270 BC), who proclaimed that the problem is not death itself, but the fear of death. In fact, Epicurus and others such as Democritus adopted a monistic perspective, which can be traced into modern times with thinkers such as Pierre Gassendi (1592–1655), Baruch Spinoza (1632–1677) and Ludwig Feuerbach (1804–1872). Despite other fundamental differences, present in diverse philosophical traditions has been the idea of a lawful understanding of the world, at times emphasizing a mechanistic explanation, which largely characterizes the scientific paradigm until today. Indeed, one can draw parallels in today’s neuroscience, which is largely grounded in frameworks, which focus on

describing the underlying mechanisms of a phenomenon, e.g., distinct neurobiological mechanisms underlying consciousness.

Various roles have been considered for the 'self' and consciousness in (more) modern science and philosophy as well. For instance, John Locke (1632–1704) focused on the relations between basic physical/mental elements, emphasizing sameness: "[…] in this alone consists personal identity, i.e., the sameness of a rational being: And as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person; it is the same self now it was then; and it is by the same self with this present one that now reflects on it, that that action was done." (Locke 1694; cf. Barresi & Martin, 2011). On the other hand, David Hume (1711–1776) claimed that the 'self' is an illusion, as there "are the successive perceptions only, that constitute the mind; nor have we the most distant notion of the place, where these scenes are represented, or of the materials, of which it is compos'd" (Hume, 1739; cf. Barresi & Martin, 2011). After all, Friedrich Nietzsche (1844–1900) not only famously argued that "God is dead", but also noted that the 'self' is dead as well (cf. Barresi & Martin, 2011). So, what is the 'self'? Barresi and Martin (2011) answer that the concept of 'self' in today's literature appears divided in a number of different roles, such as 'self-image', 'self-conception', 'self-discovery', 'self-confidence' etc.

In this article, we will approach the multi-fragmented paradox of the self, through an integrative perspective, adopting a dialectical and historical perspective. In line with dialectical cultural-historical theories (cf. Vygotsky, 1930–1935/1978), we will try to motivate a shift of focus from being to becoming, along multiple temporal scales. In doing so, we will move beyond the individual in the question of the (a-)typical 'self', in both conceptual and empirical regards. More concretely, we will argue that the 'self' lies beyond the static individual, namely in the unfolding

of social relations, as a dialectic of internalization/externalization, over multiple temporal scales. Along these lines, autism and other psychiatric conditions have been recently revisited as processes of cumulative misattunement between persons, rather than mere brain disorders (Bolis *et al.*, 2016, 2017). Subsequently, we will delineate an empirical research framework for scientifically validating relevant questions, i.e., two-person psychophysiology and multi-level analyses of intersubjectivity. Finally, putting this approach into a broader context, we will discuss why challenging the concept of the self is important anyway by describing the practical implications of our approach across various fields of research and practice. Here, we will consider aspects ranging from ethics and pedagogy to psychiatry, neuroscience and artificial intelligence.

A Dialectical Perspective

To begin, we will make a case for the use of dialectics as a powerful tool for science. To this end, we will first provide a brief introduction to the method and present concrete dialectical insights for the discussion of the self. Dialectics can be thought of as an evolving school of thought, met in various historical and cultural contexts (e.g., Greek, Chinese, Indian, German dialectic; Wong, 2006; Dafermos, 2015). It asserts that phenomena cannot be meaningfully understood by reducing them into single levels of description or by assuming a metaphysical independence between levels of description. It rather states that phenomena should be studied as processes in their wholeness, inner contradiction and movement. In this light, the self cannot be understood in isolation from the body, social interaction and society (Bolis *et al*., 2017). More concretely, primarily leaning on views of Vygotsky and colleagues on the dialectical nature of human thought and development, we will try to overcome traditional dichotomies, such as

object/subject and organism/environment, by viewing them as both a result and a cause of reciprocal adjustments, or individual/society by considering the whole and the part as, albeit partially autonomous, highly interdependent levels of organization. Along these lines, the self is not to be taken as a static entity bounded by the individual, but rather as the interplay of dynamically and reciprocally interacting factors. More specifically, we will consider it as a process of circular causality among different levels of organization (**Fig. 2.1**; e.g., physical, biological, psychophysiological and sociocultural) unfolding over different time frames (e.g., evolutionary, cultural, developmental, psychophysiological and microbiological scales; Vygotsky, 1930–1935/1978; Bolis *et al.*, 2017).

In a nutshell, dialectical thought emphasizes change over sameness and becoming over being, by viewing reality as dynamic processes rather than static entities (see also process philosophy; Seibt, 2017). As Nietzsche (1844–1900) noted, citing Heraclitus (ca. 535–475 BC): he [Heraclitus] altogether denied being. [...] Louder than Anaximander, Heraclitus proclaimed: "I see nothing other than becoming. Be not deceived. It is the fault of your short-sightedness, not of the essence of things, if you believe you see land somewhere in the ocean of becoming and passing-away. You use names for things as though they rigidly, persistently endured; yet even the stream into which you step a second time is not the one you stepped into before" (Nietzsche, 1872/1999, p. 51–52). Along these lines, Nietzsche strictly denies a dichotomy of object/subject: "[the subject is but a] term for our belief in a unity underlying all the different impulses of the highest feeling of reality". There is no such unity, only "the fiction that many similar states in us are the effect of one substratum: but it is we who first created the "similarity" of these states; our adjusting them and making them similar is the fact, not their similarity, which ought rather to be denied." (Nietzsche, 1901/2017; see also Barresi & Martin, 2011). Nietzsche then goes on to

criticize an “absurd overestimation of consciousness” which had been transformed “into a unity, an entity: ‘spirit’, ‘soul’, something that feels, thinks, wills”, provocatively characterizing this as one of the “tremendous blunders” intellectual culture had created (Nietzsche 1901/2017; see also Barresi & Martin, 2011). In other words, Nietzsche here rejects the idea of an ‘artificial’ unity of consciousness (or the self). This brings us to a cardinal concept of dialectics, the ‘unity of opposites’.

Put simply, ‘unity of opposites’ defines a phenomenon by its internal oppositions: “All things come into being by conflict of opposites, and the whole flows like a stream” (Diogenis on Heraclitus, ca. Third century BC; cf. Magnus, 1970). Later Hegel (1770–1831) elegantly elaborated: “[…] every actual thing involves a coexistence of opposed elements. Consequently to know, or, in other words, to comprehend an object is equivalent to being conscious of it as a concrete unity of opposed determinations, [whereas] the old metaphysic, as we have already seen, when it studied the objects of which it sought a metaphysical knowledge, went to work by applying categories abstractly and to the exclusion of their opposites” (cited in Blunden, 2000). In brief, Hegel claimed that ideas and concepts can be only understood in historical terms, as when abstracted, they become meaningless (Grossmann, 2018). Importantly, a dialectical account does not merely focus on interpreting a harmonic development of internal contradictions, but also unveils dramatic tensions, conflicts and struggle of opposites. In fact, within dialectical thinking, such inner contradictions are the ones that drive change. Gradual change, in turn, is thought to lead to ‘crises’, which are overcome by qualitative leaps. Taken together dialectics, therefore, assume a constant evolution of phenomena, where change is periodic but not returning to the same point.

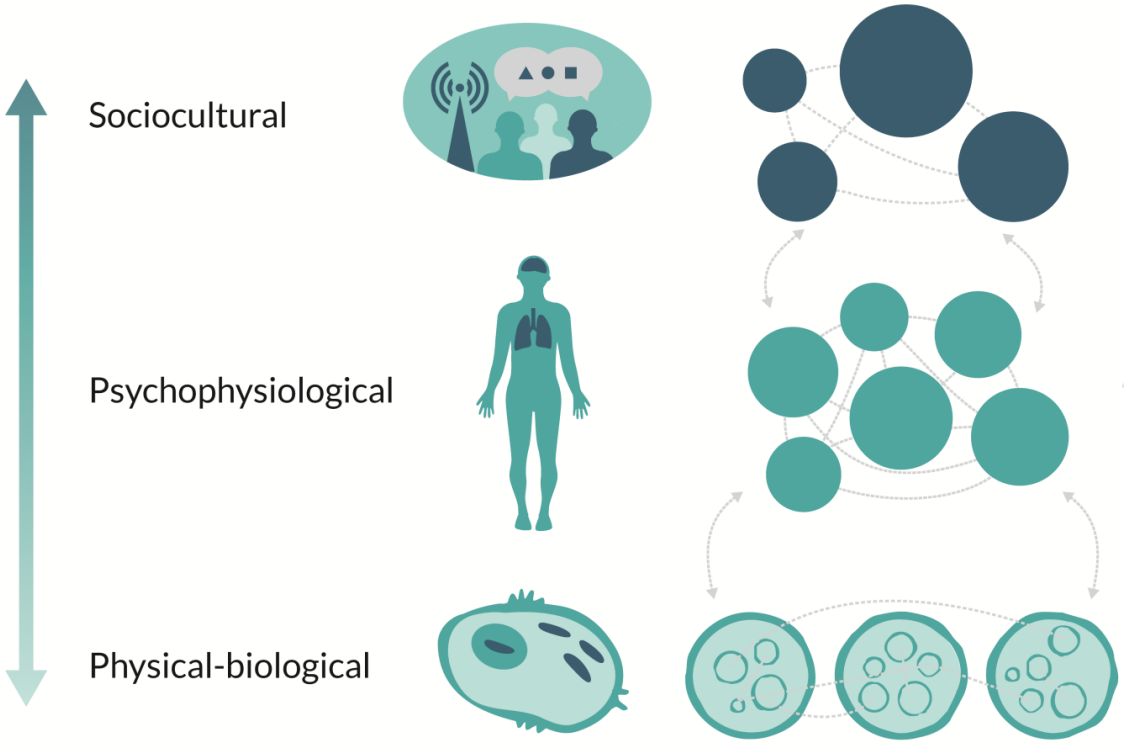

Fig. 2.1 Schematic depiction of dynamic interrelationships in the evolution of matter organization across several time scales.

With regard to the topic at hand, Hegel suggested that self-consciousness does not emerge through passive and individualistic introspection, but through dynamic and reciprocal relations with others (cf. Barresi & Martin, 2011). In fact, Karl Marx (1818–1883), whose work leaned on but also criticized Hegel's work, proclaimed: "[…] the human essence is no abstraction inherent in each single individual. In its reality it is the ensemble of the social relations" (Marx and Engels, 1888). The primacy of the social realm has later been stressed by so-called cultural-historical approaches (cf. Roth 2016). A prime example of this can be found in the work of Lev Vygotsky (1896–1934), who directly applied dialectical thinking to developmental psychology and proclaimed that "through others we become ourselves". He further suggested that all 'higher'

mental processes within an individual result from an internationalization of prior social interactions between people. But dialectical thinking, as described here, should not be exclusively attributed to Western philosophy. For instance, according to African Ubuntu “a person is a person through other persons” (Birhane, 2017). We also meet forms of dialectical thinking in Buddhism and Taoism (cf. Grossmann, 2018). Taken together, in the formation of the self, the social can be assumed to dialectically precede the individual.

In this line of thought, we suggest that interpersonal statistical regularities shape multiscale hierarchical models on an individual level and vice versa. For instance, at the level of perceptual awareness and everyday learning (time scale of seconds to hours), others play an important role in shaping subjective feelings and decision-making. Let us imagine an illustrative scenario (Bolis & Schilbach, 2017b): a person, in the process of deciding what is the most appropriate clothing for tonight’s walk, checks current weather out at the balcony. She feels a cold breeze, which initially makes her think that a warm coat might be a good idea; yet a glance down the road makes her change opinion, as all people outside this day are lightly dressed. Such kind of decisions, especially when reinforced by persistent cultural norms (time scale of weeks to years), is possible to form even more stable personal habits. For instance, people in ancient Greece were accustomed to exercising without clothes. In modern societies, despite objective conditions that might call for such a habit sometimes (e.g., warm weather) such a behavior would be considered uncomfortable by most people. Here we see how a socially constructed statistical regularity is internalized at the level of the individual – at such an extent so that its violation directly evokes certain subjective feelings.

Across longer time scales, the cumulative internalization of such interpersonal regularities directly shapes who we are becoming, literally changing our bodies and brains. Both ‘higher

level' mental functions and 'automatic' processes can be thought of as emerging due to and through social interactions across the life span. Let us examine a simple example inspired by Vygotsky (1930–1935/1978, p. 103 of the Greek translation): a child in an effort to maintain interoceptive balance unsuccessfully tries to reach for food with the index finger extended. The caregiver, who observes the effort, helps with bringing the food toward the hand of the child. After a number of repetitions, this kind of interpersonal process, and the statistical regularities associated with it, is internalized by the child in such a way that the extension of the index finger eventually represents a call for attention to a pointed object. Intriguingly, it is not only higher symbolic functions that are culturally shaped, but also more 'fundamental' ones, such as eating and drinking. For instance, while babies eat and drink when they feel the need for it, adults regularly do so not for covering direct survival needs but rather social ones (e.g., eating as a part of a break from work or drinking alcohol when socializing). Along these lines even interoceptive control can be thought of having social origins, being developed in this way already from infancy onwards (Ciaunica & Fotopoulou, 2017; Fotopoulou & Tsakiris, 2017). To probe this further, interaction processes can be thought of as ontologically primary to entities on an ultimately basic level. In this line of thinking, entities actually emerge through interactions (intra-actions for emphasis; Barad 2003) and not *vice versa*, from within their relationship and not outside. In this light, we view the multifaceted construction of the self as an active process of culturally mediated internalization of social interactions along multiple time scales.

Here, it is crucial to note that internalization plays an important albeit partial role in the formation of the self. It is the dialectic between internalization and externalization that provides a more complete picture of the co-construction of individual and social reality. Internalization can be thought of as the active reconstruction and synthesis of incoming information and past

experiences, while externalization can be thought of as the tool-mediated translation of inner processes into collectively transforming the world, including others. The dialectic between internalization and externalization becomes apparent when examining the simple example of holding and manipulating an object (Leontyev, 1975/1983; Stetsenko & Arievitch, 2004). In this very moment a person transforms not only the world but also herself, as in her effort to act on the environment, she embodies its structure and dynamics. Tools are not to be confused only with conventional material objects. The term here is used to also broadly encompass 'intellectual objects' in the service of communication, such as language and art (cf. Vygotsky, 1930–1935/1978; Dafermos, 2002). In other words, humans change themselves through changing the environment in a socioculturally mediated procedure of mutual adjustment. More broadly, evolution (or development[3]) of species, societies, persons and concepts should not be viewed as an one-way adjustment, but rather as a dialectical, namely a dynamic, reciprocal and cumulative process (cf. Levins & Lewontin, 1985). We will come back to this crucial insight later, but will now review these cardinal concepts through a Bayesian lens, which will allow us to operationalize our suggestions formally.

A Bayesian Perspective

The main premises of the "Bayesian brain hypothesis" rest on the idea that the brain represents information accessed via the sensory organs in the form of probability densities, as opposed to single numbers, which are continuously updated, as if following a specific set of mathematical

---

[3] Hereafter we will use the term development to broadly imply change across various scales, emphasizing our historical standpoint.

formulas based on Bayes theorem (cf. Bolis *et al.*, 2017). Interestingly, such a perspective brings together under a common umbrella diverse putative cognitive processes of major importance, such as optimal information integration both in time and space, optimal multimodal cue integration, as well as flexible information manipulation without the need to commit to particular decisions at an early stage of processing (Knill & Pouget, 2004). In other words, through a Bayesian lens one can view the brain as an organ which calculates and maintains probabilities about events in the world or about the organism itself, via a combination of already gained experience and newly sensed information. Importantly, the more confidence (i.e., precision) is placed on the validity of experience (i.e., prior beliefs) the less beliefs are updated based on new incoming information (i.e., evidence). Notably, a Bayesian 'belief' should not be confused with an everyday meaning of the word belief which might be taken to refer to a conscious representation. On the contrary, a Bayesian belief can be thought of as a dynamic state, either conscious (e.g., determination not to eat meat) or unconscious (e.g., glucose levels).

A concrete and prominent implementation of the Bayesian brain hypothesis can be found in predictive processing (i.e., predictive coding and active inference; Friston, 2010; 2013; Clark, 2013). Within this framework a biological system is essentially viewed as a prediction machine and action generator, which actively tries to align reality with internalized models of reality, as precisely as possible. As noted above, reality embraces both the world and the organism itself. According to such a perspective, the brain's ultimate goal is the long-term minimization of free energy, by calculating (under certain simplifying assumptions) prediction errors, i.e., the discrepancy between incoming information and generated predictions, based on prior experience. Importantly, this is thought to be accomplished via two main avenues, namely either via updating the (Bayesian) beliefs one holds for aligning them with the environment (i.e., predictive coding;

cf. internalization), or through action, which can help to experience the environment in accordance with prior beliefs (i.e., active inference; cf. externalization). Put simply, to survive, an organism obeys the following straightforward rule: adjust yourself to reality or change the reality itself (cf. Friston *et al.*, 2010).

In this framework, the updating of beliefs is accomplished across various hierarchical levels at the same time. More concretely, two processes run in parallel: prediction errors ascend the hierarchy reconfiguring the organism for optimizing predictions, while in parallel predictions descend the hierarchy explaining away prediction errors. The hierarchical organization of this scheme is of immense importance, as it allows for the consideration of multiple levels of increasing abstraction. For instance, social relations along development are not merely stored and represented as concrete memories, but are perhaps more crucially, internalized at higher levels of the hierarchy as generalized cultural norms. The latter can, in turn, be utilized to guide behavior across a multitude of contexts.

As noted above, a process of belief updating should be always thought of in relation to action. Importantly, such a dialectic of internalization and externalization can take either 'adaptive' or 'maladaptive' forms along various time scales, leading to a cascade of interpersonal (mis-)attunement (Bolis *et al*., 2017). To give a simple example, abusive interactions with care-givers in early life could influence the way an individual forms relations later, which may help to explain personal tendencies and so-called personalities, but also symptoms across different psychiatric and psychological conditions. In other words, growing up in an interpersonally aversive environment may lead to expectations about how social interactions unfold, which will modulate how future interactions actually play out. From our standpoint, such an example illustrates how the Bayesian perspective may be able to capture and express the inextricable

linkage of social and individual reality. Seen through a Bayesian and dialectical lens at the same time, we can, therefore, view the 'self' as a non-linear dynamic process, rather than as a static and unified entity.

Notably, predictive coding and active inference can be thought of as a dialectical framework in and of itself. Perception and action become two dialectical facets of the same process, i.e., the minimization of prediction error. Current internal (e.g., perceptual) states inform future actions, while informed interaction with the environment (including others) greatly modulates internal states. Furthermore, the interrelation between the environment and the 'self' is controlled by the synthesis of an organism's current state and incoming information, either through updating current beliefs or the environment itself. In these terms the 'self' can be considered as the dialectic between predictive coding (cf. internalization) and active inference (cf. externalization) processes (Bolis *et al.*, 2017). Taken together, multilevel computational frameworks grounded in predictive processing (cf. Bolis & Schilbach, 2017b; Ramstead *et al*., 2017) can, therefore, serve as a formal bridge between philosophical arguments and neuropsychological evidence for revisiting the 'self' as a historical product of dialectical attunement.

2.1.2. The Dynamic Self in Action

The Dialectic of Internalization/Externalization: Insights from Evolutionary & Developmental Psychology, Neuroscience & Psychiatry

In the following section, we selectively review results and insights from different disciplines in order to add empirical findings to the argument that the self can be regarded as a (historical or developmental) dialectic of internalization/ externalization over multiple scales.

Across an evolutionary scale, the change to upright position comprises perhaps one of the most important qualitative leaps. In fact, bipedal walking has been crucial to the evolution of the self for various reasons. Perhaps, most importantly, walking on two feet allowed the development and use of sophisticated tools. The latter revolutionized the way humans adapt to the environment, allowing them to actively and dialectically transform the world they inhabit according to their needs. That is, it is not only humans who change the environment, but the environment in turn changes them in face of their impact on it (cf. Levins & Lewontin, 1985). In brief, contrary to a perhaps common belief, humans (and other organisms) do not evolve via passive adaptation, but they fundamentally change themselves via socioculturally mediated transformations of the environment. However, having said that, this development has not come without compromises.

It has been hypothesised that bipedal walking has imposed certain constraints on the birth canal, which does not allow the birth of a fetus much older (and thus bigger) than 9 months. Additionally, according to the 'metabolic crossover hypothesis' (Dunsworth *et al.* 2012; Ellison, 2001) the mother may not be able to support an older and more energetically demanding fetus. Consequently, while apes and other animals quickly master basic skills that grant them relatively early independency after birth, human infants are born unable to survive on their own. Indeed, the brain size of newly born infants is only a quarter of its fully developed size. This means that major development occurs after birth in direct interaction with the environment and others: "Maybe human newborns are adapted to soaking up all this cultural stuff and maybe being born earlier lets you do this […] Maybe being born earlier is better if you're a cultural animal" (Karen Rosenberg on Adolf Portman; cited in Wong, 2012). Such a compromise between early

independency and optimal development might actually, in and of itself, define the timing of birth.

Another major evolutionary leap with regard to human cognition is the change from individual to shared intentionality (Tomasello, 1999; 2014; Tomasello *et al*., 2005; Tomasello & Carpenter, 2007; Vygotsky 1930–1935/1978), which can be broken down to more intermediate leaps (e.g., from individual to joint and from joint to collective intentionality; Tomasello, 2014). The question here is: How did we go from relatively competitive great ape societies to (possibly) cooperative human cultures? It might have been a huge leap if there had not been an intermediate link between our common ancestor and humans. The needs for cooperation (e.g., for foraging) in the early human societies may have led to the transformation of individual to joint intentionality, involving two (or a small number of) individuals (Tomasello, 2014). According to this hypothesis, this development has allowed for the coordination of roles and perspectives toward joint objectives, resulting in new forms of perspectival and symbolic representations, socially recursive inference and self-monitoring (regulating one's own actions from the perspective of a cooperative partner). The practical need for coordination might have actually prompted the development of bodily structures, which subsequently supported more abstracted cognitive functions beyond the 'here and now'. One tempting line of thought here would be to consider human body (e.g., eye and face) and brain evolution as reciprocally driven in the context of collaborative social interaction (cf. Dobson, 2012; Kobayashi & Kohshima, 2001; Powell *et al*., 2010). From a Bayesian perspective, ascending the hierarchy of a neural network, information gets more and more abstracted (e.g., from dealing with the probability of an event, to dealing with volatility, volatility of volatility and so forth; cf. Mathys *et al*., 2011). Taken together, we hypothesize that such a kind of evolution, which have allowed for abstracting beyond the

concreteness of real-time social interactions, might have been toward the direction of extended bodily hierarchies.

Similarly to development at the scale of phylogeny, development at the scale of ontogeny can be also thought of as unfolding in socioculturally mediated interaction with the environment and others, undergoing a series of qualitative leaps along the lifespan (e.g., from individual to collective intentionality; Tomasello & Rakoczy, 2003). More concretely, the acquisition of language, which can be considered as a particularly transformative leap for social cognition and interaction, is thought to emerge out of various pre-speech communicative acts (cf. Bruner, 1974). An initial basic form of dyadic interaction (between the infant and the caregiver) could serve as the substratum for the development of joint attention, as well as more complex forms of interaction. For instance, dyadic (face-to-face) and triadic (including an object) interactions have found to be developmentally linked (Striano & Rochat, 1999). Furthermore, joint attention, which is observed before fully developed social-cognitive awareness (Brooks & Meltzoff, 2005), can predict future linguistic ability (Morales *et al*., 2000; Mundy *et al*., 2007). Additionally, maternal sensitivity (Hobson *et al*., 2004) and synchronicity (Carpenter *et al*., 1998) have found to correlate with infants' propensity to engage in social interactions and language development respectively.

Also in so-called psychiatric disorders, here thought of as disorders of social interaction or cases of so-called atypical social interaction, we find an interrelation between the manifestation of the organic condition and interpersonal difficulties (Bolis *et al*., 2017; Schilbach 2016; Vygotsky 1930–1935/1978). When it comes to autism, synchronicity in earlier play interactions between the child and the caregiver was found to correlate with the development of subsequent communicative forms, such as joint attention and language (Siller & Sigman 2002). In fact, it has

been suggested that autism can be viewed not as a mere brain disorder, but rather as an evolving interpersonal misattunement encompassing various levels of description (Bolis *et al*., 2017). An attunement between the child and the caregiver along development is crucial in language acquisition. Yet, even when an autistic individual becomes able to talk, in most of the cases they achieve a propositional attunement (knowing that), as opposed to a pragmatic attunement (knowing how), a fact which largely prevents an intuitive participation in interactions with others. This alone, we suggest, might have direct implications in the formation of the self in autism due to the crucial dialectical nature of language.

Our discussion on tool mediated evolution holds also for individual development: language can be viewed as a communicative tool used for transforming the (social) world, but also the self itself (Vygotsky, 1934/1962). This dialectical nature of language becomes evident when examining its dual role, in speech (interpersonal) and thought (intrapersonal), which should be thought in unity, rather than in external (even tight) relation (Vygotsky, 1934/1962). In other words, contrary to a common assumption that speech is merely an enacted thought, speech and thought unfold together, inextricably entangled. Indeed, recent evidence demonstrates neural coupling during production and comprehension of real-life speech (Silbert *et al*., 2014). Importantly, the interpersonal aspect of language should be still thought of as temporally and conceptually preceding the intrapersonal one. That is, in contrast to a Piagetian perspective, we adhere to the Vygotskian idea that it is social interaction that drives development and not *vice versa*.

In sum, basic forms of interpersonal sensorimotor contingencies gradually evolve into more complex forms of interactions, such as joint attention and multi-person interactions. This kind of initial social interactions might be exactly what (reciprocally) drives development of social

cognition for dealing with beyond 'the now and here' (cf. Theory of Mind; Baron-Cohen, 1991; Tomasello, 1995). At the neural level, it has been suggested that joint attention might be the outcome of two interacting systems, namely the posterior and the anterior attention system (Mundy & Newell, 2007). The posterior system, which is relatively involuntary and common to many primates, begins to develop during first months of life and can be, simply speaking, thought of as serving for an understanding of "where others' eyes go, their behaviour follows" (Jellema *et al*., 2000; Mundy & Newell, 2007). The anterior system, which is considered volitional and goal-directed, develops later and can be, along similar lines, thought of as serving an understanding of "where my eye's go, my behaviours follows" (Mundy & Newell, 2007). We take this as suggestive of a claim that the 'self' develops tightly connected to the understanding of the 'other' and that in fact the latter precedes.

It might actually be the case that it is exactly in our effort to understand others that we develop an understanding of ourselves. Here, three tangled modeling loops are considered: (i) the inner loop, dealing with the prediction of internal bodily processes (cf. interoception), (ii) the perception–action loop, which involves the anticipation of the consequences of one's actions on the world and (iii) the self-other loop, which deals with modelling other minds (Timmermans *et al*., 2012). Exactly the latter loop, through social interactions, might be what ontogenetically forge sophisticated bodily structures that are later deployed for reflective social cognition (e.g., Theory of Mind; Frith & Frith, 2012; Schilbach *et al*., 2010; 2013), via neural reuse (Anderson, 2010). There is empirical evidence suggesting that unconstrained cognition, emotional processing and social cognition might all share common neural networks in the dorsomedial prefrontal cortex and in the precuneus (Schilbach *et al*., 2012). Interestingly, the latter brain networks partially comprise the Default Mode Network, which is putatively activated more when

a person does not directly focus on the outer world. Such a neural overlap between 'social cognition' and 'introspection' can be taken to suggest that not only thinking about others (either implicitly or explicitly), but even thinking about ourselves is driven by social interactions.

Taken together, we construe the self as a historical process of dialectical attunement unfolding over various time scales (**Fig. 2.2**). More concretely, we view two cardinal groups of processes dialectically interconnected, namely internalization and externalization. These processes are thought of unfolding along different time scales, e.g., (i) in the time frame of evolution, involving genetic and environmental adaptations, (ii) across generations, as cultural practices, or (iii) during individual development, including bodily and world reconfigurations, such as perception, action and learning. Put simply we view both low- and high-level attunement. Low-level attunement emerges during collective behaviour, when people are coupled together or when they coordinate (cf. De Jaegher & Di Paolo, 2007). However, while people interact, and thus act and perceive each other, they mutually co-construct internal models across multiple levels of bodily hierarchies. As we saw before, the construction of such hierarchies allows for consideration of increasingly higher levels of abstraction and thus temporal scales. That means that people in social interactions co-construct each other not only in the 'here and now', but also beyond, via co-configuring higher-level abstracted beliefs and patterns of action, on hand in future instances across a variety of interactive contexts or privately (cf. Theory of Mind; **Fig. 2.2**). Simply speaking, poetry (from the Greek "poiesis", literally meaning "making") can be thought of as an active externalization of internalized social interactions.

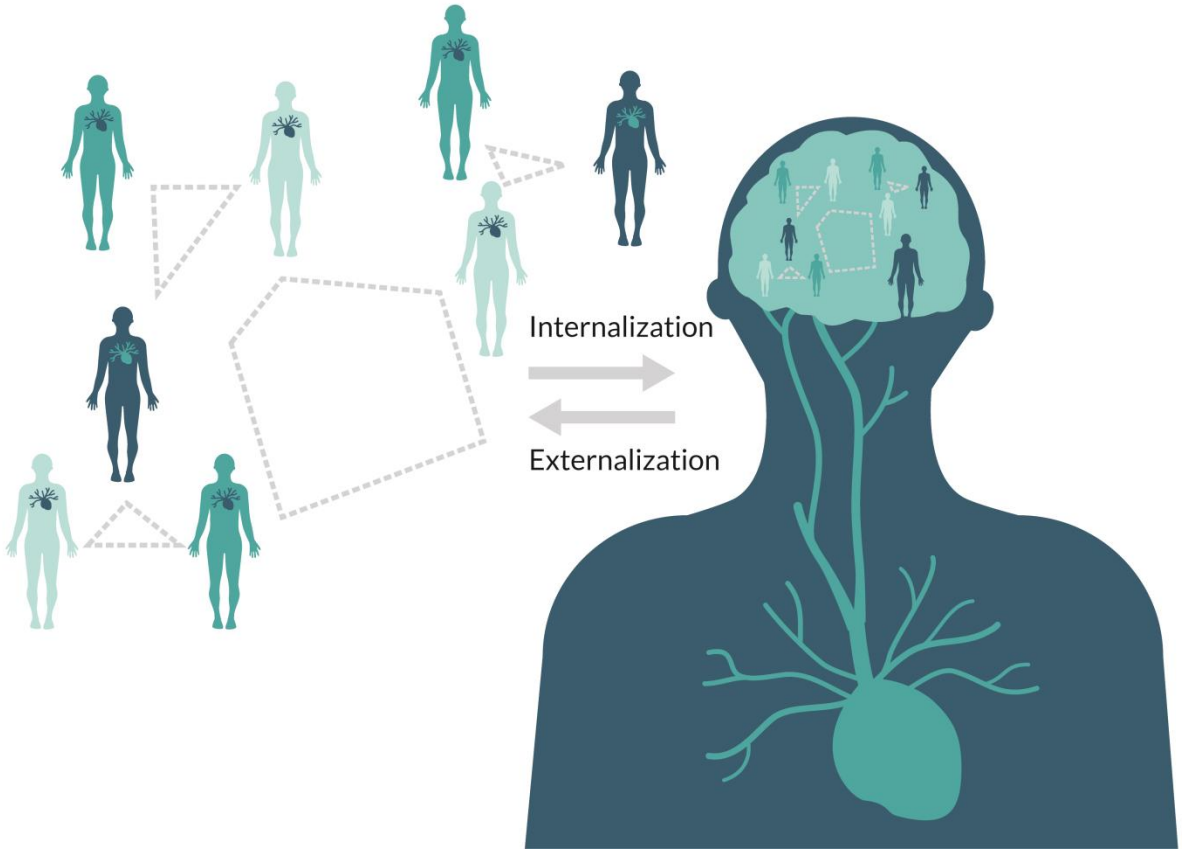

Fig. 2.2. Dialectical attunement: Environmental structure (cf. social relations) is transformed within an individual via internalization processes (cf. predictive coding; rightward arrow). Internalized structures serve for co-regulating the external (social) world via externalization processes (cf. collective activity; leftward arrow). Internalization and externalization processes are thought of as unfolding dialectically, that is in a dynamic, reciprocal and cumulative interrelation. Please note, here schematic focus is put on the brain only for convenience; in reality the body participates in the dialectic of internalization/externalization as a whole.

Internalization is the set of processes via which the structure of the environment (e.g., social relations) is actively transformed and implemented within an individual. From a Bayesian perspective, internalization entails the creation and maintenance of dynamic hierarchical models of the world in an effort to effectively predict future changes and act accordingly. We consider internalization as being accomplished across various time scales, from genetic information

encoding and cultural adaptation, to bodily reconfiguration across development and real-time perception. For instance, in the evolutionary scale, the human visual system is attuned to the peak of the solar radiation spectrum that reaches the surface of the earth. In other words, human species has bodily internalized the environment in terms of electromagnetic conditions. Interestingly, similar attunement to environmental condition is also observed along developmental scales. For instance, experiments have demonstrated that extreme exposure to a restricted range of visual stimuli (e.g., exclusively vertical visual orientation), early in development, modifies the morphology of neurons in visual cortex accordingly (e.g., Tieman & Hirsch, 1982). Furthermore, with regard to shorter time scales, perception and action can be seen as real-time bodily attunement to the environment. Finally, undeniably people are also culturally attuned in multiple aspects. For instance, what is considered beautiful or delicious seems to be different across sociocultural contexts, both across time and space.

In fact, humans used cultural models for describing, predicting and manipulating the environment already in the cradles of civilization. For instance, ancient societies have construed natural phenomena, such as weather or earthquakes, as behavioural expressions of personified deities. At first sight, this might appear as a rather naive approach. However, we consider this as an ingenious tactic that might have allowed pre-scientific communities to recruit powerful cognitive capacities, originally developed for dealing with the undoubtedly complex social realm. Any level of abstraction can be considered as a model of the world. To come back to the example of language, a word can be thought of as a sociocultural model in and of itself, which of course presupposes the evolution of both the necessary biological apparatus across evolution and an interpersonal attunement across development. For instance, the word ‘animal’ or ‘wave’ practically captures and summarizes higher level similarities being met in a plethora of diverse

natural processes. Here, we should stress that we do not consider the construction of internal models as a passive accumulation of representations.

The construction of internal models allows not only for the prediction of the world, but also the (socioculturally) transformation of it for meeting survival needs, through collective externalization. In other words, dialectical attunement does not merely imply a single-sided adjustment of the individual into the environment, but also transforming thereof across multiple scales: from cooking food, building shelters and developing technology, to transforming social structures and domesticating animal species. The activity of an individual in everyday life is decisively modulated by evolutionary, cultural and developmental factors. For instance, the use of a tool is defined by human anatomy, accumulated collective knowledge and individual learning. As discussed above, though, a change of the environment inherently entails a reconfiguration of the self as well. Externalization directly impacts on internalized models (cf. the interplay between active inference and predictive coding), as well as indirectly via the feedback of a transformed world. For example, learning to use a tool is fundamentally different when it is enacted rather than being merely theoretical, even though in both situations an internal model is developed. Additionally, both mechanical and conceptual tools (see the example of ‘wave’ from above) have helped the construction of modern technology, which in turn continuously modifies humans in multiple aspects and scales (from everyday behaviour to cultural habits and genetics in the long run). Crucially, when it comes to humans, transforming the world is fundamentally social, both with regard to our impact on others and the environment: the former is inherently social, while the latter becomes such via the mediation of sociocultural tools. In sum, we view the self exactly as the dialectic of the abovementioned internalization and externalization processes.

We will come back to this point and its scientific and societal implications during our concluding remarks, after first describing how our hypotheses could be put to the test scientifically. To this end, we will describe experimental and data analytic means for studying the dialectic of internalization and externalization in real-time social interactions and beyond.

Two-Person Psychophysiology & Multi-level Accounts of Intersubjectivity

Due to conceptual and methodological constraints, research has largely focused on either intrapersonal (e.g., neurobiological and psychological), or interpersonal (e.g., socio-cultural) processes. Here we emphasize the importance of studying intrapersonal and interpersonal processes in their inherent interrelation, as they unfold during social interactions. In what follows, we describe an experimental framework, namely two-person psychophysiology and an analysis scheme, namely multi-level analysis of intersubjectivity that could help us do so.

Two-person psychophysiology appears as a promising avenue for empirical research, which while offering great experimental control, also preserves adequate degrees of ecological validity (Bolis & Schilbach 2017; 2018a). Traditionally, psychophysiology has enabled the empirical investigation of the relation between physiological and psychological processes (e.g., through physiological monitoring and introspection), offering important insights about individual mechanisms. However informative this kind of approach may have been, the concept of the (a-)typical 'self' will remain largely misconstrued until dynamic interpersonal processes are systematically considered, as social cognition might be fundamentally different when we are in interaction with others rather than merely observing them (Schilbach *et al*., 2013). It has been argued that the most important experience of the other comes from face-to-face situations; that this is the archetypic situation of social interaction, while all other situations are products of it

(Berger & Luckmann, 1967). It is exactly in this kind of situation that the 'here and now' of each other's subjectivity come together and possibly form an inextricable intersubjective unity (Berger & Luckmann 1967; Bolis & Schilbach, 2018a; De Jaegher & Di Paolo 2007).

Building upon empirical frameworks of interpersonal research (e.g., Barišic *et al*., 2013; Dumas *et al*., 2010; Froese *et al*., 2015; Koike *et al*., 2016; Liu *et al*., 2016; Montague *et al*., 2001; Schilbach *et al*., 2006), two-person psychophysiology crucially allows for the empirical investigation and systematic manipulation of face-to-face social interaction, across various modalities and temporal scales. In such a framework (Bolis & Schilbach, 2018a), participants sit opposite each other, working on tasks either individually or collectively, while being able to interact, either in real-time or offline, through a micro-camera communication system. Such a two-person framework allows for systematic control and monitoring of processes that live in different levels of description, from (epi-)genetics and culture to interpersonal behaviour and psychophysiology. In fact, via controlling the synchronicity of social interaction and composition of dyads, cardinal aspects of the self can be put into scientific test: Emerging contextual and interpersonal differences in social interactions might prove equally, or even more important than individual traits in defining the becoming of the (a-)typical self (Bolis *et al*., 2017).

Interpersonal frameworks for empirical research might be an important tool for moving beyond the individual as the unit of analysis, yet not sufficient on their own. Conceptual and experimental practices should be developed hand-in-hand with methods of analysis (e.g., Abney *et al*., 2014; Bahrami *et al*., 2010; Bolis & Schilbach, 2017; Dumas *et al*., 2014; Friston & Frith, 2015; Froese *et al*., 2015; Fusaroli & Tylén, 2016; Konvalinka & Roepstorff, 2012; Schilbach *et al*., 2013; Sevgi *et al*,. 2016; Zapata-Fonseca *et al*., 2016). Here, we suggest a shift from an exclusive focus on the (Bayesian) brain in isolation, toward a multilevel understanding of

intersubjectivity and psychopathology. In this framework of analysis, principled accounts of brain function (e.g., predictive processing) are employed for describing crucial neurobiological mechanisms, while being connected to real-life phenomena, which by definition live in an interpersonal space. More concretely, grounded in established models (e.g., Bolis *et al*., 2015; Daunizeau *et al*., 2010; Mathys *et al*., 2011), a two-level modelling scheme could be used for capturing both individual processes (Bayesian level) and collective behaviour (meta-Bayesian level). Put simply, in this scheme intrasubjective parameters will be deployed for capturing individual mechanisms (e.g., neuromodulation), while intersubjective ones to describe emergent processes on the collective level (e.g., interpersonal coupling). Collective parameters refer to sociocultural tools, such as artefacts, communication mediating factors, and generally any co-constructed and commonly held convention. For instance, the efficacy of a communication channel might strongly modulate interpersonal coupling in social interaction (Bolis & Schilbach, 2018a).

Such an intersubjective scheme could be exploited for considering emergent phenomena on higher levels of description, such as for instance questions about the autonomy of a dyad or a group of people. To give a more specific example, in the context of collective externalization a non-linear model might explain observed behaviour optimally, thus, providing evidence that the group is different than the sum of individuals. Inversely, this framework could address questions about how collective processes, in turn, shape individual reality. For instance, one could differentially study the potentially distinct impact that a competitive or individualistic versus a collaborative structure might exert upon an individual (Bolis *et al*., 2017). Collective activity and societal structure are thought of being capable in shaping individual levels (from neurobiology to phenomenology) via internalizing mechanisms. In other words, it is not only lower-level

mechanisms that result in emergent collective ones, but internal processes are treated, here, as dynamically internalized interpersonal processes.

Notably, a meta-Bayesian framework can consider observable activity in any level of description, such as neural activity, motor responses or collective behaviour. With regard to social interactions, an interesting avenue for future research might involve studying whether interpersonal coordination on the behavioural level might actually, serve as a prior and modulate, or even relax, the need for inferences about the hidden causes of social behaviour. Furthermore, at a neurobiological level, we hypothesize that activity of different neuromodulators could be related to a subject's ability of tracking different levels of interpersonal regularities. In short, a Bayesian account of intersubjectivity intends to offer a principled and quantitative description of the dialectic between internalizing and externalizing processes across different levels of description, as discussed above.

The Dialectical Self: Scientific and Societal Relevance

Our approach shares common ground and most importantly brings together under a dialectical umbrella two seemingly disparate perspectives, i.e., interactionist-enactivist (e.g., De Jaegher & Di Paolo, 2007; Maturana & Varela, 1980) and computational-Bayesian accounts of cognition (e.g., Clark, 2013; Friston, 2013). Enactivist accounts have constructively put their focus on the fundamental role of interaction and coupling with the environment, including others. Bayesian accounts of cognition have provided important computational tools for describing individual cognition, mainly through hierarchical models. Our dialectical suggestion, on one hand emphasizes the primacy of (social) interactions. More concretely, it states that for a

comprehensive understanding of the (a-)typical self, we will need to move beyond the individual, to the historical unfolding of (social) interactions over multiple scales. On the other hand, our approach extends Bayesian accounts of cognition by situating them in the context of real-time social interaction and providing a description of internalization and collective externalization processes beyond the individual. More precisely, it connects internalization to predictive coding and collective externalization to active inference. By doing so, it describes perception, learning and collective action as a unified process that allows for aligning personal (psycho-physiological) and interpersonal (coupling and synchrony) states with environmental (nature and others) conditions. Taken together, via integrating levels of description and time scales such an approach provides a unifying and principled way for studying the self beyond the individual.

In this article we have described the self as the dialectic of internalization and externalization and more concretely as a historical product of dialectical attunement over various temporal scales (see **Fig. 2.2**). According to this view, low-level attunement is achieved largely automatically (beyond awareness) during embodied interactions, via mechanisms of collective externalization. High-level attunement is achieved through mechanisms of internalization. For instance, low-level attunement captures human action as an emergent collective phenomenon (cf. interpersonal bodily coupling, coordination and synergy) in the 'here and now'. High-level attunement captures human mind as an active environmental reflection. In a cultural frame, this takes the form of internalized values and conventions in a society, generalized across multiple temporal and contextual frames. In sum, low- and high-level attunement is dynamically and cumulatively interrelated, via internalization and collective externalization processes, forming the dialectical self.

Yet still one might wonder why even question the question of the self. We believe that any thesis on the self is inherently implicated in numerous fields of science and the society. A dialectical perspective, as the one described here, points toward specific directions that acknowledge the primacy of the social, without neglecting the importance of the individual in their interrelation, co-construction and tension. Additionally, it points toward the necessity of adopting an empirical and principled approach to studying the self. To this end, formal approaches of predictive processing and dynamical systems appear as most promising. Approaching the formation of the self under the unifying umbrella of the dialectic of internalization/externalization might allow formal integration and re-description of seemingly disparate mechanisms across different scales. Yet the implications of such a dialectical approach reach further than the realm of scientific research.

In pedagogy, this is translated into an educational system that would promote collective problem solving as compared to mainstream competitive individual tests. Put simply, taking such an approach seriously, it would make no sense to isolate inherently limited individual cognitive capacity and reward merely the most relevant to a given task. On the contrary, promoting collective problem solving and decision making via active participation and interaction would enhance both cognitive and motivational aspects, yielding superior pedagogical but also practical achievements. In psychiatry, one would not be merely focused on diagnosing and 'fixing' individual impairments, but also tuning interpersonal communication and enhancing social inclusion (**Fig. 2.3**; Bolis *et al*., 2017). Within a clinical context, such an approach would suggest the monitoring of not only individual progress, but also interpersonal coupling between a 'therapist' and the 'individual', as well as between multiple persons during group therapy. In fact, not every therapist might be optimally suited for every patient and therefore matching of

therapist and patient might need to be assessed in order to predict whether therapy will eventually work. Within a societal context, ‘tuning’ will not target only the individual with a psychiatric condition, but also her social environment. For instance, anti-stigma and informational campaigns will target tuning of social expectations of others as well, effectively resulting in a reciprocal amelioration of existing interpersonal misattunement. Such developments might help bring a redefinition of what a psychiatric disorder is, situating it back into the social realm within which it emerges.

In the field of ethics and law, seriously assimilating the idea that the self goes beyond the static individual, a juridical system would not only focus on individual intentionality and responsibility, but also take into account collective factors and societal structure. Along similar lines, confronting social problems such as racism will not merely address educating individuals, but also dealing with social structures, which potentially instigate and maintain such patterns of behavior. Finally, such a perspective would suggest developing artificial intelligence and robotics, not via static pre-configuration, but via allowing interaction for co-constructing and internalizing knowledge. This should be expected to yield not only more robust artificial systems, but insightful conclusions on cardinal questions about human cognition as well.

More concretely, in line with cultural historical and enactivist perspectives, we suggest that the role of social interaction and active participation in the co-construction of a culturally shaped self should be taken more seriously, in both research and social practice, as paraphrasing Descartes: ‘we interact, therefore I become’, or put simply ‘I interact, therefore I am’.

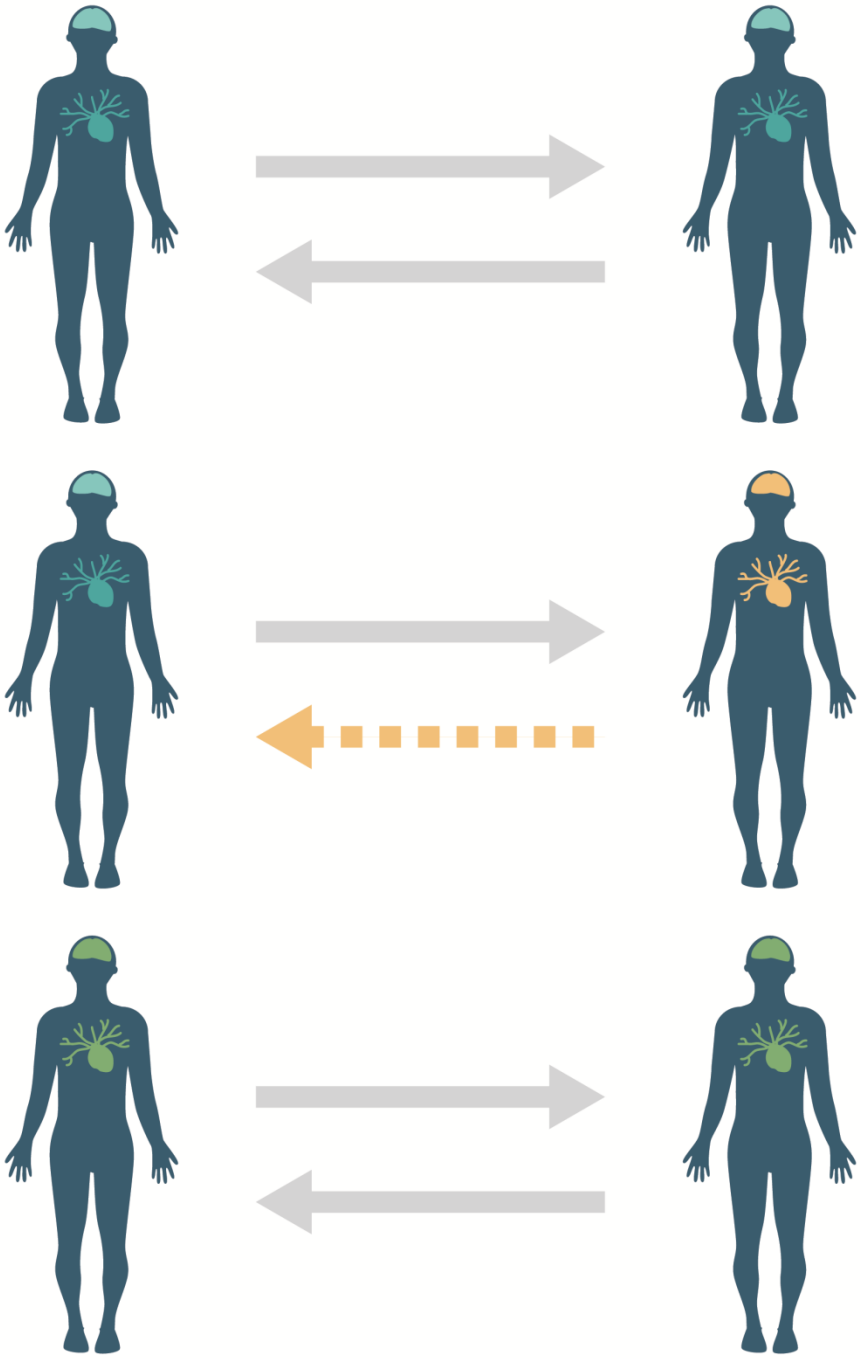

Fig. 2.3. Dialectical (mis-)attunement and interpersonal re-tuning: (top) a homogeneous dyad interacting 'smoothly', (middle) a heterogeneous dyad interacting less effectively, (bottom) retuned interaction via not only targeting a person with a condition, but also others, as well as the interaction itself (cf. *Bolis et al.*, 2017).

2.2 Dialectical misattunement

Original publication details of this chapter (submitted 2016; accepted 2017)

Bolis, D., Balsters, J., Wenderoth, N., Becchio, C., & Schilbach, L. (2017). Beyond autism: introducing the dialectical misattunement hypothesis and a Bayesian account of intersubjectivity. Psychopathology, 50(6), 355-372.

Contribution of authors

DB did the initial conceptualization work, planned the manuscript and organized its argumentative structure. DB wrote the initial draft of the manuscript. LS revised the manuscript, provided critical feedback and overall supervision. JB, NW, CB provided helpful comments. LS, NW provided funding.

Beyond Autism: Introducing the dialectical misattunement hypothesis and a Bayesian account of intersubjectivity

Dimitris Bolis[a-c,4], Joshua Balsters[c], Nicole Wenderoth[c], Cristina Becchio[d,e], Leonhard Schilbach[f,g]

Affiliations

[a] Independent Max Planck Research Group for Social Neuroscience, Max Planck Institute of Psychiatry, and
[b] International Max Planck Research School for Translational Psychiatry (IMPRS-TP), Munich, Germany;
[c] Neural Control of Movement Lab, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland;
[d] Department of Psychology, University of Turin, Turin, and
[e] Robotics, Brain and Cognitive Sciences, Fondazione Istituto Italiano di Tecnologia, Genoa, Italy;
[f] Department of Psychiatry, Ludwig Maximiliam University, and
[g] Outpatient and Day Clinic for Disorders of Social Interaction, Max Planck Institute of Psychiatry, Munich, Germany

Keywords

Autism · Dialectical misattunement · Social interaction · Intersubjectivity · Cultural historical activity theory · Enactivism · Predictive processing/coding · Active inference · Dialectics · Vygotsky · Bayes

[4] Corresponding author: Dimitris Bolis. E-mail: dimitris_bolis@psych.mpg.de

Abstract

Drawing on sociocultural theories and Bayesian accounts of brain function, in this article we construe psychiatric conditions as disorders of social interaction to fully account for their complexity and dynamicity across levels of description and temporal scales. After an introduction of the theoretical underpinnings of our integrative approach, we take autism spectrum conditions (ASC) as a paradigm example and discuss how neurocognitive hypotheses can be translated into a Bayesian formulation, i.e., in terms of predictive processing and active inference. We then argue that consideration of individuals (even within a Bayesian framework) will not be enough for a comprehensive understanding of psychiatric conditions and consequently put forward the dialectical misattunement hypothesis, which views psychopathology not merely as disordered function within single brains but also as a dynamic interpersonal mismatch that encompasses various levels of description. Moving from a mere comparison of groups, i.e., "healthy" persons versus "patients," to a fine-grained analysis of social interactions within dyads and groups of individuals will open new avenues and may allow to avoid an overly neurocentric scope in psychiatric research as well as help to reduce social exclusion.

δὶς ἐς τὸν αὐτὸν ποταμὸν οὐκ ἂν ἐμβαίης.
You could not step twice into the same river.

τὰ ὄντα ἰέναι τε πάντα καὶ μένειν οὐδέν.
All flows, nothing stays.

Heraclitus (ca. 535–475 BC)

Through others, we become ourselves.

Lev Vygotsky (1896–1936)

2.2.1. A Synthesis of Dialectical and Computational Perspectives

Psychiatry through a Dialectical Lens

In this paper, we will put forward an integrative approach for revisiting psychiatric conditions, taking dialectics as a point of departure. The latter could be considered as an evolving school of thought, met in various historical contexts (e.g., Greek, Chinese, Hegelian, and Marxian

dialectics; Dafermos, 2015; Wong, 2006) critical to both reductionism and dualism. It asserts that phenomena cannot be meaningfully understood by reducing them into single levels of description (cf. reductionism) or assuming a metaphysical independence between levels (cf. dualism), but should be rather studied in their wholeness, inner contradiction, and movement (Table 1). In this light, human mind and psychopathology cannot be understood in isolation from society, the body, and social interaction. To quote Hegel *"to know, or, in other words, to comprehend an object is equivalent to being conscious of it as a concrete unity of opposed determinations"* (Bidell, 1988; Lawler, 1975). We will, therefore, try to overcome traditional dichotomies, such as organism/environment, by viewing them as both a result and a cause of reciprocal adjustments, or individual/society by considering the whole and the part as, albeit partially autonomous, highly interdependent levels of organization. In this effort, we will also draw upon accounts of intersubjectivity, which emphasize that single levels of analysis or cutting off the part from the whole may severely limit our understanding of a phenomenon. We will emphasize viewing psychiatric conditions not as static conditions driven by a single cause, but rather as the outcome of an interplay of multiple and diverse factors (**Fig. 2.4**) and to be more specific as a process of circular causality among different levels of description (e.g., biological, cognitive-behavioral, and sociocultural) as well as multiple functions within a level (e.g., action and perception within the cognitive-behavioral level), unfolding over different temporal frames (e.g., evolutionary, cultural, social, individual-psychological, subindividual-biological developing scales; based on Lev Vygotsky and colleagues' views on human development; Vagenas, 2003; Vygotsky, 1930-1935/1978).[5]

---

[5] Please note the specific definition and distinction between levels, functions, and temporal frames, as put

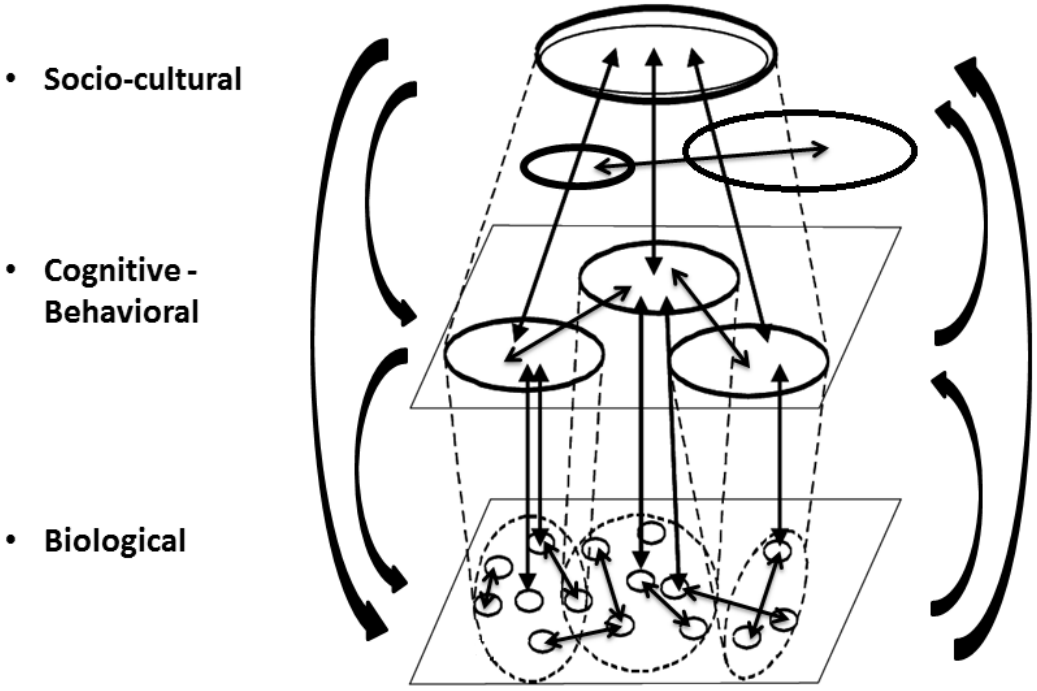

Fig. 2.4. Schematic depiction of dynamic interrelationships: between multiple levels (e.g., biological, cognitive-behavioral, sociocultural) and functions (e.g., including but not limited to the functionality of multiple neuromodulators or bacteria at the 1st level, body-/ neurosystemic, and phenomenological aspects at the 2nd level, and social structure, institutions, or cultural practices at the 3rd level) interacting in several temporal scales. Please note arrows may appear static on the image, but we interpret them as representations of developing interrelationships reflecting both quantitative and qualitative changes (cf. dialectics). Certain additional core levels of description, i.e., the (micro-/macro-)physical levels, have been omitted from this illustration.

Indeed, contrary to a common assumption that a full description on a micro-spatiotemporal level is causally complete, it has been suggested that a genuine causal emergence on a macro-level might also be possible (Hoel *et al*., 2013). Importantly, such an emergence is not to be solely

---

forward here, are made for intelligibility purposes only and it should not be taken as implying dichotomies; processes and their interrelationships appear complex, continuous, and overlapping in reality (Pessoa, 2008).

attributed to a weakness of experimental means to fully grasp the micro-phenomena but rather due to inherent characteristics of systemic processes themselves. For example, coarser mechanisms on a higher level might appear more robust in terms of causality than relevant stochastic micro-processes. Thus, a genuine causal emergence on a macro-level is necessary for a complete description. In fact, this is a conclusion from physics where the circular causality between the microscopic and the macroscopic is well established in terms of concepts such as the slaving principle and the center manifold theorem. In brief, these theorems suggest the emergent macroscopic (order parameters) that describe the whole enslave the microscopic components that constitute the whole. This induces a circular causality that lies at the heart of synergetics (cf. Kelso, 1995; Haken, 2004). It also speaks to the circular causality to which enactivism and embodied (situated) cognition approaches appeal (**Table 1**). Following such a line of thought, this paper will argue that while considering neurobiological and phenomenological processes is an important step toward the understanding of psychiatric conditions, it may remain incomplete as further levels of analysis, such as sociocultural processes and generally social structure, are neglected. For instance, structures promoting social exclusion or competitiveness, as opposed to communication and collaboration, could distinctly shape individual behavior, mental reality, and biological mechanisms. Here, our approach heavily leans on work from the cultural historical activity theory (Table 1), which re-interpreted human development across a variety of conditions as a dynamic interplay between biological and sociocultural forces (Vygotsky, 1934/1962; cf. Vagenas, 2003; Dafermos, 2002 on the work of Lev Vygotsky and colleagues). Notably, the aforementioned variety of conditions was not limited to what one could think of "social conditions" but rather included individuals who were both deaf and blind, to give an example. The organic condition can of course still affect the construction of the social self via atypical

development if amelioration of social exclusion is not taken into account. As Vygotsky, pointed out:

> The confusion and failure to differentiate the organic from the cultural, the natural from the historical and the biological from the social […], inevitably leads to a fundamentally incorrect understanding and interpretation of the data (observations; excerpt from Vygotsky's work; translated in Vagenas, 2003).

Psychiatry through a Computational Lens

In our effort to adopt an integrative perspective, we will use Bayesian accounts of cognition and behavior as powerful tools of analysis within the level of the individual, but most importantly we will suggest ways of going beyond the individual as the unit of analysis and eventually overcoming limitations of a single-level approach (see the last two chapters of this study). Computational psychiatry can be thought of as lying on the interface between computational neuroscience and clinical psychiatry. It deploys computational (e.g., Bayesian) modeling in order to mechanistically describe psychiatric conditions (Huys *et al*., 2016; Stephan & Mathys, 2014). A more specific hierarchical Bayesian approach to perception and action, which we will focus on here, has been described as the predictive coding (also mentioned as predictive processing; a term which we will be using in this article) and active inference account (**Table 1**). In brief, according to such a perspective, the brain's ultimate goal is the long-term minimization of free energy, which (as we will explain later under simplifying assumptions) can be thought of as the "prediction error," i.e., the discrepancy between incoming information and generated predictions, based on consolidated experience (**Table 1**). Importantly, this is thought to be accomplished through two main avenues, namely either via updating the beliefs one holds for aligning them with the environment (i.e., predictive processing) or through action, which can help to experience the environment in accordance with prior beliefs (i.e., active inference). Here, it

should be noted that Bayesian beliefs inherent in any Bayesian approach to cognition should largely be thought of as subpersonal. In other words, the experience subtended by predictive processing is not necessarily a conscious experience but more like a percept (or possibly a causative experience; i.e., qualia), embracing also other "automatic" processes such as homeostatic control. One of the many interesting aspects of this account is that perception, learning, and action are not considered as isolated and passive processes, but they constitute interconnected processes, which an organism actively deploys for making sense or (to put it in computational terms) "model" the world in order to maintain its current living form (Friston, 2010).

The Dialectical Misattunement Hypothesis and a Bayesian Account of Intersubjectivity

Taken together, we suggest that formally considering (both quantitative and qualitative) dynamically changing interrelationships between and within levels of description (**Fig. 2.4**) as well as temporal scales will be essential for a comprehensive understanding of complex psychiatric conditions, such as autism spectrum conditions (ASC). In light of this, the purpose of this paper will be threefold: Firstly, to consider the integration of diverse within-level (i.e., neurocognitive) processes embedded in a common framework, i.e., the predictive processing and active inference account. Secondly, to outline the importance of taking into account interrelationships across levels (i.e., the individual and the collective) via putting forward the "dialectical misattunement" hypothesis. Thirdly, to ultimately motivate the development of a "Bayesian account of intersubjectivity" rather than of individual brains. Importantly, we also highlight the practical implications of our theoretical approach (i.e., ethical, research, clinical and pedagogical). Taking ASC as a paradigm case, we will give a description of the general

framework of our approach. More concretely, we will first review the field of autism research with emphasis on recent interest in providing a Bayesian formulation of ASC. Based on this, we will argue in favor of adopting the Bayesian accounts of brain function as a framework to integrate seemingly contradictory neurocognitive hypotheses. Then, we will discuss different accounts of intersubjectivity, which share a common ground by stating that individual level analyses do not suffice for a comprehensive understanding of social perception and cognition. Bringing together a dialectical perspective to human communication and Bayesian (i.e., predictive processing and active inference) accounts of individual mechanisms (Bolis *et al*., 2016), we will introduce the dialectical misattunement hypothesis which emphasizes the interdependence of individual and collective levels of description.

More concretely, the dialectical misattunement hypothesis rethinks ASC not merely as a disorder of the individual brain but also as cumulative misattunement between persons. Misattunement across persons can be thought of as disturbances of the dynamic and reciprocal unfolding of an interaction across multiple time scales, resulting in increasingly divergent prediction and interaction styles. Consequently, with regard to neuroscientific research, we propose moving from focusing only on comparing groups of individuals to considering types of interaction between persons (e.g., homogeneous dyads consisted of either only neurotypical persons or only of persons with a certain condition, as well as heterogeneous dyads; including both tuned and non-tuned interactions[6], Table 1). Here the hypothesis holds clear predictions: Interactions within homogeneous dyads are expected to appear smoother compared to heterogeneous dyads.

[6] The term "tuned" here refers to multiple aspects: tuning expectations of either or both the interactors, as well as facilitating the interaction via tuning the communication medium (e.g., social conventions, as well as the cultural or technological environment, in which the interaction is embedded).

Additionally, tuned interactions of either homogeneous or heterogeneous dyads should appear as most effective. If these hypotheses are valid, the definition of a psychiatric condition as ASC can be thought of as relative to the "other" and generally the social context. Such an approach, will eventually allow us to escape an overly neurocentric research scope in psychiatry. Along similar lines, we suggest that clinical and pedagogical practices should move beyond the individual to monitoring, evaluating, and facilitating processes at the interpersonal level. Also, reviewing ASC as a misattunement between people, and not as disorder of the brain *per se*, may help to alleviate social stigma and reduce social exclusion.

We will end by outlining a Bayesian account of intersubjectivity, referred to as the "observing-the-interactors" scheme, which will allow us to computationally describe the interplay of individual and collective levels of activity during social interactions. Subsequent papers will delineate a practical approach for testing the misattunement hypothesis of social interaction based upon hierarchical models of interpersonal interactions (Bolis & Schilbach, 2017) and 2-person psychophysiology (Bolis & Schilbach, 2018). In what follows, we focus on autism, but the proposed approach more generally applies to any process evolving at the interface between the intra- and interpersonal level (**Table 1**), including social exclusion across different conditions.

2.2.2. Traditional Views on ASC

Although sparse references about resembling cases may have existed before (Wing, 1997), it was not until the 1940s that Hans Asperger and Leo Kanner described the condition of autism. Today, autism is considered as a neurodevelopmental disorder spanning a spectrum characterized by impairments in social interaction and communication as well as restricted, repeated behaviors

and interests. It is also not uncommon for ASC individuals to show enhanced abilities for specific cognitive aspects including perception (Dakin & Frith, 2005), attention (Plaisted Grant & Davis, 2009), and memory (Treffert, 2009). While some approaches have focused on the impairments, other accounts encompass both impaired and enhanced skills (Frith & Happé, 1994; Baron–Cohen, 2000), especially when it comes to the so-called "high- functioning" end of the spectrum. In the past half century, a number of different cognitive hypotheses have been pursued in order to understand core aspects of ASC. Although several important ideas have helped to shed light on specific facets, there is still no consensus about a single theory that could offer a universal and yet specific explanation of the condition. We will primarily focus on the "5 big ideas" about autism, as suggested by Uta Frith (2008):

Firstly, Baron-Cohen et al. (Baron-Cohen *et al*., 1985) proposed that ASC individuals lack a specific meta-representational capacity, namely a "theory of mind," which prevents them from inferring upon other people's mental states. As a consequence of this, ASC individuals –so it is assumed– cannot know about other people's beliefs, emotions, desires, perceptions, and intentions. In light of findings that ASC individuals can make a conscious effort to think about others' mental states, it has been suggested that implicit, namely spontaneous, mechanisms of mentalizing might be the ones that are primarily linked to relevant difficulties in ASC, rather than explicit ones, which might be easier compensated for through learning (Senju *et al*., 2009; Schilbach *et al*., 2012).

The second big idea focuses on a special category of neurons, the so-called "mirror neurons" (Di Pellegrino *et al*., 1992; Rizzolatti & Craighero, 2004), which are active both when an action is performed and observed. The broken mirror neuron (BMN) hypothesis proposes the explanation of impaired social skills in ASC on the basis of a dysfunctioning mirror neuron system (MNS)

(Oberman *et al*., 2005; Ramachandran & Oberman, 2006). A number of studies offered supportive evidence for the involvement of MNS (Dapretto *et al*., 2006; Oberman *et al*., 2005; Perkins *et al*., 2010). However, both the validity of a broken MNS and a direct, causal relationship between MNS and social skills in ASC have been challenged by other reports (Fan *et al*., 2010; Southgate & Hamilton, 2008). Differences in MNS activation between neurotypical and ASC individuals could be alternatively traced back to earlier modulatory effects of the mentalizing system as well (Dumas *et al*., 2014; Wang & Hamilton, 2012).

Alternatively, the social motivation hypothesis focuses on motivational rather than "purely cognitive" aspects (Chevallier *et al*., 2012). It proposes that people with ASC lack the inherent social drive, which would assist them in exploiting the necessary learning opportunities for developing expertise in social cognition. More precisely, the hypothesis is settled upon the fact that social orienting, social seeking and liking, as well as social maintaining appear to be affected in ASC. On a biological level, the focus is placed on the human reward system, where either specific social impairments or more general reward-related dysfunction could explain the behavioral findings. A suboptimal oxytocin regulation has also been implicated in ASC, which could, for example, reflect differences in relating social stimuli to rewarding values (Bartz *et al*., 2011; Gordon *et al*., 2013; Modahl *et al*., 1998; Modi & Young, 2012).

The fourth idea, namely the weak central coherence hypothesis, considers ASC as a different, detail-oriented cognitive style (Dakin & Frith, 2005; Frith, 2003; Happé, 1999; Happé & Frith, 2006). More precisely, it claims that people with an ASC tend to process information locally rather than globally. It predicts that people with ASC will have difficulties in perceiving information in context. According to this idea, ASC individuals perceive the world differently in a number of aspects, including visual, auditory, and linguistic functions. Later, the enhanced

perceptual functioning hypothesis attributed this local bias to superiority of detail processing *per se* and not due to inferiority of global information processing (Mottron, 2006). Meanwhile, the monotropism hypothesis proposed a generalization from the tendency to focus on a local level to a need of focusing on a single source level of information (Murray *et al*., 2005).

Finally, the executive dysfunction (ED) hypothesis focuses on the difficulties that ASC individuals face when it comes to executive functions, i.e., problems primarily associated with functions such as planning, flexibility, inhibition, and working memory (Hill, 2004; Ozonoff *et al*., 1991; Rumsey, 1985; Steel *et al*., 1984). For instance, difficulties related to dealing with novel situations and improvising, as well as perseverative stereotyped behavior in ASC, can be explained by ED. This hypothesis has been taken to suggest that the study of frontal cortex function should be particularly relevant for a neurofunctional understanding of ASC.

To conclude this brief introduction of various accounts of ASC, it can be said that a number of different hypotheses have provided important insights into specific aspects of ASC; still, none of them is considered to provide a global explanation. In fact, it has been argued that a single explanation at the cognitive, neural, or genetic level might be intractable (Gallagher & Varga, 2015; Happé, 2003; Happé *et al*., 2006). However, interest in a potentially unifying account has recently re-emerged while making reference to and drawing upon the Bayesian brain hypothesis and particularly the predictive processing and active inference scheme (Bolis & Schilbach, 2018a; Brock, 2012; Burr, 2012; Lawson *et al*., 2014; Pellicano & Friston *et al*., 2013; Sevgi *et al*., 2019; Teufel *et al*., 2013; Van Boxtel & Lu, 2013; Van de Cruys *et al*., 2013; Van de Cruys *et al*., 2014). In the following, we direct our attention to the discussion of this approach and its relevance for ASC.

Bayesian Approaches

The Bayesian Brain Hypothesis

The main premise of the Bayesian brain hypothesis rests on the idea that the brain represents information accessed via the sensory organs in the form of probability densities, as opposed to single numbers, which are continuously updated, as if following a specific set of mathematical formulas based on the Bayes theorem. Crucially, this allows for optimal information integration both in time and space, multimodal cue integration, as well as flexible information manipulation without the need to commit to particular decisions at an early stage of processing (Knill & Pouget, 2004). To put it simply, through a Bayesian lens one can view the brain as an organ which calculates and maintains probabilities about events in the environment or about the self via a combination of already gained experience and newly sensed information. Crucially, the more confidence (i.e., precision) is placed on the validity of experience (i.e., prior beliefs) the less the latter is updated in the face of new incoming information (i.e., evidence).

To make it more intuitive, let us imagine a young woman, Penelope, living in Southern Greece, wakes up on a summer morning late for her work. The blinds are shut down, and there is no time to check the weather outside the window. Will she take her umbrella on the way out? Based on her experience (i.e., prior beliefs: it rarely rains in Southern Greece in the summer), she decides not to take her umbrella with her. However, in the evening it happens to rain (i.e., evidence). The next day, Penelope, bringing together experience and the previous day's facts, thinks there might be a slightly higher probability of raining (i.e., posterior belief), but this is still not high enough to persuade her that carrying an umbrella might be a good idea. After several days of raining, she eventually decides to put the umbrella in her bag. She has come to believe that the probability of raining is high enough these days despite her opposing experience of previous years. Perhaps not

surprisingly from a Bayesian point of view, Penelope still keeps the umbrella with her for a few days after the weather has been sunny and dry again. Before concluding our example, it is worthwhile to introduce the concept of precision, which can be generally thought of as the confidence about a certain belief. Let us imagine a second scenario, where Penelope wakes up on a summer morning in Japan, where she has been travelling for a few days. She has heard that the weather is generally dry in summer in the city she stays. Yet, on the first day, it does happen to rain. Interestingly, already from the next day, she decides to take an umbrella with her. Why did she change her mind so quickly in this case? Adopting a Bayesian perspective, one could argue that Penelope, although holding a high prior belief about not raining, changes her mind quickly due to the relatively low confidence (i.e., precision) she places on these prior beliefs of her, which have been the result of rumors and not her own experience.

The Hypo-Prior Hypothesis of Autism

Coming back to our main example of ASC, Pellicano and Burr (2012) adopted a Bayesian standpoint to argue that nonsocial features of ASC might be well explained in reference to attenuated Bayesian priors (i.e., priors of relatively low precision, so-called hypo-priors). This hypothesis anticipates a relatively more "precise" perception in ASC, driven primarily by perceptual evidence as opposed to prior knowledge, as well as a sense of being overwhelmed by this information, a common complaint of persons with ASC. Moreover, the hypo-prior hypothesis predicts the impedance of performance in ambiguous situations when prior knowledge is crucial for optimally solving a perceptual problem of inference. Finally, it was considered that a different learning style, namely one resembling overfitting in machine learning, and differences in adaptation can also be explained by this hypothesis (cf. Harris *et al*., 2015).

The hypo-prior hypothesis was then reformulated (Friston *et al*., 2013; Van Boxtel & Lu, 2013) within the predictive processing scheme, a more specific Bayesian account (Clark, 2013; Friston, 2005; Friston, 2008; Mumford, 1992), while considering social aspects of individual cognition and behavior (Lawson *et al*., 2014; Van de Cruys *et al*., 2014). It is worth noting that the importance of difficulties related to predictions had been noted in the autism literature in the past as well (Gomot & Wicker, 2012). However, the more recent shift towards focusing on predictive processing and particularly on the concept of precision described above can offer a potentially unifying explanation of autistic symptoms and directly relate computational findings with tractable neurobiological mechanisms. Before explaining how a predictive processing and active inference framework could, therefore, facilitate research into autism, we will first present the underlying basic ideas.

Predictive Processing and Active Inference

The general idea of predictive processing and active inference is not new. For instance, one can find indications in Hermann von Helmholtz (Von Helmholtz, 1867), who spoke about "unconscious inference" in the 19th century, drawing on ideas going back to ancient philosophers. Additionally, relevant traces can be found in ideas such as the reafference and ideomotor principles (Herbort & Butz, 2012; James *et al*., 1890; Stock & Stock, 2004). To put it simply, within a predictive processing and active inference framework, the brain is essentially viewed as a "prediction machine" whose ultimate goal is the minimization of "prediction error" by deploying hierarchical generative models. More precisely, higher levels of a hierarchy continuously produce predictions, which are tested against the input information of the immediate lower levels. The discrepancy between predictions and incoming information, i.e., the

"prediction error," is propagated to higher levels, reconfiguring the system to optimize its next predictions. Notably, propagating only the error and not the actual incoming information to higher levels is an efficient and resource-oriented way of reducing the bandwidth of the processed information, which is also exploited in data compression techniques, such as the common JPEG format. In short, two processes take place at the same time in opposite directions; predictions are propagated backward from higher to lower levels, trying to explain away prediction errors, and prediction errors are propagated forward from lower to higher levels, updating predictions (**Fig. 2.5**).[7] The hierarchical structure of the model is of immense importance because it enables the brain to optimize its own (empirical) priors on the fly. Additionally, it allows for effective representations of increasing abstraction. From a neurobiological perspective, forward connections may arise in superficial pyramidal cells, whereas the sources of backward connections are assumed to reside in deep pyramidal cells (Felleman & Van, 1991; Friston & Kiebel, 2009).

At this point, it is important to place the predictive processing in the more general context of active inference (a corollary of the free energy principle). Crucially, active inference takes predictive processing beyond the domain of perceptual inference and provides an account of action. The brain can be seen as inferring upon the causal structure of the world by updating "beliefs", which are represented as probability densities. Most simply, the latter would take the form of Gaussian distributions, fully defined by their mean (i.e., expectation) and variance (i.e., inverse precision). Under this simplifying assumption (i.e., the Laplace assumption), the

[7] Please note the new perspective, which is introduced with the predictive processing definitions of "backward" and "forward" connections, contrasted with the "feedback" and "feedforward" ones, since in the context of predictive processing the backward connections are the ones providing feedback via prediction error information on the forward stream of predictions (Clark, 2013).

generalization of prediction error minimization to "free energy" minimization becomes mathematically more evident.[8] The latter then takes the form of a difference between the predictions of a model and the representations to be predicted (Friston, 2013). Indeed, free energy had been originally formulated for confronting the difficult problem of exact inference, transforming it into an easy problem of optimization. It could be possible that a similar trick is used by the brain in order to efficiently approximate the inference problem in a quasi-optimal Bayesian way.

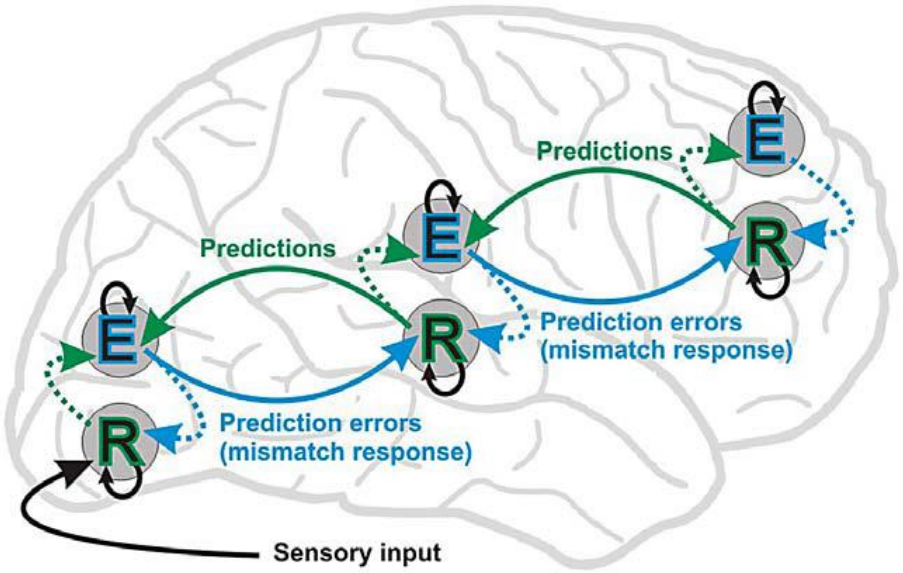

Fig. 2.5. A simplified representation of the predictive processing idea (taken from: Stefanics *et. al*., 2014): representation units (R; deep pyramidal cells) receive inputs (blue arrows) from error units (E; superficial pyramidal cells) of the same (dotted line) and lower levels, while error units receive inputs (green arrows) from the same (dotted lines) and higher levels. Black arrows represent inhibitory intrinsic connections.

[8] In this setting, free energy can be regarded as an approximation, namely an upper bound, to Bayesian model evidence, which is the probability of observing the data given a specific model.

Interestingly, the free energy principle has been proposed as a potentially unifying brain theory, accounting for action, perception, and learning. In short, an agent has two options for suppressing free energy: first by selectively sampling the environment for fulfilling its own expectations (i.e., through acting referred to as active inference) and second by optimizing these expectations for better matching with its sensations (i.e., through perception and learning referred to as predictive processing; Friston & Stephan, 2007; Friston *et al*., 2010). More broadly, one could sketch a path which, starting from the existence of life (as a process leading to a restricted number of states), passes through entropy (referring to a tendency to resist the 2nd law of thermodynamics), surprise (viewing entropy here as a mean value of surprise over time), free energy (as an upper bound of surprise), and eventually leads to prediction error, which, as we pointed out, can be considered as the free energy under certain simplifying assumptions. As provocatively put by Karl Friston (Friston, 2013), "the motivation for minimizing free energy has hitherto used the following sort of argument: systems that do not minimize free energy cannot exist [...]".

Crucially, in the setting of predictive processing and active inference, the degree of prediction updating (i.e., the learning rate) is controlled by the relative precision of successive levels. More precisely, it is proportional to a relative precision-weighted prediction error. This makes sense, since it would be generally desirable for an agent to update their beliefs first when the prediction error is large and second when they are unsure (low precision or confidence) about their prior beliefs compared to incoming information of lower levels in the hierarchy (about the importance of precision, see Clark, 2013). Importantly, the idea of an updating rule proportional to the precision-weighted prediction error is a potentially neurobiologically plausible account, where precision is assumed to be represented by the gain of superficial pyramidal cells calculating

precision errors (Feldman & Friston, 2010; Moran *et al*., 2013; Shipp *et al*., 2013). Psychologically, increases and decreases in the precision of sensory prediction errors have been associated with sensory attention and attenuation, respectively. In other words, attending to (or attenuating) a sensory stream is (under predictive processing) mediated by affording more (or less) precision to that stream (Friston, 2009).

Before concluding this introduction to predictive processing and active inference, it is worth noting that this scheme could be considered as a dialectical framework in and of itself. Firstly, it defines action and perception as the interplay between two closely intertwined avenues for minimizing prediction error. New perceptual states can inform future actions, while informed adjustment and sampling of the environment (i.e., action) decisively contributes to updating perception. Essentially, perception and action become here two dialectical facets of the same process, namely minimization of free energy. Additionally, prediction updating and activity can be viewed as dialectical processes in time between prior experience and incoming information, whose confrontation yields adjusted relations between environment and the self either through updating current beliefs or the perceived environment itself. We again see here a circular causality that is central to enactive (Bayesian) inference and speaks to related notions in enactivism and embodied cognition (see Integrating Individual and Collective Levels of Analysis). After having provided a general introduction to the predictive processing and active inference framework, their putative roles in understanding autism will be presented in the following.

The Aberrant Precision Hypothesis of Autism

It has been suggested that considering the role of precision in cognitive and behavioral processes could be important for understanding differences between neuro- typical persons and ASC individuals: Indeed, there is preliminary neurobiological evidence with regard to the functionality of certain neuromodulators that is suggestive of aberrant precision in ASC (Lawson *et al*., 2014). Additionally, several, psychological findings in ASC could be putatively attributed to aberrant precision estimation (Hohwy, 2013; Van de Cruy *et al*., 2014). For instance, hypersensitivity to sound and visual stimuli is typically observed in ASC individuals (Mottron *et al*., 2006). Through a predictive processing and active inference lens, consideration of irrelevant information due to increased precision can possibly lead to perceptual overload or, in other words, perceptual hypersensitivity. Furthermore, stereotypies, repetitive behaviors, and self-stimulation, all commonly observed in ASC, could be viewed as efforts for creating scenarios of reduced prediction error, because other pathways fail to do so. Finally, another core attribute of ASC, i.e., withdrawal to one's own self, might constitute an alternative strategy of generally keeping prediction errors low. This kind of behavior could also be linked to an attenuation of motivational factors due to a persistent inefficiency to trigger reward through decreasing prediction errors (Van de Cruys & Wagemans, 2011; Joffily & Coricelli, 2013).

Intriguingly, certain predictions made by the aberrant precision hypothesis can be formally tested via deploying predictive processing modeling. The latter approach al- lows for the tracking of potentially critical processes of the hypothesized "predictive brain" and may, therefore, have the potential to become an invaluable tool for revisiting the condition of autism. To date, a number of different theoretical and computational predictive processing and active inference models have been put forward, covering a variety of levels, functions, and temporal scales. In the next

section, we will suggest modeling examples of potential relevance to the autism research at the individual level. More specifically, we will view here predictive processing and active inference as a common framework for re-addressing traditional ideas about ASC. The "5 big ideas," which rest on diverse cognitive functions, will motivate and help to structure our suggestion.

Table 1. Glossary of terms, as they either appear in the bibliography (Adorno 2017; Clark, 2013; Friston, 2010; Friston & Kiebel, 2009; Haken, 1989; Hirsch *et al*., 2002; O'Connor & Wong, 2015; Robinson, 2017; Toomela, 2014; Vygotsky, 1930-1935/1978) or were introduced in this article.

Active inference	An account of action according to which (biological) systems sample the environment in accordance with prior beliefs for minimizing free energy (see Friston, 2010)
Bayesian account of intersubjectivity	The "Bayesian account of intersubjectivity" is considered here as a Bayesian account of human activity that takes into account both intra- and interpersonal processes (see this article)
Biofeedback	A training technique by which a person learns how to regulate certain body functions, such as heart rate, blood pressure, or brain wave patterns, that are normally considered to be involuntary (see Hirsch *et al*., 2002)
Cultural-historical psychology	Theory aiming at accounting for the inseparable unity of mind, brain, and culture (see Toomela, 2014; Vygotsky, 1930-1935/1978)
Dialectics	The dialectical method states that phenomena can be understood only in their wholeness, inner contradiction, and movement (see Adorno 2017)
Dialectical misattunement	The "dialectical misattunement hypothesis" rethinks a psychiatric condition, such as autism spectrum conditions (ASC), not merely as a disorder of the individual brain but also as cumulative misattunement between persons, which can be thought of as disturbances in the dynamic and reciprocal unfolding of an interaction across multiple time scales, resulting in increasingly divergent prediction and (inter-)action styles (see this article)
Dualism	Theory stating that for some particular domain, there are two fundamental kinds or categories of things or principles (e.g., the physical and the mental) (see Robinson, 2017)
Emergence	Emergent entities (properties or substances) "arise" out of more fundamental entities and yet are "novel" or "irreducible" with respect to them (see O'Connor & Wong, 2015)

Free energy	An information theory measure that bounds or limits (by being greater than) the surprise on sampling some data, given a generative model Put simply, with regard to an organism free energy minimization can be thought of as a process of maintaining current living form by being restricted in a limited number of possible states (see Friston, 2010)
Heterogeneous dyads	"Heterogeneous dyads" are considered here dyads which consist of persons with different conditions, such as a neurotypical person and a person with ASC (see this article)
Homogeneous dyads	"Homogeneous dyads" are considered here dyads which consist of either only neurotypical persons or only of persons with a certain condition, such as ASC (see this article)
Interaction tuning	"Interaction tuning" here refers to tuning of expectations of either or both interactors as well as facilitating a social interaction via tuning the communication medium (see this article)
Intrapersonal	(Processes) being unfolded within the person
Interpersonal	(Processes) being unfolded between persons
Precision	A statistical term defined as the inverse variance and can be thought of as the confidence a (biological) system places upon its beliefs (see Clark, 2013; Friston & Kiebel, 2009)
Prediction error	The discrepancy between incoming information and (biological) system-generated predictions (see Clark, 2013; Friston & Kiebel, 2009)
Predictive coding/ processing	Theory that states that (biological) systems are constantly generating and updating hypotheses about the causal structure of the environment and the self along different levels of abstraction for ultimately minimizing free energy (see Clark, 2013; Friston & Kiebel, 2009)
Sociofeedback	"Sociofeedback" is considered here a (future) training technique by which a person, a dyad, or a group of people will learn how to (co-)regulate certain social interaction processes, such as interpersonal coupling and coordination; the concept also applies to automatic adjustment of the interaction medium based on social interaction monitoring (see this article)
Synergetics	An interdisciplinary field of research that studies the spontaneous, i.e., self-organized, formation of structures in systems far from thermal equilibrium (see Haken, 1989)

2.2.3. Individual Level: Predictive Processing and Active Inference as a Common Framework for Integrating Diverse Neurocognitive Hypotheses

Theory of Mind – as described above – can be viewed as an inference problem (Baker *et al*., 2011), where the brain tries to understand "invisible" mental states through observable human behavior. Koster-Hale and Saxe review evidence that relates theory of mind to predictive processing formulations (Koster-Hale & Saxe, 2013). To that end, they consider how relevant brain regions such as the superior temporal sulcus, temporoparietal junction and medial prefrontal cortex might be involved in mental state inference across different time scales. To be more specific, the superior temporal sulcus has been implicated in neural reactions to face and body action in the scale of seconds, while the temporoparietal junction has been related to assessing desires and beliefs of other people, which can last from minutes to years, and the medial prefrontal cortex has been thought to contribute to the evaluation of temporally more stable traits of other people.

The social motivation hypothesis of autism focuses on how a lack of motivation for processing and learning about social aspects could be relevant for understanding ASC or how difficulties in social cognition could decrease interest in social cues. Interestingly, Heyes has argued that social learning shares the same basic cognitive mechanisms with nonsocial learning (2012). In line with this, Behrens *et al.* indicated that standard reward-based associative processes guide the acquisition of social information, too (2008). More specifically, they showed activation of the anterior cingulate cortex (ACC) gyrus and ACC sulcus for reward-based and social learning, respectively. At the level of decision-making, it was found that the ventromedial prefrontal cortex encodes both probabilities about social and nonsocial sources, appearing to integrate information from ACC sulcus and ACC gyrus in a subject-specific fashion. Consequently, the

above-mentioned brain regions could potentially play an important role in the investigation of ASC-related differences in multimodal cue integration and contextualization of precision in social and nonsocial cues (Apps *et al*., 2013; Balsters *et al*., 2016).

As previously discussed, the so-called "mirror neuron system" has also been implicated in ASC via the BMN hypothesis. According to the BMN hypothesis, difficulties in ASC in understanding others' actions and intentions may arise from a defective functioning of the MNS. However, precisely how mirror neurons contribute to action/intention understanding is still unclear (Kilner, 2011). Kilner *et al.* suggested that the brain deploys a mirror neuron predictive processing model and minimizes prediction error at all levels (Kilner *et al*., 2007). More specifically, they considered a hierarchy that consists of 4 levels of decreasing abstraction descending the hierarchy; the (1) intention, (2) goal, (3) kinematic level, and (4) muscular level, respectively (Hamilton & Grafton, 2007). These levels of behavior are generally assumed to be independent of each other (Hamilton & Grafton, 2007). This assumption, however, appears not to be true as recent evidence indicates that the kinematics of a performed movement already reflect the agent's intention and makes it distinguishable (Becchio *et al*., 2012). This raises the intriguing possibility that intentions may be decoded from movement kinematics (Ansuini *et al*., 2015). A reasonable framework for integrating different sources of prediction is that a range of possible intentions is first estimated from the spatial and temporal context, e.g., in predictive areas outside the mirror system (Kilner, 2011). This prior prediction can impact on action understanding, constraining the number of possible intentions. Early movement-discriminant kinematic features of the observed motor act can lead then to the selection of the most probable intention. Studying such inference problems in light of predictive processing and active inference could provide further insights into the implications of a BMN account for understanding ASC.

Visual processing and particularly the extraction of spatiotemporal regularities might also be related to specific theories about ASC, such as the weak central coherence hypothesis. Natural images tend to be correlated both in space and time. That is, natural scenes usually consist of finite regions of relatively uniform attributes and tend to reflect region-specific uniform intensity values (Huang & Rao, 2011). For example, a stable object, being viewed from a constant perspective, appears to emit relatively similar intensity values over time. These regular spatiotemporal characteristics can be exploited by the visual system to predict intensity values in advance based on neighboring and historical information. Indeed, Rao and Ballard (1999) proposed that the brain predicts this kind of regularities via a predictive processing model embodied in neural loops of increasing receptive fields with ascending hierarchy (e.g., the lateral geniculate nucleus-primary visual cortex-secondary visual cortex feedback loop; Huang & Rao, 2011). Such a family of models could be exploited in the future for an investigation of aspects related to a weak central coherence in ASC and more precisely the extraction of perceptual regularities. For instance, quantifying autism-specific styles in extracting such regularities could yield further insights about facts as perceptual hypersensitivity and differences in perceiving certain kinds of illusion (Happé, 1996).

The ED hypothesis focuses on executive cognition and behavior. Kopp has recently stressed the relevance of executive function for predictive processing theories (Kopp, 2012). More precisely, drawing on the latter and self-terminating operating units (Miller, 1960), Kopp proposed a theoretical hierarchical model for dealing with ED, especially focusing on brain regions as the medial, orbital, and lateral prefrontal cortex. Indeed, there is evidence speaking for a hierarchical organization of the rostrocaudal axis of the prefrontal cortex based on the level of abstraction (Badre & D'Esposito, 2007; Badre & D'Esposito, 2009). We suggest such kind of models could

prove to be fruitful in studying putative ED through the hierarchical inference entailed by predictive processing and active inference in ASC.

Lawson *et al.* have recently put forward several suggestions with regard to potentially aberrant predictive processing processes relevant for understanding ASC at a neurobiological level, too (2014). For instance, plasma oxytocin, which has been suggested to control the relative salience of social and nonsocial stimuli (Gordon *et al*., 2013), has been found to be reduced in children with ASC (Modahl *et al*., 1998). These can be linked to an aberrant precision hypothesis under the assumption that oxytocin is involved in contextualizing precision of social as compared to nonsocial stimuli (Quattrocki & Friston, 2014).

Taken together, we suggest that a multitude of aspects in ASC can be integrated under the predictive processing and active inference perspective. By doing so, ASC can be revisited as a different prediction and (inter-)action style, as opposed to a set of a priori impaired neurocognitive functions that reside in specific brain regions. This exact shift of perspective, however, begs the question of how does such a different style emerge? In the next section, we tackle this question by leaning on sociocultural historical theories, which emphasize the social construction of the (a-)typical self, and Bayesian accounts of brain function, which provide a powerful toolbox for the investigation of underlying mechanisms.

2.2.4. Integrating Individual and Collective Levels of Analysis: The Dialectical Misattunement Hypothesis

We open this section by discussing different approaches which – although following distinct lines of argument – converge on the idea that focusing on individual brains will not be enough to fully understand the human mind and psychopathology. In particular, we will argue against

considering only biological mechanisms, since, in our view, the latter reductionist approach covers only part of the dialectical interplay between individual processes and the collective level. In fact, cultural historical activity theories have strongly emphasized the importance of considering the interrelationship between individual and sociocultural processes in psychological and psychopathological research: For instance, Vygotsky already distinguished social interaction as a key factor in the formation of consciousness and "higher" human psychological processes, which he argued are developed through and due to social interactions (Vagenas, 2003). Additionally, he claimed that every function appears twice in a child's development, first at a social level (i.e., "intermind") and then at an individual level (i.e., "intramind"): "All the higher functions originate as actual relationships between individuals" (Vygotsky, 1930-1935/1978). In other words, he suggested that through communication, through the direct social interaction with others, a child internalizes active cultural values in society (as cited in Vagenas, 2003), realizing that the (a-)typical self is dialectically and socially constructed.

Interestingly, recent developments in accounts of social cognition and intersubjectivity have also focused on the enabling or even constitutive role of social interaction (Becchio *et al*., 2012; Bolis & Schilbach 2017; Bolis & Schilbach 2018a; Dale *et al*., 2013; De Jaegher & Di Paolo, 2007; De Jaegher *et al*., 2010; Di Paolo & De Jaegher, 2012; Dumas, 2011; Dumas *et al*., 2012; Froese & Di Paolo, 2011; Froese & Fuchs, 2012; Gallagher, 2001; Hari & Kujala, 2009; Koike *et al*., 2015; Konvalinka & Roepstorff, 2012; Oullier *et al*., 2008; Reddy, 2008; Schilbach *et al*., 2010; Schilbach *et al*., 2006; Schilbach *et al*., 2013; Schilbach, 2014; Schilbach, 2016; Timmermans *et al*., 2012). More specifically, mainstream accounts of social cognition have been criticized for neglecting the interactive dimension of social situations and for adopting an individualistic view of (social) cognition (e.g., specifically on the example of autism; Gallagher, 2004), philosophical

considerations; De Jaegher & Di Paolo, 2007), and neuroscientific research; Schilbach *et al*., 2013). With regard to psychiatric conditions, it has also been suggested that transdiagnostically observed social impairments are more likely or may only manifest under conditions of real-time social interaction, whereas situations of social observation might be less problematic (Schilbach, 2016). Furthermore, several accounts have been critical toward core assumptions of contemporary cognitivist paradigms, which have been thought of as viewing the brain, or more generally the organism, merely as a passive "consumer" of external stimuli (Froese, 2015). Despite each account's distinct commitments, these kinds of approaches are usually positioned under the umbrella of the "4Es" (Menary, 2010; Ward & Stapleton, 2012), which described cognition as enactive (Hurley, 1998; Thompson, 2010; Varela *et al*., 2017), embodied (Clark 1998; Gallagher, 2006; Haugeland, 1998), embedded (Clark 1998; Hurley, 1998, Haugeland, 1998), and extended (Clark & Chalmers, 1998; Hurley, 1998), but also affective (Colombetti, 2007; Ratcliffe, 2009). In line with these accounts, using scenarios of higher ecological validity, which do not neglect the critical role of the body, the environment and interactions in cognition could offer a more suitable framework to study brain function and behavior (Bolis & Schilbach, 2018a; Schilbach *et al*., 2013).

On top of providing a naturalistic scenario, interactive situations also potentially allow for the consideration of turn taking (Schilbach *et al*., 2010) and emergent social phenomena at higher levels of description, which otherwise might remain intangible (Bolis & Schilbach, 2017). In neuroscience, cognition has generally tended to imply a dynamic interaction between brain areas merely within a single skull. However, there is no theoretical reason to a priori exclude other body parts, and generally other people, as well as mediating cultural tools, as cultural historical activity theories would emphasize. In line with an enactivist or dynamical system perspective,

two or more communicating agents can be seen as a coupled system, being driven by nonlinear interactions (Dumas *et al*., 2014; Froese & Di Paolo, 2011; Froese & Fuchs, 2012). However, investigating individual predictive processing mechanisms in order to understand communicative processes between agents could also be particularly informative. Notably, a formal account of addressing communication as reciprocal exchange of predictions about the other's behavior has recently been put forward (Friston & Frith, 2015a; Friston & Frith, 2015b). This account, which rests on predictive processing, considers both perceptual updating and action expression within a closed loop between two agents. Here, simulations were used to illustrate how two agents, which model each other, could in theory converge into a system of generalized synchrony (i.e., synchronization of chaos), thereby effectively embodying a single shared model. In contrast to this 'solipsistic' understanding of communication, we argue that by adopting a dialectical perspective we will look for such synchronization dynamics across different levels of description and do not assume that my understanding of another is realized entirely in my own head.

To be more specific, we suggest that a "dialectical misattunement" constitutes one of the defining factors of ASC and other psychiatric conditions. Communication misalignments and weak interpersonal coupling in social interactions might be the result of increasingly divergent predictive and (inter-)action styles across individuals (cf. Predictive Processing and Active Inference). From an ontogenetic perspective, such a misattunement could result in impoverished opportunities for acquiring socioculturally mediated knowledge and skills. In other words, we view two potentially cardinal processes that are tightly intertwined in a dialectical relationship: at the collective level weak coupling, crucially modulated by sociocultural factors, might lead to greater interindividual incompatibilities in prediction and (inter-)action styles, while at the individual level, diverging prediction and (inter-)action styles might lead to weak communicative

coupling with others in social interaction.[9] In short, "dialectical misattunement" refers to an imbalance between individual and collective levels rather than exclusively considering single levels. This view particularly highlights the critical role of social interaction into human development and the social construction of the (a-)typical self. Consequently, the interactive nature of social situations can help to enhance or decrease differences in prediction and (inter-)action style in a feedback loop fashion (cf. the circular causality introduced above). That is, small initial differences at the individual level are thought of cumulatively enhancing (or weakening) interpersonal coupling during social interaction and *vice versa*. Schematically, an initial communicative gap could yield incompatible prediction and (inter-)action styles and *vice versa* (**Fig. 2.6**).

Notably, such communicative misattunement could be expected to unfold across multiple temporal scales; for example, this could take place during the course of a dialogue (scale of minutes), during a human relationship (scale of months or years), or along development (scale of a lifetime). Additionally, with regard to groups of people (e.g., the so-called psychopathological groups or, generally, any other social group), this kind of misattunement could even take on a cultural form, spanning a scale of several generations. For instance, culturally cultivated beliefs in a given society about a specific group of people (e.g., stereotypes) might modulate the communication efficacy between in- and out-group persons.

---

[9] Please note misattunement encompasses both aspects of dissimilarity (e.g., social misalignment) and noncomplementarity (e.g., dysregulated coupling).

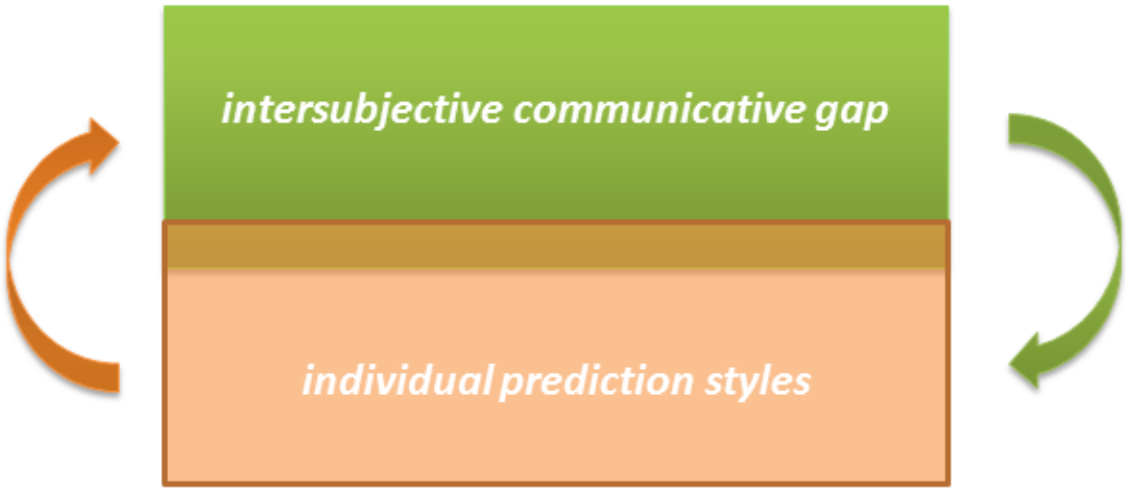

Fig. 2.6. Dialectical misattunement: increasing communicative gap (collective level) yields increasingly different prediction and (inter-)action styles (individual level) and *vice versa.*

More broadly, we believe that for gaining a complete understanding of conditions such as ASC, a shift in focus from the individual brain to the interaction between people is essential. Intriguingly, as we will argue in the next and final section, such an approach could yield formal insights into both individual and collective mechanisms (Bolis & Schilbach, 2017), as well as intra- and intercondition communication characteristics. Additionally, in psychiatry, it could facilitate research at both diagnostic and treatment levels. In short, we view the future of relevant theoretical research and clinical practice not only as an investigation of "disordered" brain mechanisms but also of a "misattunement" between persons. In line with the dialectical misattunement hypothesis, which highlights intersubjectivity as an indispensable factor of human development, we also suggest the enrichment of approaches which exclusively aim at "tuning" the ASC person. To this end, we suggest considering tuning also the "other" (i.e., the neurotypical person with whom the ASC person interacts), as well as the social interaction

medium (i.e., sociocultural framework, such as social expectations and stereotypes, as well as the technological medium, such as educational social robotics; Bolis & Schilbach, 2018a).

More precisely, in a clinical setting, one could, therefore, pay attention not only to the potentially "maladaptive" processes within the diseased individual but also to the coupling dynamics of the dyad (for instance during psychotherapy or group sessions) and critically the interaction between the individual and the collective. Additionally, our approach also motivates an alternative pedagogical program. The latter would primarily aim at tuning not merely individual behavior but crucially the interaction between people. Here, the pedagogical procedure would move beyond the traditional classroom, focusing on cognitive and behavioral aspects of not only the person with a specific condition (e.g., ASC) but also their interactors (e.g., parents, educators, or peers) and, most importantly, communication and mediating factors (**Fig. 2.7**).

This could be achieved by developing adjustable frameworks both to the individual and the interaction itself. A promising solution could be found in the form of "smart" technology, which could track and guide traditional educational practice, taking into account real-time activity but also historically relevant aspects (Vittorias *et al*., 2008). Crucially, while biofeedback techniques have been fruitfully used for monitoring and constructively exploiting individual activity (e.g., physiological factors), our approach would further point toward an extended notion of feedback, here referred to as "sociofeedback" (**Table 1**), including relational parameters (e.g., interpersonal coupling), too. Furthermore, the proposed shift in attention could not only be beneficial in clinical and pedagogical practice but also more broadly with regard to societal practice.

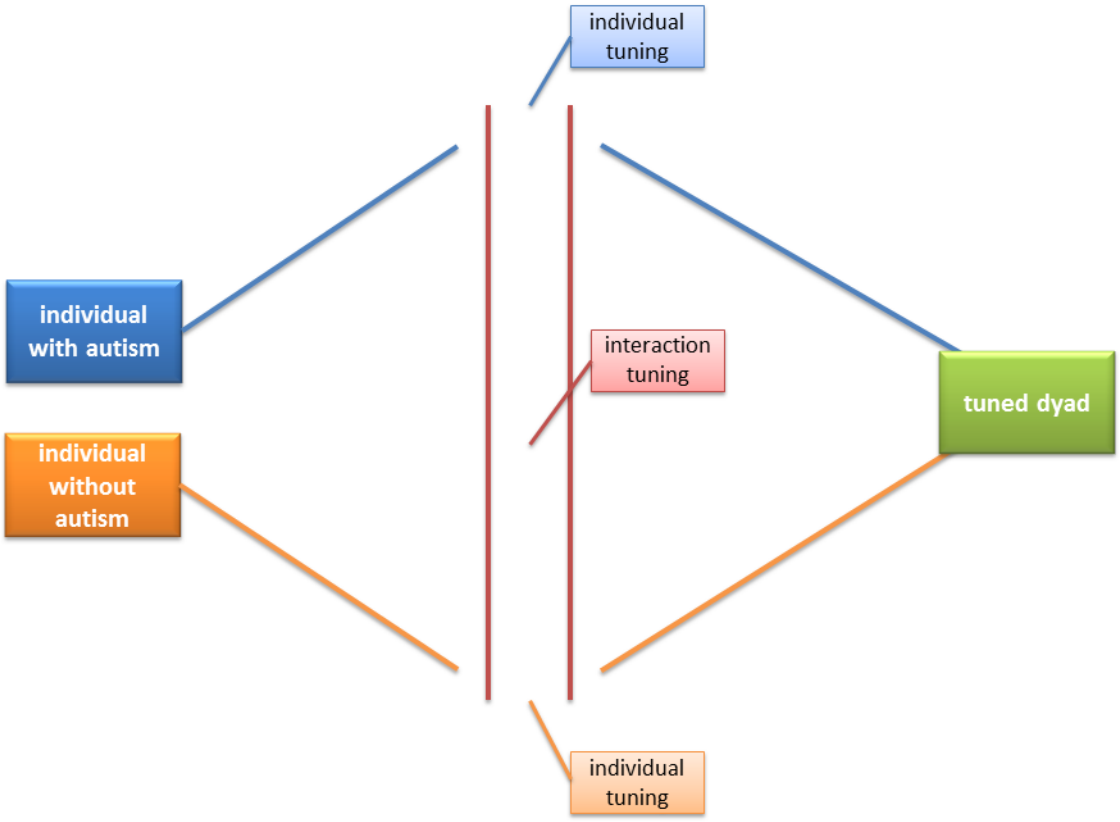

Fig. 2.7. Schematic presentation of a misattunement amelioration: by intervening both at the individual (e.g., cognitive and behavioral training of both interactors) and the collective level (e.g., adjustments of cultural/technological tools, sociofeedback). Blue, individual trajectory of a person with autism spectrum conditions (ASC); orange, individual trajectory of a person without ASC (trajectories here represent multiple temporal scales, from minutes in the course of a conversation to years across development).

For instance, by diffusing ideas in society about viewing psychiatric conditions as disorders of social interaction (Schilbach, 2016) rather than disorders of individuals, psychiatric stigma could be attenuated. As Vygotsky used to highlight, simply speaking, aspects of specific difficulties related to psychiatric conditions can be thought of as falling into two main categories: first aspects which are directly related to a biological level and second aspects which are related to relevant beliefs and practices in society. Although social processes play a decisive role in

shaping a person's mental reality, emphasis is usually only given to biology. Notably, being a social product to a large extent, such difficulties could be historically (along both social-historical and individual-developmental trajectories) alleviated. Furthermore, our approach emphasizes the dialectical relation of the collective and the individual (e.g., interrelations between culture and individual persons, as in interactions between "patient" and "examiner," or "patient" and "non-patient"). The broadened scope of effective treatment could encompass both personal and interpersonal parameters. In this light, the relativity of psychiatric diagnosis, which is usually the outcome of a communicative procedure between a potential patient and a culturally tuned examiner (e.g., psychiatrist or psychologist), also becomes more evident (Bolis & Schilbach, 2018a; Dafermos, 2012; Di Paolo & De Jaegher, 2012; Laing *et al*., 1966; Schilbach, 2016). In technical terms, our approach could be reframed as studying potential dynamic and recurrent feedback loops across and within different levels of description, as well as temporal scales, driving both quantitative and qualitative changes (cf. dialectics). We believe that computational modeling, such as Bayesian accounts, as well as dynamical system approaches can prove to be fruitful tools for scientifically testing the potentials of such a perspective. In fact, in our closing section, we will motivate a Bayesian account of intersubjectivity, which will aim at formally accommodating both individual and collective mechanisms.

2.2.5. Summary and Outlook: From a Synthesis of Dialectical and Computational Approaches to a Bayesian Account of Intersubjectivity

In this article, taking dialectics as a point of departure and drawing upon insights from multiple areas of research, we have argued that considering inherent interrelations as well as integrating findings from diverse levels of description, within-level processes and multiple temporal scales will be essential in future autism research. Such a holistic development, we claim, will help to

unveil the intrinsic units of analysis for reconstructing the critical dimensions of a multilevel and multidimensional condition such as ASC: thus, it is here thought of as an "autism space" rather than a spectrum. In particular, we discussed how a framework such as predictive processing and active inference could be used to bring traditional hypotheses at the level of the individual (e.g., neurobiology, cognition, and behavior) together and re-address them under a common umbrella. By doing so, ASC was revisited as a different prediction and (inter-)action style, as opposed to a set of *a priori* impaired neurocognitive functions that reside in specific brain regions. Then, we argued that such an approach is not sufficient on its own but needs to be directed towards the relevant real-life phenomena that take place during social interaction. Consequently, we propose an approach for integrating a computational and a dialectical perspective to psychiatric conditions for scientifically studying both intra- and interpersonal processes by introducing the "dialectical misattunement" hypothesis. Misattunement across persons is thought of as disturbances in the dynamic and reciprocal unfolding of an interaction across multiple time scales, resulting in increasingly divergent prediction and (inter-)action styles (ways of generating and expressing expectations about the [social] world and the self). This thesis does not consider psychiatric conditions, such as ASC, merely as disordered function within individual brains but rather as an interactive mismatch between persons.

In a forthcoming paper, we will use the conceptual arguments introduced above to illustrate the dialectical misattunement hypothesis formally. Specifically, we will analyze two-person simulations and experiments (Bolis & Schilbach, 2018a) with dual hierarchical Gaussian filters (Mathys *et al*., 2011) as a formal (computational) model of dyadic exchange (Bolis & Schilbach, 2017). This provides a quantitative and principled description of the dialectical misattunement hypothesis, and how it could be verified empirically using relatively simple paradigms and

analyses. In concrete terms, we suggest that established techniques of multilevel computational modeling (Bolis *et al*., 2015; Mathys *et al*., 2011) can be used to investigate the interrelation of individual brain mechanisms and interpersonal processes. Intrasubjective parameters (e.g., on the dynamics of belief updating) will be used for modeling individual brain processes of two (or more) brains, while intersubjective parameters will be introduced on a second meta-Bayesian level for capturing dyadic (or group collective) processes, such as interpersonal coupling (Bolis & Schilbach, 2017). The latter scheme will thus move beyond current neuromodeling approaches by also considering emergent phenomena on higher levels of description, such as questions about the autonomy of a dyad or a group of people and the individuality of the mind. To give a more specific example, in the context of collective decision-making or joint action, a nonlinear model might optimally explain observed behavior, thus, providing evidence that the dyad or the group is different than the sum of individuals. Inversely, this framework could address questions about how mechanisms of societal structure and, in general, collective processes, in turn, shape individual reality. For instance, one could differentially study the potentially distinct impact which a competitive versus a collaborative structure might exert upon an individual. Notably, this kind of modeling architecture will not be merely able to model multiple levels of description but interlevel processes as well (e.g., internalization and externalization mechanisms).

Moving the focus from the observation of individual observers toward a multilevel observation of dyads and groups of interactors could help to explore whether and how interpersonal coordination might actually serve as a prior and modulate the need for inferences about hidden causes of social behavior. Such an intersubjectively Bayesian approach, we claim, will provide a formal characterization of subject-specific as well as dyad and group level dynamics. It will, thereby, significantly advance our understanding of ASC and other psychiatric conditions

thought of as disorders of social interaction. As we provocatively state in the title of this article, we suggest we need to go beyond autism — not by neglecting the existence of the condition but by adopting a holistic approach which will embrace the individual with autism as well as the socioculturally mediated interactions with other people. The ultimate goal of such an approach will be to go beyond current diagnostic and treatment practice by promoting a reciprocal alignment of individual and societal practices as opposed to a single-sided adjustment of individual behavior and brain function into the “normal”.

3. Methodology — *'Two-person psychophysiology'*

"The most important experience of others takes
place in the face-to-face situation, which is
the prototypical case of social interaction.
All other cases are derivatives of it."

Peter Berger & Thomas Luckmann

3.1. Interpersonal attunement in *real-time* social interaction

3.1.1. Framework

After discussing the theoretical underpinnings of our approach, throughout this chapter we present *two-person psychophysiology*, demonstrating how it quantifies *real-time* social interactions (**Section 3.1**) and *real-life* social relations (**Section 3.2**).

With regard to real-time social interactions, this thesis, building upon seminal social neurobehavioral frameworks (e.g., Barisic *et al*., 2013; Dumas *et al*., 2012; Koike *et al*., 2019; Leong *et al.*, 2017; Montague *et al*., 2002; Murray & Trevarthen, 1985; Pfeiffer *et al.*, 2013), establishes an integrative platform for the high-resolution monitoring of two persons interacting in the lab, across various modalities, including brain function, gaze behavior, facial expressions and decision-making. In this framework, participants sit opposite each other, trying to accomplish perceptual tasks individually, while interacting via gaze behavior in real-time through a micro-camera communication system. This dual functionality is pivotal for relating multiscale processes (Bolis & Schilbach, 2018a), ranging from the level of the individual (e.g., psychophysiological reaction to perceptual information; **Fig. 3.1**: blue and green arrows) to the level of the collective (e.g., interpersonal processes during real-time social interaction; **Fig. 3.1**: red arrows).

The social interaction loop is mediated by a Skype-like communication system. However, it parts from such a commercial solution in crucial ways. First, it deploys micro-cameras, fixed at the center of the screen and more precisely between the eyes of each person, the ideal position for enabling realistic eye-contact (Kuster *et al.,* 2012). This was made possible via manufacturing and placing in front of each screen a transparent Plexiglas frame, at the center of which a special

3D-printed micro-camera holder was adjusted (**Fig. 3.2**: item #8). Second, our video link system was established on an analog basis so that signal delay is kept to a minimum, primarily defined by the physical characteristics of the video sensors and the electronic transmission line. Additionally, this platform is equipped with devices for the real-time manipulation of the social interaction synchronicity (cf. Koike *et al.*, 2019), as well as equipped with an extended version of an established graphical environment for psychophysiology research (cf. Bahrami *et al.*, 2010). Additionally, a high-resolution, multimodal and precisely-synchronized recording system was brought together. For optimizing experimental conditions, the two-person psychophysiology platform is built within two attached but separate rooms, i.e., the *experimenter's room*, where the major controlling devices are located and the *participants' room*, where the actual study takes place. Below the current technical design of the platform is presented in more detail (**Fig. 3.2**).

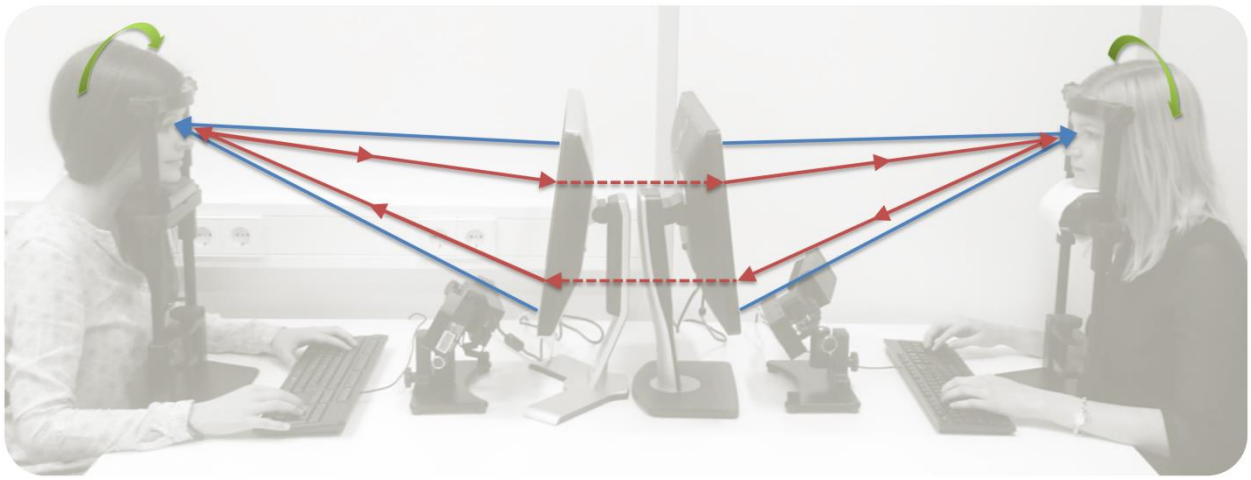

Fig. 3.1. Two-person psychophysiology in action: This platform allows for an investigation of the interrelation of individual psychophysiological and interpersonal processes in direct, gaze-based social interaction between two persons; **blue arrows**: incoming perceptual information from systematically controlled visual stimuli; **green arrows**: intra-personal bodily processes; **red arrows**: social interaction loop mediated by a micro-camera communication system. Please note the chin rests ensure optimal data acquisition quality but their use is optional (cf. **Appendix A.2.2**; Bolis & Schilbach, 2018a).

To begin with, in the experimenter's room two presentation PCs (**1**) generate the task graphics, which are presented in real-time in both the experimenter's and the participants' room on two pairs of monitors (**2** and **3** respectively). Participants register their decisions via two keyboards, which are placed on the table in front of them. The presentation PCs also serve for sending time stamps, via the parallel port, to the recording devices, such as the 1,000Hz eye-trackers (**4**) and the 1,000Hz EEG systems (**5**). Additionally, the presentation PCs control two red LEDs (**6**), in order to physically mark each trial phase on the video recordings. Precise time-stamping is crucial for ensuring the synchronization of the multimodal signals across the two persons during data analysis.

Furthermore, analog video mixers (**7**) combine the two types of video signal, i.e., the task graphics and the video stream of the co-participant captured by the analog 25Hz micro-cameras (**8**), projecting them on each participant's monitor (**3**). For addressing the issue of infrared interference by the active eye-trackers, the micro-camera lenses are protected by passive optical filters. Importantly, the micro-camera video stream is transmitted with a systematically controlled time delay, which ranges from zero to several seconds (cf. Takahiko *et al.*, 2019). This delay is set by the presentation PCs via two analog video delay devices (**9**). As we will see below this is crucial for studying the *real-time-specific* effects of social interaction: it defines the two cardinal conditions, i.e., *online* (real-time, thus two-way interaction) and *offline* (pseudo-interaction with a delayed video replica of the other, thus one-way interaction). Furthermore, the micro-camera video stream is simultaneously directed through an analog video splitter to a 'video-grabber' (**10**), which digitalizes it, projects it and stores it in two laptops (**11**) for later analysis, such as interpersonal facial expression analysis.

For facilitating participants' mobility, the EEG signals are transmitted through a wireless Bluetooth system to the two laptops (cf. Leong *et al.*, 2017), where they are presented in real-time and stored for later analysis. Similarly, the eye-tracking signals are presented on two tracker monitors in the experimenter's room in real-time (**12**), while being stored on two tracker PCs (**13**), exclusively assigned with this role for minimizing time delay. Additionally, PAL/NTSC and analog/digital converters were used in parts of the system for addressing cross-compatibility issues with the various devices. Finally, the experimenter simultaneously controls both sides of the setup with a single push button (**14**; cf. Koike *et al.*, 2019). This is crucial for enabling precise synchrony between the two sides of the platform. Notably, this two-person psychophysiology platform is fully compatible with further recording modalities, such as galvanic skin response (GSR), electromyography (EMG) and electrocardiography (ECG), as well as virtual reality frameworks.

The task graphics projected by the presentation PCs (cf. *study #1*, **Section 4.1**) are based on a collective low-level perceptual decision-making task introduced by Bahrami and colleagues (2010). However, it differs in various ways. Most importantly, although participants in both frameworks engage with versions of an established perceptual task, in Bahrami's task participants first decide individually about the correct option and in a second phase collectively, after having discussed their individual impressions. On the contrary, in the task introduced here, the individual and the collective dimension is synthesized from the very beginning: participants are asked to solve the task for maximizing their *individual* score, while another person, *face-to-face* with them, is doing the same.

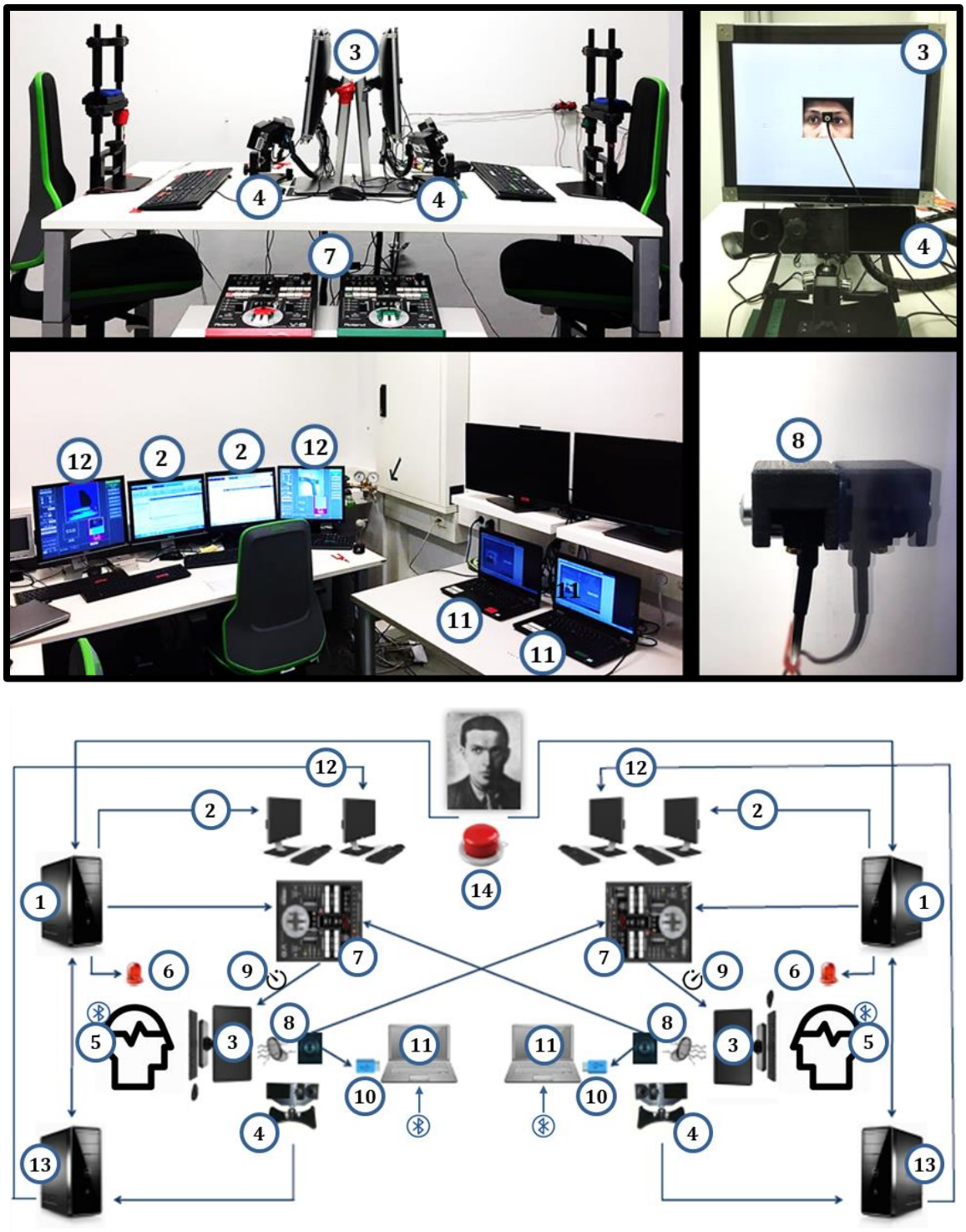

Fig. 3.2. The current design of the two-person psychophysiology platform: It includes among other elements, two eye-trackers, two micro-cameras, two analog video delay devices, two Bluetooth EEG recorders, two video mixers, four PCs, two laptops, six monitors and one dual button for the simultaneous control of the task for the two persons.

More precisely, participants were informed that they would play a difficult game in which they would see various stimuli (four sets of four Gabor patches, **Fig. 3.3**) and that they would have to select "the odd one out" (i.e. the target stimulus of slightly higher contrast), while their psychophysiological data will be being recorded. Participants were also informed that another person would be playing exactly the same game with them, which might only differ in difficulty across trials and across persons. Additionally, participants were told that at the center of the screen a live video feed of the other person's eye region while they were doing the same task will be presented and that the task graphics would appear mirrored on the two monitors. This detail was mentioned so that players feel, like in real-life, free from performing mental rotations when engaging with interpreting the other's gaze behavior. Indeed, to make the task even more intuitive, we motivated participants to imagine a real-life situation where, sitting face-to-face with another person, they observe various real objects placed between them. Additionally, they were told that while playing the game they were not allowed to speak or communicate with each other via any form of sound, ensuring that the visual channel constituted the exclusive mode of communication.

Each experimental session consisted of 4 blocks, with each block consisting of 30 trials. Each trial was divided in 3 phases, i.e., the stimulus (20 sec), the decision (5 sec) and the confidence (5 sec) phase. In the first phase of each trial, participants had the opportunity to observe the stimuli on the screen, as well as to interact with each other — in case they choose to do so. In the second phase participants were required to register their decision (i.e., left, right, up or down) via a button press of the arrow keys on the keyboard. Finally, in the third phase participants were required to register the confidence about their decision on a scale from 1 to 6, by pressing the corresponding number key on the keyboard. Before starting the actual task, participants had the

opportunity to carry out a 10-trial training task of varying difficulty. Between each block, participants could rest for a few minutes, but they were not allowed to discuss about aspects of the task.

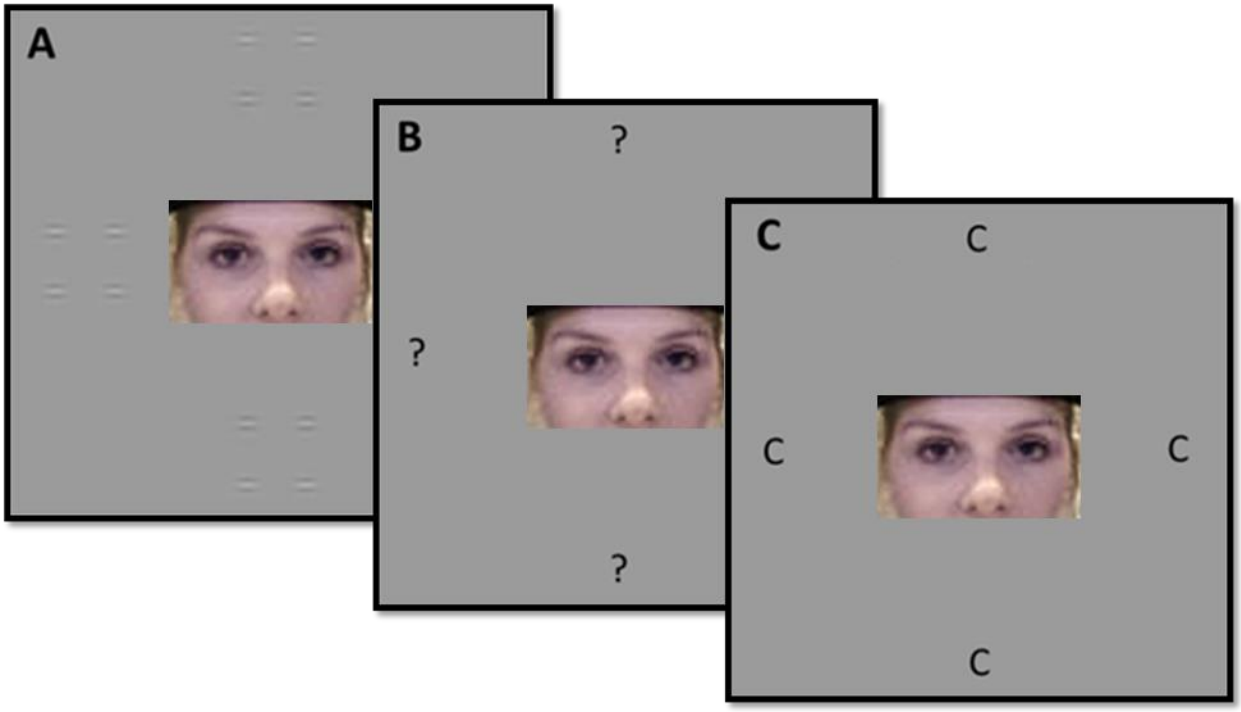

Fig. 3.3. Depiction of a trial evolution within the interpersonal decision-making task, including a video image of the co-participant, as viewed by one of the participants: the stimulus (**A**), the decision (**B**) and the confidence (**C**) phase.

To accomplish the principal aim of this study, which was to investigate the impact of the *real-time* dynamics of social interaction, we deployed two task manipulations. First, participants were informed that they will be always watching a real-time video feed of the other, although this was not always the case. In fact, only in half the blocks participants genuinely interacted with their partner (*online condition*). In the other half, without knowing, participants 'interacted' with a delayed version of the other's video — with a delay corresponding exactly to the duration of a single trial. In other words, participants in half of the trials watched the other person's behavior

exactly during the previous trial (*offline condition*; **Fig. 3.4**). This manipulation was essential, as while it completely decoupled the two persons from each other –by transforming a real-time two-way interaction into an unintentionally observational one-way pseudo-interaction– it still retained all other aspects of real-time interaction; from task context to top-down expectations about interactivity. In order to exclude potential ordering effects, each experiment's order of blocks was randomly selected from either: (i.) *Online-Offline-Online-Offline* or (ii.) *Offline-Online-Offline-Online*.

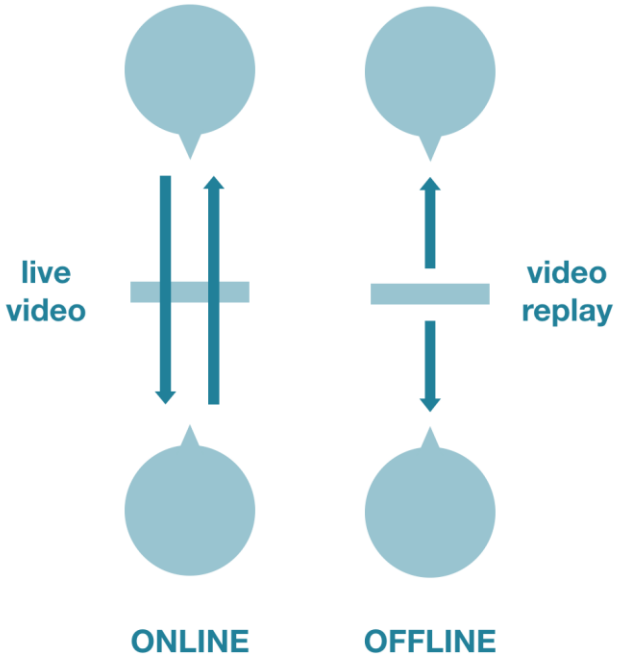

Fig. 3.4. The two social interaction conditions, namely the **online** (two-way) and the **offline** (one-way): The dynamically controlled video delay system allows for the systematic manipulation of interaction's reciprocity, by either enabling a real-time transmission of the video signal or reproducing the video feed recorded in the last trial.

Second, while participants were informed that they should identify a stimulus of slightly higher contrast, i.e., a target stimulus, in reality all 16 stimuli were always identical. This important

aspect ensured that potential effects between conditions should be exclusively attributed to the 'broken' dynamics of the social interaction, which was the only changing factor. Additionally, via this manipulation, an incidental coordination due to salient external driving factors was excluded. To make this more intuitive, let us imagine two persons on a balcony discussing face-to-face, at which point a loud firework explosion is heard: both persons turn towards the site of the explosion. Interpersonal synchrony is likely to increase at the moment of the explosion (e.g., due to simultaneous head turning), but this is just an incidental type of synchronicity, not emerging through the genuinely reciprocal dynamics of the real-time social interaction.

The results of the two-person psychophysiology study (#1) on *real-time social interaction* are presented in **Section 4.1**. In this study 62 healthy adult persons were recruited, among which 5 persons either reported at the end of the task that they might have realized the online/offline manipulation or failed to comply with the instructions and therefore the 5 respective dyads were excluded from the analysis. Thus, the analyzed sample consists of 52 persons, forming 26 same-gender dyads. Participants' ages ranged from 19 to 44 years ($M = 26.9$, $SD = 5$) with 65% identified as female. In a nutshell, this study sought to quantitatively address the role of the real-time dynamics of social interaction across various scales — from aspects of interpersonal attunement to individual cognition. More concretely, to exclude potential external driving factors, such as shared human anatomy and task structure we compared genuine real-time social interaction with a pseudo-condition of randomly paired participant data. In order to control for strictly reciprocity-specific effects we deployed the one-way experimental condition, discussed above. In the analyses presented here, we consider interpersonal attunement in gaze behavior and decision-making, as well as individual metacognition, namely confidence about one's own decision and collaborative intentions.

3.1.2. Analysis

Here, we concisely present the data analysis pipeline used in Study #1 (cf. **Section 4.1**), which operationalize *real-time* social interaction across scales, quantifying interpersonal gaze behavior and decision-making, as well as confidence and collaboration intention. Starting with the shortest time scale, we firstly quantify interpersonal attunement in dyadic gaze behavior (msec), as 'peak multi-lagged correlation' averaged across task trials. We unpack this in detail below.

1. The eye-tracking signals are preprocessed, i.e., they undergo:
    a. epoching for isolating the first 20 secs of each trial (stimulus presentation phase).
    b. z-scoring (i.e., subtracted by the mean and divided by the standard deviation).
    c. missing value interpolation (e.g., eye-blinking), using a spring metaphor (D'Errico, 2019).
    d. smoothing with a Gaussian filter for a 100msec window, which is the estimated duration of a minimum gaze fixation (cf. Manor and Gordon, 2003).
2. The gaze trace of Person-A and Person-B is correlated across multiple lags with a step of 1msec (from -1 to +1sec) along the horizontal (r_x) and vertical (r_y) dimension, as gaze cueing effect typically disappears after approximately 1 sec (cf. Friesen & Kingstone, 1998).
3. The average value across the two dimensions is calculated (r_{xy}), for every lagged (i.e., shifted) correlation.
4. The maximum value (peak correlation) of all averaged-across-dimensions correlations is calculated (r_p).
5. The average value of the peak correlation across trials is calculated for each condition (r_a).This is the metric which operationalizes interpersonal attunement in gaze behavior.

Secondly, we quantify interpersonal attunement in decision-making (scale of seconds) as the percentage of common decisions between Person-A and Person-B across all trials of the task.

Please note that expected value of chance for this metric should be approximately 25%, as there are 4 possible answers for each person (i.e., left, right, top, and bottom). That means that the two persons have the possibility to agree in 4 out of the total 16 joint options (i.e., left-left, right-right, top-top and bottom-bottom). Thirdly, we quantify individual metacognition as the degree of confidence (from 1 to 6) the person places on each decision (scale of seconds), averaged across trials. Fourthly, we quantify dyadic collaborative intentions (from 1 to 6) at the end of the task (scale of minutes).

Then, we compare our conditions, *i.e., a. online, b. offline and c. pseudo interaction,* across modalities, by means of a one-sided Wilcoxon signed rank test –if not otherwise specified– expecting the highest interpersonal attunement and confidence in the online and the lowest at the pseudo condition. Finally, we examine whether gaze coordination –the primary communication channel– related to the rest of the modalities, by means of one-sided bivariate Spearman correlation. Finally, comparisons of correlations are implemented using the Fisher's Z test, while Bonferroni's method is used to correct for multiple comparisons within modalities when determining respective significance thresholds.

3.2. Interpersonal misattunement in *real-life* social relations

3.2.1. Framework

Two-person psychophysiology, as presented in the sections above, enables a high-resolution analysis of *real-time* social interactions, across a range of relatively short time scales (from msec to minutes). Thus, however excellent an experimental control such a lab-based platform

provides, it still captures a certain part of the multiscale real-life whole of human social interaction. Therefore, in this section, we describe means for shining light in relational processes unfolding over larger time scales (months to decades) in *real-life*. Below we will describe how we operationalize *real-life* social relations, considering autism as a paradigm example.

While social interaction difficulties, as discussed in **Chapter 2**, lie at the core of psychiatric conditions such as autism, prominent research paradigms remain methodologically individualistic, even when it comes to research on inherently relational aspects, such as friendship quality (e.g., Baron-Cohen & Wheelwright, 2003). Departing from an individualistic paradigm, two-person psychophysiology aims at situating psychopathology back where it unfolds and manifests itself, in social relations *between* persons. To this end, it moves from studying *individual* autistic traits (cf. Baron-Cohen *et al.*, 2006) to *interpersonal mismatch* thereof and from *abstract individual* impressions about friendship (cf. Baron-Cohen & Wheelwright, 2003) to *concrete* impressions of *one another* within specific relations through adapting established questionnaires (i.e., Friendship Quality Scale; Thien *et al.*, 2012).

The results of the two-person psychophysiology study (#2) on *real-life social relations* are presented in **Section 4.2**. In this study 144 healthy adults were recruited, forming 72 same-gender dyads. Participants' ages ranged from 19 to 50 years (M =25.5, SD =4.8) with 53% identified as female. This study provides a direct validation of the dialectical misattunement hypothesis, asking: Is it the autistic traits *per se* –or a mismatch thereof– that primarily predict core aspects of interpersonal relations in real life?

3.2.2 Analysis

In order to quantitatively address this question, we quantify interpersonal attunement in real-life relations (**Section 4.2**; Study #2; scale of months to decades), via calculating dyadic factors of friendship. To this end, we adapt the Friendship Quality Scale (Thien, *et al*., 2012), so that it can be used for measuring the impressions of a specific relation — as the original version targeted impressions of an individual about their friendships in general. More concretely, we calculate the averaged relational impressions of two friends about one another with regard to safety, closeness, acceptance and help. Then, to test the dialectical misattunement hypothesis (**Section 2.2**; Bolis *et al.*, 2016, 2017), we relate these dyadic relational factors not only to dyadic autistic traits (i.e., averaged within dyads), but also the mismatch thereof (i.e., absolute value of the AQ difference between the two persons of a dyad), by means of one-sided Bivariate Spearman correlations. Bonferroni's method to correct for multiple comparisons is also used when determining respective significance thresholds.

All our studies received approval from the institutional ethics committee of the medical faculty at the Ludwig-Maximilian University of Munich. Additionally, all participants read and signed a declaration of consent. Participants also read detailed information about the course of the studies, including potential side effects and safety aspects, as well as written instructions. In what comes next, we present the results of our two empirical studies on real-time social interactions (**Section 4.1**) and real-life social relations (**Section 4.2**). Please note that these studies present only a part of the dataset collected during this thesis. Ongoing research directions will be presented in forthcoming research articles (cf. **Section A.1.5**).

4. Empirical work — *'It takes two to tango'*

"The vast majority of contemporary psychological investigations write out the last decimal point with great care and precision in answer to a question that is stated fundamentally incorrectly."

Lev Vygotsky

4.1. Interpersonal attunement study (#1) — *Interaction matters: the whole is more than the sum of its parts in real-time social interaction*

4.1.1. Multiscale effects: Stronger interpersonal attunement in gaze behavior and decision-making as well as higher confidence during *real-time* social interactions

Based on our theoretical and methodological discussion (cf. **Section 2.1 & 3.1**), we hypothesized that *real-time* social interactions entail processes inherently different from observational social situations, across various time scales and modalities, even beyond individual awareness. To test this, we, firstly, quantified and compared interpersonal attunement in gaze behavior across the three conditions, i.e., *a. online, b. offline and c. pseudo* interaction.

Interpersonal attunement in gaze behavior (**Fig. 4.1**: left) was found to be significantly higher in the online compared to the offline condition: $Z = 3.45$, $p < 0.001$ (*online*: $M = 0.183$, $SD = 0.029$; *offline*: $M = 0.169$, $SD = 0.022$). Furthermore, a comparison of interpersonal attunement in gaze behavior of the *pseudo* condition ($M = 0.155$, $SD = 0.011$) with the *online* and *offline* conditions revealed a significant difference with both the *online* and *offline* conditions showing higher interpersonal attunement compared to the *pseudo* condition: $Z = 4.01$, $p < 0.001$ and $Z = 3.00$, $p = 0.001$ respectively.

Secondly, we investigated interpersonal attunement in decision-making as the percentage of common decisions, i.e., decisions, in which both participants selected the same target. Interpersonal attunement in decision-making (**Fig. 4.1**: right) was –similarly to gaze behavior– found to be significantly higher in the online compared to the offline condition: $Z = 3.10$, $p < 0.001$ (*online:* $M = 38.5\%$, $SD = 19.8\%$; *offline:* $M = 30.1\%$, $SD = 9.8\%$). Furthermore, a comparison of interpersonal attunement in decision-making of the *pseudo* condition ($M = 24.3\%$,

SD = 2.2%) with the *online* and *offline* conditions, similarly, revealed a significant difference with both the *online* and *offline* conditions showing higher interpersonal attunement compared to the *pseudo* condition: Z = 3.90, p < 0.001 and Z = 3.09, p = 0.001 respectively.

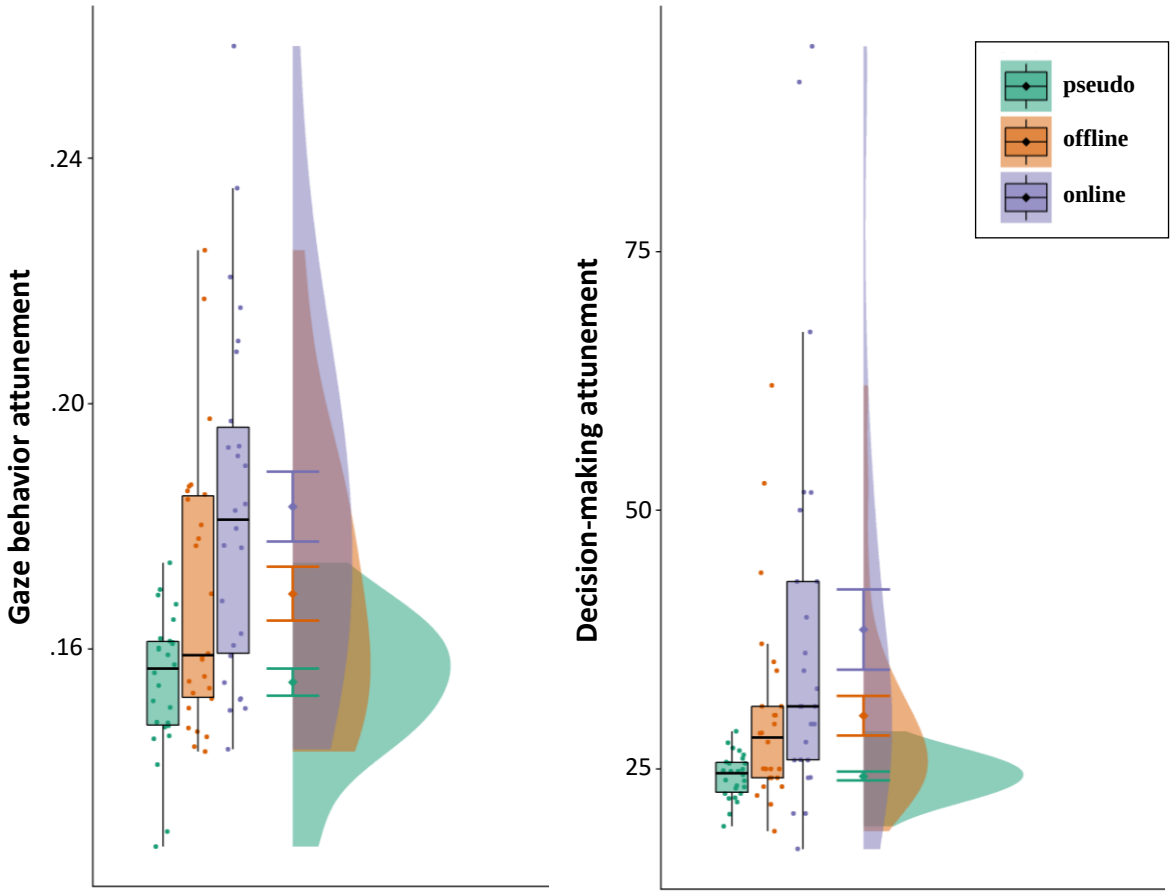

Fig. 4.1. Multilevel effects: Interpersonal attunement in gaze behavior and decision-making was significantly higher in online (**purple**) than offline (**orange**) interaction, while these kinds of interpersonal attunement in both conditions were significantly higher than chance, that is any incidental attunement within pseudo-pairs (**green**). Dot plots represent raw data, boxplots display sample median alongside interquartile range, 'whisker' error-bars depict mean and standard error, while distribution 'clouds' portray a smoothed version of the histogram (cf. rain cloud plots; Allen *et al*., 2018).

Notably, results in the pseudo-condition did not significantly differ from theoretical chance (i.e., 25%; Z = 1.51, p = 0.13; two-sided test). Additionally, the subjective confidence of participants

about the correctness of their own decisions was found to be significantly higher in the online as compared to the offline condition: Z = 1.87, p = 0.03 (*online:* M = 2.193, SD = 1.067, *offline:* M = 2.052, SD = 0.918).

4.1.2. Interscale effects: Dyadic gaze behavior interrelates with interpersonal decision-making, confidence levels and collaborative intentions

Subsequently, we examined whether interpersonal attunement in gaze behavior –as gaze coordination was the sole channel of interaction between participants– correlated with decision-making attunement, individual confidence ratings and collaborative intentions by means of bivariate Spearman correlation analyses. Interpersonal attunement in gaze behavior was found to positively and significantly correlate with interpersonal attunement in decision-making in both online and offline condition (**Fig. 4.2**; *online:* r(24) = 0.82, p < 0.001, *offline:* r(50) = 0.43, p < 0.001). Notably, the correlation was significantly higher in the online than in the offline condition, as tested by means of Fisher's Z test (p = 0.003).

Additionally, interpersonal attunement in gaze behavior was positively and significantly correlated with individual confidence ratings in the online condition (r(50) = 0.29, p=0.018), but not in the offline (r(50) = 0.20, p = 0.075). Finally, the reported level of collaborative intention at the end of the task, positively and significantly correlated with interpersonal attunement in gaze behavior in both the online and offline condition (r(49) = 0.42, p = 0.001 , *offline:* r(49) = 0.43, p < 0.001).

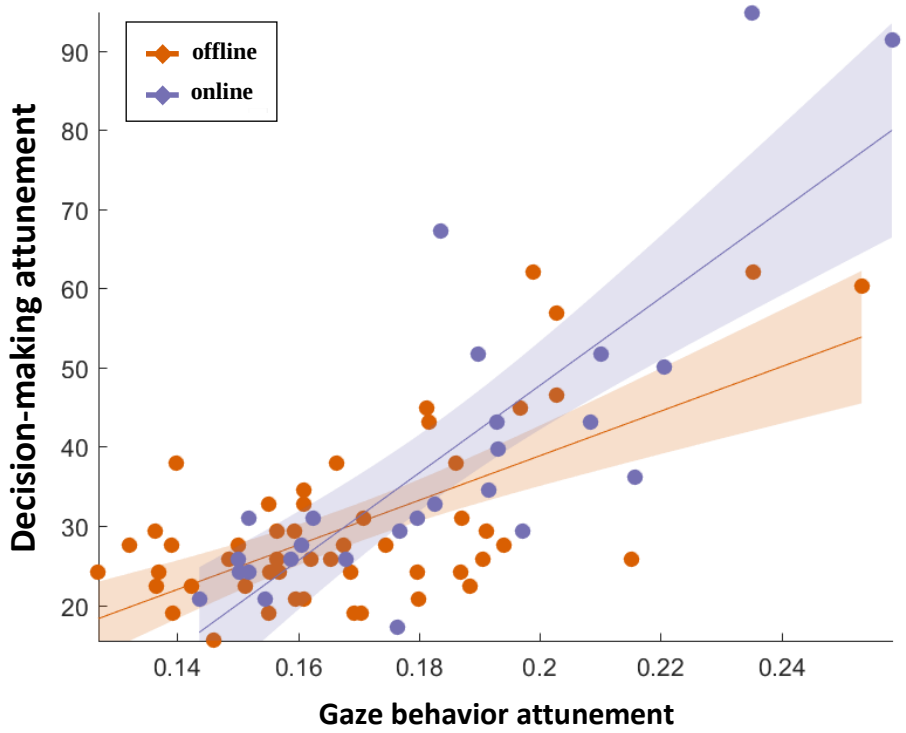

Fig. 4.2. Interscale effects: Interpersonal attunement in gaze behavior during the online condition (**purple**) is positively and more strongly correlated with interpersonal attunement in decision-making as compared to the offline condition (**orange**). Regression lines and the 95% confidence bands were drawn with the bounded line toolbox (Kearney, 2020).

4.2. Interpersonal misattunement study (#2) — *Beyond the individual: mismatch of autistic traits –not traits per se– predicts interpersonal misattunement in real-life social relations*

After examining interpersonal attunement in real-time social interaction across time scales and modalities, we continued with investigating it within *real-life* social relations. More concretely, we aimed at examining whether *autistic traits per se* or the *mismatch* thereof –as the dialectical misattunement hypothesis predicts– is what most optimally predicts the degree of interpersonal attunement in real-life social relations (cf. **Section 2.2; 3.2**). To this end, we quantified and

tested the two above-mentioned cases across four relational factors, namely interpersonal *safety, closeness, acceptance and help*.

Bivariate Spearman correlations between *dyadic autistic traits*[10] and *relational factors* while all yielded negative, none achieved statistical significance after Bonferroni correction (**Fig. 4.3**: left column, 1st to 4th row respectively; interpersonal *safety:* $r(70) = -0.21$, $p = 0.04$, *closeness:* $r(70) = -0.13$, $p = 0.14$, *acceptance:* $r(70) = -0.21$, $p = 0.04$, *help:* $r(70) = -0.25$, $p = 0.018$).

On the contrary, the *interpersonal difference* of autistic traits found to significantly and negatively correlate with the interpersonal *closeness, acceptance* and *help*, but not *safety*, which did not survive Bonferroni correction (**Fig. 4.3**: right column, 1st to 4th row respectively; *safety:* $r(70) = -0.24$, $p = 0.024$, *closeness:* $r(70) = -0.30$, $p = 0.005$, *acceptance:* $r(70) = -0.32$, $p = 0.003$, *help:* $r(70) = -0.37$, $p < 0.001$. Taken together, while all examined correlations yielded negative, it was not autistic traits *per se,* but the mismatch thereof that it was significantly correlated to relational factors.

[10] autistic traits averaged within dyads

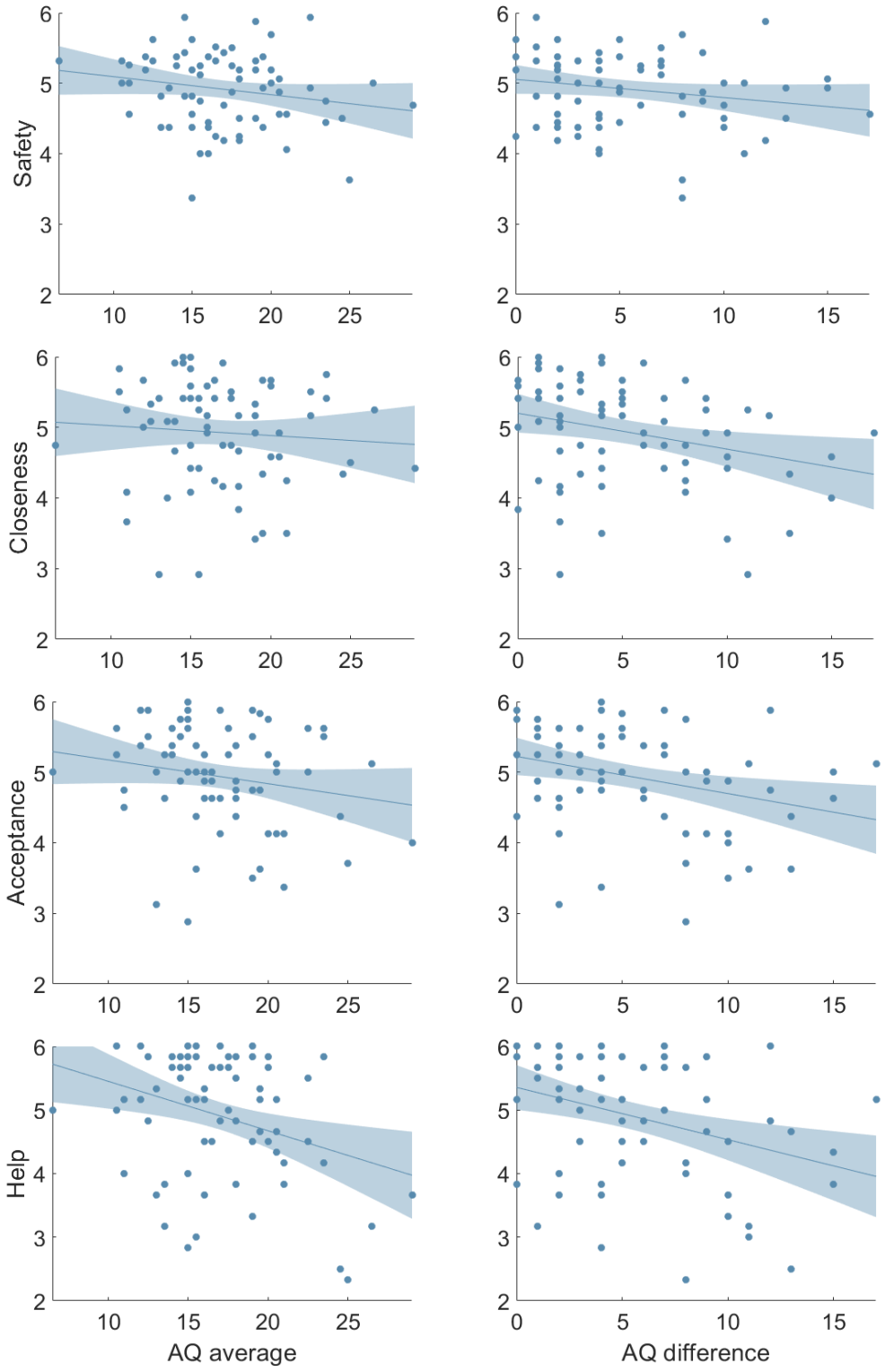

Fig. 4.3. Relational factors (yy') and autistic traits (xx'): correlation of relational factors, i.e., safety (1st row), closeness (2nd row), acceptance (3rd row) and help (4th row) with dyadic autistic traits (left column) and absolute difference thereof (right column).

5. Discussion — *'The truth is the whole'*

"What is meaningful cannot in fact be isolated."

"We achieve understanding within a circular movement
from particular facts to the whole that includes them
and back again from the whole thus reached to
the particular significant facts."

Karl Jaspers

5.1. Interpersonal attunement in and through social interaction

After completing our theoretical, methodological and empirical analyses, we now revisit and address them integratively. We commenced our theoretical analyses by presenting the *dialectical attunement* account (cf. **Section 2.1**; Bolis & Schilbach, 2018b), which heavily leans on the Vygotskian approach and principally the commitment on a dialectical, naturalistic and systematic study of the human mind, with the ultimate goal of tracing back its social origins. Dialectical attunement construes human becoming as the multiscale interplay between (social) internalization and (collective) externalization in and through culturally mediated social interaction. More concretely, we operationalized our conceptualization leaning on theories of predictive processing and active inference, attempting to unveil their entangled interrelation through a dialectical prism and while doing so, situate them in the sociocultural realm. Indeed, modern computational neuroscientific approaches, while having been developing independently, speak to a Vygotskian perspective, which had, interestingly, anticipated such integrative developments, e.g., "no specific function is ever connected with the activity of one single brain center. It is always the product of the integral activity of strictly differentiated, hierarchically interconnected centers" (Vygotsky, 1917-1934/1997, p. 140). What is core to the Vygotskian cultural historical approach, though, is the integrative view on human mind development beyond the individual: “a primacy of the social such that all higher psychological functions are social relations between people before they are functions. As a result, human development occurs in and as sociogenesis” (Roth, 2016).

In this light and leaning on recent second-person (e.g., Schilbach *et al.*, 2013) and enactivist proposals (e.g., De Jaegher & Di Paolo, 2007), we emphasized the fundamental and constitutive role of *real-time* social interaction, but based on hierarchical Bayesian accounts we also

demonstrated how the '*here and now'* concreteness of the prototypical face-to-face encounter can be abstracted and extended across higher scales. In a nutshell, dialectical attunement illustrates how we co-construct each other and create common (cultural) ground *in and through* social interaction, which literally shapes our hierarchical bodily –also interbodily– structures, on hand in future interactions or privately.

Dialectical attunement, is inherently in line with philosophical traditions which have emphasized in the past the indispensable role of the 'other' in human experience, such as the broad field of phenomenology. A prime example can be found in the work of Merleau-Ponty, cf. his notion of intercorporeality, "between this phenomenal body of mine and that of another … there exists an internal relation which causes the other to appear as the completion of the system" (as cited in Gallagher, 2014; Merleau-Ponty, 1962, p. 352).

Furthermore, dialectical attunement resonates, at least in part, with various other evolving perspectives to the free energy principle that attempt to go beyond the individual brain. For instance, Friston and Frith (2015) interpreted dyadic communication in predictive coding and active inference terms, as a repeated exchange of predictions about each other, via sequentially attending (cf. listening) and attenuating (cf. talking) sensations. On the other hand, Bruineberg and colleagues (2018), viewing the free energy principle from an enactive and ecological perspective, proclaimed 'the anticipating brain is not a scientist', suggesting a more holistic perspective. Furthermore, Ramstead and colleagues (2018) articulated a multiscale account of the free energy principle, grounded in spatially and temporally nested Markov blankets. Fotopoulou and Tsakiris (2017), drawing on predictive processing and active inference, discussed the social origins of interoception (for a detailed comment see **Appendix A.2.3**; Bolis & Schilbach, 2017). Additionally, Boonstra and Slagter (2019) extended the dialectical

discussion of the free energy principle focusing on a Hegelian perspective. More recently, Veissière *et al.* (2019) situated the free energy principle within anthropology, providing a discussion of cognition and culture which nicely aligns with our dialectical attunement proposals (for a detailed comment see **Appendix A.2.4**; Bolis & Schilbach, 2019). Such conceptual developments are of immense importance with far-reaching implications in various fields of research and practice. However, they will remain an incomplete effort unless they are empirically grounded in reality.

To this end, this thesis establishes a novel *two-person psychophysiology* platform (**Section 3.1**; also **Appendix A.2.2**; Bolis & Schilbach, 2018a), which allows for two individuals to perform perceptual and decision-making tasks, while interacting with each other in –systematically controlled– synchrony. The social interaction loop is mediated by micro-cameras, fixed between the eyes of each person. The possibility of direct eye-contact highly impacts several communicative factors, such as interest, trust, interaction patterns, social feedback and presence, as well as the overall communication quality (Kuster *et al.*, 2012). A study showed that a video-mediated system providing the possibility of eye-contact enables interpersonal behavior similar to face-to-face communication, as opposed to one that does not (Mukawa *et al*., 2005). Indeed, in a series of pilot experiments we confirmed the intuition that a slightly misplaced camera, such as the ones standard laptops are equipped with, fails to invoke the intense feeling of mutuality that eye contact inflicts.

Two-person psychophysiology builds upon and synthesizes various prominent empirical setups, while in doing so it complements and goes beyond them in various ways. First of all, it draws inspiration from seminal ‘double video’ works (e.g., Murray & Trevarthen, 1985), which enabled interaction of an infant either in real-time or with a videotape of her mother. Our platform retains

the principal idea, which is the systematic ‘breaking’ of interpersonal contingencies for studying its fundamental mechanisms. Yet, it dynamically controls the offline condition of these ‘broken dynamics’ to equalize it as much as possible to the online one. That is, the two non-verbally communicating participants are positioned face-to-face while they see each other in exactly the previous trial, via an analog video-delay system (cf. Koike *et al.*, 2019). This is crucial for making the manipulation transparent to our adult sample, thereby strictly controlling for top-down effects due to a potential awareness of the situation. The face-to-face arrangement is also pivotal for enhancing social presence, “the degree to which a person is perceived as “real” in mediated communication”, which is a strong predictor of interaction outcomes (Gunawardena & Zittle, 1997).

At the same time the platform performs precise recordings in synchrony across persons and modalities, such as eye tracking (cf. Barisic *et al*., 2013), video recording and EEG (cf. Dumas *et al*., 2012; Leong *et al.*, 2017), while embedding an established graphical environment for psychophysiological research (cf. Bahrami *et al.*, 2010) within real-time social interaction. Taken together, this two-person psychophysiology platform achieves integrating multiple modalities, being recorded in excellent quality and sampling rate. For instance, it allows for interpersonal gaze behavior analysis at 1,000Hz, a rate about 20-30 times faster compared to the latest developments in the area (e.g., Hessels *et al.*, 2018; Rogers *et al.*, 2018), which is important for a fine-grained analysis of real-time social interaction.

Yet, one might still wonder, *why all the fuss anyways?* Indeed, the field has not decisively concluded the importance of *real-time* social interaction in studying human behavior and cognition. For instance, Schönherr and Westra (2017) claim to have (conceptually) shown that “ersatz interactivity works just as well as the real thing”, by ‘real thing’ implying genuine, real-

time social interaction. In other words, they doubt about "how much we really learn when we try to directly compare interactive and non-interactive contexts", in fact, arguing that such an endeavor is not informative. Schönherr and Westra get more specific when suggesting "at least in gaze paradigms, it is more significant whether a subject believes that she engaged in an interaction rather than whether she is actually engaged in an interaction". In what follows, we discuss the results of our two-person psychophysiology study (#1) on *real-time* social interaction to clarify these very points, strongly contradicting such skepticism empirically.

In short, our results (**Section 4.1**) demonstrate that the reciprocal dynamics of the interaction, in and of themselves, enhance interpersonal attunement, in both gaze behavior and decision-making — even in situations where individuals are not explicitly asked to coordinate with each other. Additionally, individuals appear statistically more confident about their own decisions when interacting in *real-time* with another person. Crucially, such an enhancement in interpersonal attunement and confidence extends beyond individual awareness.

Furthermore, our results quantitatively manifest that different scales of social interaction are tightly interrelated. That is, interpersonal attunement in gaze behavior during real-time social interaction is statistically interrelated with interpersonal attunement in decision-making, as well as individual confidence and collaborative intentions. More concretely, in our experiments, the higher the dyadic gaze coordination was, the more common decisions the dyad made, the more confident the individuals appeared and the more collaborative they reported to be. Importantly, interpersonal gaze behavior was found to drive interpersonal decision-making in a higher degree during the online (two-way real-time) than the offline (one-way) social interaction. Additionally, interpersonal attunement in gaze behavior was found to significantly correlate with individual

confidence only in the real-time condition. Finally, interpersonal attunement in gaze behavior was found to significantly correlate with collaborative intentions in both conditions.

After quantitatively presenting the empirical evidence we will now qualitatively discuss the two types of interpersonal attunement in our study, namely attunement in gaze and in decision-making. First, with regard to gaze behavior, in certain cases, participants reported a feeling of uneasiness during parts of the task, related to an impression that the other was not responsive to them anymore. Participants, typically, attributed this feeling to disturbed dynamics internal to the interaction, such as relational and psychological factors, e.g., along the lines *'the other got tired, bored or uninterested in me'*. While in our case, this judgment was incorrect, it was still, perhaps, the most rational one — considering that externally manipulated interpersonal contingencies are rare in real life, to say the least. In fact, this observation might be crucial to an eventual multiscale understanding of interpersonal attunement. As we discussed, all participants remained unaware of the use of both online and offline conditions. So, how was the higher interpersonal attunement in gaze behavior during the real-time interaction achieved? Here, we view a dual role of the interpersonal dynamics, consisting of both *bottom-up* and one *top-down* processes that are tightly interrelated (cf. Frischen *et al.*, 2007).

On one hand, the bottom-up or implicit role of interpersonal dynamics includes subliminal mechanisms of attention orientation. For instance, gaze observation can trigger both covert and overt automatic orienting responses, as well as a tendency to execute saccades similar to the observed behavior, following another person's gaze (Frischen *et al.*, 2007). With regard to neural function, such type of exogenous control of attention orienting has been thought of in terms of a posterior attention system, including subcortical structures such as the pulvinar and the superior colliculus. Typically, these gaze cueing phenomena have been studied within the context of clear

roles of an initiator and a responder (cf. Frischen *et al.*, 2007; Schilbach *et al.*, 2013). However, people use gaze not only to make sense of others, but also to interact, while various gaze-based processes are interactively constituted (cf. Schilbach, 2015). Crucially, in our naturalistic setting which enables interpersonal reciprocity, the initiating and responding processes can be inextricably linked. This can potentially lead to interactively enhanced gaze orientating process in a feedback-loop fashion between the two persons.

On the other hand, the explicit role of interpersonal dynamics includes consciously perceived impressions and voluntarily controlled gaze-based processes. For instance, an 'initiating' saccade may involve closer tracking of the interactional process, whereas a 'responding' reaction might elicit stronger attention to the emotional dimension of the interaction (Schilbach *et al.*, 2013). When it comes to brain function, such processes are expected to dynamically recruit areas of the so called mentalizing and reward neural networks. Additionally, endogenous control of attention orienting is expected to rely more on cortical areas in anterior and posterior regions of the brain (Frischen *et al.*, 2007). Additionally, the reciprocal and dynamic interpretation, adoption and modification of the other's gaze behavior might gradually allow for a more effective communication and jointly constructed symbolic patterns, intentions, motivations and eventually decisions.

This brings us to the second point of our qualitative discussion, namely the interpersonal attunement in decision-making. Participants, as discussed, appeared more attuned to each other and more confident about their own choices when interacting in real-time, despite the absence of an objectively correct option. So, how was that possible? We propose two potential mechanisms: (i.) *interpersonal construction of consensus* and (ii.) *interpersonal construction of experience*. These mechanisms are not thought of as being mutually exclusive, but rather as interlinked.

Interpersonal construction of consensus refers to processes of informational conformity (cf. Sherif, 1935), which involve the reliance on the opinions of others, when facing ambiguous situations. Notably, participants were informed that the difficulty might differ across persons. Considering that participants placed minimal confidence in their own perceptual impressions due to the nature of the task (no target to identify), from a Bayesian perspective, it does make sense for one to take into account the other's opinion. This kind of *reciprocal informational conformity* we call here *interpersonal construction of consensus*.

However, this reciprocal influence may, indeed, go deeper than the decision-making level, into the actual perceptual impression formation. It has been suggested that an attended object might appear to have a higher contrast (cf. Carrasco, 2004; Helmholtz, 1866; James, 1890). In fact, an experiment using visual stimuli almost identical to that in our study has demonstrated that (auditory) cueing of attention boosted the perceived contrast of a target stimulus, accompanied with an enlarged response in contralateral visual cortex, within 100 ms after target onset. Taken together, this experiment suggests that attention does enhance perceived contrast of visual stimuli and, in fact, via boosting early sensory processing in the visual cortex (Stoermer *et al.*, 2009). Bringing together our discussion about the endogenous and exogenous attention orientation due to gaze cueing and the attention-based enhancement of perceived contrast, we suggest that our participants may indeed have experienced higher contrast in their commonly attended patches, which in turn may have boosted the respective interpersonal attunement in decision-making. This kind of *reciprocal contrast enhancement* is what we call here *interpersonal construction of experience*.

Now, let us slightly reformulate our core finding integratively: at certain spatiotemporal scales, individuals converged in a relatively confident consensus –in and through social interaction–

even in the absence of ‘objective truth’. Here, one can possibly draw parallels with sociological definitions of real-life phenomena, pertinent to culture, ethics, religion and science. Taken together, these are significant empirical findings as they show that interpersonal dynamics in one modality not only drive the dynamics in another but also feed-back to the level of individual metacognition, with relevance to real-life phenomena. In a nutshell, these results unequivocally demonstrate that genuine social interaction is more than simply the belief one is doing so.

This two-person psychophysiology study has also certain limitations that should be considered. First, while the quantitative and qualitative discussion presented above, constitutes a crucial step toward a mechanistic understanding of the multiscale dynamics of social interaction, to quantitatively elucidate the actual mechanisms, further control experiments and computational analyses will be needed. To this end, we have recently run an EEG version of this study. Second, as this study conveyed within the context of a research lab, relevant real-life experiments will complement our conclusions. Additionally, while the two persons do sit face-to-face in close proximity, the social interaction was non-verbal and mediated through monitors. While this was indispensable to the question this study asked, unmediated face-to-face interactions, which might involve verbal communication and/or touching, might yield further insights. Additionally, only same gender dyads of adult individuals and of relatively homogeneous cultural background were recruited. To address these points, we have planned similar experiments in Japan, including the use of cross-gender dyads. Furthermore, only a part of the wide-ranging time scale was captured within the 2 hour duration of the experiment.

These empirical results nicely resonate and go beyond interactive psycholinguistic accounts, such as the ‘interactive alignment’ model, which postulates that “in dialogue, the linguistic representations employed by the interlocutors become aligned at many levels, as a result of a

largely automatic process" (Garrod & Pickering, 2009; Pickering & Garrod, 2004) and 'interpersonal synergy', which emphasizes dialog as an "emergent, self-organizing, interpersonal system capable of functional coordination" (Fusaroli *et al.*, 2014). Taken together, this two-person psychophysiology study, demonstrates that interpersonal and intrapersonal processes, inextricably linked, do differ in *real-time* social interaction due to its reciprocal dynamics. This constitutes a comprehensive empirical validation and extension of second-person and enactivist approaches to the study of human behavior and cognition (cf. De Jaegher & Di Paolo, 2007; Schilbach *et al.*, 2013). In doing so, this study opens up completely new avenues toward a principled understanding of human becoming as interpersonal attunement (cf. Bolis & Schilbach, 2018b).

5.2. Psychopathology as interpersonal misattunement

After discussing interpersonal attunement in and through social interaction, we now turn to our analyses of situations where such an attunement may unfold atypically. In our theoretical considerations, taking autism as a paradigm example, we put forward the *dialectical misattunement hypothesis*, which views psychopathology as a dynamic and cumulative interpersonal mismatch, rather than an *a priori* disorder of the individual brain (cf. **Section 2.2**). In other words, dialectical misattunement is thought of as "disturbances in the dynamic and reciprocal unfolding of an interaction across scales, resulting in increasingly divergent prediction and (inter-)action styles" (Bolis *et al.*, 2017). Here, we viewed two inextricably linked processes: at the collective level, weak interpersonal coupling leads to increasing interindividual incompatibilities in generating and expressing (social) expectations; while at the individual level,

such diverging expectations lead to weak coupling with others in social interactions. Critically, these processes can increasingly enhance (or diminish) each other in a feedback loop, where even minor changes at the individual result in corresponding changes at the collective level and *vice versa*.

Importantly, the dialectical misattunement hypothesis makes concrete prognoses with regard to social interaction and interpersonal relations in real life, amenable to empirical validation. In doing so, it motivates a novel research line for experimentally studying psychopathology within two-person (and generally collective) psychophysiology across scales and contexts. That is, the dialectical misattunement hypothesis proposes moving beyond merely contrasting –*a priori* and largely arbitrarily defined– groups of individuals toward scientifically studying the multiscale dynamics of social interaction between persons. In order to push beyond the 'healthy' vs 'patient' dichotomy, as a first step, it considers interactions not only within neurotypical, but also mixed dyads or groups and crucially between individuals with a certain condition.

A focus on interactions *between persons with a psychiatric condition*, such as autism, will have a dual benefit. Firstly, tapping into interpersonal mismatch processes might result in a more precise analysis of communication breakdown mechanisms, beyond an exclusive neurobiological etiology. Such a research approach would align with more holistic perspectives of impairment as "profoundly bio-social, that is, shaped by the interaction of biological and social factors, and ... bound up with processes of socio-cultural naming" (Thomas, 1999, p.43; cited in Graby, 2012). Secondly, by taking atypical social interaction seriously, in terms of both research and practice, voice is given to the most relevant part of the population, namely those with the psychiatric condition themselves. As Milton (2014) powerfully put it, “autistic people will need to be

utilising their voices in, claiming ownership of the means of autistic production, and potentially celebrate the diversity of dispositions within and without the culture".

Nevertheless, the dialectical misattunement hypothesis eventually questions *a priori* dichotomies altogether, aiming at eventually breaking free from acknowledged weaknesses of prominent nosological disease models[11]. As van Praag inspiringly questioned (2000), "are the diagnostic constructs we are used to working with valid and clinically relevant or, rather, pseudo-entities; artefacts of a rigidly applied nosological doctrine". Dialectical misattunement attempts to overcome such pitfalls by studying psychopathology in terms of interpersonal mismatch along a continuum. With regard to autism, a straightforward way to operationalize this is via studying interpersonal difference of autistic traits, rather than merely individual traits (cf. **Section 3.2**). Ultimately, such a research line points toward studying psychiatric conditions, transdiagnostically, as interpersonal distance in a multidimensional feature space. Here, the hypothesis predicts that in certain scales "interactions within homogeneous dyads are expected to appear smoother compared to heterogeneous dyads. Additionally, tuned interactions of either homogeneous or heterogeneous dyads should appear as most effective" (Bolis *et al.*, 2017). If these hypotheses prove accurate, the definition of a psychiatric condition should be –not only theoretically but also practically– reconsidered as relative to the 'other' and broadly the social context.

The dialectical misattunement hypothesis by definition aligns with accounts that emphasize the social dimension of the human condition. First of all, it draws on the Vygotskian approach to

[11] *"This disease model conceives psychiatric conditions as discrete entities, with a particular pathophysiology and predictable relations between phenomenology, course and outcome"* (van Praag, 2000).

psychopathology, as well as second-person neuropsychiatry, which both emphasize the critical role of social interaction in the development, manifestation and clinically unveiling of psychopathology. Dialectical misattunement draws also inspiration from the philosophical and methodological underpinnings of enactivism and the focus on reconsidering psychopathology beyond the individual in both conceptual and empirical regards. Additionally, dialectical misattunement, underlining the critical role of social expectations in human communication, shares common ground with various sociological approaches such as ethnomethodology, which for instance suggests that "interactions between individuals involve assumptions of normalcy, such as the assumption that others will behave in expected ways, and that when ambiguous meanings are found, they will either be deemed irrelevant to the interaction, or will be immanently explained" (Milton, 2013; cf. Garfinkel, 1967; Cicourel, 1974).

Dialectical misattunement also resonates with the social model of disability in acknowledging systemic asymmetries and social exclusion as constitutive factors, as well as the distinction between impairment and disability (cf. Paley, 2002). The social model thinks of impairment as the straightforward result of psychophysiological variation, which is transformed into a disability only when it is not sufficiently accommodated by society. The dialectical misattunement hypothesis nicely aligns here, drawing directly from Vygotsky's notion of primary and secondary difficulties in psychopathology: "it is not the primary difficulties, which are directly linked to the physical condition, that are crucial to a child's development, but rather the secondary ones, which relate to an exclusion from sociocultural activities, which other children freely participate in" (Bolis *et al.*, 2017).

Nevertheless, we still maintain that considering exclusively the social dimension of psychopathology of course leads to an incomplete understanding and accommodation and that

psychophysiological mechanisms should be addressed in parallel. Here, dialectical misattunement resonates with the neurodiversity paradigm which, acknowledging the need to address certain aspects on an individual basis, it still views psychopathology as a human variation rather than an *a priori* disorder: "We need a world that is much more inclusive of a broader range of individuals. Does this mean that we should not be engaged in trying to ameliorate the many challenges associated with being autistic? Of course not. What it does mean is that, first, we should target our efforts towards the real challenges we face, rather than towards a broader, nebulous concept of 'curing' autism that is offensive to many of the people that it aims to benefit" (Ne'eman, 2010). With regard to the case of autism, the most relevant account to our hypothesis within the sociological field is the 'double empathy problem', which questions the ontological status of autism as articulated in prominent cognitivist accounts in favor of an interactional and relational one: "a disjuncture in reciprocity between two differently disposed social actors which becomes more marked the wider the disjuncture in dispositional perceptions of the lifeworld – perceived as a breach in the 'natural attitude' of what constitutes 'social reality' for 'non-autistic spectrum' people and yet an everyday and often traumatic experience for 'autistic people'" (Milton, 2012).

Furthermore, dialectical misattunement, leaning on predictive processing and active inference, shares commonalities with computational psychiatry accounts. For instance, with regard to autism, the 'hypo-priors' hypothesis (Pellicano & Burr, 2012) discusses sensory and other non-social atypicalities as a result of overlying on incoming information as opposed to prior experience. The HIPPEA (Van de Cruys *et al.*, 2014; for High Inflexible Precision of Prediction Errors) and aberrant precision (Lawson *et al*., 2014) accounts position the 'hypo-priors' hypothesis within the predictive processing framework, accommodating further facets of the

condition. That is, these accounts attempt to redefine autism as a deficit of domain-general information processing, explaining difficulties such as relevant to theory of mind, executive dysfunction and central coherence under a common umbrella. Notably, while touching upon social aspects, they still view the condition from an individualistic perspective: the deficit or difference lies exclusively *in* the autistic individual.

Taken together, arguably, due to various conceptual and methodological, but also societal constraints, sociological and psychophysiological processes have been largely studied in isolation (cf., Bolis & Schilbach, 2019; **Appendix A.2.4**). The dialectical misattunement hypothesis, aims at dialectically synthesizing the levels of the individual and the collective through a principled approach. In a nutshell, adopting a Vygotskian perspective, it considers the historical and social construction of the atypical self, while adhering to a scientific understanding of not only interpersonal but also interrelated neurobiological mechanisms. Taken together, the dialectical misattunement hypothesis hopes to serve as a tool for the theoretical, methodological and empirical study of the multiscale dynamics of psychopathology, pushing beyond descriptive accounts.

To provide with a first validation of the hypothesis, we conveyed a two-person psychophysiology study on *real-life* social relations (cf. **Section 4.2**). To this end, pairs of friends were invited to complete questionnaires with regard to their individual autistic traits and the quality of their mutual friendship. Subsequently, pairwise correlational analyses performed between measures of autistic traits and aspects of friendship. Crucially, as anticipated by the dialectical misattunement hypothesis, it was not the autistic traits *per se*, but rather the interpersonal mismatch thereof that predicted core aspects of interpersonal attunement in real life. More concretely, this study shows that the higher the mismatch of autistic traits between

friends, the lower the mutual perceived closeness, acceptance and help. These results are significant because they shed light on a previously researched topic from a completely new angle.

So far, the interrelation of autistic traits and interpersonal relations has been –surprisingly– primarily studied from an individualistic perspective. For instance, Baron-Cohen and Wheelwright (2003) "tested adults with HFA/AS [i.e., high functioning autism / Asperger Syndrome] on the FQ [i.e., Friendship Questionnaire] to explore the notion that autism is an extreme form of the male brain. The extreme male brain (EMB) theory of autism predicts that on any test of "empathizing," unaffected males will score lower than unaffected females, and performance by individuals with an autism spectrum condition will be even lower than unaffected males". Indeed, the study showed that neurotypical females scored highest in the friendship questionnaire, followed by neurotypical males, while autistic males scored the lowest. These results were taken by the authors as evidence supporting the extreme male brain hypothesis.

Contrasting the two-person psychophysiology and the extreme male brain studies, the relativity of empirical data becomes apparent. That is, a slightly differently posed question might yield different interpretation of the measurements. Abstracting interpersonal relations and restricted the unit of analysis to the individual resulted in empirical evidence suggestive to the extreme male brain hypothesis, which describes autistic persons as extreme cases of a 'male phenotype'. On the contrary, studying friendship within concrete interpersonal relations and including the collective as a core unit of analysis, resulted in empirical evidence –within the neurotypical sample– suggestive to the dialectical misattunement hypothesis, which describes autism as an interpersonal mismatch, rather than a mere function of the brain. As Vygotsky insightfully put it,

“a psychology concerned with the study of the complex whole must comprehend this. It must replace the method of decomposing the whole into its elements with that of partitioning the whole into its units. Psychology must identify these units in which the characteristics of the whole are present, even though they may be manifested in altered form. Using this mode of analysis, it must attempt to resolve the concrete problems that face us” (Vygotsky 1934/1987, p. 47; as cited in Dafermos 2018).

Having said that, there are certain limitations in our two-person psychophysiology study that should be considered. First, our sample did not include autistic persons. Therefore, the study should be extended to cover the whole spectrum. Moreover, only same gender dyads of adult individuals were recruited of a relatively homogeneous cultural background. Furthermore, a time scale in the range of human relationships was captured. Additionally, exclusively explicit measures were analyzed. In other words, a static snapshot of the multiscale dynamics of relevant individual and interpersonal processes was captured here. The two-person psychophysiology platform introduced above (cf. **Section 3.1**) and second-person neuropsychiatry appear as promising translational frameworks for further testing the hypothesis both scientifically and clinically across finer time scales and along development. We are indeed currently conveying a clinical study on real-time social interaction, recruiting neurotypical, autistic and mixed dyads.

Additionally, the two-person psychophysiology study presented here deployed the AQ questionnaire for quantifying autistic traits. Yet, in line with critical views on traditional personality theory, we do not consider autistic or other traits as fundamental attributes of an individual. On the contrary, our first dialectical misattunement study aims at demonstrating a straightforward way of breaking away from a trait-based research tradition focused on a static individual. As we discussed in **Section 2.1**, human essence should not be considered as an

abstraction inherent within the individual. With regard to human development, as Vygotsky (1930-1935/1978, p.85) insightfully claimed, “what children can do with the assistance of others might be in some sense even more indicative of their mental development than what they can do alone”.

Taken together, the dialectical misattunement hypothesis, along with the first relevant empirical evidence points toward a definition of psychiatric conditions, such as ASC, relative to the other and generally the social context. Such an approach might eventually allow us to break free from an overly neurocentric research scope in psychiatry on a solid evidence-based ground. To this end, we need to move beyond deeply rooted dichotomies in current research and clinical practice by promoting a reciprocal attunement of the individual and the collective. The view of psychopathology as interpersonal misattunement rather than an individual deviance from the ‘normal’ might help to alleviate social stigma and reduce social exclusion.

5.3. Scientific and societal relevance

Questioning *human becoming*, independent on the answer one gives, carries inherent implications for both scientific and societal practice. In what follows, we delineate certain broad directions, which this project points toward, as well as real-life implications. The concrete ongoing lines of empirical research are listed in **Appendix A.1.5**.

5.3.1. Toward a multiscale synthesis of dynamical systems and active inference

Dialectical attunement by providing a formal account of the interplay between cognition and culture, points toward new avenues for empirical research. For instance, it anticipates

experimental and computational approaches grounded in a synthesis of dynamical systems theories for formally grasping the real-time interpersonal dynamics on one hand and Bayesian accounts of cognition for formally grasping the intrapersonal bodily processes on the other hand. Such an approach could formally show how collective dynamics in social interactions (quantified by dynamical systems trajectories) are potentially tracked and enacted by the individual (quantified by active inference states) *and vice versa*. Put simply, here, we view both 'low-level attunement', which plays out in relatively short spatiotemporal scales (e.g., interpersonal coupling in real-time social interaction) and 'high-level attunement', which is achieved through the interpersonal alignment and deepening of (inter-)bodily structures toward increasing abstraction (e.g., formation of cultural norms). Having said that, low- and high-level attunement should not be viewed as parts of a dichotomy, but rather within their dynamic interrelation.

A concrete implementation could be found in a meta-Bayesian framework, embedded in dynamical systems theory (**Fig. 5.1**; cf. **Appendix A.2.3**; Bolis & Schilbach, 2017, Brandi *et al*., 2019). This scheme aims at capturing individual cognition and action via distinct Bayesian models, but crucially, integrating them on a collective (meta-Bayesian) level of behavior. Here, different generative models within each individual track states of dynamic bodily and environmental states (upper part of red panels). At the level of the individual, autonomic and motor control, are modeled as probabilistic translations of the aforementioned states predictions (lower part of red panels). Collective behavior on the other hand, i.e., the non-linear fusion of interacting motor processes across persons and the environment, can be cast as a coupled dynamical system (orange panel). Notably, the environment is taken not only as a perceptual source, but also as a mediator of the social interaction (green panel).

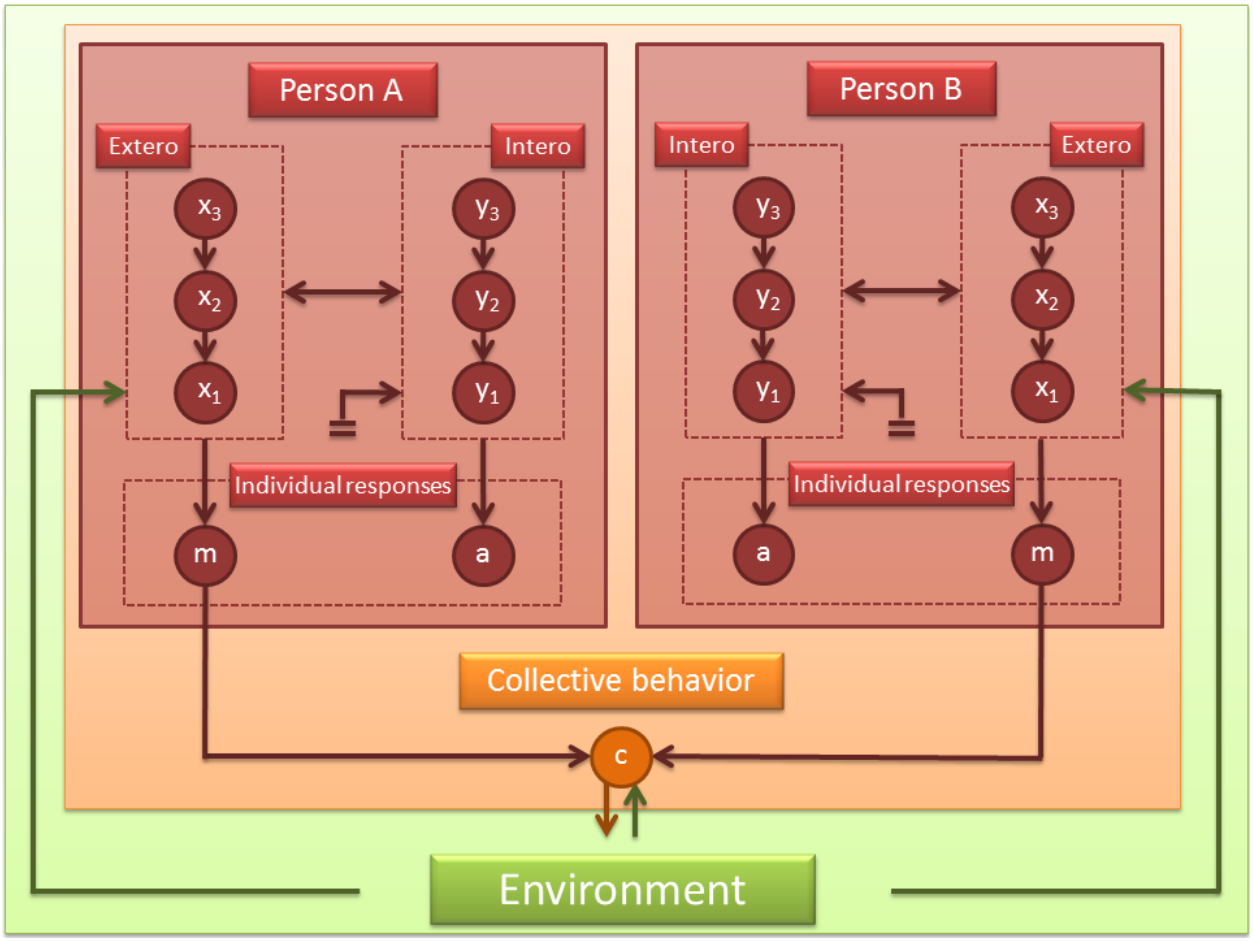

Fig. 5.1. A dynamical-system embedded Bayesian framework for studying the interrelation of *intra*personal and *inter*personal processes during social interaction (cf. **Appendix A.2.3**; Bolis & Schilbach, 2017).

5.3.2. Autism spectrum or autism space?

In this light, the dialectical misattunement hypothesis considered psychopathology as a process unfolding between the level of the individual and the collective. Here, we extend this discussion for motivating further conceptual, empirical and computational directions which might allow for a formal redefinition of *spectrum conditions* as *space conditions*, making justice to their multidimensional and multiscale nature. To this end, we again take autism research as a

paradigm example and apply an extended version of generative embedding (Brodersen *et al*., 2011).

Generative embedding is a data analysis approach consisted of two steps, namely the *generative modeling step*, which aims at modeling mechanisms of phenomena and the *discriminative step*, which aims at capturing discriminative information in the modeled data. The generative modeling step serves as a meaningful dimensionality reduction from the measurement to a latent space. The discriminative step deploys machine learning for group classification (e.g., autistic and non-autistic) and feature selection (i.e., selection of crucial parameters for distinguishing between groups). Critically, this step makes it also possible to adopt an unsupervised scheme, i.e., learning directly based on unlabeled information, overcoming *a priori* group definitions.

Based on generative embedding we now delineate a research line consisted of four core steps (**Fig. 5.2**). In the first step, the units of analysis are explicitly defined and subsequently multi-person psychophysiology data is acquired in different social interaction contexts, aiming at probing different mechanisms. In a second step, by repeatedly applying generative modeling, the raw data is projected onto several low-dimensional parameter spaces, one for each experiment. In a third step, the separate parameter spaces are concatenated to form a single hyperspace. In the fourth step, discriminative approaches are performed for identifying the crucial independent dimensions of the hyperspace, yielding a formalized definition of the feature space. By repeating this cycle of experiments and data analyses an increasingly informed *'autism space'* (within a broader inter-condition space) is constructed, being motivated by and resulting in increasingly sophisticated units of analyses and experimental designs. This pipeline is thought of performing repeated cycles: from the definition of the units of analysis, to computational modeling, machine

learning, data interpretation and back to the redefinition of the units of analysis. In other words, this research line performs a periodic movement, but never returning to the same point.

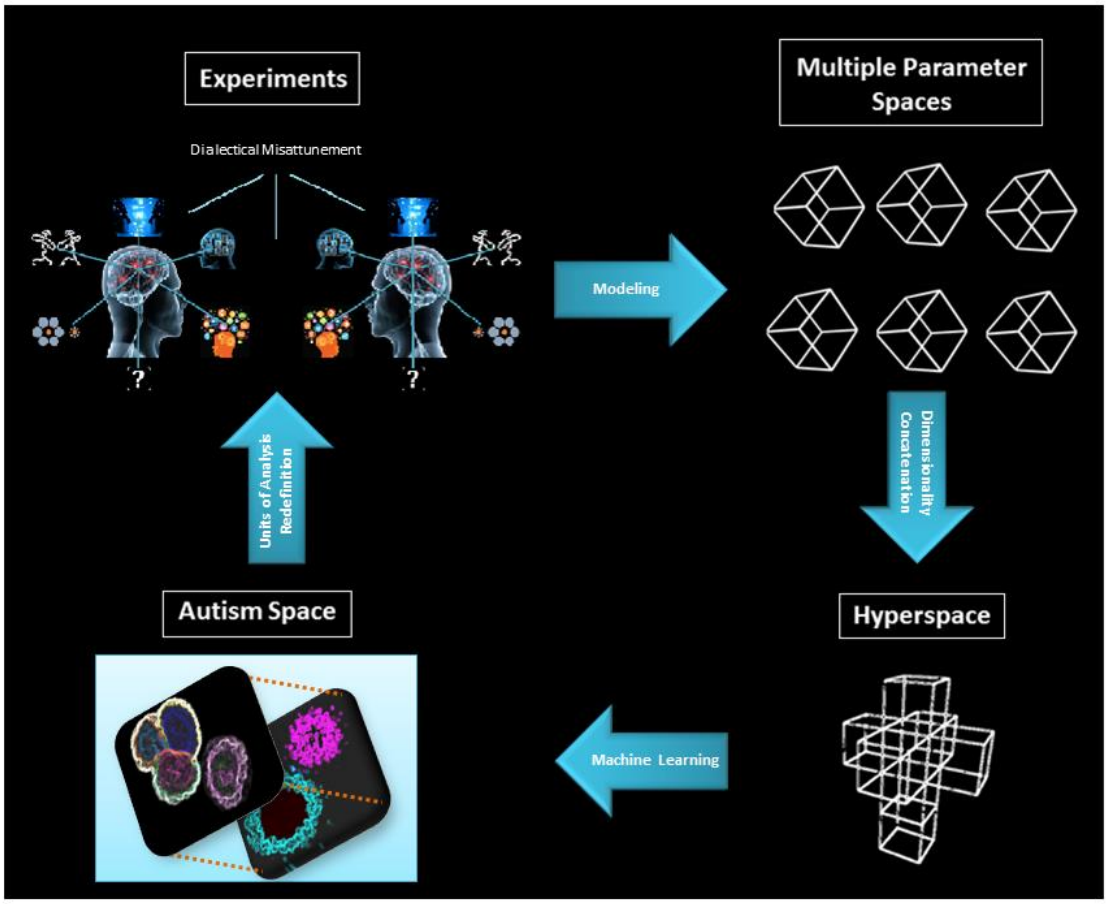

Fig. 5.2. Proposed spiral pipeline (from top-left to bottom-left): a) hypothesis-specific multimodal data is acquired. b) computational modeling yields multiple parameter spaces. c) hypothesis-specific spaces are merged into a single hyperspace. d) dimensionality reduction yields a multilevel autism space including both individual and relational factors. Please note on the upper-left part of the image, people's bodies and other communicative factors have been omitted for simplicity in presentation.

In short, the proposed procedure is expected to delineate a dynamic and multiscale space of conditions that is populated by both individual and collective parameters. A rather continuous space is expected, as well as fuzzy clusters within conditions. Crucially, such an approach is not

necessarily grounded in current diagnostic criteria, but on the contrary, by following an unsupervised and partially data-driven approach, inherent biases of diagnostic manuals could be avoided (cf. Stephan & Mathys, 2014).

Before concluding this section, it is important to emphasize the dynamicity of such a definition of an *autism space*. Taking into account that people, their interrelationships within society, as well as the concept of psychiatric conditions itself are all dynamic processes, such a procedure does not aim at concluding with a fixed definition, but on the contrary to allow for capturing the essential dynamics of the co-development of biological processes, people themselves, their interactions and related concepts in their historical movement and inherent contradiction.

5.3.3. Dialectical attunement in society

Perhaps, most importantly, the implications of such a dialectical approach reach further than the field of research. As Vygotsky splendidly proclaimed, "we cannot master the truth about personality and personality itself as long as mankind has not mastered the truth about society and society itself" (Vygotsky, 1917-1934/1997, p. 342; as cited in Dafermos, 2014)". Taking the collective dimension of human becoming –in its interrelation with the individual– seriously, it points toward concrete directions of societal practice. For instance, with regard to pedagogy, dialectical attunement speaks to an interactive and collaborative learning framework as opposed to a commonly deployed hierarchical and competitive one. The dialectical attunement account is also in line with a juridical system which takes into account not only individual but also collective responsibility, while readily rejects certain rehabilitation practices, such as solitary confinement, as literally dehumanizing.

When it comes to artificial intelligence (AI), it promotes a dynamic and interactive mode of developing. This will not only potentially provide with smarter assistive technology, but could also serve as a dynamic reflection of current societal biases, which might help identify and address them more efficiently. Indeed, with regard to the recently debated social biases inherent in current AI (cf. Birhane & Cummins, 2019), it draws attention to primarily addressing the systemic factors which instigate and perpetuate them, rather than merely focusing on correcting AI itself — in the same way that we do not fix reality by merely correcting a mirror. Additionally, truly interactive and embodied AI could also help overcome challenges of static experimental approaches in social cognition research (cf. Wykowska *et al*., 2016; Kompatsiari *et al*., 2018; Chevalier *et al*., 2019).

With regard to clinical practice, the dialectical misattunement hypothesis suggests systematically monitoring, evaluating and treating not only intrapersonal (e.g., psychophysiological and phenomenological), but also genuinely interpersonal processes across various contexts. This could include interpersonal relations with significant others, within the family, school or work, but also the broad link to society. It also embraces the relation between psychotherapist and patient, whose interpersonal match also needs to be evaluated, as not every psychotherapist might be effective for every patient (Bolis & Schilbach, 2018b). Additionally, it points, from current individualistic treatment options, such as *biofeedback*, toward interpersonal ones, such as *sociofeedback* — from learning to regulate *intrapersonal* functioning to learning to regulate *interpersonal* functioning (Bolis *et al*., 2017).

Furthermore, the dialectical misattunement perspective emphasizes the interrelation between psychological and socioeconomic processes in and through social interactions such as the generation and perpetuation of social stigma. Social stigma may not function only as a cause to

inequality, but also as a result thereof. Therefore, a pragmatic approach to psychopathology should aim –in parallel to the clinical treatment– at balancing structural asymmetries within actual society. This could include, but not limited to, reducing social exclusion, as well as facilitating housing, employment and relational seeking. Finally, as the causality between social dysfunction and poor mental health is largely circular (cf. Schilbach, 2016), social dysfunction and social withdrawal could serve as a helpful transdiagnostic domain (cf. Porcelli *et al.*, 2019), across various psychiatric conditions, from autism and schizophrenia to depression and dementia. Taken together, a *mutually* informed understanding of dysfunction in social interactions –across scales and contexts– and its biological correlates may be exactly the key to a pragmatically efficient research and treatment strategy (cf. Bolis *et al*., 2017; Schilbach, 2016).

5.3.4. Psychopathology as quantomechanics

Before concluding this thesis, we emphasize the deep responsibility of the researcher when conceptualizing and empirically studying human phenomena, such as psychology and psychopathology. This is not because we consider the mind as separate from the matter. This is because the study of mental phenomena resembles quantomechanics with regard to the ‘observer effect’, which is the idea that the mere observation of a phenomenon inevitably changes that very phenomenon. In other words, the endeavor of psychological and psychopathological conceptualization and experimentation does not merely ‘unearth’ given phenomena, but, in doing so, it rather actively co-constructs them.

6. Conclusion — *'Does social interaction matter after all?'*

"True, we love life,
not because we are used to living,
but because we are used to loving.
There is always some madness in love, but
there is also always some reason in madness."

Friedrich Nietzsche

Taken together, we argue, social interaction does indeed matter. In fact, we consider the question of social interaction as the fundamental philosophical question and our argument, here, is a genuinely existential one; without social interaction we do not exist as humans — we interact to live and we live to interact.

References

Abney, D. H., Paxton, A., Dale, R., Kello, C.T. (2014). Complexity matching in dyadic conversation. J Exp Psychol General 143(6):2304

Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *Behav Brain Sci* 33:245–266. https://doi. org/10.1017/S0140525X10000853

Adorno, T. W. (2017). *An Introduction to Dialectics*. John Wiley & Sons.

Ansuini, C., Cavallo, A., Bertone, C., & Becchio, C. (2015). Intentions in the brain: the unveiling of Mister Hyde. *The Neuroscientist*, *21*(2), 126-135.

Apps, M. A., Green, R., & Ramnani, N. (2013). Reinforcement learning signals in the anterior cingulate cortex code for others' false beliefs. *Neuroimage*, *64*, 1-9.

Badre, D., & D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of cognitive neuroscience*, *19*(12), 2082-2099.

Badre, D., & D'esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical?. *Nature Reviews Neuroscience*, *10*(9), 659.

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science*, 329(5995), 1081-1085.

Baker, M. (2016). Is there a reproducibility crisis? A Nature survey lifts the lid on how researchers view the 'crisis rocking science and what they think will help. *Nature*, 533(7604), 452-455.

Baker, C., Saxe, R., & Tenenbaum, J. (2011). Bayesian theory of mind: Modeling joint belief-desire attribution. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 33, No. 33).

Balsters, J. H., Apps, M. A., Bolis, D., Lehner, R., Gallagher, L., & Wenderoth, N. (2016). Disrupted prediction errors index social deficits in autism spectrum disorder. *Brain*, *140*(1), 235-246.

Barad, K. (2003). Posthumanist performativity: toward an understanding of how matter comes to matter. *Signs J Women Cult Soc* 28:801– 831. https://doi.org/10.1086/345321.

Barišic, I., Timmermans, B., Pfeiffer, U. J. et al (2013). Using dual eye-tracking to investigate real time social interactions. *CHI2013.*

Baron-Cohen, S. (1991). Precursors to a theory of mind: understanding attention in others. *Nat Theor Mind* 1:233–251.

Baron–Cohen, S. (2000). Is Asperger syndrome/high-functioning autism necessarily a disability?. *Development and psychopathology*, *12*(3), 489-500.

Baron-Cohen, S., Hoekstra, R. A., Knickmeyer, R., & Wheelwright, S. (2006). The autism-spectrum quotient (AQ) — adolescent version. *Journal of autism and developmental disorders*, 36(3), 343.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition*, *21*(1), 37-46.

Baron-Cohen, S., & Wheelwright, S. (2003). The Friendship Questionnaire: An investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Journal of autism and developmental disorders*, 33(5), 509-517.

Barresi J, Martin R (2011) *History as prologue.* Oxford University Press, Oxford.

Bartz, J. A., Zaki, J., Bolger, N., & Ochsner, K. N. (2011). Social effects of oxytocin in humans: context and person matter. *Trends in cognitive sciences*, *15*(7), 301-309.

Becchio, C., Manera, V., Sartori, L. et al. (2012). Grasping intentions: from thought experiments to empirical evidence. *Frontiers in human neuroscience*, *6*, 117.

Behrens, T. E., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. (2008). Associative learning of social value. *Nature*, *456*(7219), 245.

Berger, P. L., Luckmann, T. (1967). *The social construction of reality: a treatise in the sociology of knowledge.* Doubleday, New York.

Bidell, T. (1988). Vygotsky, Piaget and the dialectic of development. *Human Development, 31*(6), 329-348.

Birhane, A. (2017). Descartes was wrong: "a person is a person through other persons" | *Aeon Ideas.* https://aeon.co/ideas/descartes-was-wrong-a-person-is-a-person-through-other-persons. Accessed 1 Nov 2017.

Birhane, A., & Cummins, F. (2019). Algorithmic Injustices: Towards a Relational Ethics. *arXiv preprint* arXiv:1912.07376.

Blunden, A. (2000). Getting to know Hegel. *Talk at Hegel, Marx and Derrida Summer School*, University of Melbourne.

Bolis, D., Balsters, J., Wenderoth, N. et al (2017). Beyond autism: introducing the dialectical misattunement hypothesis and a bayesian account of intersubjectivity. *Psychopathology*. https://doi. org/10.1159/000484353.

Bolis, D., Becchio, C., & Schilbach, L. (2016). Revisiting psychological definitions at the interface of sociocultural historical theories and predictive coding. *WORLDING THE BRAIN: Patterns, Rhythms, Narratives in Neuroscience and the Humanities.* p, 10.

Bolis, D., Heinzle, J., Mathys, C., Stephan, K. E. (2015). Inferring cognitive traits of individual subjects through gaze controlled video games. In: Bernardis, P., Fantoni, C., Gerbino, W. (eds.) "TSPC2015: *proceedings of the Trieste Symposium on Perception and Cognition*, November 13rd 2015", Trieste, EUT Edizioni Università di Trieste, 2015, p 19.

Bolis, D., Schilbach, L. (2017). Beyond one Bayesian brain: modeling intra- and inter-personal processes during social interaction: commentary on "mentalizing homeostasis: the social origins of interoceptive inference" by Fotopoulou & Tsakiris. *Neuropsychoanalysis* 19:35–38. https://doi.org/10.1080/15294145.2017.1295215.

Bolis, D., & Schilbach, L. (2018a). Observing and participating in social interactions: action perception and action control across the autistic spectrum. *Developmental cognitive neuroscience, 29*, 168-175. https://doi.org/10.1016/j. dcn.2017.01.009.

Bolis, D., & Schilbach, L. (2018b). 'I interact therefore I am': The self as a historical product of dialectical attunement. *Topoi*, 1-14.

Bolis, D., & Schilbach, L. (2020). 'Through others we become ourselves': The dialectics of predictive processing and active inference.

Boonstra, E. A., & Slagter, H. A. (2019). The dialectics of free energy minimization. *Frontiers in systems neuroscience*, 13, 42.

Brandi, M. L., Kaifel, D., Bolis, D., & Schilbach, L. (2019). The Interactive Self–A Review on Simulating Social Interactions to Understand the Mechanisms of Social Agency. *i-com, 18*(1), 17-31.

Brock, J. (2012). Alternative Bayesian accounts of autistic perception: comment on Pellicano and Burr.

Brodersen, K. H., Schofield, T. M., Leff, A. P., Ong, C. S., Lomakina, E. I., Buhmann, J. M., & Stephan, K. E. (2011). Generative embedding for model-based classification of fMRI data. *PLoS computational biology*, 7(6), e1002079.

Brooks, R., Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Dev Sci* 8:535–543. https://doi.org/10.11 11/j.1467-7687.2005.00445.x.

Bruineberg, J., Kiverstein, J., & Rietveld, E. (2018). The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese*, 195(6), 2417-2444.

Bruner, J. S. (1974). From communication to language—A psychological perspective. *Cognit* 3(3):255–287.

Carpenter, M., Nagell, K., Tomasello, M. et al (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr Soc Res Child Dev* 63:i. https://doi. org/10.2307/1166214.

Carrasco, M., Ling, S., Read, S. (2004). Attention alters appearance. *Nat Neurosci* 7:308–313.

Chevalier, P., Kompatsiari, K., Ciardo, F., & Wykowska, A. (2019). Examining joint attention with the use of humanoid robots-A new approach to study fundamental mechanisms of social cognition. *Psychonomic Bulletin & Review*, 1-20.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in cognitive sciences*, *16*(4), 231-239.

Ciaunica, A., Fotopoulou, A. (2017). The touched self: psychological and philosophical perspectives on proximal intersubjectivity and the self. In: *Embodiment, enaction, and culture investigating the constitution of the shared world*. MIT Press, New York, pp 173–192.

Cicourel, A. (1974). *Cognitive Sociology.* Harmondsworth: Penguin.

Clark, A. (1998). *Being there: Putting brain, body, and world together again*. MIT press.

Clark, A., & Chalmers, D. (1998). The extended mind. *analysis*, *58*(1), 7-19.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181-204.

Clark, A. (2013). The many faces of precision (Replies to commentaries on "Whatever next? Neural prediction, situated agents, and the future of cognitive science"). *Frontiers in psychology*, *4*, 270.

Colombetti, G. (2007). Enactive appraisal. *Phenomenology and the Cognitive Sciences*, *6*(4), 527-546.

Dafermos, M. (2002). *The Cultural–Historical Theory of L. Vygotsky.* Athens: Atrapos.(in greek).

Dafermos, M. (2014). Vygotsky's analysis of the crisis in psychology: Diagnosis, treatment, and relevance. *Theory & Psychology*, 24(2), 147-165.

Dafermos, M. (2015). Rethinking the crisis in social psychology: A dialectical perspective. *Social and Personality Psychology Compass*, *9*(8), 394-405.

Dafermos, M. (2018). Developing a Dialectical Perspective on Vygotsky's Theory. In: *Rethinking Cultural-Historical Theory*. Perspectives in Cultural-Historical Research, vol 4. Springer, Singapore.

Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2013). The self-organization of human interaction. In *Psychology of learning and motivation* (Vol. 59, pp. 43-95). Academic Press.

Dakin, S., & Frith, U. (2005). Vagaries of visual perception in autism. *Neuron*, *48*(3), 497-507.

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., & Iacoboni, M. (2006). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature neuroscience*, *9*(1), 28.

Daunizeau, J., den Ouden, H., E., M., Pessiglione, M., et al (2010) Observing the observer (I): meta-Bayesian models of learning and decision-making. *PLoS ONE.* https://doi.org/10.1371/journal.pone.00155 54.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making. *Phenomenology and the cognitive sciences*, *6*(4), 485-507.

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition?. *Trends in cognitive sciences*, *14*(10), 441-447.

D'Errico J., (2019), inpaint_nans (https://www.mathworks.com/matlabcentral/fileexchange/4551-inpaint_nans), *MATLAB Central File Exchange.* Retrieved December 11, 2019.

Di Paolo, E. A., & De Jaegher, H. (2012). The interactive brain hypothesis. *Frontiers in human neuroscience*, *6*, 163.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental brain research*, *91*(1), 176-180.

Dobson, S. (2012). Face to face with the social brain. Philos Trans R Soc Lond B 367:1901–1908. https://doi.org/10.1098/rstb.2011.0224 Dumas G, Nadel J, Soussignan R et al (2010) Inter-brain synchronization during social interaction. PLoS ONE 5:e12166. https://doi.org/10.1371/journal.pone.0012166.

Dumas, G., Soussignan, R., Hugueville, L., Martinerie, J., & Nadel, J. (2014). Revisiting mu suppression in autism spectrum disorder. *Brain research*, *1585*, 108-119.

Dumas, G. (2011). Towards a two-body neuroscience. *Communicative & integrative biology*, *4*(3), 349-352.

Dumas, G., de Guzman, G. C., Tognoli, E., & Kelso, J. S. (2014). The human dynamic clamp as a paradigm for social interaction. *Proceedings of the National Academy of Sciences*, *111*(35), E3726-E3734.

Dumas, G., Martinerie, J., Soussignan, R., & Nadel, J. (2012). Does the brain know who is at the origin of what in an imitative interaction?. *Frontiers in human neuroscience*, *6*, 128.

Dunsworth, H., M., Warrener, A., G., Deacon, T., et al (2012). Metabolic hypothesis for human altriciality. *Proc Natl Acad Sci USA* 109:15212–15216.

Elms, A. C. (1975). The crisis of confidence in social psychology. *American Psychologist*, 30, 967–976.

Ellison, P., T. (2001). *On fertile ground.* Harvard University Press, Cambridge.

Fan, Y. T., Decety, J., Yang, C. Y., Liu, J. L., & Cheng, Y. (2010). Unbroken mirror neurons in autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *51*(9), 981-988.

Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in human neuroscience*, *4*, 215.

Felleman, D. J., & Van, D. E. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, *1*(1), 1-47.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic bulletin & review*, 5(3), 490-495.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin*, 133(4), 694.

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, *360*(1456), 815-836.

Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, *159*(3), 417-458.

Friston, K. (2008). Hierarchical models in the brain. *PLoS computational biology*, *4*(11), e1000211.

Friston, K. (2009). The free-energy principle: a rough guide to the brain?. *Trends in cognitive sciences*, *13*(7), 293-301.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211-1221.

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature reviews neuroscience*, 11(2), 127.

Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biological cybernetics*, *102*(3), 227-260.

Friston, K. (2013). Life as we know it. *Journal of the Royal Society Interface*, *10*(86), 20130475.

Friston, K., & Frith, C. (2015a). A duet for one. *Consciousness and cognition*, *36*, 390-405.

Friston, K. J., & Frith, C. D. (2015b). Active inference, communication and hermeneutics. *cortex*, *68*, 129-143.

Friston, K. J., Lawson, R., & Frith, C. D. (2013). On hyperpriors and hypopriors: comment on Pellicano and Burr. *Trends in cognitive sciences*, *17*(1), 1.

Frith, U., & Happé, F. (1994). Autism: Beyond "theory of mind". *Cognition*, *50*(1-3), 115-132.

Frith, U. (2003). *Autism: Explaining the enigma*. Blackwell Publishing.

Frith, U. (2008). *Autism: A very short introduction* (Vol. 195). Oxford University Press.

Frith CD, Frith U (2012) Mechanisms of social cognition. *Annu Rev Psychol 63*:287–313. https://doi.org/10.1146/annurev-psych-120710-100449.

Froese T.: Overview. Dr. Tom Froese. https://froese.wordpress.com/research/overview/.

Froese, T., & Di Paolo, E. A. (2011). The enactive approach: Theoretical sketches from cell to society. *Pragmatics & Cognition*, *19*(1), 1-36.

Froese, T., & Fuchs, T. (2012). The extended body: a case study in the neurophenomenology of social interaction. *Phenomenology and the Cognitive Sciences*, *11*(2), 205-235.

Froese, T., Iizuka, H., & Ikegami, T. (2014). Embodied social interaction constitutes social cognition in pairs of humans: a minimalist virtual reality experiment. *Scientific reports*, 4, 3672.

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, 32, 147-157.

Fusaroli R, Tylén K (2016) Investigating conversational dynamics: interactive alignment, interpersonal synergy, and collective task performance. Cogn Sci 40:145–171. https://doi.org/10.1111/ cogs.12251.

Gallagher, S. (2001). The practice of mind. Theory, simulation or primary interaction?. *Journal of Consciousness Studies*, 8(5-6), 83-108.

Gallagher, S. (2004). Understanding interpersonal problems in autism: Interaction theory as an alternative to theory of mind. *Philosophy, Psychiatry, & Psychology*, 11(3), 199-217.

Gallagher, S. (2006). *How the body shapes the mind*. Clarendon Press.

Gallagher, S. (2014). The cruel and unusual phenomenology of solitary confinement. *Frontiers in Psychology*, 5, 585.

Gallagher, S., & Varga, S. (2015). Conceptual issues in autism spectrum disorders. *Current Opinion in Psychiatry*, *28*(2), 127-132.

Garfinkel, H. (1967) *Studies in Ethnomethodology.* Englewood Cliffs: Prentice-Hall.

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, 1(2), 292-304.

Ghaemi, S. N. (2018). After the failure of DSM: clinical research on psychiatric diagnosis. *World Psychiatry*, 17(3), 301.

Gomot, M., & Wicker, B. (2012). A challenging, unpredictable world for people with autism spectrum disorder. *International Journal of Psychophysiology*, *83*(2), 240-247.

Gordon, I., Vander Wyk, B. C., Bennett, R. H., Cordeaux, C., Lucas, M. V., Eilbott, J. A., ... & Pelphrey, K. A. (2013). Oxytocin enhances brain function in children with autism. *Proceedings of the National Academy of Sciences*, *110*(52), 20953-20958.

Graby S (2012) To be or not to be disabled: autism, disablement and identity politics. Paper presented at the *'Theorising Normalcy and the Mundane' conference, University of Chester, Chester*, 27 June.

Grassian, S. (2006) *Psychiatric effects of solitary confinement.* Wash. UJL & Pol'y, 22, 325.

Grossmann, I. (2018). Dialecticism across the lifespan.

Gunawardena, C. N., & Zittle, F. J. (1997). Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American journal of distance education*, 11(3), 8-26.

Haken, H. (1989). Synergetics: an overview. *Reports on Progress in Physics*, *52*(5), 515.

Haken, H. (2004). Advanced Topics. In *Synergetics* (pp. 389-758). Springer, Berlin, Heidelberg.

Hamilton, A. F., & Grafton, S. T. (2007). The motor hierarchy: from kinematics to goals and intentions. *Sensorimotor foundations of higher cognition*, *22*, 381-408.

Happé, F. G. (1996). Studying weak central coherence at low levels: children with autism do not succumb to visual illusions. A research note. *Journal of Child Psychology and Psychiatry*, *37*(7), 873-877.

Happé, F. (1999). Autism: cognitive deficit or cognitive style?. *Trends in cognitive sciences*, *3*(6), 216-22.

Happé, F. (2003, January). Cognition in autism: one deficit or many?. In *Autism: Neural Basis and Treatment Possibilities: Novartis Foundation Symposium 251* (Vol. 251, pp. 198-212). Chichester, UK: John Wiley & Sons, Ltd.

Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of autism and developmental disorders*, *36*(1), 5-25.2.

Happé, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. *Nature neuroscience*, *9*(10), 1218.

Hari, R., & Kujala, M. V. (2009). Brain basis of human social interaction: from concepts to brain imaging. *Physiological reviews*, *89*(2), 453-479.

Harris, H., Israeli, D., Minshew, N., Bonneh, Y., Heeger, D. J., Behrmann, M., & Sagi, D. (2015). Perceptual learning in autism: over-specificity and possible remedies. *Nature neuroscience*, *18*(11), 1574.

Haugeland, J. (1998). *Having thought*. Harvard University Press.

Heidegger, M. (1962). *Being and Time.* Trans. J. Macquarrie and E. Robinson. New York: Harper and Row.

Helmholtz, H. V. (1866). *Treatise Psychological Optics* (Optical Society of America. Rochester, NY).

Herbort, O., & Butz, M. V. (2012). Too good to be true? Ideomotor theory from a computational perspective. *Frontiers in psychology*, *3*, 494.

Hessels, R. S., Holleman, G. A., Cornelissen, T. H., Hooge, I. T., & Kemner, C. (2018) Eye contact takes two – autistic and social anxiety traits predict gaze behavior in dyadic interaction. *Journal of Experimental Psychopathology, 9*(2), jep-062917.

Heyes, C. (2012). What's social about social learning?. *Journal of Comparative Psychology*, *126*(2), 193.

Hill, E. L. (2004). Executive dysfunction in autism. *Trends in cognitive sciences*, *8*(1), 26-32.

Hirsch, E. D., Kett, J. F., & Trefil, J. S. (2002). *The new dictionary of cultural literacy*. Houghton Mifflin Harcourt.

Hobson, R., P., Patrick, M., P., H., Crandell, L., E. et al (2004). Maternal sensitivity and infant triadic communication. *Psychol Learn Motiv-Adv Res Theory 45*:470–480. https://doi.org/10.111 1/j.1469-7610.2004.00238.x.

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

Huang, Y., & Rao, R. P. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 580-593.

Hume, D. (1739/1888). Treatise of Human Nature. In: *Selby-Bigge LA* (ed). Clarendon Press, Oxford.

Hurley, S. L. (1998). *Consciousness in action*. Harvard University Press.

Huys, Q. J., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature neuroscience*, *19*(3), 404.

Hoel, E. P., Albantakis, L., & Tononi, G. (2013). Quantifying causal emergence shows that macro can beat micro. *Proceedings of the National Academy of Sciences*, *110*(49), 19790-19795.

James, W., Burkhardt, F., Bowers, F., & Skrupskelis, I. K. (1890). *The principles of psychology* (Vol. 1, No. 2). London: Macmillan.

James, W. (2001). Psychology: The briefer course. Toronto, Canada: General Publishing Company. (Original work published in 1892).

Jellema, T., Baker, C., I., Wicker, B., Perrett, D., I. (2000). Neural representation for the perception of the intentionality of actions. Brain Cogn 44:280–302. https://doi.org/10.1006/brcg.2000.1231.

Joffily, M., & Coricelli, G. (2013). Emotional valence and the free-energy principle. *PLoS computational biology*, *9*(6), e1003094.

Kearney, K. (2020). boundedline.m (https://www.github.com/kakearney/boundedline-pkg), GitHub. Retrieved November 2, 2019.

Kelso, J. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. MIT press.

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in cognitive sciences*, *15*(8), 352-357.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive processing*, *8*(3), 159-166.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, *27*(12), 712-719.

Kobayashi, H., Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. J Hum Evol 40(5):419–435.

Koike, T., Sumiya, M., Nakagawa, E., Okazaki, S., & Sadato, N. (2019). What makes eye contact special? Neural substrates of on-line mutual eye-gaze: a hyperscanning fMRI study. *eNeuro, 6*(1).

Koike, T., Tanabe, H., C., Okazaki, S. et al (2016). Neural substrates of shared attention as social memory: a hyperscanning functional magnetic resonance imaging study. Neuroimage 125:401–412. https://doi.org/10.1016/j.neuroimage.2015.09.076.

Koike, T., Tanabe, H. C., & Sadato, N. (2015). Hyperscanning neuroimaging technique to reveal the “two-in-one” system in social interactions. *Neuroscience research, 90*, 25-32.

Kompatsiari, K., Pérez-Osorio, J., De Tommaso, D., Metta, G., & Wykowska, A. (2018). Neuroscientifically-grounded research for improved human-robot interaction. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 3403-3408). IEEE.*

Konvalinka, I., & Roepstorff, A. (2012). The two-brain approach: how can mutually interacting brains teach us something about social interaction?. *Frontiers in human neuroscience*, *6*, 215.

Kopp, B. (2012). A simple hypothesis of executive function. *Frontiers in Human Neuroscience*, *6*, 159.

Koster-Hale, J., & Saxe, R. (2013). Theory of mind: a neural prediction problem. *Neuron, 79*(5), 836-848.

Kraut, R. (2017). *Plato.* In: Zalta EN (ed) Stanford encyclopedia of philosophy. Stanford University, Stanford.

Kuhn, T. S. (1970). *The structure of scientific revolutions.* Chicago, IL: University of Chicago Press.

Kuster, C., Popa, T., Bazin, J. C., Gotsman, C., & Gross, M. (2012). Gaze correction for home video conferencing. *ACM Transactions on Graphics (TOG), 31*(6), 174.

Laing, R. D., Phillipson, H., & Lee, A. R. (1966). Interpersonal perception: a theory and a method of research.

Lawler, J. (1975). Dialectical philosophy and developmental psychology: Hegel and Piaget on contradiction. *Human Development*, *18*(1-2), 1-17.

Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in human neuroscience*, *8*, 302.

Leong, V., Byrne, E., Clackson, K., Georgieva, S., Lam, S., & Wass, S. (2017). Speaker gaze increases information coupling between infant and adult brains. *Proceedings of the National Academy of Sciences, 114*(50), 13290-13295.

Leontyev, A., N. (1983). Activity, consciousness, and personality. In: *Davydov V, Zinchenko V, Leontiev AA, Petrovskij A* (eds) A.N. Leontiev. Izbrannie psihologicheskie proizvedenija [A.N. Leon- tiev. Selected psychological works], vol 2. Pedagogika, Moscow, pp 94–231 (Original work 1975).

Levins, R., Lewontin, R., C. (1985) *The dialectical biologist.* Harvard University Press, Cambridge.

Liu, N., Mok, C., Witt, E., E. et al (2016). NIRS-based hyperscanning reveals inter-brain neural synchronization during cooperative Jenga Game with face-to-face communication. *Front Hum Neurosci 10*:82. https://doi.org/10.3389/fnhum.2016.00082.

Locke, J. (1690–4/1979) *An essay concerning human understanding.* In: Nidditch PH (ed) Oxford University Press, Oxford.

Magnus, B. (1970). *Heidegger's metahistory of philosophy: Amor Fati, being and truth.* Springer, Dordrecht.

Manor, B. R., & Gordon, E. (2003). Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *Journal of neuroscience methods, 128*(1-2), 85-93.

Marx, K., Engels, F. (1888). *The German ideology: including theses on Feuerbach and introduction to the critique of political economy.* Prometheus Books, New York.

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in human neuroscience, 5*, 39.

Maturana, H. R., Varela, F. J. (1980). *Autopoiesis and cognition: the realization of the living.* Springer, Dordrecht.

Mead, G. H. (1912). The mechanism of social consciousness. *The Journal of Philosophy, Psychology and Scientific Methods, 9*(15), 401-406.

Menary, R. (2010). Introduction to the special issue on 4E cognition. *Phenomenology and the Cognitive Sciences*, *9*(4), 459-463.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). Plans and the structure of behavior.

Milton, D. E. (2012). On the ontological status of autism: the 'double empathy problem'. *Disability & Society*, *27*(6), 883-887.

Milton, D. (2013). 'Filling in the gaps': A micro-sociological analysis of autism. *Autonomy, the Critical Journal of Interdisciplinary Autism Studies*, *1*(2).

Milton, D. (2014). Embodied sociality and the conditioned relativism of dispositional diversity. *Autonomy, the Critical Journal of Interdisciplinary Autism Studies*, *1*(3), 1-7.

Modahl, C., Green, L. A., Fein, D., Morris, M., Waterhouse, L., Feinstein, C., & Levin, H. (1998). Plasma oxytocin levels in autistic children. *Biological psychiatry*, *43*(4), 270-277.

Modi, M. E., & Young, L. J. (2012). The oxytocin system in drug discovery for autism: animal models and novel therapeutic strategies. *Hormones and behavior*, *61*(3), 340-350.

Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., ... & Fisher, R. E. (2002). Hyperscanning: simultaneous fMRI during linked social interactions.

Morales, M., Mundy, P., Delgado, C. E. F. et al (2000). Responding to joint attention across the 6-through 24-month age period and early language acquisition. *J Appl Dev Psychol 21*:283–298. https://doi. org/10.1016/S0193-3973(99)00040-4.

Moran, R. J., Campo, P., Symmonds, M., Stephan, K. E., Dolan, R. J., & Friston, K. J. (2013). Free energy, precision and learning: the role of cholinergic neuromodulation. *Journal of Neuroscience*, *33*(19), 8227-8236.

Morrison, K. E., Pinkham, A. E., Kelsven, S., Ludwig, K., Penn, D. L., & Sasson, N. J. (2019). Psychometric evaluation of social cognitive measures for adults with autism. *Autism Research*, *12*(5), 766-778.

Mottron, L., Dawson, M., Soulieres, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of autism and developmental disorders*, *36*(1), 27-43.

Mukawa, N., Oka, T., Arai, K., & Yuasa, M. (2005). What is connected by mutual gaze?: user's behavior in video-mediated communication. In *CHI'05 Extended Abstracts on Human Factors in Computing Systems* (pp. 1677-1680). ACM.

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological cybernetics*, *66*(3), 241-251.

Mundy, P., Block, J., Delgado, C. et al (2007). Individual differences and the development of joint attention in infancy. *Child Dev 78*:938– 954. https://doi.org/10.1111/j.1467-8624.2007.01042.x.

Mundy, P., Newell, L. (2007). Attention, joint attention, and social cognition. *Curr Dir Psychol Sci 16*:269–274. https://doi.org/10.111 1/j.1467-8721.2007.00518.x

Murray, D., Lesser, M., & Lawson, W. (2005). Attention, monotropism and the diagnostic criteria for autism. *Autism*, *9*(2), 139-156.

Murray, L, & Trevarthen, C. (1985), Emotional regulation of interactions between two-month-olds and their mothers. In: *Social Perception in Infants, ed. T. Field & N. Fox. Norwood,* NJ: Ablex, pp. 177–197.

Ne'eman, A. (2010). The future (and the past) of Autism advocacy, or why the ASA's magazine, The Advocate, wouldn't publish this piece. *Disability Studies Quarterly*, *30*(1).

Nietzsche, F. (1999). *The birth of tragedy and other writings.* Cambridge University Press, Cambridge (Original work 1872).

Nietzsche, F. (2017). *The will to power.* Penguin, Cambridge (Original work 1901).

Oberman, L. M., Hubbard, E. M., McCleery, J. P., Altschuler, E. L., Ramachandran, V. S., & Pineda, J. A. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Cognitive brain research*, *24*(2), 190-198.

O'Connor, T., Wong, H.Y. Emergent Properties; in Zalta EN (ed): *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition). https://plato.stanford.edu/

Oullier, O., De Guzman, G. C., Jantzen, K. J., Lagarde, J., & Scott Kelso, J. A. (2008). Social coordination dynamics: Measuring human bonding. *Social neuroscience, 3*(2), 178-192.

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: relationship to theory of mind. *Journal of child Psychology and Psychiatry, 32*(7), 1081-1105.

Paley, J. (2002). "The Cartesian melodrama in nursing". *Nursing Philosophy.* 3 (3): 189–192. doi:10.1046/j.1466-769X.2002.00113.x.

Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence?. *Perspectives on Psychological Science, 7*(6), 528-530.

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends in cognitive sciences, 16*(10), 504-510.

Perkins, T., Stokes, M., McGillivray, J., & Bittar, R. (2010). Mirror neuron dysfunction in autism spectrum disorders. *Journal of clinical neuroscience, 17*(10), 1239-1243.

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature reviews neuroscience, 9*(2), 148.

Pfeiffer, U. J., Vogeley, K., & Schilbach, L. (2013). From gaze cueing to dual eye-tracking: novel approaches to investigate the neural correlates of gaze in social interaction. *Neuroscience & Biobehavioral Reviews, 37*(10), 2516-2528.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences, 27*(2), 169-190.

Plaisted Grant, K., & Davis, G. (2009). Perception and apperception in autism: rejecting the inverse assumption. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1522), 1393-1398.

Porcelli, S., Van Der Wee, N., van der Werff, S., Aghajani, M., Glennon, J. C., van Heukelum, S., ... & Posadas, M. (2019). Social brain, social dysfunction and social withdrawal. *Neuroscience & Biobehavioral Reviews, 97*, 10-33.

Powell, J. L., Lewis, P. A., Dunbar, R. I. M. et al (2010). Orbital prefrontal cortex volume correlates with social cognitive competence. *Neuropsychologia 48*:3554–3562. https://doi.org/10.1016/J.NEURO PSYCHOLOGIA.2010.08.004.

Quattrocki, E., & Friston, K. (2014). Autism, oxytocin and interoception. *Neuroscience & Biobehavioral Reviews, 47*, 410-430.

Ramachandran, V. S., & Oberman, L. M. (2006). Broken mirrors: A theory of autism. *Scientific American, 295*(5).

Ramstead, M. J. D., Badcock, P. B., & Friston, K. J. (2018). Answering Schrödinger's question: A free-energy formulation. *Physics of life reviews, 24*, 1-16.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, *2*(1), 79.

Ratcliffe, M. (2009). Existential feeling and psychopathology. Philosophy, Psychiatry, & Psychology, 16(2), 179-194.

Reddy, V. (2008). *How infants know minds*. Harvard University Press.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci., 27*, 169-192.

Robinson, H. (2017). Dualism; in Zalta EN (ed): *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/ archives/fall2017/entries/dualism.

Rogers, S. L., Speelman, C. P., Guidetti, O., & Longmuir, M. (2018). Using dual eye tracking to uncover personal gaze patterns during social interaction. Scientific reports, 8(1), 4271.

Roth, W. M. (2016). The primacy of the social and sociogenesis. Integrative Psychological and Behavioral Science, 50(1), 122-141.

Rumsey, J. M. (1985). Conceptual problem-solving in highly verbal, nonretarded autistic men. *Journal of autism and developmental disorders*, *15*(1), 23-36.

Schilbach, L. (2014). On the relationship of online and offline social cognition. *Frontiers in human neuroscience*, *8*, 278.

Schilbach, L. (2015). Eye to eye, face to face and brain to brain: Novel approaches to study the behavioral dynamics and neural mechanisms of social interactions. *Current Opinion in Behavioral Sciences*, *3*, 130-135.

Schilbach, L. (2016). Towards a second-person neuropsychiatry. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150081.

Schilbach, L., Bzdok, D., Timmermans, B., et al (2012). Introspective minds: using ALE meta-analyses to study commonalities in the neural correlates of emotional processing, social & unconstrained cognition. *PLoS ONE 7*:e30920. https://doi.org/10.1371/journ al.pone.0030920

Schilbach, L., Eickhoff, S. B., Cieslik, E. C., Kuzmanovic, B., & Vogeley, K. (2012). Shall we do this together? Social gaze influences action control in a comparison group, but not in individuals with high-functioning autism. *Autism*, *16*(2), 151-162.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience 1. *Behavioral and brain sciences*, *36*(4), 393-414.

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., ... & Vogeley, K. (2010). Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. *Journal of cognitive neuroscience*, *22*(12), 2702-2715.

Schilbach, L., Wohlschlaeger, A. M., Kraemer, N. C., Newen, A., Shah, N. J., Fink, G. R., & Vogeley, K. (2006). Being with virtual others: Neural correlates of social interaction. *Neuropsychologia*, *44*(5), 718-730.

Schneider, D. , Slaughter, V. P., Bayliss, A. P., Dux, P. E. (2013). A temporally sustained implicit theory of mind deficit in autism spectrum disorders. *Cognition 129*, 410–417.

Schönherr, J., & Westra, E. (2017). Beyond 'Interaction': How to Understand Social Effects on Social Cognition. *The British Journal for the Philosophy of Science, 70*(1), 27-52.

Seibt, J. (2017). Process philosophy. In: Zalta EN (ed) *Stanford encyclopedia of philosophy.* Stanford University, Stanford.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science*, *325*(5942), 883-885.

Sevgi, M., Diaconescu, A. O., Tittgemeyer, M., & Schilbach, L. (2016). Social Bayes: Using Bayesian modeling to study autistic trait–related differences in social cognition. *Biological psychiatry*, *80*(2), 112-119.

Sherif, M. (1935). A study of so socimeal factors in perception. *Archives of Psychology*, *27*(187).

Shipp, S., Adams, R. A., & Friston, K. J. (2013). Reflections on agranular architecture: predictive coding in the motor cortex. *Trends in neurosciences*, *36*(12), 706-716.

Sihvola, J. (2008). Aristotle on the individuality of self. In: Remes P, Sihvola J (eds) *Ancient philosophy of the self.* Springer, Dordrecht, pp 125–137.

Silbert, L. J., Honey, C. J., Simony, E. et al (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proc Natl Acad Sci USA 111*:E4687–E4696. https://doi.org/10.1073/pnas.1323812111.

Siller, M., Sigman, M. (2002). The behaviors of parents of children with autism predict the subsequent development of their children's communication. J Autism Dev Disord 32:77–89. https://doi. org/10.1023/A:1014884404276.

Southgate, V., & Hamilton, A. F. D. C. (2008). Unbroken mirrors: Challenging a theory of autism. *Trends in cognitive sciences*, *12*(6), 225-229.

Steel, M. J. G., Gorman, R., & Flexman, J. E. (1984). Neuropsychiatric testing in an autistic mathematical idiot-savant: Evidence for nonverbal abstract capacity. *Journal of the American Academy of Child Psychiatry*, *23*(6), 704-707.

Stefanics, G., Kremláček, J., & Czigler, I. (2014). Visual mismatch negativity: a predictive coding view. *Frontiers in human neuroscience*, *8*, 666.

Stephan, K. E., & Mathys, C. (2014). Computational approaches to psychiatry. *Current opinion in neurobiology*, *25*, 85-92.

Stetsenko, A., Arievitch, I. M. (2004). The self in cultural-historical activity theory reclaiming the unity of social and individual dimensions of human development. *Theory Psychol 14*:475–503. https://doi. org/10.1177/0959354304044921

Stock, A., & Stock, C. (2004). A short history of ideo-motor action. *Psychological research*, *68*(2-3), 176-188.

Störmer, V. S., McDonald, J. J., & Hillyard, S. A. (2009). Cross-modal cueing of attention alters appearance and early cortical processing of visual stimuli. *Proceedings of the National Academy of Sciences*, *106*(52), 22456-22461.

Striano T., Rochat, P. (1999). Developmental link between dyadic and triadic social competence in infancy. *Br J Dev Psychol 17*:551–562. https://doi.org/10.1348/026151099165474.

Teufel, C., Subramaniam, N., & Fletcher, P. C. (2013). The role of priors in Bayesian models of perception. *Frontiers in computational neuroscience*, *7*, 25.

Thien, L. M., Razak, N. A., & Jamil, H. (2012). Friendship Quality Scale: Conceptualization, Development and Validation. *Australian Association for Research in Education (NJ1).*

Thomas, C. (1999) *Female Forms: Experiencing and Understanding Disability*, Buckingham, Open University Press.

Thompson, E. (2010). *Mind in life: biology, phenomenology, and the sciences of mind*. Belknap Press of Harvard University Press.

Tieman, S. B., Hirsch, H. V. B. (1982). Exposure to lines of only one orientation modifies dendritic morphology of cells in the visual cortex of the cat. *J Comp Neurol 211*:353–362. https://doi.org/10.1002/ cne.902110403.

Timmermans, B., Schilbach, L., Pasquali, A., & Cleeremans, A. (2012). Higher order thoughts in action: consciousness as an unconscious re-description process. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1412-1423.

Tomasello, M. (1995). Joint attention as social cognition. In: *Moore C, Dunham PJ (eds) Joint attention: its origins and role in development*. Lawrence Erlbaum Associates, Hillsdale, pp 103–130.

Tomasello, M. (1999). *The cultural origins of human cognition.* Harvard University Press, Cambridge.

Tomasello, M. (2014). *A natural history of human thinking.* Harvard University Press, Cambridge.

Tomasello, M., Carpenter, M., Call, J. et al (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav Brain Sci 28*:675-91-735. https://doi.org/10.1017/S0140525X05000129.

Tomasello, M., Carpenter, M. (2007). Shared intentionality. *Dev Sci 10*:121–125. https://doi.org/10.1111/j.1467-7687.2007.00573.x

Tomasello, M., Rakoczy, H. (2003). What makes human cognition unique? From individual to shared to collective intentionality. *Mind Lang 18*:121–147.

Toomela, A. (2014). There can be no cultural-historical psychology without neuropsychology. And vice versa. *The Cambridge handbook of cultural-historical psychology*, 315-349.

Treffert, D. A. (2009). The savant syndrome: an extraordinary condition. A synopsis: past, present, future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1522), 1351-1357.

Vagenas, E. (2003). The soviet psychology in child's development. *Communist Review*, *6*.

Van Boxtel, J. J., & Lu, H. (2013). A predictive coding perspective on autism spectrum disorders. *Frontiers in psychology*, *4*, 19.

Van de Cruys, S., & Wagemans, J. (2011). Putting reward in art: a tentative prediction error account of visual art. *i-Perception*, *2*(9), 1035-1062.

Van de Cruys, S., de-Wit, L., Evers, K., Boets, B., & Wagemans, J. (2013). Weak priors versus overfitting of predictions in autism: Reply to Pellicano and Burr (TICS, 2012). *i-Perception*, *4*(2), 95-97.

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological review*, *121*(4), 649.

van Praag, H. M (1992). *Make-believes in psychiatry or the perils of progress.* New York: Brunner Mazel.

Varela, F. J., Thompson, E., & Rosch, E. (2017). *The embodied mind: Cognitive science and human experience*. MIT press.

Veissière, S. P., Constant, A., Ramstead, M. J., Friston, K. J., & Kirmayer, L. J. (2019) Thinking Through Other Minds: A Variational Approach to Cognition and Culture. *Behavioral and Brain Sciences*, 1-97.

Vittorias, J., Petrantonakis, P., Bolis, D., Tsiligkyri, A., Kosmidou, V., & Hadjileontiadis, L. J. (2008). Noesis: An enhanced educational environment for kids with autism spectrum disorders. In *2008 Eighth IEEE International Conference on Advanced Learning Technologies* (pp. 1019-1020). IEEE.

Von Helmholtz, H. (1867). *Handbuch der physiologischen Optik* (Vol. 9). Voss.

Vygotsky, L., S. (1962). Thinking and speech (E. Hanfmann & G. Vakar, eds. and Trans). MIT Press, Cambridge, MA. (Original work 1934).

Vygotsky, L., S. (1978). *Mind in society: the development of higher psychological processes.* Harvard University Press. (Original work 1930–1935). Translated in Greek, (1997) by A. Bibou and S Vosniadou, Gutenberg.

Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber, & A. S. Carton (Eds.), The collected works of L. S. Vygotsky (Vol. 1, pp. 39–285). New York: Plenum Press. (Original work 1934).

Vygotsky L. S. (1997). The collected works of LS Vygotsky: Problems of the theory and history of psychology. Vol. 3 (New York and London: Springer Science & Business Media, Plenum Press). (Original work 1917-1934).

Vygotsky, L. S. (1997). *The historical meaning of the crisis in psychology: A methodological investigation.* The collected works of LS Vygotsky, 3, 233-343. (Original work 1927).

Wang, Y., & Hamilton, A. F. D. C. (2012). Social top-down response modulation (STORM): a model of the control of mimicry in social interaction. *Frontiers in human neuroscience*, *6*, 153.

Ward, D., & Stapleton, M. (2012). Es are good. *Consciousness in interaction: The role of the natural and social context in shaping consciousness*, 86-89.

Wing, L. (1997). The history of ideas on autism: legends, myths and reality. *Autism*, *1*(1), 13-23.

Wong, K. (2012) Why humans give birth to helpless babies - *Scientific American Blog Network.* https://blogs.scientificamerican. com/observations/why-humans-give-birth-to-helpless-babies/. Accessed 1 Nov 2017.

Wong, W. C. (2006). Understanding dialectical thinking from a cultural-historical perspective. *Philosophical Psychology*, *19*(2), 239-260.

Wykowska, A., Chaminade, T., & Cheng, G. (2016). Embodied artificial agents for understanding human social cognition. Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1693), 20150375.

Yaroshevsky, M. G. (1989). Lev Vygotsky. Moscow, Russia: Progress.

Zapata-Fonseca, L., Dotov, D., Fossion, R., Froese, T. (2016). Time-series analysis of embodied interaction: movement variability and complexity matching as dyadic properties. Front Psychol 7:1940. https://doi.org/10.3389/fpsyg.2016.01940.

Appendix

A.1. PhD output

A.1.1. Peer reviewed publications

1. **Bolis, D.**, Schilbach, L. (2020). 'Through others we become ourselves': The dialectics of predictive coding and active inference. BBS commentary on “Thinking Through Other Minds: A Variational Approach to Cognition and Culture”.
2. Brandi, M. L., Kaifel, D., **Bolis, D.**, & Schilbach, L. (2019). The Interactive Self–A Review on Simulating Social Interactions to Understand the Mechanisms of Social Agency. i-com, 18(1), 17-31.
3. **Bolis, D.**, & Schilbach, L. (2018). ‘I interact therefore I am’: The self as a historical product of dialectical attunement. Topoi, 1-14.
4. **Bolis, D.**, & Schilbach, L. (2018). Observing and participating in social interactions: action perception and action control across the autistic spectrum. Developmental cognitive neuroscience, 29, 168-175.
5. **Bolis, D.**, Balsters, J., Wenderoth, N., Becchio, C., & Schilbach, L. (2017). Beyond autism: introducing the dialectical misattunement hypothesis and a bayesian account of intersubjectivity. Psychopathology, 50(6), 355-372.
6. **Bolis, D.**, & Schilbach, L. (2017). Beyond one Bayesian brain: Modeling intra-and inter-personal processes during social interaction: Commentary on “Mentalizing homeostasis: The social origins of interoceptive inference” by Fotopoulou & Tsakiris. Neuropsychoanalysis, 19(1), 35-38.
7. **Bolis, D.**, Becchio, C., & Schilbach, L. (2016). Revisiting psychological definitions at the interface of sociocultural historical theories and predictive coding. Welcome to WORLDING THE BRAIN: Patterns, Rhythms, Narratives in Neuroscience and the Humanities. p, 10.

A.1.2. Talks & posters

1. **Bolis D**, Schilbach L (2019) 'I interact therefore I am': Two person psychophysiology and multilevel analysis of intersubjectivity to study interpersonal attunement in social interactions.

Inaugural Workshop of the Social-Cultural Computational Neuroscience and Psychiatry Network. Toward a social-cultural computational psychiatry: Studying human interaction through hyperscanning and Bayesian modelling. Montreal, Canada.

2. **Bolis D**, Padalkar B, Lahnakoski JM, Seidel D, Folz J, Schilbach L (2019) Interaction matters: Elucidating the interpersonal mechanisms underlying joint decision-making in dyadic social interactions. Aegina Summer School 2019 - Norms & Biases in Social Interactions, Aegina, Greece.
3. **Bolis D**, Schilbach (2019) 'I Interact Therefore I Am': I interact therefore I am: Two-person psychophysiology to study interpersonal misattunement in social interaction. Norihito Sadato group, Okazaki, Japan.
4. **Bolis D**, Schilbach L (2019) Beyond the Individual: Multilevel Intersubjectivity to Test the Dialectical Misattunement Hypothesis in Social Interactions. Paris, France.
5. **Bolis D**, Schilbach L (2019) 'I Interact Therefore I Am': The Self as a Historical Product of Dialectical Attunement. Paris, France.
6. **Bolis D**, Lahnakoski J, Seidel D, Tamm J, Folz J, Schilbach L (2019) It takes two to tango: Interpersonal attunement in social interactions. Peter Dayan group, Freiburg, Germany.
7. **Bolis D**, Lahnakoski J, Seidel D, Tamm J, Folz J, Schilbach L (2019) It takes two to tango: Two-person psychophysiology for studying interpersonal attunement in social interaction. Thomas Fuchs group, Heidelberg, Germany.
8. **Bolis D**, Schilbach L (2019) It takes two to tango: Two person psychophysiology to study interpersonal attunement in social interaction. Victoria Leong group, Cambridge, UK.
9. **Bolis D**, Schilbach L (2018) Multidimensional psychiatry from the Bayesian brain to multiperson psychophysiology and multilevel intersubjectivity. DGPPN symposium, Computational and predictive data analysis approaches in psychiatry: Too difficult for psychiatrists or the way to go towards personalized treatment? Berlin, Germany.
10. **Bolis D**, Lahnakoski J, Seidel D, Tamm J, Folz J, Schilbach L (2018) It takes two to tango: Two-person psychophysiology for studying interpersonal attunement in social interaction. School of Advanced Science on Social & Affective Neuroscience. Sao Paulo, Brasil.
11. **Bolis D** & Schilbach (2018) Two-person psychophysiology to study social interaction. The Future of Social Cognition. Norwich, United Kingdom.

12. **Bolis D**, Lahnakoski J, Folz J, Schilbach L (2018) Two-person psychophysiology for capturing the emergence of interpersonal dynamics in social interaction. Aegina Summer School 2018 - New Perspectives & Methods on Social Cognition, Aegina, Greece.
13. **Bolis D**, Lahnakoski J, Schilbach L (2018) It takes two to tango: Two-person psychophysiology for capturing the emergence of interpersonal dynamics in social interaction. SANS, NY, USA.
14. **Bolis D** (2018) Predictive psychiatry: from the Bayesian brain to multiperson psychophysiology. Seminar course for the PhD students of IMPRS-TP, Munich, Germany.
15. **Bolis D**, Lahnakoski J, Schilbach L (2018) Beyond the individual in autism research: Two-person psychophysiology for testing the dialectical misattunement hypothesis. 11th Scientific Meeting for Autism Spectrum Conditions. Frankfurt, Germany.
16. **Bolis D**, Scilbach L (2018) It takes two to tango: Two-person psychophysiology for studying interpersonal attunement in social interaction. Presentation to Karl Friston, Munich Germany.
17. **Bolis D**, Foltz J, Schilbach L (2018) It takes two to tango: Two-person psychophysiology for studying interpersonal attunement in social interaction. Ophelia Deroy group, philosophy of mind chair, LMU, Munich, Germany.
18. **Bolis D**, Lahnakoski J, Schilbach L (2017) Beyond the individual in psychiatry: Studying intra- & interpersonal processes in social interaction. Studying Complex Behaviour Conference. Weizmann Institute of Science. Rehovot, Israel.
19. **Bolis D**, Schilbach L (2017) Beyond the individual: Two-person psychophysiology to investigate intra- and interpersonal processes in social interaction. The Social Brain: Embodiment & Culture. Aegina, Greece.
20. **Bolis D**, Schilbach L (2017) Beyond the individual: Two person psychophysiology & intersubjective Bayesian analysis for studying social interaction. Presentation to Chris and Uta Frith. Munich, Germany.
21. **Bolis D**, Schilbach L (2016) When Vygotsky meets Bayes: Introducing the dialectical misattunement hypothesis, a two-person setup & intersubjective Bayesian analysis. Social Cognition: From interactions to intersubjectivity, Aegina, Greece.
22. **Bolis D**, Schilbach L (2016) Beyond the individual in psychiatry: A two person setup for studying social interactions. Max Plank Institute of Psychiatry, Munich, Germany.
23. **Bolis D** (2015) Predictive Coding and Applications. Max Plank Institute of Psychiatry, Munich, Germany.

A.1.3. Supervised MSc, MD & BSc students

1. Alana Darcher (MSc research)
2. Beril Nisa Can (MSc research)
3. Bhagyashree Padalkar (MSc research)
4. Carolin Nafziger (BSc research)
5. Daniela Seidel (MSc thesis)
6. Dorontina Ismajli (MSc research)
7. Ebru Ecem Tavacioglu (MSc research)
8. Elena Wang (MD research)
9. Emre Yavuz (BSc research)
10. Jeanette Tamm (MSc thesis)
11. Julia Folz (MSc thesis)
12. Kamile Giedraityte (BSc research)
13. Lioba Enk (MSc research)
14. Magdalena Seethaler (MD research)
15. Milena Aleksic (MSc research)
16. Tannaz Mostafid (MSc research)
17. Uta Schneider (MSc research)

A.1.4. Awards

- **Travel grant from the São Paulo School of Advanced Science on Social & Affective Neuroscience**

  Project - It takes two to tango: Two-person psychophysiology for studying interpersonal attunement in social interaction. São Paulo, Brasil, 2018.

- **Prize (#2) from the Society for research in autism spectrum conditions**

  The 11th Scientific Meeting for Autism Spectrum Conditions (WTAS). Project - Beyond the individual in autism research: Two-person psychophysiology for testing the dialectical misattunement hypothesis. Frankfurt, Germany, 2018.

- **Prize (#2) from the Neurocognitive psychology program in LMU**

  Supervised MSc research project (student - Beril Nisa Can) Project - Interpersonal attunement in real time social interaction through facial expressions. Munich, Germany, 2018.

- **Prize (#2) from the Neurocognitive psychology program in LMU**

  Supervised MSc research project (student - Julia Folz) Project - Interpersonal attunement in real time social interaction through eye blinks. Munich, Germany, 2017.

A.1.5 Ongoing empirical research directions

1. Validation of the dialectical misattunement hypothesis in the autistic spectrum
2. EEG Hyperscanning (collaboration with the University of Cambridge, UK)
3. fMRI Hyperscanning (collaboration with NIPS, Japan)
4. Intercultural modulation of interpersonal attunement (collaboration with NIPS, Japan)
5. The impact of perceptual incongruency on interpersonal attunement (collaboration with IIT, Italy)
6. The impact of social motivation and affective bond on interpersonal attunement
7. The interrelation of personal traits and impressions with interpersonal attunement
8. Interpersonal facial expression attunement
9. Social metacognition
10. Multiagent predictive processing

A.2. Text of additional first-authored articles

A.2.1. Revisiting Psychological Definitions at the Interface of Sociocultural Historical Theories and Predictive Coding

Original publication details (submitted 2016; accepted 2016)

Bolis, D., Becchio, C., & Schilbach, L. (2016). Revisiting psychological definitions at the interface of sociocultural historical theories and predictive coding. *WORLDING THE BRAIN: Patterns, Rhythms, Narratives in Neuroscience and the Humanities*. p, 10.

Revisiting Psychological Definitions at the Interface of Sociocultural Historical Theories and Predictive Coding

Dimitris Bolis[a], Cristina Becchio[b,c] & Leonhard Schilbach[a,d]

Affiliations

[a] Independent Max Planck Research Group for Social Neuroscience, Max Planck Institute of Psychiatry, Kraepelinstrasse 2-10, 80804 Munich-Schwabing, Germany
[b] Department of Psychology, University of Torino, Torino, Italy
[c] Robotics, Brain and Cognitive Sciences, Fondazione Istituto Italiano di Tecnologia, Genova, Italy
[d] Department of Psychiatry, University Hospital Cologne, Kerpener Strasse 62, 50924 Cologne, Germany

Here, primarily leaning on sociocultural-historical theories and predictive coding we attempt to revisit psychological and psychiatric definitions through a dialectical prism. More specifically, we emphasize on viewing human condition, not as a static one driven by linear causality, but rather as the outcome of a dialectical interplay of multiple and diverse factors spanning different levels (e.g., biological, cognitive-behavioral and socio-cultural), as well as multiple functions

within a level, being eventually unfolded within several temporal frames (e.g., evolutionary, cultural, social, individual-psychological, subindividual-biological developing scales). In this light, taking autism as a case example, the purpose of this talk is three-fold. Firstly, we consider the embedment of diverse neurocognitive functions in a common framework, namely predictive coding and active inference. Secondly, we underline the importance of taking into account interrelationships across levels. Drawing from seemingly diverse approaches, ranging from engineering to philosophy, we consider the so-called psychopathology not merely as disordered function within single brains, but rather as "misaligned communication" between persons. Subsequently, we present the "Observing-the-Interactors" scheme, which based on a meta-Bayesian framework, aims at accommodating computational modeling of multiple (inter-)level processes (i.e., residing in both individual and collective levels). This will allow us to move beyond the individual as the unit of analysis. Thirdly, we describe a recursive pipeline of successive experimental and computational stages, aiming at facilitating and monitoring an integrative dialogue between hypotheses. Eventually, such an approach could yield a dynamic (re-)definition of multidimensional autism spectrum, here referred to as autism space, as well as a generalized inter-condition space.

A.2.2. Observing and participating in social interactions: Action perception and action control across the autistic spectrum

Original publication details (submitted 2016; accepted 2017)

Bolis, D., & Schilbach, L. (2018). Observing and participating in social interactions: action perception and action control across the autistic spectrum. *Developmental cognitive neuroscience, 29*, 168-175.

Observing and participating in social interactions: Action perception and action control across the autistic spectrum

Dimitris Bolis[a,b] & Leonhard Schilbach[a-d]

Affiliations

[a] Independent Max Planck Research Group for Social Neuroscience, Max Planck Institute of Psychiatry, Munich, Germany
[b] International Max Planck Research School for Translational Psychiatry (IMPRS-TP), Munich, Germany
[c] Department of Psychiatry, Ludwig Maximilian University, Munich, Germany
[d] Graduate School of Systemic Neuroscience (GSN), Munich, Germany

Keywords

Autism · Social interaction · Two-person psychophysiology · Multilevel account · Predictive coding

Abstract

Autism is a developmental condition, characterized by difficulties of social interaction and communication, as well as restricted interests and repetitive behaviors. Although several important conceptions have shed light on specific facets, there is still no consensus about a universal yet specific theory in terms of its underlying mechanisms. While some theories have exclusively focused on sensory

aspects, others have emphasized social difficulties. However, sensory and social processes in autism might be interconnected to a higher degree than what has been traditionally thought. We propose that a mismatch in sensory abilities across individuals can lead to difficulties on a social, i.e., interpersonal level and vice versa. In this article, we, therefore, selectively review evidence indicating an interrelationship between perceptual and social difficulties in autism. Additionally, we link this body of research with studies, which investigate the mechanisms of action control in social contexts. By doing so, we highlight that autistic traits are also crucially related to differences in integration, anticipation and automatic responding to social cues, rather than a mere inability to register and learn from social cues. Importantly, such differences may only manifest themselves in sufficiently complex situations, such as real-life social interactions, where such processes are inextricably linked.

Autism: is it a social or a sensory condition?

Autism is a pervasive developmental condition, which is characterized by difficulties in social interaction and communication, as well as restricted interests and repetitive behaviors. This short definition already suggests that autism's cardinal characteristics fall into two broad categories, first, a collection of social aspects and, second, a group of non-specifically or less social (hereafter, for simplicity, non-social) aspects (e.g., Huerta et al., 2012; Fitzgibbon et al., 2013). Indeed, the vast majority of hypotheses during the last decades have mainly focused on facets either belonging to the one or the other of these two categories.

For instance, on the non-social side, the weak central coherence hypothesis considers autism as a different, detailed-oriented cognitive and perceptual style (Frith, 1989; Happé and Frith, 2006). More precisely, it claims that people with an autism spectrum disorder (ASD) tend to process information locally, rather than globally. According to this idea, people with ASD perceive the world differently in a number of aspects such as visual and auditory information. Similarly, the executive dysfunction hypothesis (e.g., Hill, 2004) focuses on difficulties that people with ASD

face when it comes to executive functions, i.e., problems with functions such as planning, flexibility, inhibition and working memory. On the other hand, one of the first theories focusing on specifically social aspects of the condition, the Theory of Mind hypothesis (Baron- Cohen et al., 1985) proposed that individuals with autism lack a specific meta-representational capacity, namely a "theory of mind", which prevents them from inferring other people's mental states, such as beliefs, emotions or desires. Later, it was suggested that implicit and spontaneous mechanisms of mentalizing might be the ones that are primarily linked to relevant difficulties in autism, rather than explicit processes as initially believed, which might be more easily compensated for through strategic learning (Senju et al., 2009; Schilbach et al., 2011). A second theory focusing on the social dimension emphasizes a special category of neurons, which are thought to be active both when an action is performed and observed (Di Pellegrino et al., 1992; Rizzolatti and Craighero, 2004). This broken mirror neuron hypothesis of autism proposes that impaired social skills in autism are related to dysfunctions in the putative human mirror neuron system making it difficult for individuals with autism to simulate and thus understand others' behavior (MNS; Altschuler et al., 2000; Ramachandran and Oberman, 2006). Some studies have offered supportive evidence for the involvement of the MNS (e.g., Perkins et al., 2010). However, both the validity of a broken MNS and a direct, causal relationship between the MNS and social skills in autism, have been challenged by other reports (e.g., Southgate and Hamilton, 2008). Differences in MNS activation between neurotypical individuals and persons with an ASD could be alternatively traced back to potential modulatory effects of the so called "mentalizing system", a set of brain regions known to subserve explicit mental state attribution (e.g., Wang and Hamilton, 2012; Cook and Bird, 2012; Dumas et al., 2014a). Yet again, the social motivation (SM) hypothesis focuses on motivational rather than cognitive aspects

(Chevallier et al., 2012). It proposes that people with autism lack the social drive inherent to non-autistic individuals, which would assist them in exploiting the necessary learning opportunities in social interactions in order to develop relevant expertise in social cognition. More precisely, this hypothesis is settled upon the fact that the propensity to initiate social contacts, social orienting, social seeking and liking, appears to be diminished in ASD. This idea, however, is brought into ques- tion by evidence, which suggests that individuals with autism are in fact interested in social interaction and exchange, but only when the interaction is structured in such a way that it suits their needs (Wing and Gould 1979; Schilbach 2016a).

In short, several important theories on autism have advanced our understanding in crucial facets of the condition; however, there is still no established unified account, which could explain social and sensory aspects of autism in the context of their inherent inter- relationship. In fact, it has even been suggested that a single theory might be intractable (Happé, 2003; Happé et al., 2006; Gallagher and Varga, 2015). However, recent developments centered around the idea of the human brain organized around principles of Bayesian inference and predictive coding have recently refueled interest in a unifying account of autism: For instance, Pellicano and Burr (2012), adopted a standpoint to argue that non-social features of autism might be explained in reference to attenuated Bayesian priors (so-called hypo-priors), which suggests that previ- ous experiences might be less important when processing current sensory input for individuals with autism. This hypothesis predicts the more accurate and acute perception in autism, driven primarily by perceptual evidence as opposed to prior knowledge, as well as the sense of being overwhelmed by this information, which is commonly reported by individuals with autism. The hypo- priors hypothesis was then reformulated (Friston et al., 2013; Van Boxtel and Lu, 2013) within the predictive coding scheme, a more specific Bayesian account (Mumford, 1992; Friston,

2005; Friston, 2008; Clark, 2013), while considering social aspects of cognition and behavior as well (Lawson et al., 2014; Van de Cruys et al., 2014). The predictive coding framework relies on the idea that sensory information is processed hierarchically in levels of increasing abstraction. In this setting, prediction errors (i.e., the discrepancy between predictions and incoming information) ascend the processing hierarchy for optimizing neural configuration in generating accurate predictions, which descending the hierarchy, are contrasted to sensory input. More concretely, higher levels of the hierarchy produce predictions, which are tested against the input information of the immediate lower levels. Propagating only the prediction error and not the actual incoming information to higher levels is an efficient and resource-oriented way of reducing the bandwidth of the processed information. The neural processes and computations needed to extract regularities in the environment can be described in terms of Bayesian inference. In this regard, the brain is thought to represent information accessed via the sensory organs in the form of probability densities; these probabilities are maintained via a combination of already gained experience (so-called priors) and newly sensed information (evidence). The more confidence (precision) is placed on the validity of experience the less the latter is updated in the face of new incoming information. The ultimate goal of such a predictive system is the effective minimization of the prediction error, through perception, learning and action (for a comprehensive review of traditional theories and a future integrative direction in autism research see Bolis *et al*., under review).

Such endeavors of developing a more unified account of autism are further supported by evidence that social and non-social domains are not as independent as once might have been assumed in research practice. For instance, Linkenauger et al. (2012) showed that deficits of individuals with autism in relating information about their own bodies' action capabilities to

visual information specifying the environment, strongly predicted the degree of social and communicative difficulties. Additionally, MacDonald et al. (2013) demonstrated that children with autism that showed weaker motor skills had greater social communicative skill difficulties. Moreover, Leekam et al. (2007) linked the distinct sensory processing in autism with higher-level social processes. Having said that, focusing on 'internal' (i.e., within individual brains) dynamics has, indeed, yielded informative insights, such as providing insights into the relevance of a dysbalance of inhibitory/excitatory neurotransmission in autism (e.g., Robertson et al., 2016). Additionally, considering 'external' (i.e., collective socio-cultural) dynamics, such as the role of collaborative morality (Spikins et al., 2016) or social expectations of others (Jensen et al., 2016), can prove to be crucial in achieving a comprehensive account of autism. However, studying 'internal' and 'external' dynamics in isolation and thus neglecting the dialectics between the individual and the collective (Vygotsky, 1930–1935/1978; views of Vygotsky and colleagues in Dafermos, 1930–1935/2002), which are inherently intertwined across multiple temporal scales (i.e., from evolutionary and cultural to developmental and daily learning processes), might result in misconstruing the essence of a condition such as autism (Bolis *et al*., under review).

More specifically, Vygotsky and colleagues argued that the development of the human mind has its origin at the interaction between the individual and society, viewing culture and social interaction as the major developmental driving forces (e.g., Vygotsky, 1934/2008; 1930–1935/1978). When it comes to children with certain "disabilities", one of the main propositions of the so-called cultural historical approach was the recognition of primary and secondary difficulties. It was suggested that it is not the primary difficulties, which are directly linked to the physical condition, that are crucial to a child's development, but rather the secondary ones,

which relate to an exclusion from sociocultural activities, which other children freely participate in. Different factors can contribute to this, such as unsuitable cultural and technological environment or social expectations of others. Interestingly, such lines of thought lend support to alternative avenues of research and intervention, which will not exclusively target the person of interest (i.e., an individual with autism in our case), but the social environment as well, e.g., via personalizing education (Vittorias *et al*., 2008) or facilitating communication between people (Bolis *et al*., under review), bringing focus back to social interaction (Schilbach et al., 2013).

Taken together, we suggest that it may be the coupling and the inextricable interplay of sensory and motor functions (within an individual) rather than selective deficits thereof – and whether or not a given partner in a social interaction is more or less similar to the autistic person (at the interpersonal level), which play an important role in the development and manifestation of cardinal characteristics of autism.

In this article, we, therefore, review evidence, that addresses the integration of sensory processing and individual as well as interpersonal aspects of action control to suggest that individuals with high autistic traits are not 'blind' to social information in the environment, but the extent to which they update their beliefs and they are, thus, influenced by this information when making decisions and executing actions is lower than in individuals with low autistic traits. Furthermore, we review findings to suggest that a more comprehensive understanding of autism (and other psychiatric disorders) will have to rely upon studying it in the context of ecologically valid real-time social interactions as social difficulties are known to be more pronounced (or may only manifest) under such conditions (Schilbach *et al*., 2013; Schilbach, 2016a). Here, we propose that a mismatch in autistic trait-related perceptual abilities across individuals can lead to

difficulties on a social level and briefly describe a two-person experimental setup and avenues for future research to formally investigate this.

From action observation to interaction in autism

An important initial suggestion for understanding and studying the biological basis of autistic symptomatology has been to focus on the ability of individuals with autism to observe and interpret actions in others. With regard to general visual abilities and the idea of an eagle eye hypothesis some studies have provided evidence for superior performance of individuals with autism on visual tasks (Dakin & Frith, 2005). Furthermore, evidence has been found to document an increased reliance on visuospatial information by ASD individuals, which was related to increased parietal brain activation (e.g., DeRamus *et al*., 2014). Such differences in visuo-spatial processing have also been related to differences at the level of neurotransmission: binocular rivalry, a visual function that is thought to rely on the balance of excitation/inhibition in visual cortex, has been shown to be tightly linked to GABAergic signaling in healthy controls, while this link was shown to be completely and specifically absent in autism (Robertson *et al*., 2016). Also, alterations of serotonergic functioning have been discussed as a contributory factor in autism (e.g., Cook & Leventhal, 1996). Interestingly, serotonergic modulation is also known to change the balance between different sources of neural activity in sensory systems (Lottem *et al*., 2016).

With regard to visual processing of others' actions, disruptions in the visual perception of biological motion have been discussed as a potential hallmark of ASD (Kaiser and Pelphrey, 2012). Here, an impaired sensitivity for processing the actions of others (as compared to observing objects) has been demonstrated (e.g., Blake *et al*., 2003; Kaiser *et al*., 2010). These

findings were also paralleled by neuroimaging results indicative of atypical neural response patterns to biological motion perception with point-light stimuli (e.g., Herrington *et al*., 2007). Such point-light displays are created by attaching markers to a person's body and head and then recording that person's movements so that only the point-lights are visible (cf. Johansson, 1973). These stimuli are easily manipulated and predispositions to process their social aspects can be inferred by detecting enhanced behavioral sensitivity to displays of biological motion relative to animal (Pinto & Shiffrar, 2009), object (Kaiser *et al*., 2010) or other kinds of biological motions (Manera *et al*., 2010, 2011). Importantly, differential responses to point-light displays of human movement are already observed in typically developing infants (Yoon & Johnson, 2009), which has been taken to suggest that biological motion perception relies on an early emerging, evolutionary conserved brain system (Kaiser & Pelphrey, 2012). A large body of work has demonstrated responses of the so-called "social brain" to point-light displays of biological motion (for a review see Blake & Shiffrar, 2007). Here, the posterior superior temporal sulcus (pSTS) has been discussed as a key component of the neural system that supports social perception.

Studies of the visual perception of biological motion have, thus, been thought to provide a window into social dysfunction in ASD. This was based upon the idea that autism is related to an early and initial failure to develop the specialized brain mechanisms for social perception which, in turn, results in abnormal development and the phenotypic expression of ASD (Pelphrey *et al*., 2011). Consequently, different studies have investigated behavioral responses to point-light displays of biological motion in individuals with autism. While some studies do find impairments, the overall picture appears to be rather mixed. Interestingly, several studies have documented that when asked to verbally describe point-light displays of biological motion,

children and adults with ASD exhibit impairments in emotion perception, but intact action perception (Hubert *et al*., 2007; Moore *et al*., 1997; Parron et al., 2008). Recently, Cusack *et al.* (2015) have systematically addressed this question by using an extensive test battery of point light displays in adult individuals with autism. Here, it was consistently shown that action perception is intact when autistic individuals are motivated to perform the relevant task under controlled conditions. This finding was replicated by von der Lühe *et al.* (2016), who demonstrated that, when prompted and explicitly asked to assess point-light displays of individual and even communicative actions between two agents, participants with autism do not perform worse than a group of matched controls. Only when the complexity of the task is increased, however, by including noise signals, do individuals with autism show an impairment of interpersonal action prediction, i.e., no modulation in behavioral sensitivity for the detection of a second agent in light of a first agent, who generates a communicative action, is found. Interestingly, this ability to predict action sequences across two agents shows an inverse relationship with increasing autistic traits across the entire spectrum (von der Lühe *et al*., 2016).

Apart from the ability to perceive the relevant aspects of the social environment and their interrelation, it is, of course, imperative to be able to quickly and adaptively generate adequate behavior to respond to it. In this respect, a recent suggestion has been that social perception and cognition may be fundamentally different when we are actively engaged in interaction with others as compared to merely observing others. This difference may be particularly relevant in autism and has been taken to suggest that interactive situations rely more heavily on the integration of perception- and action-based processes (Schilbach *et al*., 2013). While a vast literature exists on the behavioral and neural correlates of social observation, much less is known about the behavioral and neural mechanisms of social interaction (Schilbach, 2015).

In direct social interaction the unconscious imitation of the actions of others (often described as mimicry) is a powerful and ubiquitous behavior. Facial mimicry, for instance, is thought to be a form of "physiological linkage" or socio-emotional contagion between individuals (Dimberg, 1982; Dimberg *et al*., 2000) and assumed to be of considerable importance for interpersonal communication (Niedenthal *et al*., 2005; Schilbach, 2016b). Conversely, alterations of involuntary facial reactions (e.g., due to conditions resulting in facial paralysis) may have a detrimental effect on the quality of interpersonal communication (Cole, 2001; Oberman *et al*., 2007). In a series of studies, Schilbach and colleagues investigated the behavioral and neural correlates of facial mimicry by using anthropomorphic virtual characters that showed either self- or other-directed facial expressions (e.g., smiling) as compared to arbitrary facial movements (e.g., puckering; Schilbach *et al*., 2006; Mojzisch *et al*., 2006; Schilbach *et al*., 2008). In these studies, it was shown that attention allocation, as assessed by fixation duration, was specifically related to the perception of self-directed stimuli. EMG measurements demonstrated that facial activity was influenced by the perception of socially relevant facial expressions and showed spontaneous, involuntary facial responses irrespective of whether the facial expression was directed towards the observer or not (Mojzisch et al., 2006).

In a separate fMRI study, it was demonstrated that specific brain activations are related to the occurrence of involuntary facial movements in human observers in response to the perception of socially relevant facial expressions shown by perceived others. These activations comprise but extend beyond classical motor regions (i.e., face motor area) and include other regions of the social brain, such as the mentalizing system. While activity in motor cortex might help to generate a representation of the action which may, in fact, translate to mimicking that behavior oneself, involvement of the mentalizing system might contribute to social cognition by

processing the differentiation of self and other. In dyadic interaction both mechanisms are crucially important as a facial expression might highlight someone else's internal state, but could also refer to some object or might be expressive of the assessment of the vis-à-vis behavior or the process of interacting itself. Whether mimicry responses are abnormal in individuals with ASD has been subject of long-standing debate. Seminal studies by Heyes and colleagues have demonstrated that individuals with ASD, in fact, do show automatic imitative behavior (e.g., Bird *et al*., 2007; see also Southgate & Hamilton, 2008). More recently, Hamilton and co-workers (Forbes *et al*., 2016) showed that individuals with ASD also show mimicry, but they do not show an enhancement of mimicry responses in a gaze-based social context as neurotypical individuals do; hence providing support for the social top-down response modulation (STORM) model of mimicry in autism, in that the mimicry response is not top-down modulated by social context.

In another series of studies conducted by Schilbach *et al.* (2010, 2011), an action control paradigm was used, which required that participants performed spatially congruent or incongruent but- ton presses to respond to an unpredictable change in the visual stimulus. In order to investigate the effect of social context on behavioral and neural mechanisms of action control, the stimulus type was varied to include both social (face stimulus) and non- social (geometric shape) stimuli. Using this paradigm it was shown that the so-called incongruency costs, i.e., longer reaction times necessary to generate a spatially incongruent as compared to a congruent response, were significantly smaller for the social as compared to the non-social stimulus in healthy controls. At the neural level this effect reflected by a differential increase of neural activity in subcortical structures relevant for the habitual performance of actions (caudate nucleus) as well as regions involved in action monitoring (ACC), action preparation (IFG) and

social cognition (dorsomedial prefrontal cortex). These findings indicate that even a minimal, gaze-based social context significantly and automatically changes the neural networks relevant for action control, which might explain the effortless and prompt behavioral adaptations that non-autistic individuals make in the presence of others. As expected, a group of individuals with high-functioning autism did show incongruency effects, but those were not modulated by the social context.

While the above described studies are informative, they do not provide an account of the computational mechanisms that may underlie differences in behavior and brain activity. As described in the introduction, recent developments in cognitive neuroscience have embraced a perspective that describes the brain as a "prediction machine" whose ultimate goal it is to construct a model of the environment in order to predict the causes of sensory input, to anticipate future states and to minimize the resulting 'prediction error'. Specific theoretical commitments aside, all these accounts tend to agree in suggesting that perception in individuals with ASD may be more strongly driven by perceptual evidence as compared to prior knowledge. This may explain a stronger reliance on perceptual input, but may render ambiguous situations, in which sensory input is not sufficient to disambiguate, problematic, because they may require stronger reliance on prior knowledge. Consequently, such accounts can be seen to align with suggestions made by Wang and Hamilton (2012) that autistic symptomatology may result from difficulties in the context-sensitive, top down modulation of perceptual processing. In order to test, whether autistic traits are, in fact, related to differences in Bayesian inference, Sevgi *et al*. (2016) used computational modeling to investigate autistic trait-related differences in the weighting of social and non-social information during reward-based learning and decision-making. In this study, it was shown that individuals with higher autistic traits are not 'blind' to social information in the

environment and, in fact, do learn about it, but the extent to which they update their beliefs and are, thus, influenced by this information when making decisions is lower than in individuals with lower autistic traits. This again emphasizes that what might be different in autism is not merely the ability to process social information, but a propensity to actually use this information when generating actions oneself.

Until today, no consistent neurochemical, neurophysiological, or neuroanatomical abnormality has been detected that could be directly used to inform the diagnosis of ASD. Based on the above described findings, it may be important in future research to assess the neural mechanisms of how sensory processing reaches cortical hubs and influences decision-making and action control (Sepulcre *et al*., 2015) in more ecologically valid social situations (Schilbach, 2015, 2016a,b). Here, it will also be important to investigate the relevant motivational mechanisms, which might be key to understanding why and when individuals actually use their social perceptual skills (e.g., Schilbach *et al*., 2010). Such studies should also include an investigation of the neurochemical basis of social reward (e.g., Dölen *et al*., 2013; Dölen 2015) and could address how differences and/or similarities across interacting individuals may impact whether individuals are motivated and able to interact with one another. Here, clinical intuition and anecdotal evidence suggests that individuals with autism might be more motivated and/or able to interact with other individuals with autism (Schilbach, 2016a). To investigate these questions systematically, we have generated an experimental platform, in which behavior and its underlying computational mechanisms can be assessed individually and collectively while two persons perform the same task. It is to the description of this setup that we now turn.

Two-person psychophysiology and computational modeling: a framework for studying the interrelation of sensory and social processes in social interaction

In this section, we present an experimental framework and analysis schemes for future research of multilevel mechanisms of social interaction, i.e., from (sub-)individual to interpersonal processes, and most importantly their interrelationships. The importance of considering the dialectics between individual and socio-cultural processes, with an emphasis on social interaction, in psychological and psychopathological research has been emphasized in the past (e.g., in cultural historical theories; Vygotsky, 1930–1935/1978; views of Vygotsky and colleagues in Dafermos, 1930–1935/2002). However, due to conceptual and methodological constraints, neuropsychiatric research has not until recently addressed these issues (for some recent attempts of the field to consider more interactive scenarios see Montague *et al*., 2002; Schilbach *et al*., 2006, 2010; Pfeiffer *et al*., 2011; Dumas *et al*., 2012; Müller *et al*., 2013; Konvalinka & Roepstorff, 2012; Froese *et al*., 2014; Dumas *et al*., 2014b; Bilek *et al*., 2015). Such studies have been successful in tapping into some intra-, as well as interpersonal aspects of social interactions, such as joint attention, imitation, interpersonal synchronization and joint decision-making. In this article, we have emphasized the importance of interplay between sensory and social processes for the development and manifestation of cardinal characteristics of autism. Consequently, we are now describing a two-person setup, which allows for a formal investigation of this interplay during gaze-based social interaction. In this setup, participants are sitting opposite each other, trying to accomplish perceptual tasks individually, while interacting via gaze behavior in real time through a micro-camera communication system. This dual functionality will be pivotal for relating multilevel processes, ranging from the level of the individual (e.g., observation and reaction to perceptual information; **Fig. A.1**: blue and green arrows) to the level of the collective (e.g., social interaction loop; **Fig. A.1**: red arrows).

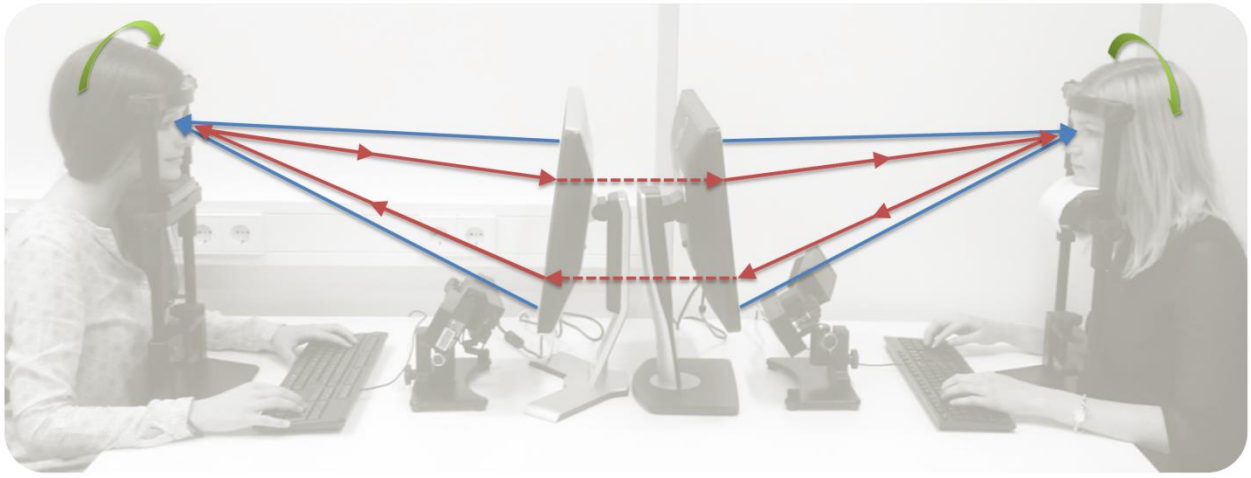

Fig. A.1. Two-person psychophysiology setup: This setup allows for an investigation of the interrelation of sensory and social processes in direct, gaze-based social interaction between two persons (Person A, Person B); blue arrows: incoming sensory information from non-social stimuli; green arrows: intra-personal processes; red arrows: social interaction loop mediated by a micro-camera communication system.

Crucially, such a setup will enable the consideration of all possible types of dyads (i.e., dyads consisting of neurotypical persons, dyads of persons with autism, as well as neurotypical-autistic dyads). Furthermore, its design allows for using tasks that comprise both free viewing and strictly structured tasks that can be performed individually, cooperatively or competitively. This com- promise will allow for formal interpretation of the data, while preserving adequate degrees of ecological validity. To this end, the two-person setup transparently obtains high-resolution empirical data via infrared eye-trackers, micro-cameras, biological motion sensors and (electro-)physiological recorders, thus, providing multiple behavioral and (electro-)physiological readouts, such as gaze position, facial expression, heart rate variability and brain activity. Collection of biological measurements will enable the specification of lower level biological mechanisms, such as genetic and epigenetic mechanisms, while consideration of cultural factors, e.g., through priming or targeted selection of participants, might allow for touching upon higher

level processes at the level of interpersonal exchange. Such a multi-level approach could complement enactivist and multi-scale approaches (e.g., De Jaegher & Di Paolo, 2007; Froese *et al*., 2011; Dumas *et al*., 2014c), as well as facilitate an integration of seemingly disparate perspectives (e.g., Frith, 1996; Gordon, 2016; Harris, 2016).

Importantly, the acquisition of high dimensional two-person data sets will allow for multi-level analyses via advanced computational methods (e.g., Bayesian and cross-recurrence approaches; Montague *et al*., 2012; Stephan & Mathys, 2014; Marwan *et al*., 2007). Firstly, intra-individual processes (**Fig. A.1**; green arrows) could be modelled on the basis of Bayesian integration (**Fig. A.2a**). Here, studying mechanisms for combining non-social (**Fig. A.1**; blue arrows) with social (**Fig. A.1**; red arrows) information will be mostly relevant (see Sevgi *et al*., 2016). More specifically, non-social information could take the form of perceptual stimuli, while social information may be constituted by both individual social cues, such as other person's gaze, and collective parameters, such as interpersonal coupling. Secondly, inter-individual processes (**Fig. A.1**; red arrows) could be modelled on top of intra-individual ones (**Fig. A.1**; blue and green arrows) via an intersubjective Bayesian analysis (**Fig. A.2b**; Bolis *et al*., under review). The latter scheme aims at modeling both individual mechanisms (orange boxes; e.g., from neurobiological to cognitive-behavior levels) and processes of collective behavior (green box; e.g., joint decision making and interpersonal synchronization). More concretely, in **Fig. A.2b**, μ and w symbolize individual beliefs and confidence placed on these beliefs respectively. The latter are translated and interlinked on a collective level of description via transformative (e.g., q and ω) and coupling parameters (e.g., p), resulting in observable activity (for more details see Bolis *et al*., under review).

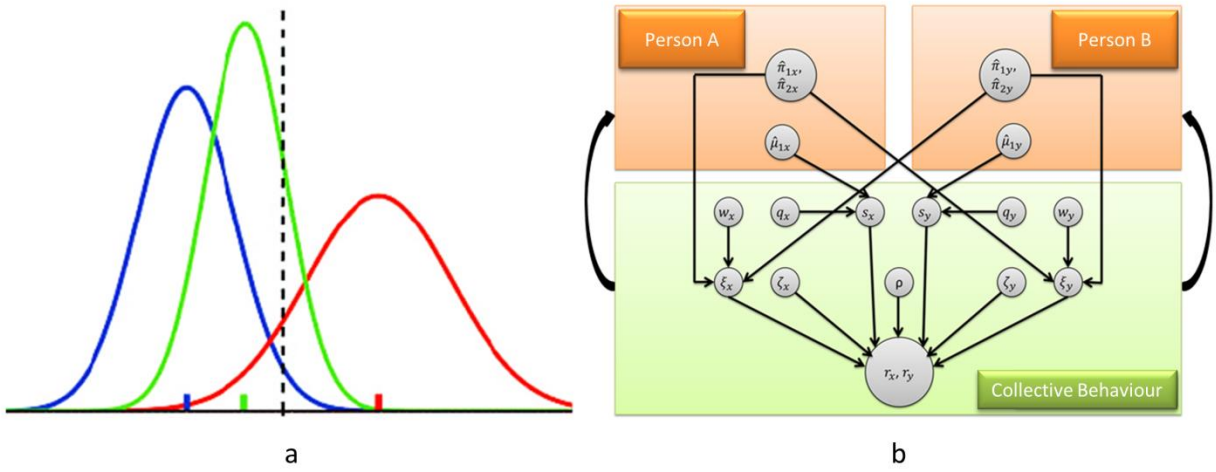

Fig. A.2. (a) An example of statistically optimal Bayesian integration (green) between non-social (blue) and social information (red). (b) A schematic representation of inter-subjective Bayesian analysis for capturing both individual mechanisms (orange boxes) and processes of collective behavior (green box); adjusted from Bolis *et al*., 2015; under review). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Conclusions

In this article, we have emphasized the importance of studying sensory-motor integration at the individual as well as at the dyadic level to more fully understand sensory and social aspects of autism. This line of research parts from more traditional theories, which have focused on specific cardinal aspects of autism that fall either into the social realm, such as social motivation, or a non- social domain, such as central coherence, while being restricted to the level of the individual. In order to extend those previous attempts of investigating the integration of sensory and social cues at the individual level, recent developments of predictive coding accounts appear most promising and allow for an investigation of the computational mechanisms of autistic trait-related differences in social cognition. Furthermore, we have reviewed empirical evidence, which indicates that autistic individuals might not face the greatest difficulties during passive

observation of either social or non-social stimuli, but rather when it comes to real-time interaction, which make cue integration necessary and the generation of adequate behavioral responses in the context of ongoing and reciprocal social interactions. Taking social interaction seriously, however, we believe, also means that hypotheses and empirical studies to test them will have to go beyond the individual as the level of analysis. Here, an important objective will be to explain how processes of interpersonal coordination can emerge from and are reciprocally connected to the functioning of individual sensory-motor processes. In order to realize such truly social studies in autism (and other psychiatric disorders), we have developed a two- person setup, which allows for the study of mechanisms at both the individual and the collective level during real-time social interaction. This development opens up completely new avenues for autism research and might help to arrive at a more comprehensive understanding of how exactly autism might be neither an exclusively sensory nor an exclusively social condition.

References

Altschuler, E.L., Vankov, A., Hubbard, E.M., Roberts, E., Ramachandran, V.S., Pineda, J.A., 2000. Mu wave blocking by observation of movement and its possible use as a tool to study theory of other minds. Soc. Neurosci. 26, Abstracts 67.23.

Baron-Cohen, S., Leslie, A.M., Frith, U., 1985. Does the autistic child have a theory of mind? Cognition 21 (1), 37–46.

Bilek, E., Ruf, M., Schäfer, A., Akdeniz, C., Calhoun, V.D., Schmahl, C., ... Meyer-Lindenberg, A., 2015. Information flow between interacting human brains: identification, validation, and relationship to social expertise. Proc. Natl. Acad. Sci. 112 (16), 5207–5212.

Bird, G., Leighton, J., Press, C., Heyes, C., 2007. Intact automatic imitation of human and robot actions in autism spectrum disorders. Proc. R. Soc. Lond. B: Biol. Sci. 274 (1628), 3027–3031.

Blake, R., Shiffrar, M., 2007. Perception of human motion. Annu. Rev. Psychol. 58, 47–73.

Blake, R., Turner, L.M., Smoski, M.J., Pozdol, S.L., Stone, W.L., 2003. Visual recognition of biological motion is impaired in children with autism. Psychol. Sci. 14 (2), 151–157.

Bolis, D., Heinzle, J., Mathys, C., Stephan, K.E., 2015. Inferring Cognitive Traits of Individual Subjects Through Gaze Controlled Video Games.

Bolis D., Balsters J., Wenderoth N., Becchio C., Schilbach L., Beyond autism: Introducing the dialectical misattunement hypothesis and a Bayesian account of intersubjectivity, (under review).

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E.S., Schultz, R.T., 2012. The social motivation theory of autism. Trends Cogn. Sci. 16 (4), 231–239.

Clark, A., 2013. The many faces of precision. Front. Theor. Philos. Psychol. 4, 270. Cole, J., 2001. Empathy needs a face. J. Conscious. Stud. 8 (5–6), 51–68.

Cook, J.L., Bird, G., 2012. Atypical social modulation of imitation in autism spectrum conditions. J. Autism Dev. Disord. 42 (6), 1045–1051.

Cook Jr., E.H., Leventhal, B.L., 1996. The serotonin system in autism. Curr. Opin. Pediatr. 8 (4), 348–354.

Cusack, J.P., Williams, J.H., Neri, P., 2015. Action perception is intact in autism spectrum disorder. J. Neurosci. 35 (5), 1849–1857.

Dölen, G., Darvishzadeh, A., Huang, K.W., Malenka, R.C., 2013. Social reward requires coordinated activity of nucleus accumbens oxytocin and serotonin. Nature 501 (7466), 179–184.

Dölen, G., 2015. Autism: oxytocin, serotonin, and social reward. Soc. Neurosci. 10 (5), 450–465.

Dafermos M., 1930–1935/2002. The cultural historical theory of Vygotsky: philosophical, psychological and pedagogical aspects, Atrapos (in Greek).

Dakin, S., Frith, U., 2005. Vagaries of visual perception in autism. Neuron 48, 497–507.

De Jaegher, H., Di Paolo, E., 2007. Participatory sense-making. Phenomenol. Cognit. Sci. 6 (4), 485–507.

DeRamus, T.P., Black, B.S., Pennick, M.R., Kana, R.K., 2014. Enhanced parietal cortex activation during location detection in children with autism. J. Neurodev. Disord. 6 (1), 1.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G., 1992. Understanding motor events: a neurophysiological study. Exp. Brain Res. 91, 176–180.

Dimberg, U., Thunberg, M., Elmehed, K., 2000. Unconscious facial reactions to emotional facial expressions. Psychol. Sci. 11 (1), 86–89.

Dimberg, U., 1982. Facial reactions to facial expressions. Psychophysiology 19 (6), 643–647.

Dumas, G., Martinerie, J., Soussignan, R., Nadel, J., 2012. Does the brain know who is at the origin of what in an imitative interaction? Front. Hum. Neurosci. 6, 128.

Dumas, G., Soussignan, R., Hugueville, L., Martinerie, J., Nadel, J., 2014a. Revisiting mu suppression in autism spectrum disorder. Brain Res. 1585, 108–119.

Dumas, G., de Guzman, G.C., Tognoli, E., Kelso, J.S., 2014b. The human dynamic clamp as a paradigm for social interaction. Proc. Natl. Acad. Sci. 111 (35), E3726–E3734.

Dumas, G., Kelso, J.S., Nadel, J., 2014c. Tackling the social cognition paradox through multi-scale approaches. Front. Psychol. 2014 (5), 882.

Fitzgibbon, B.M., Segrave, R.A., Fitzgerald, P.B., Enticott, P.G., 2013. Can studies of pain help to bridge the gap between sensory and social impairments in autism? Proc. Natl. Acad. Sci. USA 108, 6270–6275.

Forbes, P.A., Wang, Y., Hamilton, A.F.D.C., 2016. Stormy interactions: gaze and the modulation of mimicry in adults on the autism spectrum. Psychon. Bull. Rev., 1–7.

Friston, K., Lawson, R., Frith, C.D., 2013. On hyperpriors and hypopriors: comment on Pellicano and Burr. Trends Cogn. Sci. 17 (1), 1.

Friston, K., 2005. A theory of cortical responses. Philos. Trans. R. Soc. B: Biol. Sci. 360, 815–836.

Friston, K., 2008. Hierarchical models in the brain. PLoS Comput. Biol. 4, e1000211. Frith, U., 1989. Autism: Explaining the enigma. Blackwell Scientific Publications, Oxford.

Frith, U., 1996. Cognitive explanations of autism. Acta Paediatr. 85 (s416), 63–68.

Froese, T., Paolo, D., Ezequiel, A., 2011. The enactive approach: theoretical sketches from cell to society. Pragmat. Cognit. 19 (1), 1–36.

Froese, T., Iizuka, H., Ikegami, T., 2014. Embodied social interaction constitutes social cognition in pairs of humans: a minimalist virtual reality experiment. Sci. Rep. 4.

Gallagher, S., Varga, S., 2015. Conceptual issues in autism spectrum disorders. Curr. Opin. Psychiatry 28 (March (2)), 127–132.

Gordon, J.A., 2016. On being a circuit psychiatrist. Nat. Neurosci. 19 (11), 1385.

Happé, F., Frith, U., 2006. The weak coherence account: detail-focused cognitive style in autism spectrum disorders. J. Autism Dev. Disord. 36 (1), 5–25.

Happé, F., Ronald, A., Plomin, R., 2006. Time to give up on a single explanation for autism. Nat. Neurosci. 9 (10), 1218–1220.

Happé, F., 2003. Cognition in autism: one deficit or many? In: Autism: Neural Basis and Treatment Possibilities: Novartis Foundation Symposium 251. John Wiley & Sons Ltd., pp. 198–212.

Harris, J.C., 2016. The origin and natural history of autism spectrum disorders. Nat. Neurosci. 19 (11), 1390–1391.

Herrington, J.D., Baron-Cohen, S., Wheelwright, S.J., Singh, K.D., Bullmore, E.T., Brammer, M., Williams, S.C., 2007. The role of MT+/V5 during biological motion perception in Asperger Syndrome: an fMRI study. Res. Autism Spectr. Disord. 1 (1), 14–27.

Hill, E.L., 2004. Executive dysfunction in autism. Trends Cognit. Sci. 8 (1), 26–32. Hubert, B., Wicker, B., Moore, D.G., Monfardini, E., Duverger, H., Da Fonseca, D.,

Deruelle, C., 2007. Brief report: recognition of emotional and non-emotional biological motion in individuals with autistic spectrum disorders. J. Autism Dev. Disord. 37 (7), 1386–1392.

Huerta, M., Bishop, S.L., Duncan, A., Hus, V., Lord, C., 2012. Application of DSM-5 criteria for autism spectrum disorder to three samples of children with DSM-IV diagnoses of pervasive developmental disorders. Am. J. Psychiatry 169 (10), 1056–1064.

Jensen, C.M., Martens, C.S., Nikolajsen, N.D., Gregersen, T.S., Marx, N.H., Frederiksen, M.G., Hansen, M.S., 2016. What do the general population know, believe and feel about individuals with autism and schizophrenia: results from a comparative survey in Denmark. Autism 20 (4), 496–508.

Johansson, G., 1973. Visual perception of biological motion and a model for its analysis. Percept. Psychophys. 14 (2), 201–211.

Kaiser, M.D., Pelphrey, K.A., 2012. Disrupted action perception in autism: behavioral evidence, neuroendophenotypes, and diagnostic utility. Dev. Cognit. Neurosci. 2 (1), 25–35.

Kaiser, M.D., Delmolino, L., Tanaka, J.W., Shiffrar, M., 2010. Comparison of visual sensitivity to human and object motion in autism spectrum disorder. Autism Res. 3 (4), 191–195.

Konvalinka, I., Roepstorff, A., 2012. The two-brain approach: how can mutually interacting brains teach us something about social interaction? Front. Hum. Neurosci. 6, 215.

Lawson, R.P., Rees, G., Friston, K.J., 2014. An aberrant precision account of autism. Front. Hum. Neurosci. 2014 (8), 302.

Leekam, S.R., Nieto, C., Libby, S.J., Wing, L., Gould, J., 2007. Describing the sensory abnormalities of children and adults with autism. J. Autism Dev. Disord. 37 (5), 894–910.

Linkenauger, S.A., Lerner, M.D., Ramenzoni, V.C., Proffitt, D.R., 2012. A perceptual–motor deficit predicts social and communicative impairments in individuals with autism spectrum disorders. Autism Res. 5 (5), 352–362.

Lottem, E., Lörincz, M.L., Mainen, Z.F., 2016. Optogenetic activation of dorsal raphe serotonin neurons rapidly inhibits spontaneous but not odor-Evoked activity in olfactory cortex. J. Neurosci. 36 (1), 7–18.

Müller, V., Sänger, J., Lindenberger, U., 2013. Intra- and inter-brain synchronization during musical improvisation on the guitar. PLoS One 8 (9), e73852.

MacDonald, M., Lord, C., Ulrich, D.A., 2013. The relationship of motor skills and social communicative skills in school-aged children with autism spectrum disorder. Adapt. Phys. Activ. Q 30 (3), 271–282.

Manera, V., Schouten, B., Becchio, C., Bara, B.G., Verfaillie, K., 2010. Inferring intentions from biological motion: a stimulus set of point-light communicative interactions. Behav. Res. Methods 42 (1), 168–178.

Manera, V., Becchio, C., Schouten, B., Bara, B.G., Verfaillie, K., 2011. Communicative interactions improve visual detection of biological motion. PLoS One 6 (1), e14594.

Marwan, N., Romano, M.C., Thiel, M., Kurths, J., 2007. Recurrence plots for the analysis of complex systems. Phys. Rep. 438 (5), 237–329.

Mojzisch, A., Schilbach, L., Helmert, J.R., Pannasch, S., Velichkovsky, B.M., Vogeley, K., 2006. The effects of self-involvement on attention, arousal, and facial expression during social interaction with virtual others: a psychophysiological study. Soc. Neurosci. 1 (3–4), 184–195.

Montague, P.R., Berns, G.S., Cohen, J.D., McClure, S.M., Pagnoni, G., Dhamala, M., Fisher, R.E., 2002. Hyperscanning: simultaneous fMRI during linked social interactions. Neuroimage 16 (4), 1159–1164.

Montague, P.R., Dolan, R.J., Friston, K.J., Dayan, P., 2012. Computational psychiatry. Trends Cogn. Sci. 16 (1), 72–80.

Moore, D.G., Hobson, R.P., Lee, A., 1997. Components of person perception: an investigation with autistic, non autistic retarded and typically developing children and adolescents. Br. J Dev. Psychol. 15 (4), 401–423.

Mumford, D., 1992. On the computational architecture of the neocortex II. Biol. Cybern. 66, 241–251.

Niedenthal, P.M., Barsalou, L.W., Winkielman, P., Krauth-Gruber, S., Ric, F., 2005. Embodiment in attitudes, social perception, and emotion. Pers. Soc. Psychol. Rev. 9 (3), 184–211.

Oberman, L.M., Winkielman, P., Ramachandran, V.S., 2007. Face to face: blocking facial mimicry can selectively impair recognition of emotional expressions. Soc. Neurosci. 2 (3–4), 167–178.

Parron, C., Da Fonseca, D., Santos, A., Moore, D., Monfardini, E., Deruelle, C., 2008. Recognition of biological motion in children with autistic spectrum disorders. Autism 12 (3), 261–274.

Pellicano, E., Burr, D., 2012. When the world becomes ‘too real’: a Bayesian explanation of autistic perception. Trends Cogn. Sci. 16 (10), 504–510.

Pelphrey, K.A., Shultz, S., Hudac, C.M., Vander Wyk, B.C., 2011. Research review: constraining heterogeneity: the social brain and its development in autism spectrum disorder. J. Child Psychol. Psychiatry 52 (6), 631–644.

Perkins, T., Stokes, M., McGillivray, J., Bittar, R., 2010. Mirror neuron dysfunction in autism spectrum disorders. J. Clin. Neurosci. 17 (10), 1239–1243.

Pfeiffer, U.J., Timmermans, B., Bente, G., Vogeley, K., Schilbach, L., 2011. A non-verbal turing test: differentiating mind from machine in gaze-based social interaction. PLoS One 6 (11), e27591.

Pinto, J., Shiffrar, M., 2009. The visual perception of human and animal motion in point-light displays. Soc. Neurosci. 4 (4), 332–346.

Ramachandran, V.S., Oberman, L.M., 2006. Broken mirrors: a theory of autism. Sci. Am. 295, 62–69.

Rizzolatti, G., Craighero, L., 2004. The mirror-neuron system. Annu. Rev. Neurosci. 27, 169–192.

Robertson, C.E., Ratai, E.M., Kanwisher, N., 2016. Reduced GABAergic action in the autistic brain. Curr. Biol. 26 (1), 80–85.

Schilbach, L., Wohlschlaeger, A.M., Kraemer, N.C., Newen, A., Shah, N.J., Fink, G.R., Vogeley, K., 2006. Being with virtual others: neural correlates of social interaction. Neuropsychologia 44 (5), 718–730.

Schilbach, L., Eickhoff, S.B., Mojzisch, A., Vogeley, K., 2008. What's in a smile? Neural correlates of facial embodiment during social interaction. Soc. Neurosci. 3 (1), 37–50.

Schilbach, L., Wilms, M., Eickhoff, S.B., Romanzetti, S., Tepest, R., Bente, G., Vogeley, K., 2010. Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. J. Cogn. Neurosci. 22 (12), 2702–2715.

Schilbach, L., Eickhoff, S.B., Cieslik, E.C., Kuzmanovic, B., Vogeley, K., 2011. Shall we do this together? Social gaze influences action control in a comparison group, but not in individuals with high-functioning autism. Autism, 1362361311409258.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., Vogeley, K., 2013. Toward a second-person neuroscience. Behav. Brain Sci. 36 (04), 393–414.

Schilbach, L., 2015. The neural correlates of social cognition and social interaction.

In: Brain Mapping: An Encyclopedic Reference. Elsevier, pp. 159–164.

Schilbach, L., 2016a. Towards a second-person neuropsychiatry. Phil. Trans. R. Soc. B 371 (1686), 20150081.

Schilbach, L., 2016b. The Neuroscience of Mimicry During Social Interactions.

Emotional Mimicry in Social Context. Cambridge University Press.

Senju, A., Southgate, V., White, S., Frith, U., 2009. Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. Science 325 (5942), 883–885.

Sepulcre, J., Sabuncu, M.R., Gon˜i, J., 2015. Hubs and pathways. In: Arthur Toga, W. (Ed.), Brain Mapping, vol. 2015. Academic Press, Waltham, pp. 441–447, http://dx.doi.org/10.1016/B978-0-12-397025-1.00023-3, ISBN 9780123973160.

Sevgi, M., Diaconescu, A.O., Tittgemeyer, M., Schilbach, L., 2016. Social bayes: using bayesian modeling to study autistic trait?related differences in social cognition. Biol. Psychiatry 80 (2), 112–119.

Southgate, V., Hamilton, A.F., 2008. Unbroken mirrors: challenging a theory of Autism. Trends Cogn. Sci.

Spikins, P., Wright, B., Hodgson, D., 2016. Are there alternative adaptive strategies to human pro-sociality? The role of collaborative morality in the emergence of personality variation and autistic traits. Time Mind 9 (4), 289–313.

Stephan, K.E., Mathys, C., 2014. Computational approaches to psychiatry. Curr. Opin. Neurobiol. 25, 85–92.

Van Boxtel, J.J., Lu, H., 2013. A predictive coding perspective on autism spectrum disorders. Front. Psychol. 4, 19.

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., Wagemans, J., 2014. Precise minds in uncertain worlds: predictive coding in autism. Psychol. Rev. 121 (4), 649.

Vittorias, J., Petrantonakis, P., Bolis, D., Tsiligkyri, A., Kosmidou, V., Hadjileontiadis, L.J., 2008. Noesis: an enhanced educational environment for kids with autism spectrum disorders. In: Eighth IEEE International Conference on Advanced Learning Technologies, 2008. ICALT’08, IEEE, pp. 1019–1020.

Vygotsky L.S., (1930-1935/1978). Mind in Society: The Development of Higher Psychological Processes, Harvard University Press; Revised ed. edition (March 7, 1978).

Vygotsky L.S., (1934/2008). Thought and Language, translated by Maria Rodim Angelina Rodi, Athens : Gnosi (in Greek).

Wang, Y., Hamilton, A.F., 2012. Social Top-down Response Modulation (STORM): A model of the control of mimicry in social interaction. Front. Hum. Neurosci.

Wing, L., Gould, J., 1979. Severe impairments of social interaction and associated abnormalities in children: epidemiology and classification. J. Autism Dev. Disord. 9 (1), 11–29.

Yoon, J., Johnson, S.C., 2009. Biological motion displays elicit social behavior in 12-month-olds. Child Dev. 80 (4), 1069–1075.

von der Lühe, T., Manera, V., Barisic, I., Becchio, C., Vogeley, K., Schilbach, L., 2016. Interpersonal predictive coding, not action perception, is impaired in autism. Phil. Trans. R. Soc. B 371 (1693).

A.2.3. Beyond one Bayesian brain: Modeling intra- and interpersonal processes during social interaction

Original publication details (submitted 2016; accepted 2017)
Bolis, D., & Schilbach, L. (2017). Beyond one Bayesian brain: Modeling intra-and inter-personal processes during social interaction: Commentary on "Mentalizing homeostasis: The social origins of interoceptive inference" by Fotopoulou & Tsakiris. *Neuropsychoanalysis*, *19*(1), 35-38.

Beyond one Bayesian brain: Modeling intra- and interpersonal processes during social interaction

Commentary on "Mentalizing Homeostasis: The Social Origins of Interoceptive Inference" by Fotopoulou & Tsakiris

Dimitris Bolis[a,b] & Leonhard Schilbach[a-d]

Affiliations

[a] Independent Max Planck Research Group for Social Neuroscience, Max Planck Institute of Psychiatry, Munich, Germany
[b] International Max Planck Research School for Translational Psychiatry (IMPRS-TP), Munich, Germany
[c] Department of Psychiatry, Ludwig Maximilian University, Munich, Germany
[d] Graduate School of Systemic Neuroscience (GSN), Munich, Germany

Abstract

In their thought-provoking and integrative target article Fotopoulou and Tsakiris cut across different bodies of literature to argue for a second-person account of interoception and social cognition. More specifically, they argue for the constitutive role of embodied social interactions in the development of interoceptive abilities and the ability for self-other differentiation. Furthermore, they review evidence to suggest that social interactions have a specific role in binding together subjective feeling states with the

perception of the body and the (social) world. Along this line of reasoning, they refer to predictive coding and active inference frameworks of the ‘Bayesian brain’ function to suggest that basic inferential processes of embodied perception and action may be turned into more advanced forms of social understanding. We applaud Fotopoulou and Tsakiris for advocating a second-person account of social cognition and for connecting the predictive coding account of interoception to the dynamics of social interaction conceptually. We discuss these contributions in light of the existing literature and encourage the authors to be more precise about the computational processes, which they suggest may connect social interaction and interoception at the individual level. Furthermore, we describe a multilevel Bayesian framework that could be used to formally test a proposal, such as the one discussed by the authors, but also allows for going beyond one Bayesian brain, by modeling interpersonal processes during social interaction.

In developing their view of the social origins of interoceptive inference, Fotopoulou and Tsakiris use as their starting point the challenging question of whether "mental life is initially shaped by the embodied dimensions of the individual or the interpersonal relations" that surround us. In their interdisciplinary treatise of this question, they not only argue that some of the core aspects of self, namely subjective feeling states, are shaped by embodied interactions with other people in infancy, but that it is these specifically social embodied interactions that contribute to the generation of mental models of the infant's physiological states. This generation of mental models, described as the process of "mentalization" by Fotopoulou and Tsakiris, is linked to the predictive coding framework, which views the human brain as a ‘prediction machine’ (also more broadly referred to as the ‘Bayesian brain’), to support the idea that all predictive brain processes are social in nature. Furthermore, it is proposed that interactions with others are also governed by the same principles of predictive coding and that the processing of one's own body can, thus, include signals from the body of the interaction partner.

In this commentary, we would like to start out by supporting the notion that a theory of self-awareness needs to address the social domain. In our own work, we have suggested that a tangled hierarchy of predictive loops exists, so that the brain can continuously and unconsciously learn to anticipate the consequences of action or activity on itself, on the world, and on other people (Timmermans et al., 2012): The inner loop involves the brain re-describing its own representations to itself. The second loop is the perception-action loop whereby the agent predicts the consequences of its actions on the world. The third loop is the self-other loop and links the agent with other agents, using the exact same prediction-based mechanisms as involved in the other two loops. We have argued that the existence of the third loop is constitutive of conscious experience, for it is in the attempt of modelling other minds that I develop an understanding of myself. Crucially, and in line with the proposal by Fotopoulou and Tsakiris, understanding ourselves depends on the ability to anticipate the consequences of our actions on others. Here, Fotopoulou and Tsakiris introduce the important observation that it might be due to homeostatic necessity in the absence of a fully functioning motor repertoire that infants primarily rely on such embodied encounters. Consequently, predictive processes not only apply to the inner loop or bodily activity, but also the activity of interaction partners, particularly in infancy, may be plausibly experienced as one's own.

Fotopoulou and Tsakiris then go on to discuss a modern notion of interoception, according to which interoception informs the body about “how well the body is doing” in light of certain needs. Exteroception on the other hand, in their view, informs the body about external changes in relation to such needs. In addition to this, it is highlighted that interoception is thought to be uniquely related to the generation of subjective feelings, which may arise from predictive inferences on the causes of interoceptive signals (e.g., Seth et al. 2012). Interoception, in turn,

has recently been related to bodily self-awareness, as indexed by bodily illusions, thereby creating a link to self-other differentiation.

Furthermore, Fotopoulou and Tsakiris review the interesting case of social affective touch to suggest that it involves both interoceptive and exteroceptive processing, which could make it a developmentally important phenomenon for the establishment of "physical boundaries of the psychological self". In other words, the multisensory input during embodied social interactions is thought to facilitate perceptual inferences needed for body ownership, but also helps the development of an awareness of internal needs.

While Fotopoulou and Tsakiris provide an excellent introduction to the free energy principle, the predictive coding framework and its relationship to action, the link between the above described conceptual propositions and empirical research remains relatively loose at times. The authors spell out an intriguing hypothesis, namely that specific social behaviors (i.e., social touch) may be particularly effective in "binding together subjective feeling states and external perceptions of the body and the world", thereby promoting the development of self-other differentiation. While they do argue for the need of a mechanistic explanation thereof and clearly seem to suggest that the predictive coding and active inference accounts could provide one, computational solutions and competing models are not spelled out explicitly. How for instance are interoceptive and exteroceptive inference processes thought to be computationally integrated within the abovementioned schemes? What models exist to explain the development of self-other differentiation and how could those be formally tested? Could the phenomena described in the article be reduced to intrapersonal Bayesian mechanisms? To address these questions, we propose an integration of a recently spelled out proposal for a Bayesian account of

intersubjectivity (Bolis & Schilbach, 2017; Bolis *et al*., under review) with processes of interoceptive inference (Seth *et al*., 2012; Seth, 2013; Quattrocki & Friston, 2014):

In the former, connecting cultural-historical theories, which view culture and social interaction as major developmental driving forces (e.g., Vygotsky, 1934/2008; 1930-1935/1978), Bayesian accounts of brain function, (e.g., Friston, 2010; Daunizeau et al., 2010; Mathys et al., 2012), enactivism (e.g., De Jaegher & Di Paolo, 2007), and advances in social neuroscience (e.g., Konvalinka and Roepstorff, 2012; Schilbach et al., 2013), we have argued for a dialectical approach to neuroscientific and psychiatric research. Most importantly, such an approach aims at going beyond the individual as the unit of analysis, emphasizing the inextricable, dynamic, and reciprocal interrelation between individual and collective processes. Here, we are describing an extended Bayesian framework, which aims at modeling the interrelations of interoceptive, exteroceptive (intrapersonal level), and collective (interpersonal level) processes (**Fig. A.3**). More concretely, we consider three levels of description:

1. Processes and mechanisms within an individual (person A or person B). Here we consider two broad categories, namely processes of interoceptive and exteroceptive inference.
2. Collective behavior, i.e., processes that emerge between individuals, such as interpersonal coupling during social interaction.
3. The environment, which constitutes the source of perceptual stimuli, the field of action for the interacting person A and person B.

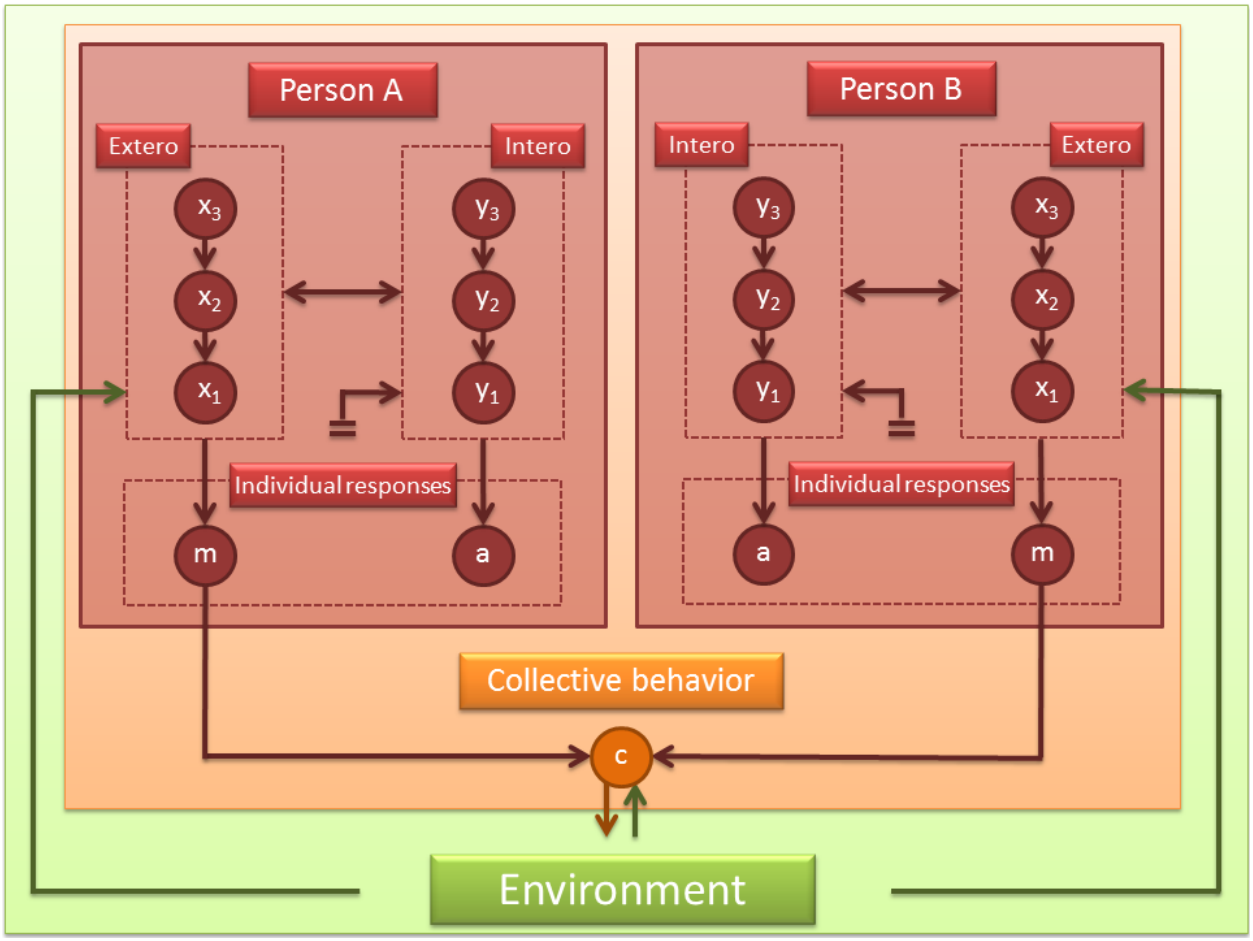

Fig. A.3. Observing-the-interactors: A Bayesian framework for studying intrapersonal and interpersonal processes during social interaction. Two generative models within each individual, namely one interoceptive and one exteroceptive, track states of dynamic bodily and environmental phenomena, respectively (upper part of red panels). Autonomic and motor control, are modeled as probabilistic translations of the aforementioned predictions on an intraindividual level (lower part of red panels). Collective behavior can be considered as a non-linear fusion of individual motor control processes (of person A and person B) via interpersonal coupling parameters (orange panel). The environment serves as a provider of perceptual stimuli, a field for action, as well as exerts modulatory effects on social interactions (green panel).

Due to space limitations, here we briefly unpack the major components of the framework and establish its relevance to existing models. Intrapersonal mechanisms within each person are

modeled with two hierarchical models, which implement exteroceptive and interoceptive inferences, via the multilevel states xi and yi respectively (upper part of red panels in **Fig. A.3**). An example of exteroceptive inference could be the prediction of next day's lowest temperature, while interoceptive inference could refer to the estimation of one's own cardiac pace. Higher levels in the hierarchy capture the (hidden) probabilistic structure of the time varying phenomenon (estimating not merely its state, but also its volatility, the volatility of the volatility, and so on; for conceptual and technical details see Friston, 2010 and Mathys et al., 2012). Importantly, these models capture not only the multilevel predictions of an individual about the states of a phenomenon, but also the precision (i.e., confidence) they place on them.

Additionally, adapting the "observing-the-observer" scheme introduced by Daunizeau et al. (2010) and models of interoceptive inference (Seth et al., 2012; Seth 2013), we model motor and autonomic control processes via the states m and a respectively, as probabilistic outcomes of an individual's exteroceptive and interoceptive states (lower part of red panels in **Fig. A.3**). For instance, motor control is related to processes of motor planning, monitoring and execution, implemented by supplementary motor and (pre-)motor cortices. Autonomic control could be linked to predictions generated within a salience network, which covers brain regions such as the anterior insular and anterior cingulate cortices, relevant to autonomic reflexes (see Seth, 2013). Notably, such generative models are not assumed to function in isolation, but as functionally coupled. Here, the most plausible coupling hypothesis could be formally identified via Bayesian model comparison, e.g., between implementations of parallel and hierarchical coupling. Furthermore, the interrelation between perceptual and motor/autonomic control states could be investigated via modeling reciprocal connections (between upper and lower parts of red panels in **Fig. A.3**). Crucially, in line with our proposal for a Bayesian account of intersubjectivity (Bolis

& Schilbach, 2017; Bolis *et al.*, under review), we model observable behavior in social interaction as a collective phenomenon via an interpersonal state c (orange panel in **Fig. A.3**). The latter in its full form will constitute a non-linear and probabilistic combination of individual motor states via interpersonal parameters. It is noteworthy that such an interpersonal coupling does not need to be taken for granted in advance, but its existence will be subjected to Bayesian model comparison against simpler models, which assume that individual processes alone are capable in explaining the data most optimally.

Finally, the environment serves not only as a perceptual stimulator of the individuals, but as a modulator of their interaction as well (green arrows in **Fig. A.3**). Consequently, such a framework will allow for devising a number of neurobiologically and socially plausible hypotheses of multimodal and multilevel (involving both interoceptive and exteroceptive, as well as interpersonal) processes, such as those in the case of social affective touch described by Fotopoulou and Tsakiris. The subsequent model comparison will therefore yield the most optimal mechanistic explanation of phenomena, such as self-other differentiation.

Taken together, Fotopoulou and Tsakiris provide an exciting and, in its scope, unrivaled interdisciplinary article that sheds new light on the topic of interoception and social interaction. In particular, they argue that an understanding of self-awareness cannot be separated from the dynamics of social interaction. Furthermore, they make a case for the use of computational modeling to mechanistically explain why and how social interactions are so very special.

References

Bolis D, Schilbach L (2017) Observing and participating in social interactions: Action perception and action control across the autistic spectrum. Developmental Cognitive Neuroscience.

Bolis D, Balsters J, Wenderoth N, Becchio C, Schilbach L (under review). Beyond autism: Introducing the dialectical misattunement hypothesis and a Bayesian account of intersubjectivity.

Daunizeau J, den Ouden HEM, Pessiglione M, Kiebel SJ, Stephan KE, Friston KJ (2010). Observing the observer (I): meta- Bayesian models of learning and decision-making. PLoS ONE 5, e15554.

De Jaegher H, Di Paolo E (2007). Participatory sense-making. Phenomenology and the cognitive sciences, 6(4), 485-507.

Friston K (2010). The free-energy principle: a unified brain theory? Nature Reviews Neuroscience 11, 127-138.

Konvalinka I, Roepstorff A (2012). The two-brain approach: how can mutually interacting brains teach us something about social interaction? Frontiers in human neuroscience, 6 (2012), p. 215

Mathys C, Daunizeau J, Friston KJ, Stephan KE (2011). A Bayesian foundation for individual learning under uncertainty. Frontiers in human neuroscience, 5, 39.

Schilbach L, Timmermans B, Reddy V, Costall A, Bente G, Schlicht T, Vogeley K (2013). Toward a second-person neuroscience. Behavioral and Brain Sciences, 36(04), 393-414.

Seth AK, Suzuki K, Critchley HD (2012). An interoceptive predictive coding model of conscious presence. Frontiers in psychology, 2, 395.

Seth AK (2013). Interoceptive inference, emotion, and the embodied self. Trends in cognitive sciences, 17(11), 565-573.

Quattrocki E, Friston K (2014). Autism, oxytocin and interoception. Neuroscience & Biobehavioral Reviews, 47, 410-430.

Timmermans B, Schilbach L, Pasquali A, Cleeremans A (2012). Higher order thoughts in action: consciousness as an unconscious re-description process. Phil. Trans. R. Soc. B, 367(1594), 1412-1423.

Vygotsky LS (1934/2008). Thought and Language, translated by Maria Rodim, Angelina Rodi. Athens: Gnosi (in Greek).

Vygotsky LS (1930-1935/1978). Mind in Society: The Development of Higher Psychological Processes. Harvard University Press; Revised ed. edition (March 7, 1978).

A.2.4. 'Through others we become ourselves': The dialectics of predictive coding and active inference

Original publication details (BBS commentary, submitted 2019; accepted 2020)

Bolis, D., Schilbach, L. (2020). 'Through others we become ourselves': The dialectics of predictive coding and active inference. BBS commentary on "Thinking Through Other Minds: A Variational Approach to Cognition and Culture".

'Through others we become ourselves': The dialectics of predictive coding and active inference

Commentary on "Thinking Through Other Minds: A Variational Approach to Cognition and Culture" by Veissière et al.

Dimitris Bolis[1,2,3] & Leonhard Schilbach[1,2,4]

Affiliations

[1] Independent Max Planck Research Group for Social Neuroscience, Max Planck Institute of Psychiatry, Kraepelinstrasse 2-10, 80804 Munich-Schwabing, Germany
[2] International Max Planck Research School for Translational Psychiatry (IMPRS-TP), Munich, Germany
[3] Munich Medical Research School (MMRS), Dekanat der Medizinischen Fakultät, Ludwig-Maximilians-Universität München, Bavariaring 19, 80336 Munich, Germany
[4] LVR Klinikum Düsseldorf / Kliniken der Heinrich-Heine-Universität Düsseldorf, Bergische Landstr. 2, 40629 Düsseldorf, Germany

Abstract

Thinking Through Other Minds (TTOM) creatively situates the free energy principle within real-life cultural processes, thereby enriching both sociocultural theories and Bayesian accounts of cognition. Here, shifting the attention from thinking to becoming, we suggest complementing such an account by

focusing on the empirical, computational and conceptual investigation of the multiscale dynamics of social interaction.

We applaud Veissière and colleagues for pursuing the ambitious goal of situating the free energy principle within the context of sociocultural processes (cf. TTOM; 2019). This is, indeed, a much needed undertaking, which has only recently started developing, holding promise for advancing not only relevant sociocultural research fields, but also computational psychiatry (cf. Friston and Frith, 2015; Bolis & Schilbach, 2017; 2018b; Gallagher and Allen, 2018; Constant *et al*., 2019). In fact, human cognition and culture have often been studied in isolation. For instance, the field of computational psychiatry has been developing rigorous experimental protocols and mathematical toolboxes to mechanistically explain human cognition and action. Yet, until recently a rather individualistic perspective has been adopted, which neglects levels of description beyond the individual (cf. De Jaegher & Di Paolo, 2007; Schilbach *et al*., 2013; Kirmayer & Crafa, 2014; Bolis *et al*., 2017). On the other hand, sociocultural fields, such as cultural anthropology, have rightfully adopted a more holistic perspective to complex phenomena of life, yet frequently lacking formal descriptions of cognitive and biological mechanisms (cf. Seligman & Brown, 2009).

An artificial dichotomy between the individual and the collective has inevitably led to a 'chicken-egg' paradox (cf. Dumas et al., 2014). However, such causality dilemmas dissolve once one considers the dialectical nature of human-becoming, which is multiscale, reciprocal, dynamic, cumulative and inherently contradictory (cf. Vygotsky 1930–1935/1978; Dumas et al., 2014; Bolis & Schilbach, 2018b; Di Paolo et al., 2018). Processes from evolution and culture to individual development, learning and sensorimotor activity, can all be viewed as mutually

interacting adjustments between the species and the environment. Here, reciprocity is deep, as “it is not only humans who change the environment, but the environment in turn changes them in face of their impact on it” (Levins & Lewontin, 1985; Bolis & Schilbach, 2018b). TTOM, therefore, constitutes an important development because it addresses how human agents learn shared expectations and how they construct their own social niches in complex interaction between the individual and the environment.

We concretely appreciate the consideration of predictive coding and active inference within a framework of circular causality. Indeed, an organism can be viewed as embedded within the dialectic between the two above-mentioned processes, which in order to survive obeys a simple, but fundamental rule: “adjust yourself to reality or change the reality itself” (Friston, 2010; Bolis & Schilbach, 2018b). When it comes to TTOM, it is not only the agent which learns environmental regularities and adjusts accordingly, but the environment in turn ‘learns’ the agents’ ‘beliefs’ through repeated and culturally regulated actions. TTOM resonates well with the dialectical attunement hypothesis (Bolis & Schilbach, 2018b), which views human-becoming as the interplay between internalization and externalization primarily within and due to culturally mediated social interaction, internalization being the “co-construction of bodily hierarchical models of the (social) world and the organism” [cf. predictive coding], while externalization the “collective transformation of the world” [cf. active inference]. In a nutshell, “interpersonal statistical regularities shape multiscale hierarchical models on an individual level and vice versa”.

To offer a formal description of how environment ‘learns’, the authors interestingly suggest twisting the modeling equations by inverting relevant quantities across actions and sensations. This offers various potential modeling scenarios about the degree of interactivity within the

system of brain-body-environment-body-brain (cf. Froese et al., 2013). Here, a multiscale meta-Bayesian scheme might nicely lend itself for modeling not only individual processes, but also collective and environmental interactions (Bolis & Schilbach, 2017; Brandi *et al*., 2019; Ramstead *et al*., 2018).

Not only are we in line with the authors on conceptual and computational grounds, but also concerning the need for empirical studies. To make this more concrete, we describe certain experimental directions: Systematically varying social structure, cultural and socioeconomic background, affective bonds and interpersonal similarity across interacting individuals will enable the mechanistic study of interpersonal attunement. With regard to psychiatric disorders, construed as disorders of social interaction (Schilbach, 2016), two-person (or indeed collective) psychophysiology allows to move beyond the individual (cf. Bolis & Schilbach, 2018a). Taking autism as a paradigm example, the dialectical misattunement hypothesis has put forward a research line, which, moving away from an exclusive study of individual differences, considers types of interacting groups: i.e., autistic, neurotypical and mixed groups, expecting smoother interactions within the more homogenous groups or dyads (Bolis *et al*., 2017). Taken together, such experiments will not only inform TTOM within the 'neurotypical social world', but also open up avenues for evaluating and updating the ontological status of conditions, such as autism, as relational and interactional (cf. double empathy problem; Milton, 2012).

Apart from praising TTOM, we would also like to point out a fundamental aspect which, in our opinion, would benefit from further elaboration. We feel that the potentially constitutive role of real-time social interaction in sense-making and human-becoming was not sufficiently taken into account within the model (cf. De Jaegher and Di Paolo, 2007; Vygotsky 1930–1935/1978). It has been suggested that thinking about and with others might be fundamentally different in real-time

interactive scenarios, as compared to passive observational situations (cf. second-person perspective; Schilbach et al., 2013; Redcay and Schilbach, 2019). Crucially, such interactive interpersonal processes have been thought of as dialectically preceding the individual both in evolutionary and developmental regards (cf. Bolis & Schilbach, 2018b; Tomasello, 2019). As Vygotsky proclaimed almost a century ago, "through others we become ourselves" (1931/1987). Yet, to do justice to the authors, the field today has not yet reached a conclusive consensus. For instance, while Di Paolo and colleagues (2018) suggest that "interactive situations present a richer, more complex set of possibilities" and "the key to our sociality is not in our heads or in our genes", Schönherr and Westra (2017) claim to have (conceptually) shown that "ersatz interactivity works just as well as the real thing", by 'real thing' denoting genuine, real-time social interaction. We, therefore, conclude our commentary with a question still desperately begging for a definite empirical answer. Does (real-time) social interaction matter… or is it all in our heads?

References

Bolis, D., Balsters, J., Wenderoth, N., Becchio, C., & Schilbach, L. (2017). Beyond autism: introducing the dialectical misattunement hypothesis and a bayesian account of intersubjectivity. Psychopathology, 50(6), 355-372.

Bolis, D., & Schilbach, L. (2017). Beyond one Bayesian brain: Modeling intra-and inter-personal processes during social interaction: Commentary on "Mentalizing homeostasis: The social origins of interoceptive inference" by Fotopoulou & Tsakiris. Neuropsychoanalysis, 19(1), 35-38.

Bolis, D., & Schilbach, L. (2018a). Observing and participating in social interactions: action perception and action control across the autistic spectrum. Developmental cognitive neuroscience, 29, 168-175.

Bolis, D., & Schilbach, L. (2018b). 'I interact therefore I am': The self as a historical product of dialectical attunement. Topoi, 1-14.

Brandi, M. L., Kaifel, D., Bolis, D., & Schilbach, L. (2019). The Interactive Self–A Review on Simulating Social Interactions to Understand the Mechanisms of Social Agency. i-com, 18(1), 17-31.

Constant, A., Ramstead, M. J., Veissière, S. P., & Friston, K. (2019). Regimes of Expectations: An Active Inference Model of Social Conformity and Decision Making. Frontiers in psychology, 10, 679.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making. Phenomenology and the cognitive sciences, 6(4), 485-507.

Di Paolo, E. A., Cuffari, E. C., & De Jaegher, H. (2018). Linguistic bodies: The continuity between life and language. MIT Press.

Dumas, G., Kelso, J. A., & Nadel, J. (2014). Tackling the social cognition paradox through multi-scale approaches. Frontiers in psychology, 5, 882.

Friston, K. (2010). The free-energy principle: a unified brain theory?. Nature reviews neuroscience, 11(2), 127.

Friston, K., & Frith, C. D. (2015). Active inference, communication and hermeneutics. Cortex, 68, 129–143.

Froese, T., Iizuka, H., & Ikegami, T. (2013). From synthetic modeling of social interaction to dynamic theories of brain-body-environment-body-brain systems. Behav. Brain Sci, 36(4), 420-421.

Gallagher, S., & Allen, M. (2018). Active inference, enactivism and the hermeneutics of social cognition. Synthese, 195(6), 2627-2648.

Kirmayer, L. J., & Crafa, D. (2014). What kind of science for psychiatry?. Frontiers in Human Neuroscience, 8, 435.

Levins, R., Lewontin, R. C. (1985). The dialectical biologist. Harvard University Press, Cambridge

Milton, D. E. (2012). On the ontological status of autism: the 'double empathy problem'. Disability & Society, 27(6), 883-887.

Ramstead, M. J. D., Badcock, P. B., & Friston, K. J. (2018). Answering Schrödinger's question: A free-energy formulation. Physics of life reviews, 24, 1-16.

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. Nature Reviews Neuroscience, 1.

Schilbach, L. (2016). Towards a second-person neuropsychiatry. Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1686), 20150081.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience 1. Behavioral and brain sciences, 36(4), 393-414.

Schönherr, J., & Westra, E. (2017). Beyond 'Interaction': How to Understand Social Effects on Social Cognition. The British Journal for the Philosophy of Science, 70(1), 27-52.

Seligman, R., & Brown, R. A. (2009). Theory and method at the intersection of anthropology and cultural neuroscience. Social cognitive and affective neuroscience, 5(2-3), 130-137.

Tomasello, M. (2019). Becoming human: A theory of ontogeny. Belknap Press.

Veissière, S. P., Constant, A., Ramstead, M. J., Friston, K. J., & Kirmayer, L. J. (2019). Thinking Through Other Minds: A Variational Approach to Cognition and Culture. Behavioral and Brain Sciences, 1-97.

Vygotsky, L. S. (1987). The genesis of higher mental functions. In R. Reiber (Ed.), The history of the development of higher mental functions (Vol. 4, pp. 97-120). New York: Plennum. Original work 1931.

Vygotsky, L. S. (1978). Mind in society: the development of higher psychological processes. Harvard University Press. Original work 1930–1935. Translated in Greek, (1997) by A. Bibou and S Vosniadou, Gutenberg.

A.3. Questionnaires

A.3.1. Collaborative intentions question (Study #1)

I played collaboratively (i.e., I tried to help my co-player)

1 2 3 4 5 6

(1= I fully disagree, 6= I fully agree)

A.3.2. Adapted Friendship Quality Scale (Study #2)

Did you know your Co-player before?

Yes No

If Yes

What kind of relationship do you have?

How long do you know each other?

How often do you see each other?

How do the following statements display the relationship between you and your Co-Player?

(1= high strongly disagree, 6 = high strongly agree)

I believe all the information given by him/her.

1 2 3 4 5 6

He/ She never breaks a promise.

1 2 3 4 5 6

I am confident that he/ she will not leak my secrets.

1 2 3 4 5 6

He/ She never lies to me.

1 2 3 4 5 6

I always listen to his/ her advice.

1 2 3 4 5 6

I feel safe when the precious belongings are kept by him/ her.

1 2 3 4 5 6

I inform other friends immediately if he or she encounters problems.

1 2 3 4 5 6

I feel safe when accompanied by him/ her.

1 2 3 4 5 6

I always joke with him/ her.

1 2 3 4 5 6

I understand his/ her mood.

1 2 3 4 5 6

I always chat with him/ her even if we are not together.

1 2 3 4 5 6

We always share our life experiences.

1 2 3 4 5 6

I understand the background of him/ her.

1 2 3 4 5 6

I would not feel shy when performing something humorous in front of him/ her.

1 2 3 4 5 6

He/ She forgives me easily.

1 2 3 4 5 6

We can overcome differences in our opinion immediately.

1 2 3 4 5 6

He/ She treats me well.

1 2 3 4 5 6

My relationship with him/ her is like brothers and sisters.

1 2 3 4 5 6

He/ She corrects my mistakes in a work, when I ask her to.

1 2 3 4 5 6

He/ She always helps me when I have problems in completing some work.

1 2 3 4 5 6

He/ She helps me to solve problems.

1 2 3 4 5 6

Full list of publications

1. **Bolis, D.**, & Schilbach, L. (in press). 'Through others we become ourselves': The dialectics of predictive coding and active inference. BBS commentary on Veissière, S. P., Constant, A., Ramstead, M. J., Friston, K. J., & Kirmayer, L. J. (2019). Thinking Through Other Minds: A Variational Approach to Cognition and Culture. Behavioral and Brain Sciences, 1-97.
2. Brandi, M. L., Kaifel, D., **Bolis, D.**, & Schilbach, L. (2019). The Interactive Self–A Review on Simulating Social Interactions to Understand the Mechanisms of Social Agency. *i-com*, *18*(1), 17-
3. **Bolis, D.**, & Schilbach, L. (2018). 'I interact therefore I am': The self as a historical product of dialectical attunement. *Topoi*, 1-14.
4. **Bolis, D.**, & Schilbach, L. (2018). Observing and participating in social interactions: action perception and action control across the autistic spectrum. *Developmental cognitive neuroscience*, *29*, 168-175.
5. **Bolis, D.**, & Schilbach, L. (2017). Beyond one Bayesian brain: Modeling intra-and inter-personal processes during social interaction: Commentary on "Mentalizing homeostasis: The social origins of interoceptive inference" by Fotopoulou & Tsakiris. *Neuropsychoanalysis*, *19*(1), 35-38.
6. **Bolis, D.**, Balsters, J., Wenderoth, N., Becchio, C., & Schilbach, L. (2017). Beyond autism: introducing the dialectical misattunement hypothesis and a bayesian account of intersubjectivity. *Psychopathology*, *50*(6), 355-372.
7. Balsters, J. H., Apps, M. A., **Bolis, D.**, Lehner, R., Gallagher, L., & Wenderoth, N. (2016). Disrupted prediction errors index social deficits in autism spectrum disorder. *Brain*, *140*(1), 235-246.
8. **Bolis, D.**, Becchio, C., & Schilbach, L. (2016). Revisiting psychological definitions at the interface of sociocultural historical theories and predictive coding. *WORLDING THE BRAIN: Patterns, Rhythms, Narratives in Neuroscience and the Humanities*. p, 10.

9. **Bolis, D.**, Heinzle, J., Mathys, C., Stephan, K. E. (2015). Inferring cognitive traits of individual subjects through gaze controlled video games. In: Bernardis P, Fantoni C, Gerbino W (eds.) "TSPC2015*: proceedings of the Trieste Symposium on Perception and Cognition*, November 13rd 2015", Trieste, EUT Edizioni Università di Trieste, 2015, p 19
10. **Bolis, D.**, Jakab, A., Göksel, O., & Székely, G. (2012). On Exploiting Connectomics for Thalamic Nuclei Localization: A Supervised Learning Approach. In *Int Conf on Machine Learning (ICML) Workshop on Statistics, Machine Learning and Neuroscience*.
11. Maronidis, A., **Bolis, D.**, Tefas, A., & Pitas, I. (2011). Improving subspace learning for facial expression recognition using person dependent and geometrically enriched training sets. *Neural Networks*, *24*(8), 814-823.
12. **Bolis, D.**, Maronidis, A., Tefas, A., & Pitas, I. (2010). Improving the robustness of subspace learning techniques for facial expression recognition. In *International Conference on Artificial Neural Networks* (pp. 470-479). Springer, Berlin, Heidelberg.
13. Vittorias, J., Petrantonakis, P., **Bolis, D.**, Tsiligkyri, A., Kosmidou, V., & Hadjileontiadis, L. J. (2008, July). Noesis: An enhanced educational environment for kids with autism spectrum disorders. In *2008 Eighth IEEE International Conference on Advanced Learning Technologies* (pp. 1019-1020). IEEE.

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Dean's Office
Medical Faculty

MMRS MUNICH MEDICAL RESEARCH SCHOOL

Affidavit

Bolis, Dimitrios

Surname, first name

Kraepelinstr. 2-10

Street

80804 München

Zip code, town

Germany

Country

I hereby declare, that the submitted thesis entitled

'I interact therefore I am': Human becoming in and through social interaction

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the submitted thesis or parts thereof have not been presented as part of an examination degree to any other university.

Munich 31.03.2020

Place, date

Dimitrios Bolis

Signature doctoral candidate

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Dean's Office
Medical Faculty

MMRS MUNICH MEDICAL RESEARCH SCHOOL

Confirmation of congruency between printed and electronic version of the doctoral thesis

Bolis, Dimitrios

Surname, first name

Kraepelinstr. 2-10

Street

80804 München

Zip code, town

Germany

Country

I hereby declare that the electronic version of the submitted thesis, entitled

'I interact therefore I am': Human becoming in and through social interaction

is congruent with the printed version both in content and format.

Munich 31.03.2020

Place, date

Dimitrios Bolis

Signature doctoral candidate