**Theory of Mind and Autism**

The mentalizing process, often referred to as "theory of mind," involves the ability to infer the mental states of others, such as their beliefs, intentions, and desires, particularly in competitive environments. This is essential for successful interaction in competitive environments, where understanding and anticipating an opponent's actions can provide a strategic advantage. The study by Hampton et al. (2008) explores the neural correlates of mentalizing during strategic interactions using a computational modeling approach. Neuroimaging studies have identified specific brain regions involved in mentalizing, including the dorsomedial prefrontal cortex (dmPFC), posterior superior temporal sulcus (pSTS), and temporal poles. However, the precise neural mechanisms by which mentalizing occurs during dynamic interactions have been less clear. They address it by proposing 3 computational models that try to capture different aspects of mentalizing in strategic interactions:

1. **Reinforcement Learning (RL) Model**: This model is based on the principle of reinforcement learning, where an individual chooses actions based on the rewards received in previous trials. It updates the value of actions based on the prediction error, which is the difference between the expected and received rewards. While effective in many decision-making scenarios, the RL model falls short in competitive environments because it does not account for the opponent's strategy.
2. **Fictitious Play Model**: The fictitious play model is a more sophisticated approach that involves predicting an opponent's next move based on their past actions. Unlike the RL model, it incorporates an element of mentalizing by considering the opponent's strategy, thus offering a better fit for competitive interactions.
3. **Influence Learning Model**: This model extends the fictitious play model by not only predicting the opponent's future actions but also incorporating how one's own actions influence the opponent's strategy. It represents a higher level of cognitive sophistication, capturing the dynamic interplay between the player's and the opponent's strategies. The influence learning model was found to best capture the subjects' behavior during the game, as it integrates both prediction and influence.

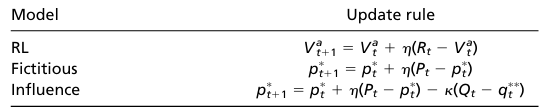


Fig 1: The Update Rules for the three models

Out of these the influence learning model is highlighted as the most comprehensive model for mentalizing in strategic interactions after comparing the fata collected from various fMRI studies.

In Corricelli 2009, a similar idea of mentalizing was discussed as the influence learning model and it was the “iterated steps of thinking/reasoning” i.e There are various levels of thinking involved while mentalizing and that different people in think at different levels. It was also argued that being able to distinguish between low- versus high-level reasoning people by their brain activity will help to explain the heterogeneity observed in human strategic behavior. They also mentioned an important difference between the Hampton et al. and this study was that in Hampton’s study subjects observed others’ behavior over time and then responded to it, whereas here the decisions required that subjects model and predict others’ choices without knowing other players’ past choices. The brain does not seem to distinguish between these 2 data sources. Essentially, the results of the two represent the first neural evidence of a close link between adaptive learning and levels of reasoning.

In the process of better understanding this phenomenon, we moved to studies done on people diagnosed with Autism. As one of the characteristics of autism spectrum dis order (ASD) is a deficit in social interaction along with an inability to understand the beliefs and intentions of others. Although various theories have been discussed regarding this there is still no common consensus over it. A few to name would be:

1. Theory of Mind hypothesis: It proposes that individuals with autism lack a specific meta-representational capacity, namely a “theory of mind”, which prevents them from mentalizing
2. Social Motivation Hypothesis: It focuses more on the motivational aspect of initiating and continuing a social interaction than cognitive aspects. Arguing that individuals with ASD 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.
3. Broke Mirror Neuron Hypothesis (MNS)

We will now be discussing two such theories in detail.

**Autism as Prediction Error Disorder**

It was proposed by Joshua H. Balsters et al (2017) which argued that disrupted prediction errors are the reason behind the profound impairments of social interaction found in people with ASD. Traditionally, prediction errors signal the discrepancy between our own expectations and the actual outcomes of an action; however, social prediction errors shift the frame of reference from the first person to the third person perspective by comparing actual outcomes with the perceived expectations of another person. They concluded that individuals with ASD have a deficit in understanding the perspectives of others in the absence of social interaction, and that understanding the first and third person perspectives in isolation could help to further inform impairments in social interaction and the second person perspective. They also linked this to decreased activity in ACCg and its deceased connectivity to vmFC regions of the brain.

Continuing with the theory that Autism is a predictive disorder, a major hypothesis that was discussed in the Sinha et al (2019) was the PIA (Predictive Impairment in Autism) hypothesis. It focuses on the fact that being able to perform well in a social interaction depends on the ability to figure out the underlying Markov System. The brain has to estimate the conditional probability P(B|A, Δt), the likelihood of transitioning to state “B” given the occurrence of “A” and elapsed temporal duration, Δt. The PIA hypothesis posits that autism may be associated with inaccuracies in estimating the P(B|A, Δt) conditional probability.

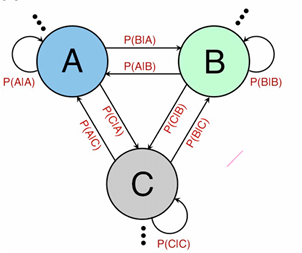


Fig 2: A fiqurative representation of a Markov System along with the

conditional probabilities

It goes on to explain few of the significant phenotypes of ASD like Insistence on Sameness (IoS), stimming, sensory hypersensitivities etc. It also explains the difficulties that people with ASD face with theory of mind. Arguing that its a prediction task—given an observation, one has to postdict or predict its antecedent or subsequent states, i.e., estimate P(past history | current behavior) or P(future actions | current observations). Hence impairments with the predictive ability would lead to issues with ToM. Then an autistic individual, not having access to such predictive relationships, is constrained to interpret behaviors without any motivating history, social or otherwise seemingly making observation in a “in the moment” manner.

**Use of Bayesian Modelling**

The predictive coding/Bayesian inference 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. 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 prior belief) 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.

In case of autism it was proposed that autistic traits might be related to higher sensory precision, i.e., a stronger reliance on (bottom-up) sensory evidence compared with (top-down) prior beliefs. Autism essentially doesnot affect the learning process of new social information but rather affects to which extend this social interaction will affect the decision making process. Here they used hierarchical generative models (to be specific it a Hierarchical Gaussian Filter) to combine the responses from perceptual model and response model to test this. This approach allows the estimation of hierarchically coupled hidden states that describe subjects’ learning about the environmental statistics based on their responses. These subjective beliefs are weighted by their precision to form the basis of a response mode. A few learning parameters are taken into account to determine the dynamics of the belief trajectories—the accuracy and volatility estimates as well as their precisions.

Their results show a learning deficit in subjects with high AQ scores, namely, the failure to take into account the social context to adapt the learning about the nonsocial stimulus to accurately predict the outcome. They also suggest that autistic traits are associated with a reduced impact of social information on the precision of higher-level prior beliefs, which has a detrimental effect on probabilistic learning about final outcomes.

The results from these two major hypothesis are in line with findings from other researches as well. Essentially proposing that autism may not impair the ability to process social information, but rather lead to differences in how the relevant representations are integrated for optimal action selection while the mentalizing process.

References:

1. [Neural correlates of mentalizing-related computations during strategic interactions in humans | PNAS](https://www.pnas.org/doi/full/10.1073/pnas.0711099105)

# [Neural correlates of depth of strategic reasoning in medial prefrontal cortex](https://www.pnas.org/doi/full/10.1073/pnas.0807721106)

1. [Observing and participating in social interactions: Action perception and action control across the autistic spectrum](https://pubmed.ncbi.nlm.nih.gov/28188104/)

4) [Disrupted prediction errors index social deficits in autism spectrum disorder](https://pubmed.ncbi.nlm.nih.gov/28031223/)

5) [Autism as a disorder of prediction | PNAS](https://www.pnas.org/doi/full/10.1073/pnas.1416797111)

6) [Social Bayes: Using Bayesian Modeling to Study Autistic Trait–Related Differences in Social Cognition - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S0006322319317901?via%3Dihub)