RIFL/SFL(2015): 173-187
DOI: 10.4396/2015SFL14

A hypothesis on the deficit of lipreading during linguistic perception in subjects with autism: observations from robotic interactions

Paola Pennisi
Università degli Studi di Messina Dipartimento di Scienze Cognitive
ppennisi@unime.it

Abstract For many years various computational studies have attempted to bring deficits into inferential deficits in integrating top-down inputs. However, in this study we aim to show via the concept of interoceptive inference that in individuals with ASD a malfunction in the processes of synchronization with others may also explain part of their language deficits. To do that, we have shown that there is an inverse correlation between the amount of time spent by children with ASD observing the interlocutor's mouth and their social skills. This data, combined with the observation (now widely accepted by scientific literature) that individuals with ASD usually do not integrate information from lipreading with the linguistic message, has led us to hypothesize that an inability to interpret as biological motion the complex articulatory movements necessary for speech production, caused by a deficit in interoceptive inferential processes, could be the effect of a dispersion of attention that impedes perceiving the audio-visual synchrony during verbal perception.

Keywords: language, lipreading, autism, interoceptive inference

Received 31 March 2016; received in revised form 02 October 2016; accepted 07 October 2016.

§ 0. Introduction

In a systematic review on Autism Spectrum Disorders (ASD) and social robotics (PENNISI et al. 2015) we found that subjects with ASD show a positive inclination toward this new technology; we observed some very interesting phenomena and here we will explore some of these:[1]

a) children with ASD show social behaviors towards robots
b) in some cases a robot can be a better stimulus than a human agent for improving social behaviors in children with ASD
c) robots improve (rather like a mediator) social behaviors of subjects with ASD towards other subjects (with or without ASD);
d) a robot helps improve language better than a human agent or other object in an interaction with ASD children.

[1] The first conclusion is the most widely supported by experimental studies, the last has less scientific data to support it. For all details, see PENNISI 2015.

For simplicity's sake, we can summarize (a), (b) and (c) in the more generic observation that social robots can help elicit several manifestations of classic intersubjective behaviors in patients with ASD. Overall (a), (b) and (c) were tested in a total of 100 subjects,[2] (d) in 46 subjects;[3] given the small number of subjects on which experimental research has focused until now, these are certainly preliminary observations. However, literature supports a general positive tendency of individuals with ASD to gravitate towards technology (PUTNAM e CHONG 2008); we all remember with astonishment Temple Grandin's anecdote where she preferred the hugs of her hug machine to those of her own mother (GRANDIN 2005).
Observations on the propensity of individuals with ASD to interact with technological devices have rarely been the subject of analytic studies. In this paper we will attempt to test the following hypothesis: is it possible that one of the benefits accruing from social robots is the lack of motor articulations related to language? In other words, is it possible that in some cases linguistic articulation negatively affects interpersonal relationships in individuals with ASD?
To answer this question we will proceed as follows:

1. We will analyze studies on different perceptive biases of subjects with ASD in scanning others' faces.
2. We will look for correlations between the different perceptive biases in scanning others' faces and the development of language.
    a. We will analyse the ASD deficit in elaborating both auditory and visual stimuli at the same time.
    b. We will analyse their deficit in lipreading.
3. Finally, we will draw our conclusions.

The final scope of the paper is to converge our observations into a more general theory that shows that the acquisition of language (aside from being determined by exteroceptive and proprioceptive processes) is profoundly influenced by what we will call (adopting ONDOBANKA et al.'s [2015] terminology) interoceptive processes.

§ 1. The concept of "interoceptive inference"

The sensory world is very uncertain, but the perceptions that we have about it are generally certain from a psychological point of view. This *gap* induced some theorists of perception to suppose that the brain, when working, must use perceptive inferences. For example, according to the *Bayesian coding hypothesis*, when for instance the subject is in front of a table and is trying to perceive its depth, his brain calculates all possible depths of the table and all relative probabilities for each value calculated by the sensory data available. In a more general perception of a whole scene, the subject will unconsciously calculate all possibilities of perception for each salient object (i.e., in the case of a table he will calculate all possible depths, all possible colors, all possible lengths) and all relative probabilities for each value of each characteristic of the objects in the scene. All these data will create a predictive model of perception (KNILL e POUGET 2004).
In 2015, Ondobaka et al. highlighted that recent views that use the predictive models of perception to explain Theory of Mind (ToM) left the role of interoception unclear.

[2] We summed the number of individuals with ASD involved in all experimental studies examined.

[3] See previous note.

For this reason they speak about *interoceptive inference* and attribute to it an important role in ToM. According to these authors «interoception or interoceptive inference can be viewed as a generalisation of active inference to the processing of interoceptive signals carrying information about visceral states (e. g, heart rate, blood pressure, temperature)» (ONDOBAKA et AL. 2015: 2). An inference is *active* when the agent actively stimulates its sensors in order to generate the sensorial consequences that the brain expects relative to the situation (CLARK 2013). To better understand the concept of *interoceptive inference* it is necessary to consider the supposed tendency of the brain (by homeostasis and allostasis) to maintain internal states relatively constant over time (ONDOBANKA et AL. 2015). Thus, we can consider an *interoceptive inference* like a predictive model of visceral states previously oriented to the balance of internal bodily states, which biases perception toward satisfying an organism's biological needs.
According to recent studies on ToM, interoceptive inference should work like exteroceptive and proprioceptive inference: i.e., to understand others' intentions from their movements, we virtually simulate the same movement. The problem posed by Ondobaka et al. (2015) is that, if exteroceptive and proprioceptive inferences can clearly use sensory input that are unavailable in relation to interoception (we cannot know our pupil dilatation or internal temperature), what data are used during the interoceptive inference? The answer proposed is that «the emotional and intentional theory of mind has to be learned through interpersonal interactions, probably at an early stage of development, in which attachments are made» (*Ivi*: 4).
If the problem is well posed, the solution – in our opinion – shows the failure of computational models. If Ondobaka et al. integrate the fundamental element of visceral states in models of perception, they are unable to explain how the interoceptive inference should work in a computational model of mind. It is not the aim of this work to determine what is missing in their model; but here I am interested in underlining that the interoceptive inference – in contrast to the proprioceptive and esteroceptive – is not a computation. As admitted by Ondobaka et al. «there may be something special about how we are particularly adept at inferring the drives and affiliative imperatives that contextualize interoception» (*Ibidem*).

§ 2. Interoceptive inference and autism

Since interpersonal relationships are the most critical area in autistic disorders, we may suppose that a predictive model of autistic cognition will show a deficit in interoceptive inference. Results from social robotics partially confirm this supposition. In fact, interaction with social robots does not require a ToM, so does not involve the interoceptive inference. Consequently, we can suppose that (a), (b), (c) and (d) would be partially explained by the unsubstantiality of interoceptive inference during robotic interaction.
It is plausible to suppose that a deficit in interoceptive inference, due to its fundamental role during ontogenesis, will cause a deficit in esteroceptive and proprioceptive inferential abilities. In fact, without interoceptive inferential abilities, interpersonal relationships will become mere computations of perceptions, insensitive to the organism's biological needs. And this will change the gestaltic rules of perception, at least for the interactions with other living beings.
We present a case of this phenomenon: specifically that the delay in language development typical of subjects with ASD could be partly linked to their difficulties in interpreting the biological motions of the mouth. Herrington et al. (2012) showed

that attention to human goals activates biological motion areas. Thus, if in order to carry out ToM inferences one needs intact interoceptive inferential abilities and if the activation of cerebral areas involved in the interpretation of the biological motion is linked to ToM inferences, we can infer that in order to interpret a movement as biological motion one needs an intact ability to carry out interoceptive inferences. Our hypothesis is that linguistic deficits related to ASD are a consequence of these deficits in processes that mediate silent intersubjective communication.[4]

In light of this hypothesis, how do we explain the apparent improvement in linguistic behaviors in the presence of social robots?

None or very few interoceptive inferences are involved in interaction with social robots since we do not attribute a ToM to them. On the contrary, in order to be integrated in the scene, lipreading requires to be seen as biological motion. If the very fine movements of lips are not integrated in a meaningful context, they risk becoming merely a distractor; all of us have experienced, at least once in our life, that feeling of discomfort and distraction watching a movie where the voice was not well-integrated and synchronized with the video. In our opinion, this sensation is in part responsible for autistic inattention during verbal interaction. The absence of biological motion in interaction with social robots could be partly responsible for the positive results reported in (d).

In the next paragraph we will attempt to show that in ecological interactions, lipreading could be a distractor for subjects with ASD.

§ 3. Do autistic subjects look more at the mouth than at the eyes?

Infants under 2 months of age tend to look at the edges of another's face (MAURER e SALAPATEK 1976; HAITH et al. 1977); later they will prefer the internal features of the face (YARBUS 1967; HUNNIUS e GEUZE 2004). Differences in these habits may reflect differences in language acquisition (LEWKOWICZ e HANSEN-TIFT 2012).

Recent studies focused their attention on face scanning of autistic subjects. Pelphrey et al. (2002) showed that ASD subjects, compared to typically developed (TD) subjects, look for a longer time at non-salient areas of faces and for a shorter time at salient areas of faces. However, van der Geest et al. (2002) found results contrasting with these last studies and showed that there are no differences in gaze behavior of scanning faces between TD and ASD subjects when faces are presented in isolation and without sound.

In this paper we will analyze the question of time spent looking at the mouth rather than the eyes during dyadic interactions.

In 2002, a prestigious psychiatric magazine published an experiment that, despite the small number of participants involved, gave rise to a widespread debate on the relationship between perception styles and social skills. Klin et al. (2002) showed that ASD subjects (n = 15; m.a. = 15.4), rather than TD subjects (n = 15; m.a. = 17.9), looked twice as long at the mouth region; half as much at the region of the eyes; twice as much at the body region and twice as much at the object region, while watching a video (30-60 s) representing naturalistic social situations. Experimenters also found a positive correlation between time looking at the mouth and social

[4] For other related works see PENNISI, P. 2014; CARROZZA et al. (in press); PENNISI 2016.

competence measured by VABS-E[5] and ADOS social scores; and conversely, there was a positive correlation between time focusing on the object region and severity of autistic symptoms. Thus, authors inferred that:

- increased focus on mouths predicted improved social skills and less autistic social impairment;
- increased focus on objects predicted decreased social skills and more autistic social impairment.

The hope of finding a biomarker for autism in perceptual biases led several research groups to replicate the experiment. Dalton et al. (2005) carried out two different experiments in which experimenters presented photographs of other human beings to participants while they underwent MRI scanning and found that ASD subjects (Exp.1 n = 14 ASD – m.a. = 15.9 vs n = 12 TD – m.a. = 17.1; Exp. 2 n = 16 ASD – m. a. = 14.5 vs n = 16 TD – m. a. = 14.5) spent less time watching eyes than did TD subjects and the same amount of time as TD subjects in watching mouths while observing static black and white images of faces.

How to explain these discrepancies? Bar-Haim et al. (2006) hypothesizes that these contrasts in results could derive from the loss of interest that occurs in longer observation. In fact, experimenters showed with a presentation of static photos of emotionally neutral faces that, like TD subjects, subjects with ASD make an initial attentional shift from the eye region and do not disengage within 400 ms from stimulus presentation. De Wit et al. (2008) proposed another interpretation. With a sample of 13 subjects with ASD (m.a. = 5.16) and 14 TD subjects (m.a. = 4,93) it was shown that there was no difference between groups in their time looking at the eye region and that children with ASD look for a shorter time than TD children at the mouth; also, the opposite trend revealed by Klin et al. (2002), poorer social and communicative skills (Autism Diagnostic Interview-Revised[6]), were correlated with a shorter time looking at the mouth region by a presentation of static, representative emotional photos (10 s). Experimenters attributed these differences to the difference in abilities of the two samples. Klin et al. (2002) included more verbal individuals than de Wit et al. (2008), so de Wit et al. concluded «predominantly verbal individuals will be expected to look preferentially at the mouth». Effectively, a more specific analysis of Klin et al. (2002)'s study reveal that indices of fixing on the mouth and eyes had a very high standard deviation; thus, if experimenters also found a statistically significant tendency to fix on the eyes in TD participants and the mouth in ASD participants, in each group it is possible that some subjects did not show a marked tendency like other members of their group. Due to the small number of subjects in their experiment, Klin et al. did not have the possibility of identifying clusters in a post-experiment analysis.

§ 4. Relations between differences in patterns of face scanning and differences in linguistic abilities

As described in the previous paragraph, some researchers suggested a correlation between high eye-mouth index (EMI[7]) and the severity of autistic symptoms (KLIN

[5] Vineland Adaptative Behavior Scales Expanded Edition.

[6] Autism Diagnostic Interview-Revised (ADI-R) (LORD et al. 1994).

[7] We use an index elaborated in ELSABBAGH et al. 2014 and that it was calculated as follows: (looking time towards the eyes – looking time towards the mouth)/ total looking time to any area of the scene. The higher the value of this index, the greater is the difference between the time spent by

et al. 2002). Due to the high occurrence of discrepant results, De Wit et al. (2008) proposed a different interpretation of the percentage of time spent looking at the mouth region, and hypothesized a correlation between time spent in looking at the mouth and linguistic abilities. Effectively, Young et al. (2009) and Elsabbagh et al. (2014) showed that the EMI in children considered at risk for developing the pathology did not relate to a future diagnosis, but they found some correlation with the development of expressive language.

Young et al. (2009) confronted 33 24-month-old infants who at 6 months of age were declared at risk for receiving a future diagnosis of autism (15 high risk, 19 low risk) with 25 TD infants in ecological 3-min interactions with the mother (the second minute in the or still face condition[8]). Only three of the 33 children at risk received an autism diagnosis. None of the 6-month-old infants' gaze behaviors predicted severity scores in ADOS or in M-CHAT. However, EMI was negatively related to expressive language at 24 months of age and to expressive language growth according to the Expressive Language subscale[9] of MSEL and Vineland. These results are perfectly in line with the study of Lewkowicz and Hansen-Fit (2012) that showed that TD infants (179 in their sample) shifted their attention from the eyes to the mouth between 4 and 8 months of age, and that afterwards they shifted back to eyes in response to native, but not non-native, language.

In a longitudinal study, in which four conditions were confronted (only eyes are moving; only mouth is moving; only hands are moving; eyes, mouth and hands are moving), Elsabbagh et al. (2014) showed that in a scene in which hands, eyes and mouth are simultaneously moving, negative EMI at 7 months predicts superior expressive language at 36 months. In contrast, EMI at 14 months or in a simplified scene in which only eyes, mouth or hands are moving did not correlate with either diagnosis, EL or receptive language.[10] Moreover, more time spent watching the mouth when it alone is moving, is associated with poor expressive language in all groups and with greater impairment in social ability (measured by ADOS) in the at-risk group. Thus, the tendency to look longer at the mouth rather than the eyes may be context-dependent and if it is present at 7 months, it is a predictor of better expressive language development, but not a predictor of a future ASD diagnosis.

Falck-Ytter et al. (2010), studying a sample of 15 children with ASD (m. a. = 5.16 years) and 15 TD children (m. a. = 4.91 years) and also with another sample of 12 children with ASD (m. a. = 6.58 years) suggest that «there is a tighter link between language and non-verbal communication skill than between language and socio-emotional skills». Experimenters showed 15 children with ASD (mean age 5 years and 2 months) and to 15 TD children (mean age 4 years and 11 months of age) 36 short videos (4 s) showing expressions of different emotions. Some of these were inverted. Experimenters used the Social Impairments[11] and the non-verbal part of the

infant in watching eyes and mouth. A positive value indicates that the child looked longer at eyes, a negative value indicates that the child spent more time looking at the mouth.

[8] The caregiver ignores the baby and is expressionless.

[9] Ability to put thoughts into words and sentences, in a way that makes sense and is grammatically accurate (subscale of Mullen Scales of Early Learning, MSEL).

[10] Ability to understand what is heard or read (subscale of MSEL).

[11] Inability to develop relationship with peers appropriate to their level of development.

Communication Impairment[12] of the ADI-R to test differences in non-verbal communication skills and socio-emotional skills and found two different cognitive profiles in ASD subjects:

- Children better at non-verbal communication look more at the mouth that at the eyes
- Children with better social skills look less at the mouth region
- There was a significant positive correlation between Social Impairment-Communication Impairment and mouth-looking time (suggesting that longer looking time at the mouth is related to the balance between Social Impairment and Communication Impairment rather than to one of their individual values).

In the same study, experimenters replicated these results with a new sample of 12 children with ASD (mean age 6 years and 7 months). None of the other subscales or other measures of intelligence correlated with data, so this result cannot be explained by IQ or other adaptive skills.

The relative independence of social and communicative abilities was also extrapolated in general to the population with ASD (HAPPÉ et al. 2006; RONALD et al. 2005, 2006b). Happé et al. (2006) sustained that – among the symptoms of the autism triad of the DSM IV – there is always prevalence and that there is no correlation between the severity of social impairment and the severity of communicative impairment. Their assumptions were based on Ronald et al. (2006a), a study conducted on 3419 twins in which it was shown that the three autistic traits described in DSM IV appear to be genetically determined but interdependent.

Thus, the theory of Flack-Ytter et al. (2010) may be useful for identifying a cognitive profile of autistic subjects: a child more focused on the mouth in face scanning would probably have poorer social skills (so could have more difficulty in developing affective relationships); on the contrary, a child more focused on the eyes during face scanning could have more problems in silent communication (and thus show more problems in the use and interpretation of eye gaze, proxemics, facial expressions, etc.).[13]

The first observation appears to be more easily explained than the second: children who focus on eye gaze for a long time are probably concentrating on the more salient stimulus, but fail to attribute a social meaning to it; thus they do not respond to it in the typical way of looking away from the eyes. On the contrary, the first question is more controversial: if looking longer at the mouth predicts poorer social abilities – how can we explain this trend?

§ 5. Lip reading in autistic subjects

In TD subjects, audiovisual input facilitates the processing of spoken syllables rather than an audio input (STEKELENBURG e VROOMEN 2007; VAN WASSENHOVE et al. 2005).

[12] Marked impairment in the use of various nonverbal behaviors, such as direct gaze, facial expression, body postures and gestures to regulate social interaction.

[13] From results shown in §2, another question arises that we have not discussed in this paper: if social attitudes are positive correlated with language skills and silent communication is uncorrelated with social attitudes, can we infer that language is uncorrelated with silent communication, as is most frequently sustained (cfr. i. e., BARA 1999)? We attempted to answer this question in PENNISI 2016.

In 1976, Harry McGurk and John MacDonald showed that if the listener could see the speaker, his visual perception of the verbal communication will influence his auditory perception, creating a confusion or a combination between them. In fact, if we show the listener a film in which the video shows lips that say "ga-ga" and the audio plays the sound "ba-ba", probably he will perceive "da-da" (MCGURK e MAC DONALD 1976). This phenomenon, known as McGurk effect, is universal.

In 2004, Williams et al. showed that ASD subjects (n = 15, m.a. = 8,81; vs TD n = 15, m. a. = 9.5) did not suffer the McGurk effect, but experimenters link these differences not to an absence of integration, but to their poor visual accuracy, measured in the experiment by a speech-reading task, whose participants' accuracy in its performance correlated with participants' performance in the British Picture Vocabulary Scale.[14]

Bebko et al. (2006) compared the reactions of three groups of children (ASD n = 16, m. a. = 5.49; DD[15] n = 15, m. a. = 4.88; TD n = 16, m. a. = 2.36) in contemporary exposure to two screens, one with correct audio-video synchrony and the other with a 3-s delay between audio and visual stimuli during three conditions: non-linguistic event; simple linguistic event and complex linguistic event. While the TD group showed greater preference for synchronized inputs, the ASD group showed this preference only in the non-linguistic conditions. The DD group manifested a preference for synchronized input in non-linguistic and simple linguistic input, and borderline preference for complex linguistic input. Experimenters proposed two possible interpretations of the ASD group's results: either ASD subjects do not perceive synchrony in linguistic stimulus, or they are slow to detect these kinds of violations and the trial is too short.

Smith & Benedetto (2007) replicated these results with a more ecological procedure in which participants (18 ASD – mean 15.84 years VS 19 TD mean 16.08 years) report captured words in the three conditions (visual, auditory and audio-visual). Results confirmed autistic audio-visual impairment, autistic deficit in lipreading and the exclusiveness of TD in improve comprehension of speech in noise with the addition of visual information.

Iarocci et al. (2010) replicated the experiment with a sample of 12 children with ASD and 12 TD children (ASD m. a. = 10.58; TD m. a.= 10.31) using photorealistic images of the mouth and nose region of a male face and found – in line with Smith e Bennetto (2007) and Williams et al. (2004) – that the ASD group performed worse in visual-only conditions.

Megnin et al. (2012) presented monosyllabic spoken words in five conditions (auditory only with face; visual only with face; audiovisual with face; visual only with scrambled face; audiovisual with scrambled face) to 14 TD and 14 ASD (both mean16.9 years of age) and ask them to detect a target while experimenters registered their ERP (they focused on three consecutive stages: word detection, so N1; transition from phonetic to lexical-semantic analysis, so P2 and semantic integration, so N4). Based on their data, experimenters inferred that neural networks that in elaborating spoken words are normally facilitated by visual features, are altered in ASD subjects. In particular, amplitude of N1 and N4 were attenuated in

[14] DUNN et al. 1997. A linguistic measure of autistic symptoms.

[15] Developmental disorders.

ASD samples during audiovisual effects[16] and the amplitude of P2 negatively correlated with scores on the Social Communication Questionnaire (Rutter et al. 2003).
Thus differences in cognitive processing of the phase of translation from phonetic to a lexical-semantic analysis are related to autistic symptoms.
According to Gardner's classic theory of multiple intelligences (GARDNER 1983), syntax and phonology are linked to the strictly linguistic intelligence; on the contrary, semantics and pragmatics are linked to personal and logical-mathematical intelligences. The data that we showed in our paragraph could be interpreted as follows: probably the semantic integration in TD subjects is supported by personal intelligence, on the contrary, in subjects with ASD it is supported by a linguistic or logical-mathematical intelligence. The pragmatic integration of the context could require an integration of visual information that the linguistic or logical-mathematical intelligence do not require or cannot include between the inputs.
However, if the movement of lips cannot be processed as lipreading, how is it considered in autistic cognition?
In the beginning, our idea was that if a salient stimulus is useless, it would have a distracting effect. We looked for confirmation of this hypothesis in the scientific and introspective literature on autism and did not find elements that strictly move in this direction.

§ 6. Absence of ToM, absence of attention, absence of lipreading

Thus, in §4 we showed a negative correlation between social abilities and the average amount of time spent by autistic subjects in looking at the mouth. In §5 we showed that subjects with ASD do not use lip reading information during linguistic interaction.
MAGNE et al. (2011) showed in an ERP study that subjects with autism are able to perform multisensory integration, but only during easy selective attention; on the contrary, TD subjects can also easily maintain multisensory integration during tasks that require divided attention. Similarly, Grossman et al. (2009) showed that adolescents with ASD can perform multisensory integration in tasks constructed with meaningful stimuli and performed in non-distracting environments.
Both these studies show that distraction is an enemy of the audio-visual integration for subjects with ASD. Thus, a subject with autism who does not integrate lipreading in the understanding of the language message is not focused on the latter. At this point the question is: what is an autistic subject doing when looking at a mouth during a verbal interaction? In the light of previous paragraphs, we may infer that he is simply not focusing on the interaction. The absence of an interoception by which to build a perceptive gestalt that makes salient biological stimuli push him to perceive a stimulus (the mouth) that is not the most salient for ToM, but is so for the classic rules of perception: the mouth is the stimulus that moves the most.

[16] Effect of seeing lip movements accompanying speech.

§ 7. Conclusions

We hypothesized in Pennisi 2016 that interoceptive inferences aids in integrating various external stimuli towards a joint interpretation of reality with other conspecifics. In light of this theory, we have attempted to explain the lipreading deficit, fixation on the mouth during linguistic interactions, and the negative correlation between fixation on mouths and social abilities We also attempted to show that in these cases, lipreading has a distractor effect in subjects with ASD during verbal interactions, but we did not find confirmation for this last hypothesis.

To summarize, in §0 we posed a question: is the attraction of subjects with ASD toward social robots reported in many studies real? How can it be explained? To answer these questions, we showed our hypothesis according to which a main problem of subjects with ASD is a deficit in interoceptive inferences, which cause their social problems from which are derived their linguistic problems. Cerebral correlates of the deficit in interoceptive inferences are localized in the reward system (deficient toward social motivation in subjects with ASD; CHEVALLIER et al. 2012) and in the salience network (UDDIN e MENON 2009; CARROZZA et al. *in press*; PENNISI 2016).

In §1 we explained the concept of *interoceptive inference* and in §2 its relation with autism.

In §3 and §4 we discussed some empirical evidence regarding the possibility of creating several clusters useful for identifying the cognitive profile of subjects with ASD on the basis of their propensity to look longer at the mouth or eyes: it seems probable that the attitude of looking for a longer period of time at the mouth is negatively correlated with social abilities. This converges with our hypothesis that subjects with ASD cannot consider the lips' movement as a biological motion, as do TD subjects. In §5 we showed some empirical evidence regardingthe lipreading deficit in subjects with ASD and attempted to explain that in terms of different cognitive styles (different intelligences used, if we consider Gardner's theory of multiple intelligences) during the interpretation of linguistic messages.

In conclusion, deficits in integrating audio-visual stimuli may be associated with atypical or delayed language acquisition (SMITH e BENNETTO 2007; IAROCCI et al. 2010).

Senju & Johnson (2009) opposed to the classic *communicative intention detector model* (BARON-COHEN 1995) their *fast-track modulator model*, which proposes that the subcortical face detection pathway (superor colliculus, pulvinar and amygdala) mediate eye contact processing in a «quick & dirty pathway» (SENJU e JOHNSON 2009:1207) and not – as Baron Cohen argues – in directly activating the theory-of-mind computation. The authors used the case of differences in salience of the mouth in subjects with ASD rather than TD subjects to compare the two models:

> Proponents of the communicative intention model could claim that lack of fixation on the eyes is a compensatory strategy to acquire more information from others' mouth, as they have modular impairment in decoding information from the eyes. For example, Klin et al. (2002) reported that in individuals with ASD greater fixation on the mouth region predicts higher levels of social adaptation and lower levels of autistic social impairment. In contrast, both the hypoarousal model and the fast-track modulator model would encourage the view that individuals with ASD are more readily attracted by visually salient features such as the speaking mouth when it is moving, because their fixation is not adequately guided to eye contact (*Ivi*: 1209).

From our study it appears to emerge that different styles of scanning faces are linked to different cognitive compensations for a salience deficit typical of ASD. The curious phenomenon by which some individuals with ASD apparently have a positive interaction with robots will likely depend on the fact that this type of activity may also be performed successfully only through exteroceptive inferences. Moreover, in some cases the absence of complex behaviors such as the extremely fine lip movements needed for linguistic articulation may be a lesser element of complexity for the integration of the audio-visual integration of the whole scene.
If our hypothesis is correct, this would raise new and profound reflections on the inherently social nature of language: in order to develop it properly, it is not enough for the future speaker to live in a community of other speakers, but they must also able to make those interoceptive inferences.

References

BARA, Bruno G. (1999), *Pragmatica cognitiva: I processi mentali della comunicazione*, Bollati Boringhieri, Torino.

BEBKO, James M., WEISS, Jonathan A., DEMARK, Jenny L., GOMEZ, Pamela (2006), «Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism» in *Journal of Child Psychology and Psychiatry*, 47(1), pp. 88-98.

BRASS, Marcel, SCHMITT, Ruth M., SPENGLER, Stephanie e GERGELY, György (2007), «Investigating action understanding: inferential processes versus action simulation» in *Current Biology*, 17(24), pp. 2117-2121.

CARROZZA, Cristina, FAMÀ, Francesca I., LEONARDI, Elisa, PENNISI, Paola, SPADARO, Letteria (2016), «Autismo e pragmatica clinica: attenzione e percezione nella prospettiva della social motivation» (in press).

CHEVALLIER, Coralie, KOHLS, Gregor, TROIANI, Vanessa, BRODKIN, Edward S. e SCHULTZ, Robert T. (2012), «The social motivation theory of autism» in *Trends in cognitive sciences*, 16(4), pp. 231-239.

CHIEN, Miaoen E., JHENG, Cyun M., LIN, Nimiao M., TANG, Hsienhui, TAELE, Paul, TSENG, Wensheng S., CHEN, Mike Y. (2015), «iCAN: A tablet-based pedagogical system for improving communication skills of children with autism» in *International Journal of Human-Computer Studies*, 73, pp. 79-90.

CLARK, A. (2013), «Whatever next? Predictive brains, situated agents, and the future of cognitive science», in *Behavioral and Brain Sciences*, 36(03), 181-204.

DALTON, Kim M., NACEWICZ, Brendon M., JOHNSTONE, Tom, SCHAEFER, Hillary S., GERNSBACHER, Morton A., GOLDSMITH, Hill H., ALEXANDER, Andrew L., DAVIDSON, Richard J (2005), «Gaze fixation and the neural circuitry of face processing in autism» in *Nature neuroscience*, 8(4), pp. 519-526.

DE GELDER, Beatrice D., VROOMEN, Jean, VAN DER HEIDE, Lucienne (1991), «Face recognition and lip-reading in autism», in *European Journal of Cognitive Psychology*, 3(1), pp. 69-86.

DE WIT, Tessa C., FALCK-YTTER, Terje, VON HOFSTEN, Claes (2008), «Young children with autism spectrum disorder look differently at positive versus negative emotional faces» in *Research in Autism Spectrum Disorders*, 2(4), pp. 651-659.

DUNN, Lloyd (1997), *The British Picture Vocabulary Scale* (2nd ed.), NFER-Nelson, Windsor Berks.

ELSABBAGH, Mayada, BEDFORD, Rachael, SENJU, Atsushi, CHARMAN, Tony, PICKLES, Andrew, JOHNSON, Mark H. BARON-COHEN, Simon, BOLTON, Patrick, CHANDLER, Susie, FERNANDES, Janice, GARWOOD, Holly, GLIGA, Teodora, HURDY, Kristelle, PASCO, Greg, TUCKER, Leslie e VOLEIN Ágnes (2014), «What you see is what you get: contextual modulation of face scanning in typical and atypical development», in *Social cognitive and affective neuroscience*, 9(4), pp. 538-543.

FALCK-YTTER, Terje, FERNELL, Elisabeth, GILLBERG, Christopher L., VON HOFSTEN, Claes (2010), «Face scanning distinguishes social from communication impairments in autism» in *Developmental science*, 13(6), pp. 864-875.

GRANDIN, Temple, JOHNSON, Catherine (2005), *Animals in translation: Using the mysteries of autism to decode animal behavior*, Scribner, New York.

GROSSMAN, R. B., SCHNEPS, M. H., & TAGER-FLUSBERG, H. (2009), «Slipped lips: onset asynchrony detection of auditory-visual language in autism», in *Journal of Child Psychology and Psychiatry*, 50(4), pp. 491-497.

HAITH, Marshall M., BERGMAN, Terry, MOORE, Michael J. (1977), «Eye contact and face scanning in early infancy» in *Science*, 198(4319), pp. 853-855.

HAPPÉ, Francesca, RONALD, Angelica, PLOMIN, Robert (2006), «Time to give up on a single explanation for autism» in *Nature neuroscience*, 9(10), pp. 1218-1220.

HERRINGTON, John, NYMBERG, Charlotte, FAJA, Susan, PRICE, Elinora, SCHULTZ, Robert (2012), «The responsiveness of biological motion processing areas to selective attention towards goals» in *Neuroimage*, 63(1), pp. 581-590.

HOHWY, Jacob (2013), *The predictive mind*, Oxford University Press, Oxford.

HUNNIUS, Sabine, GEUZE, Reint H. (2004), «Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: A longitudinal study» in *Infancy*, 6(2), pp. 231-255.

IAROCCI, Grace, ROMBOUGH, Adrienne, YAGER, Jodi A., WEEKS, Daniel J., CHUA, Romeo (2010), «Visual influences on speech perception in children with autism», in *Autism*, 14(4), pp. 305-320.

KILNER, James M, FRITH, Chris D (2008), «Action observation: inferring intentions without mirror neurons», in *Current Biology*, 18(1), pp. R32-R33.

KLIN, Ami, JONES, Warren, H., SCHULTZ, Robert, VOLKMAR, Fred R., COHEN, Donald (2002), «Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism» in *Archives of general psychiatry*, 59(9), pp. 809-816.

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

KOSTER-HALE, Jorie, SAXE, Rebecca (2013), «Theory of mind: a neural prediction problem», in *Neuron*, 79(5), pp. 836-848.

LEWKOWICZ, David J., HANSEN-TIFT, Amy M. (2012), «Infants deploy selective attention to the mouth of a talking face when learning speech», in *Proceedings of the National Academy of Sciences*, 109(5), pp. 1431-1436.

LORD, Catherine, RUTTER, Michael L., LE COUTEUR, Ann (1994), «Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders», in *Journal of autism and developmental disorders*, 24(5), pp. 659-685.

MAURER, Daphne, SALAPATEK, Philip (1976), «Developmental changes in the scanning of faces by young infants», in *Child development*, pp. 523-527.

MAGNÉE, M. J., DE GELDER, B., VAN ENGELAND, H., & KEMNER, C. (2011), «Multisensory integration and attention in autism spectrum disorder: Evidence from event-related potentials», in *PLoS One*, 6(8), e24196.

MCGURK, Harry, MACDONALD, John (1976), «Hearing lips and seeing voices» in *Nature*, 264, pp. 746-748.

MEGNIN, Odette, FLITTON, Atlanta, RG JONES, Catherine, DE HAAN, Michelle, BALDEWEG, Torsten, CHARMAN, Tony (2012), «Audiovisual speech integration in autism spectrum disorders: ERP evidence for atypicalities in lexical semantic processing» in *Autism Research*, 5(1), pp. 39-48.

ONDOBAKA, Sasha, KILNER, James, FRISTON, Karl (2015), «The role of interoceptive inference in theory of mind» in *Brain and cognition* (in press).

PELPHREY, Kevin A., SASSON, Noah J., REZNICK, James S., PAUL, Gregory, GOLDMAN, Barbara D., PIVEN, Joseph (2002), «Visual scanning of faces in autism», in *Journal of autism and developmental disorders*, 32(4), pp. 249-261.

PENNISI, Paola (2016), *Il linguaggio dell'autismo. Studi sulla comunicazione silenziosa e la pragmatica delle parole*, Il Mulino, Bologna, (in press).

PENNISI, Paola (2014), *La percezione visiva nei soggetti con autismo: osservare i volti umani attraverso un diaframma chiuso*, in FALZONE, A., NUCERA, S., PARISI, F., (a cura di), *Le ragioni della natura*, Corisco Edizioni, 2014, ISBN: 978-88-98138-11-1, pp. 337-352.

PENNISI, Paola, TONACCI, Alessandro, TARTARISCO, Gennaro, BILLECI, Lucia, RUTA, Liliana, GANGEMI, Sebastiano, PIOGGIA, Giovanni (2015), «Autism and Social Robotics: a systematic review» in *Autism Research* (in press).

PUTNAM, Cynthia, CHONG, Lorna (2008, October), «Software and technologies designed for people with autism: what do users want?», in *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*, pp. 3-10.

RONALD, Angelica, HAPPÉ, Francesca, PLOMIN, Robert J. (2005), «The genetic relationship between individual differences in social and nonsocial behaviors characteristic of autism», in *Developmental science*, 8(5), pp. 444-458.

RONALD, Angelica, HAPPÉ, Francesca, PRICE, Thomas S., BARON-COHEN, Simon, PLOMIN, Robert (2006), «Phenotypic and genetic overlap between autistic traits at the extremes of the general population», in *Journal of the American Academy of Child & Adolescent Psychiatry*, 45(10), pp. 1206-1214.

SMITH, Elizabeth G., BENNETTO, Loisa (2007), «Audiovisual speech integration and lipreading in autism», in *Journal of Child Psychology and Psychiatry*, 48(8), pp. 813-821.

STEKELENBURG, Jeroen J., VROOMEN, Jean (2007), «Neural correlates of multisensory integration of ecologically valid audiovisual events», in *Journal of Cognitive Neuroscience*, 19(12), pp. 1964-1973.

TOMASELLO, Michael (2009), *Le origini della comunicazione umana*, Cortina, Milano.

UDDIN, Lucina Q. (2015), «Salience processing and insular cortical function and dysfunction», in *Nature Reviews Neuroscience*, 16(1), pp. 55-61.

UDDIN, Lucina Q., MENON, Vinod (2009), «The anterior insula in autism: under-connected and under-examined», in *Neuroscience & Biobehavioral Reviews*, 33(8), pp. 1198-1203.

UDDIN, Lucina Q., SUPERKAR, Kaustubh, LYNCH, Charles J., KHOUZAM, Amirah, PHILLIPS, Jennifer, FEINSTEIN, Carl, RYALI, Srikanth, MENON, Vinod (2013), «Salience network–based classification and prediction of symptom severity in children with autism», in *JAMA psychiatry*, 70(8), pp. 869-879.

VAN DER GEEST, Jos Nicolaas, KEMNER, Chantal, VERBATEN, Marinus N., VAN ENGELAND, Herman (2002), «Gaze behavior of children with pervasive developmental disorder toward human faces: a fixation time study», in *Journal of Child Psychology and Psychiatry*, 43(5), pp. 669-678.

VAN WASSENHOVE, Virginie, GRANT, Ken W., POEPPEL, David (2005), «Visual speech speeds up the neural processing of auditory speech», in *Proceedings of the National Academy of Sciences of the United States of America*, 102(4), pp. 1181-1186.

VLAMINGS, Petra H., STAUDER, Johannes E., VAN SON, I. AM, MOTTRON, Laurent (2005), «Atypical visual orienting to gaze-and arrow-cues in adults with high functioning autism», in *Journal of autism and developmental disorders*, 35(3), pp. 267-277.

WILLIAMS, Justin H., MASSARO, Dominic W., PEEL, Natalie J., BOSSELER, Alexis, SUDDENDORF, Thomas (2004), «Visual–auditory integration during speech imitation in autism», in *Research in developmental disabilities*, 25(6), pp. 559-575.

YARBUS, Alfred L. (1967), *Eye movements and vision*, Plenum Press, New York.

YOUNG, Gregory S., MERIN, Noah, ROGERS, Sally J., OZONOFF, Sally (2009), «Gaze behavior and affect at 6 months: predicting clinical outcomes and language development in typically developing infants and infants at risk for autism», in *Developmental science*, 12(5), pp. 798-814.