EXPERIENCE IN AUTISM SPECTRUM DISORDER

AS EVIDENCE FOR

THE RELATIONSHIP BETWEEN SIMULATION THEORY AND EMPATHY

HASHEM RAMADAN

BOĞAZİÇİ UNIVERSITY

2022

EXPERIENCE IN AUTISM SPECTRUM DISORDER AS EVIDENCE FOR THE RELATIONSHIP BETWEEN SIMULATION THEORY AND EMPATHY

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Philosophy

by

Hashem Ramadan

Boğaziçi University

2022

DECLARATION OF ORIGINALITY

I, Hashem Ramadan, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature……

Date ……

ABSTRACT

Experience in Autism Spectrum Disorder

as Evidence for the Relationship Between Simulation Theory and Empathy

Much of the discussion around theory of mind (ToM) puts autism spectrum disorder (ASD) in the spotlight because of its many symptoms affecting social cognition and interaction. More specifically, simulation theory in those with ASD is a focus of many mind theorists. By presenting studies and a discussion of social and cognitive deficiencies pertaining to ASD, I attempt to draw a link between simulation theory and empathy where I claim that simulation theory is the cognitive process that shows its effects through what we call empathy. After presenting a literature review on ToM and ASD, I will move on to an explanation of the nature of empathy and how people with ASD have difficulties showing empathy in comparison to control groups, taking this as evidence for my claim for the relationship between simulation theory and empathy.

ÖZET

Otizm Spektrum Bozukluğu Deneyimi

Simülasyon Teorisi ile Empati arasındaki İlişkiye dair Kanıt Sunuyor

Zihin Teorisi etrafındaki çoğu tartışma, otizm spektrum bozukluğunu ön plana koymaktadır. Bunun nedeni otizm spektrum bozukluğunun belirtilerinin sosyal bilişim ve etkileşim üzerindeki etkileridir. Daha spesifik olarak, birçok zihin teoristinin araştırmaları simülasyon teorisi üzerinedir. Bu tezde, bu konudaki calışmaları ortaya koyarak ve otizm spektrum bozukluğuna dair sosyal ve bilişsel eksiklikleri tartışarak, simülasyon teorisi ve empati arasında bir bağ kurmaya calışıyorum ve bu ikisi arasında bir korelasyon olduğunu iddia ediyorum. Zihin Teorisi ve otizm spektrum bozukluğu üzerine literatürü sunduktan sonra, empatinin doğasının açıklanmasına ve otizm spektrum bozukluğu olan insanlarin, kontrol gruplarına kıyasla, empati göstermede nasıl zorluklarla karşılaştığı konusuna geçecegim. Otizm spektrum bozukluğunun kişilerin bilişsel ve sosyal yetileri üzerindeki etkilerinin, simülasyon kapasitesine sahip olmamanın empati yoksunluğu ile doğrudan ilişkili olduğuna dair kanıtlar sunuyor olabileceğini iddia ediyorum.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my cousin Sarah for her support during the last few weeks of writing. She constantly pushed me to continue working and monitored my writing process. Big thanks also go to my family (parents, sisters, and aunt) for giving me moral support during this time. Thanks also go to the committee members, especially Professor Lucas Thorpe for his constant support and advice. Finally, I thank my friends Özcan (who helped with translating the abstract), Sanghun, Enes, and Farooq who continuously encouraged me to finish writing.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

In this thesis, I will try to draw a link between simulation theory and empathy through a discussion of social and communicative failures experienced by individuals with Autism Spectrum Disorder (ASD). I will explain how the apparent results of simulation are a group of related phenomena that we call empathy. Before discussing ASD and mindreading, which is how humans understand other minds, I shall explain the concept of theory of mind (ToM). ToM is the cognitive capacity with which humans comprehend or predict others' mental states. After attributing mental states to others, predictions of the person's thoughts or feelings would be formed, and one could act accordingly (Goldman, 2012). This is needed since to mind read is to understand and predict others' desires, thoughts, and feelings (Heyes & Frith, 2014). Such a capacity is vital for communication, whether verbal or non-verbal. Much of humans' lives are governed by social queues requiring a combination of background knowledge and induction of basic intentions that may not usually be communicated in a literal sense. If one is to solely understand the literal meaning of phrases, a big chunk of the speaker's intentions would be lost. We see this exemplified in translation errors. If a translator is not familiar with the nuances and metaphors of the language being translated, they might interpret it in a literal sense and end up conveying something that was not intended by the source language. Therefore, in order for the translator to do an adequate job, a cultural and semantic understating of both languages is essential. This is similar to mindreading in that meaning and intention are understood as a result of background knowledge and cultural conditioning. In addition, being able to efficiently mind read may require the ability to simulate other minds. Simulating other minds is one process out

of various others that may be utilized to understand, explain, and predict behaviors, emotions, and intentions.

There are four main approaches that are used to explain how mindreading functions: theory theory, simulation theory, modularity, and hybrid models. Theory theory basically states that an individual creates their own theory about others' mental states and tests it by observing whether what they had predicted has materialized or not. If that prediction is materialized, their theory is correct, and it is used in future efforts to predict, interpret, and explain. If, on the other hand, the theory turns out to be false, i.e., the prediction doesn't materialize, another theory would need to be devised and tested in its place (Goldman et al., 2005). This is akin to scientific hypotheses where tests are conducted of their validity. As for simulation theory, the individual trying to understand others' mental states goes through a process whereby their own background knowledge and experience is taken as the variable by which they would test their theory on the person being simulated. As opposed to theory theory, the simulator does not create a completely novel ToM. In this case, the simulator's biases and preconceived notions come into play and influence the outcome of the simulation. In short, the simulator puts themselves in the simulated person's shoes and tries to explain and predict what is happening and what will happen (Gordon, 1986). This entails an understanding of oneself as the starting point. Through this, it is possible to form predictions about others' behaviors. Our own mental states are the variables that are tested to come to conclusions regarding a certain situation where understanding the other person is of significance. In other words, the simulator reaches their conclusions by figuring out how they themselves would act under the same circumstances of the party being simulated. A third approach to mindreading is that of modularity or modular theories.

As first proposed by Fodor (1983), mindreading may be a function of a specific module of the mind. This would be one of many modules, each of which is responsible for a certain function. There is, consequently, a degree of autonomy between the modules, but interaction amongst some of them may be required for complex functions. There are different theories as to how much functions of the mind are modular. In other words, it may be the case that each function has its own module, or it may be that modules are only for the simplest of functions. Lastly, hybrid models explain mindreading through a combination of modularity and independent yet interconnected mechanisms. These models incorporate ideas from the first three approaches mentioned above (Nichols & Stich, 2003). Through a careful consideration of what certain processes such as mental and emotional attribution require, these models were gradually built to account for much of mindreading mechanisms.

As for ASD, it is a range of mental disorders mainly characterized by cognitive and communicative difficulties. Studying the reasons for these difficulties might give us some insight into how sufferers of ASD handle processing and understanding their own and others' emotions. The main impairment occurring with ASD that is significant to this study is that of reduced interoception. Interoception is generally defined as the ability to adequately comprehend one's own body (Craig, 2003). The lack of interoceptive qualities in people with ASD is a precursor to their reduced capacity to mind read through simulation. Since both understanding oneself and understanding others are impaired in individuals with ASD, there is a possibility that the absence of one may contribute to the absence of the other. There are some studies suggesting that another disorder co-occurring alongside ASD is the cause of the impaired interoception, namely Alexithymia (Hill, Berthoz, & Frith, 2004).

Whether the main cause of impaired interoception is ASD or Alexithymia is important since their frequent occurrence together could be grounds for investigating whether one or both are related to simulation theory. From there, I will also look at the difference in empathy between individuals with ASD and those without ASD. As a counter argument, Scheler's (1954) account of empathy will be presented. Scheler raises some concerns, but I think they are easily solved by understanding that Scheler's conception is not very distinct from what is called high-level simulation. This is one type of simulation, the other being low-level. These two types will be explained in more detail.

Finally, through showing empathy's function in a child's development and comparing that to how children with ASD lack certain aspects of ToM, I will conclude by claiming that simulation and empathy are correlated in that empathy is what we call the many emotional results of simulation. As mentioned above, social and cognitive deficiencies in people with ASD will be the factor looked at the most in arriving at that conclusion.

CHAPTER 2

ToM

In this chapter, I will explain how an individual develops a ToM. This will be achieved through first describing the different facets of a theory. An explanation of a ToM and its subtypes will follow. Lastly, and paving the way for the following chapter, I will discuss how a ToM is a precursor to self-awareness, mindreading, and mentalizing.

2.1 The structure and function of a theory

I will begin by demonstrating what a theory is by presenting Gopnik and Meltzoff's (1997) explanations of its structure and function.

2.1.1 Structural properties of a theory

Gopnik and Meltzoff (1997) defined abstractness, coherence, causality, and ontological commitment as four structural properties of theories. Abstractness can be thought of as a way of describing occurrences through techniques or claims that are not directly visible in the evidence being analyzed. Since the gravitational attraction between planets, for example, is not directly observable, physicists must develop conjectures and test their findings. Their theories could be judged adequate if the results of their hypotheses correctly anticipate the visible data — in this case, the paths of the planets. Because prediction will be a focus topic within the setting of ToM, I shall dwell on it later when addressing the functional character of theories. However, the previous example of planet trajectories demonstrates how abstractness

is a quality of theories that allows for conclusions that are not based on directly observable evidence.

Coherence is the second structural property of a theory, and it simply means that portions of a theory are interconnected and not completely independent of one another. In the past, only the deductive relationships between the entities in a theory were taken into account. Coherence is given greater weight in a more modern approach by examining relationships between the overarching theory and its evidence. These connections aren't necessarily tautological; as Gopnik and Meltzoff (1997) put it, "this more recent view, in fact, makes the coherent interrelations between parts of the theory even more important" (p. 35).

Causality refers to two features of a theory's entity relationships. The first is how the laws interact with one another; we frequently think of these laws as causally related. In evolutionary biology, for example, a causal relationship between selection and the retention of other mutations is suggested. We can see how, in evolutionary biology, the concept of selection comes before the preservation of mutations as a cause. The relationship between theoretical entities and evidence is the second factor to consider. What this relates to is our desire to generate theoretical constructions from seen evidence. The proof, on the other hand, is not the cause of the theoretical entity. The premise is that the hypothesis, or how we anticipate it presenting itself, is what leads to the evidence we observe.

Finally, there is the trait of ontological commitment within the structural nature of theories. Because the goal of a theory is to explain how something works in nature, it claims its counterfactuals. The flora and fauna of a given region of the earth, for example, might have evolved differently if it weren't for the climate. As a result, theories are tasked with not only predicting what exists but also predicting

what will exist in the future. As a result, theories are tasked with not only predicting what is, but also predicting what may be if the circumstances were different.

2.1.2 Functions of a theory

A theory has three functions, according to Gopnik and Meltzoff (1997): prediction, interpretation, and explanation. Prediction is a functional quality of a theory as a result of the structural qualities described above. When a theory is proposed, the goal is not to simply explain the observed situation. Theorizing is a technique for moving from a specific situation to a more generalized understanding of the topic under investigation. A theory that allows for the prediction of a planet's movement from its surrounding celestial bodies is formed by observing a large amount of evidence that, when compared and analyzed, leads researchers to infer a general relationship between the observed phenomena and the end result, in this case the movement of planets. Researchers can extrapolate from their original findings by examining if their theory holds true in a situation that was not part of the original study that led to that notion. The theory is deemed to be successful, at least for the time being, if their prognosis proves to be right. I say "for the time being" because theories are not flawless, and it is possible that theorists did not account for all circumstances. If this is the case, the theory will be disproven (Popper, 1934), and it will need to be changed to make more accurate predictions. Prediction is a crucial idea in ToM, so I will return to it later.

The second functional element, interpretation, has been extensively addressed, primarily by Kuhn (1962), as well as its implications for the scientific process. Scientists could find out whether evidence is worth looking at as a potential falsifier of a theory by interpreting it. The theory, on the other hand, is mostly used

as a jumping off point for interpreting data. This indicates that evidence could be analyzed to maintain the current theory. Although this may appear dogmatic at first glance, such an approach to constructing hypotheses and analyzing evidence aids in the exclusion of evidence that would be considered noise within the larger framework of a theory. In some circumstances, dramatically changing a theory might lead to the abandonment of other hypotheses that were based on it. Scientists try to test well-established theories numerous times before discarding them.

Finally, within the realm of a theory's functional aspects is explanation. According to Gopnik and Meltzoff (1997), the structural elements listed above are what give a theory its explanatory power. She even goes so far as to say that the only way to explain something is to come up with a hypothesis about it. We need to set the tone by grasping the concept of a ToM. This understanding will be necessary for conveying and connecting the themes addressed throughout this thesis. According to Goldman (2012), ToM is the act of attributing mental states to others and making predictions about their feelings or ideas as a result. This enhances human social interaction because people react to predictions in accordance with other expectations such as societal traditions. In defining ToM, Gopnik, Meltzoff, and Kuhl (1999) argues that it should be named "theoretical ToM". It entails psychological thinking about our everyday, intuitive, folk mental knowledge. ToM can be both explicit and implicit. Having an explicit ToM entails being conscious of the judgment being made about a target's thoughts and desires. This could be assessed through a request for verbal reflection of one's thoughts about said target. On the other hand, when one shows implicit ToM behavior, they typically anticipate the target's thoughts and desires without necessarily forming a theory about it in a deliberate manner. This could be apparent through eye movement (Schneider, Slaughter, Bayliss, & Dux,

2013). In the next section, I will describe the different subtypes of ToM. It is important to note that each of these subtypes can exhibit either an implicit or an explicit nature.

2.2 Subtypes of ToM

There are four main approaches that have been developed to explain how mindreading works. These are "Theory theory", "Simulation theory", “Modularity”, and “Hybrid models” (Carruthurs, 2011).

2.2.1 Theory theory

The term "Theory theory" was coined by Adam Morton (1980) to describe that commonsense thinking is made up of ideas that obey specific types of relationships. However, Gopnik and Wellman’s (1992) work extended on this understanding, claiming that human thoughts are similar to scientific hypotheses. To put it another way, when we try to understand or predict the thoughts or acts of others, we develop a theory that is then tested by their actions. Our idea would be proven right if what we had predicted matches their actions. We would then use that theory in future interactions until we need a new one. In this approach, we can develop our ToM in the same way that scientists improve their understanding of the physical world.

A commonplace psychological "theory" guides attributions of mental states to both self and others. In its most basic form, a theory is a set of propositions that includes (potential) laws or generalizations. A psychological theory is thus a collection of propositions containing psychological generalizations. Such generalizations must be intrapersonal, diachronic generalizations, explaining the transitions that a given psychological system will make from some beginning states

to successor states if they are to be useful in forecasting others' mental states. Andrews (2012) discusses the idea of model-based theories whereby idealized models are used in place of universal generalizations. A theoretical model is put in place as an explanation or prediction of a target's mental state. Models would then be manipulated if deemed insufficient at explaining or predicting. This manipulation is the core of theory theory since it is the tool by which others' mental states could be accurately understood, explained, and predicted. A mental attributor, according to theory theory, is like a scientist who analyzes other people's minds—and her own—in the same way she approaches any system (Goldman et al., 2005). She generates opinions, or possibly probability judgements, regarding the target system's current state or condition. She takes some psychological generalizations or theoretical models from her knowledge base and applies them to infer future or previous states of the system. Physical states of the system and generalizations or models about such states are the contents of a theorizer's views in physics. The contents of the attributor's beliefs in mindreading are mental states and generalizations or models about mental states. Another characteristic of the theory theory approach is the learning of commonsense theory. According to theory theory, infant learning is similar to scientific learning in that it involves testing old beliefs against new information and occasionally establishing new theories to replace old ones (Gopnik & Meltzoff, 1997).

2.2.2 Simulation theory

Simulation theory claims that all we need to know about our own minds is how they work, and then we can analyze others' actions based on that knowledge (Doherty, 2009). This works by placing ourselves in the shoes of the simulated individual.

Gordon explained this by suggesting that when we try to predict the conduct of others, we engage in a type of pretend play (Gordon, 1986). The essential difference from theory theory may be seen here. On the one hand, theory theory claims that our understanding of others is largely dependent on trial and error, with our theories evolving as we make errors in anticipating others' thoughts and actions. Simulation theory, on the other hand, presupposes a ToM in which we may rely solely on our knowledge of ourselves.

Three basic steps comprise a standard simulationist mind-reading technique. First, the attributor produces false states in herself that are meant to resemble those of the target. In other words, the attributor tries to put herself in the shoes of the target (Gordon, 1986). Perceptions, wants, beliefs, hopes, goals, sensations, and emotions are among the mental states that can be feigned. The second stage is to feed these initial pretend states into a decision-making or emotion-generating mechanism in the attributor's own psychology, and then allow that mechanism to work on the pretend states to generate one or more new states. For example, if the attributor wants to predict a target's decision, she could make up fake desires and beliefs – which she assumes the target has – and use her decision-making system to make a feigned decision. Third, the attributor assigns the target's output state as a state that the target will go through or has already undergone. The most comprehensive simulation heuristic is this three-step procedure. However, there may be shorter variants, such as a two-step method. In this case, the attributor creates a fake state and just imputes it to the target without passing it through any processing mechanisms. Clearly, the core notion of simulation theory is that pretension and attempted replication are used to facilitate mindreading. A mind reader assumes the target's mental "position" and copies or tries to mimic mental activity that

corresponds to that posture. Gordon (2004) later adopted a more radical approach, as Gallagher (2007) puts it, by stating that simulation is the mechanism that allows us to recognize that other persons have minds. On the other hand, Gallese's (2001) view of simulation is a slightly more complex one as it consists of three levels:

- The phenomenological level is the one responsible for the sense of similarity ... that we experience anytime we confront ourselves with other human beings. It could be defined also as the empathic level ....
- The functional level can be characterized in terms of simulation routines, as if processes enabling models of others to be created.
- The subpersonal level is instantiated as the result of the activity of a series of mirror matching neural circuits (pp. 45).

Gallagher (2007) attempts to object to both these accounts by utilizing a phenomenological understanding of human-human interaction. Regarding Gordon (2004), he claims that simulation can sometimes fail, which in turn leads to the simulator adopting a theory theory-based approach. As for Gallese (2001), the focus on neurological mechanisms, as seen in that of mirror neurons (described later in section 2.3), may come in conflict with the conscious nature of phenomenological understanding of others.

2.2.3 Modular theories

Modularity proposes that our minds are composed of modules, some of which are responsible for mindreading. Consequently, some, such as Baron-Cohen, Leslie, and Frith (1985), have claimed that autism is the result of damage to one such module. Fodor (1983) proposed the idea of modularity along with some characteristics that should be present for a system to count as a module. Having a module suggests that mindreading is a cognitive faculty that has a certain degree of autonomy from other faculties. This does not, however, make it completely independent. I think modules may be linked when it comes to complex functions such as language learning.

The question, though, is the degree to which cognitive faculties are separated into modules. Peripheral-systems modularity proposes that modules are used in a minimal fashion to assist with cognitive processes (Fodor, 1983). Fodor further claims that modules can be innate. This means that faculties developed from these modules are universal regardless of learning patterns. They are triggered at a certain age due to general environmental factors. At the other end of the spectrum, massive modularity suggests that all our mind is made of modules (Tooby & Cosmides, 1992). These modules are used for almost all cognitive processes from decision making to understanding desires. This would suggest, however, that there would not be any "general learning" involved (Gottschling, 2019).

2.2.4 Hybrid models

Lastly, hybrid models suggest that there are mechanisms that have characteristics of the previous three. These mechanisms can work together to form a complex system of modules and mechanisms that when incorporated together help in mindreading (Nichols & Stich, 2003). A theory of mindreading was proposed by Nichols and Stich (2003) whereby various mechanisms are split into modules. This theory developed through multiple stages where mechanisms were added to account for a multitude of human mindreading capacities. The first set of mechanisms consists of desire detection, planning, and mindreading coordination. Desire attribution is meant by desire detection in this case. It is suggested by Nichols and Stich (2003) that the ability to attribute desires is innate, though it is improved through learning. The planning mechanism's role is to decide how to achieve a certain goal. After desire attribution, a decision would normally be made taking these attributed desires into consideration. These two mechanisms need the third mechanism as an intermediary.

As a coordinator, it utilizes information collected about the target to come up with an adequate answer to the questions raised as a result of desire attribution. Nichols and Stich (2003) then move on to describing how what they call the “possible worlds box” is the main section within their proposed mindreading model that helps in belief attribution. This works in tandem with an “updating” mechanism that removes representations that are deemed incompatible with a certain belief. Lastly, inference mechanisms apply the appropriate changes to the possible worlds box. These mechanisms interface beliefs and possible worlds in a kind of back-and-forth fashion and help with inferring beliefs that are the most coherent with available information. This would lead to more robust belief attribution. All these modules and mechanisms mentioned above form a complex hybrid system in which belief and desire attribution systems interface with inference and decision-making mechanisms while utilizing procedures and strategies used in both theory theory and simulation theory.

2.2.5 Discussion

After briefly describing the different theories of ToM, I now move to a discussion of their key differences in relation to the false belief task which will be discussed in greater detail in section 2.5.1. Consequently, I will conclude that simulation theory is the most appropriate to study regarding belief attribution and empathy in ASD.

The main mechanism within simulation theory is the attributor's endeavor to imagine the target's mental states, including the initial mental pretense, through a type of theorizing that stems from pretending that one is in the target’s place. This is an important aspect of mindreading for simulation, but it has no place in theory theory. An attributor, according to theory theory, solely employs descriptions of the target's states and psychological regularities. The attributor does not attempt to

mentally imitate or impersonate the target by dressing oneself in such conditions. The attributor's whole processing is purely inferential, moving from beliefs to other beliefs about the target's states. Attributors do not need to utilize mental pretending, according to theory theory (Goldman, 2006). Analyzing a false belief task entails thinking about another person's mental state. Gordon and Cruz (2004) attempted to explain why a lack of robust simulation capabilities in the child may be the reason for failing a false belief task. They claim that the child would have to understand that there are two contradicting premises in this situation, one is the objective truth, and the other is the false, subjective belief of the target. Since the child is not able to reconcile these contradictions, it could point to undeveloped simulation capabilities. Had the child been able to put themselves in the target's shows, concluding that the subjective belief is the correct answer should be relatively easy. In other words, this is achieved through theorizing what one might themselves think or do in that situation. The individual mentalizing would put themselves in the mentalized shoes. This is a basic form of simulation whereby one's background knowledge and ideas come into play when understanding or predicting other's feelings, thoughts, desires, or intentions. Children with ASD tend to fail the false belief task which suggests a lack of simulation capabilities. I agree with Gordon and Cruz (2004) since this offers a reasonable explanation as to why a child under a certain age fails the false belief task.

A low success rate on the false belief task might suggest a deficit in a certain module within a mindreading system. It could be theorized that a module is responsible for belief attribution, and consequently, false belief attribution. Since children do not pass the false belief task until they reach around four years of age, it is possible that development of inference modules comes at around that age. It could

also be the case that false belief attribution is handled by a module independent of the one which handles true belief attribution.

Ascribing false belief attribution to a hybrid model seems more appropriate than a simpler modular model since it may account for the age difference between children who usually pass the false belief task and those who do not. A mechanism that is used to interface between belief attribution and understanding false belief could be at play as a means to infer that a subject has a false belief. This module could have a function similar to simulation. Since beliefs should be attributed to others, the mechanism within this module could be utilizing a method of analysis and attribution that uses simulation as a first step before belief attribution. For this reason and those mentioned above, I want to focus on simulation theory in relation to mentalizing in ASD.

2.3 High-level simulation vs. low-level simulation

Human simulation capabilities are split into two categories by Goldman (2006): low-level simulation and high-level simulation. In a nutshell, low-level simulation is an automatic response such as mimicking someone's facial expressions without giving it much thought, but high-level simulation incorporates some background information as well as imagining scenarios.

Low-level simulation is sometimes associated with mirror neurons. They are called as such because of their “mirroring” function. What this basically means is that they trigger a certain reaction in an individual upon witnessing or sensing a situation where someone else has an experience that would otherwise have triggered that reaction were it experienced by the first individual in question. To further elucidate the matter, take someone who witnesses a person fall and break their arm.

Upon seeing this, they are likely to cringe, i.e., squint their eyes and inhale quickly, mimicking how they would have reacted had they had the same or a similar experience (Goldman, 2006). This involuntary reaction suggests that the simulating side may not be aware that they are in fact performing a simulation. Since there is a lack of voluntary action that precedes simulation, the degree to which one can engage in low-level simulation may not be an appropriate measure of one's mindreading capabilities. It may, however, be a basis for it (Goldman, 2006).

As for high-level simulation, it operates on a more conscious level in that the simulating individual goes through a mental process of putting herself in the other person's shoes. According to Goldman (2006) high-level mindreading has the following characteristics:

> (a) it targets mental states of a relatively complex nature, such as propositional attitudes; (b) some components of the mindreading process are subject to voluntary control; and (c) the process has some degree of accessibility to consciousness. (pp. 147)

The difference here from low-level simulation is that the more the simulating person is experienced with regards to the situation at hand, i.e., she knows much about the background of the situation she is simulating, the more she will be successful in her prediction of the simulated person's actions or thoughts. What follows is that there would be a margin of error inversely proportional to that person's knowledge and experience with that situation. One's prejudices and preconceived notions may come into play and affect their judgement. Consequently, this may lead to unfavorable results. Two skills are needed to help avoid such errors as much as possible. The first is that of inhibition wherein the simulating individual inhibits some of her own desires and preconceived ideas that would probably come into conflict with what is being simulated (Goldman & Jordan, 2013, p. 452). The other ability is that of imagination. Since as previously mentioned, one's background knowledge plays an

important role in shaping their simulation's outcome, imagining how the other person would react in a certain scenario could make up for gaps in the simulator's knowledge.

Neuroscientists have discovered a wide range of mirroring mechanisms in which cognitive states of one creature are mirrored by similar or comparable cognitive states in an observer (Goldman, 2006). These are instances of interpersonal mental simulation. Wicker et al. (2003) cite disgust as an example of this type of simulational phenomenon. A mirrored event is a prospective launching platform for mindreading. Goldman (2006) supports this by demonstrating that when both mirroring and mindreading are present together, low level simulation occurs as well.

2.4 Mindreading and mind blindness

Firth (2001) explains the concept of mindreading in contrast to mind blindness by stating how individuals with ASD find it incomprehensible how those without ASD easily understand and predict others' mental states, desires, and intentions. She gives an example of how when someone is bending over a filing cabinet, someone observing them would most probably immediately understand that they have misplaced a paper. Even more, they might suggest that the person searching look somewhere else without saying anything prior to that suggestion. This phenomenon shows that people without ASD can take their target's mental state into consideration and infer what they desire without much effort. This skill is what is lacking in individuals with ASD, which is why they are said to suffer from mind blindness.

The mind blindness hypothesis was first tested by Baron-Cohen et al. (1985) and Baron-Cohen, Leslie, and Frith (1986). Leslie (1987) suggested that children with ASD should not be able to represent mental states such as beliefs. The reason

for this is that autism's social impairment is the result of a deficit in the mentalizing mechanism. Although they may have reached the proper level of verbal and cognitive development, it should be difficult for them to understand or predict actions pertaining to others' beliefs. The test was based on a false belief task developed by Wimmer and Perner (1983), who found that normally developing children aged four and up passed it (Frith, 2001). I will further elaborate on the false belief task in relation to autism in later sections.

Mindreading as an ability is developed early on around 18 months into a child's life. This is seen in the young child's enjoyment of pretense (Leslie, 1987). Since mental states could involve representation, representing others' mental states could be a kind of meta-representation. Pretense could involve meta-representation as well by using an object to represent a different kind of object. In this sense, Leslie gives the example of understanding that a mother is representing a phone by holding a banana to her ear in front of her child. He further explains that if such a mental capacity were absent, there would be major difficulties regarding intentionality, and this would consequently result in mind blindness. This means that in the absence of intentionality, the only way to understand a proposition would be by what it literally represents. In the example above, a banana is nothing more than a banana. It could not represent a phone or anything else. Baird et al. (2000) used three indicators for diagnosing children with autism: failing to follow another's gaze, failing to point at or show objects of interest, and failing to understand make-believe play. This was also studied in infants aged 18 months by Baron-Cohen et al. (1996).

2.5 Testing for ToM

Perner and Leekam (2008) discussed three experiments conducted with children with ASD, taking children without ASD as the control.

2.5.1 False-belief task

The first is a false-belief task. The false-belief task illustrates how children under a certain age, or those with autism fail to understand false beliefs in others. In this test, a child is observing a room where two adults are present. Adult A places a toy in a basket and leaves the room. Adult B then removes that toy from the basket and places it in a box without Adult A's knowledge. Adult A then returns to look for the toy. The child is then asked where they think Adult A will look. As stated above, this task and similar ones are comprehended by children ages 4 or above. Those suffering from ASD, however, were unable to specify that Adult A will look in the basket, i.e., the place where she believes the toy to be in. Baron-Cohen, Tager-Flusber, and Cohen (1999) corroborate this result by stating that children with ASD can usually comprehend a false belief task at the age of 8, as opposed to non-ASD children comprehending it at the age of four.

It is important to note that the task described above is an explicit false-belief task since the child is asked to explicitly point out the answer, in this case verbally. There is another kind of experiment related to this that employs an implicit false-belief task whereby the child's gaze is monitored as opposed to asking them to say what they think is the correct answer (Schneider, Slaughter, & Dux, 2015). Such tests have shown that neurotypical children who usually fail the explicit task demonstrate eye-gaze patterns that are consistent with implicit false-belief processing (Clements & Perner, 1994). Furthermore, adults with ASD fail to show

implicit false-belief gaze patterns while demonstrating explicit patterns (Senju, Southgate, White, & Frith, 2009).

2.5.2 False-photo task

The second experiment is the false photo task (Zaitchik, 1990). A woman in a red dress has her photo taken. She then changes into a green dress. Although she is now wearing a green dress, she was wearing a red one at the time the photo was taken, which is shown in the photo. The child is then asked what the woman in the photo is wearing. This seems to be similar to the false belief task in that the child is asked to point out a feature of a real-world object where in both cases said feature is shown differently whether in a mental state or a physical representation (photograph). Due to this similarity, a similar result is expected, i.e., for children with ASD to fail the false photo task as they failed the false belief task. However, children with ASD tend to answer the question of the false photo task correctly as opposed to children without ASD who struggle with it. This could indicate that children with ASD have a deficit that is specific to mindreading, i.e., they have a deficit when it comes to understanding and/or predicting other's mental states.

2.5.3 False sign task

The third experiment involves a false sign task whereby children are presented with images of an ice cream truck behind a church, while a sign indicates that it is behind a house opposite to it. They are then asked where someone passing by would think the truck is. Similar to the false belief task, children with ASD had difficulty answering this question correctly. Another version of the false sign task had a train go pick up cargo from a landing strip. When the plane changed signals to go in a

different direction, the train still had the old signal to go to the original location. The results were similar to the former false sign task in that children with ASD thought the train would go to the new location. Perner and Leekam (2008) suggest that this is a case of a nonmental false sign task. Subsequently, this suggests that the deficit may not be related to understanding other's mental states. I believe this is in fact a mental false sign task, and the fact that vehicles are being used to explain the situation does not make it nonmental. The child with ASD may think of the vehicles as being driven by individuals with mental states. This is why, in my opinion, there is still some evidence supporting the idea that the deficit in children with ASD is mainly related to mental states.

This further correlates with what I mentioned earlier where ASD leads to an impairment in one's mentalizing capabilities. Failing false-belief tasks suggests that simulating and predicting is impaired in children with ASD and that one cause could be an impairment of self-awareness. Through a discussion on empathy in relation to ASD, the link between simulation theory and empathy will become clearer as the former being the process leading to results perceived to be the latter. This will be elaborated upon in section 4.2.

2.6 Mentalizing in children without ASD

Children in their first year of life immediately follow another person's gaze as if they are paying attention to the other person's main interest. Other indicators of mentalizing accompany shared attention. In my opinion, shared attention could be an indicator of a child's capacity to mentalize since it might be more than mere imitation. The child could be trying to understand what the other person is gazing at. Moreover, a child with reduced shared attention might not be interested in what

others are interested in or looking at. Referential looking, for example, is when children examine their mother's expression toward a novel thing before approaching or avoiding it (Repacholi, 1998). As I explained in section 2.5.1, individuals with ASD have difficulty directing their gaze. The deficiency in implicit false-belief processing in individuals with ASD stays with them well into adulthood. Not being able to implicitly predict mental states of others corroborates the problem with directing one's gaze where it is needed. As such, this might be evidence of deficiency in mentalizing and mindreading. Another indicator of the inevitable advancement of a mentalizing capacity is the ability to copy complicated and arbitrary but intentional acts of others – as opposed to their accidental activities – which is obtained in the middle of the second year of life (Meltzoff, 1995). According to Bloom and German (2000), mentalizing plays an important role in helping children understand the meanings of words. As a result, children do not acquire words simply by associating the sound of the word with the item in front of them. Frith (2001) explains that because the speaker and listener may be looking at different items, such an association is inherently imprecise and error prone. Children instead learn through following the speaker's referential intention by observing the speaker's gaze (Baldwin et al. 1996). Concepts like false belief or deception are usually understood by children between five and eight years of age.

CHAPTER 3

ASD

3.1 The nature and diagnosis of autism

Regardless of culture, color, ethnicity, or socioeconomic status, people with ASD have essential traits in two areas: social communication and confined, repetitive sensory-motor behaviors (Khan et al., 2012). ASD is caused by a disruption in early brain development and neuronal reorganization (Bauman & Kemper, 2005; O'reily, Lewis, & Elsabbagh, 2017). However, because no reliable biomarkers exist, the diagnosis must be determined based on behavior. The Diagnostic and Statistical Manual of Mental Disorders (DSM)-5 criteria issued by the American Psychiatric Association (2013) was designed to make diagnosis of mental disorders easier. The two domains, namely social communication and restricted, repetitive, or atypical sensory-motor behaviors, have now been combined into a single ASD spectrum. Clinically unreliable subtypes such as Asperger's disorder and pervasive developmental disorder, two disorders previously not categorized under ASD, have recently been merged under the single diagnosis of ASD. Furthermore, the DSM-5 recognizes that ASD can be accompanied by other illnesses, such as fragile X syndrome and attention-deficit hyperactivity disorder. To be diagnosed with ASD, a person must show or have shown problems in two of four restricted, repetitive sensory-motor behaviors and must have or have had difficulty in each of three social communication areas. I think it would be better to list them exactly as they appear in DSM-5.

For the social communication areas, the DSM-5 (2013) states the following:

1. Deficits in social-emotional reciprocity, ranging, for example, from abnormal social approach and failure of normal back-and-forth conversation; to reduced sharing of interests, emotions, or affect; to failure to initiate or respond to social interactions.
2. Deficits in non-verbal communicative behaviors used for social interaction, ranging, for example, from poorly integrated verbal and nonverbal communication; to abnormalities in eye contact and body language or deficits in understanding and use of gestures; to a total lack of facial expressions and nonverbal communication.
3. Deficits in developing, maintaining, and understanding relationships, ranging, for ex- ample, from difficulties adjusting behavior to suit various social contexts; to difficulties in sharing imaginative play or in making friends; to absence of interest in peers. (pp. 50)

As for the repetitive pattern of behavior, DSM-5 (2013) states the following:

1. Stereotyped or repetitive motor movements, use of objects, or speech (e.g., simple motor stereotypies, lining up toys or flipping objects, echolalia, idiosyncratic phrases).
2. Insistence on sameness, inflexible adherence to routines, or ritualized patterns of verbal or nonverbal behavior (e.g., extreme distress at small changes, difficulties with transitions, rigid thinking patterns, greeting rituals, need to take same route or eat same food every day).
3. Highly restricted, fixated interests that are abnormal in intensity or focus (e.g., strong attachment to or preoccupation with unusual objects, excessively circum- scribed or perseverative interests).
4. Hyper- or hyporeactivity to sensory input or unusual interest in sensory aspects of the environment (e.g., apparent indifference to pain/temperature, adverse response to specific sounds or textures, excessive smelling or touching of objects, visual fascination with lights or movement). (pp. 50)

There are also additional proposed severity levels in DSM-5 (2013). From this as well as other studies, we can see that autism is now commonly recognized as a neurodevelopmental condition. Happé and Frith (1996), as an example, agree with this. It is a lifelong condition. It varies in severity and can affect people of all ages and abilities. Thus, it is now widely recognized that autistic disorders fall on a spectrum. Social relatedness and verbal and nonverbal communication are severely limited in people with ASD. They are generally aloof as children, and even after learning the basic norms of social interaction, they remain egocentric. They may

have delayed speech or lack speech entirely before a certain age. Some might grow to become verbally competent, but they too may have difficulty comprehending what others are saying. Other common characteristics of people with autism include restricted interests, motor stereotypies, and compulsive tendencies. They may have exceptional rote memory and savant abilities (Frith, 2001). The most important of these deficiencies in the etiology of autism is a modest but fatal deficiency in human social understanding and interaction, which is the focus of this thesis. The mind blindness hypothesis, as described above, is a term used to describe this situation (Baron-Cohen, 1997).

Autism is an illness that impairs a variety of cognitive skills, but it does not indicate a general deficiency in the ability to process information (Scheuffgen, Happé, Anderson, & Frith, 2000). While a failure of social communication is a hallmark of the illness, this does not imply a general lack of social abilities. Rather, one or more specific, i.e., restricted, cognitive abnormalities appear to be the etiology of autism. Simultaneously, such deficiencies would have consequences for general functioning (Frith & Happé, 1998). This is consistent with current beliefs concerning intrinsic domain-specific mechanisms with a limited neural base (Black, 1998). The most important of these deficiencies in the origins of autism is a loss in human social awareness.

3.2 Self-awareness in individuals with ASD

I will present in this section how being self-aware, i.e., understanding one's own mental states, is a precursor to understanding others' mental states. Mindreading is a quick, and sometimes unconscious, technique for normally developed humans. To make sense of other people's conduct or behavior, we frequently assign their

thoughts, feelings, and perceptions to them. Individuals with ASD are known to have difficulties detecting mental states in others, particularly epistemic states such as belief and knowledge (Williams, 2010). Self-awareness refers to the act of becoming aware of one's own mental states including perceptions, sensations, attitudes, intentions, and emotions (Morin, 2004). There are four levels of consciousness according to Morin (2006): unconsciousness, consciousness, self-awareness, and meta-self-awareness. Individuals with ASD are among various clinical groups that have been studied regarding self-awareness. As previously stated, people with ASD generally have difficulties with language and communication, reciprocal social interaction, repetitive behaviors, intense interests, and sensory dysfunction (American Psychiatric Association, 2013). A person's self-awareness is unique to his or her experience of the disorder due to its heterogeneous nature (Elmose, 2016).

Elmose (2016) defines a theoretical basis for understanding self-awareness in those with ASD. She states that:

> the sense that you do not know what you do not know and therefore have difficulty judging when it would be relevant to get more information; difficulty distinguishing between your own or others' preferences and moods when you are together with them; perceiving your own actions (e.g., not being able to get out of the door in time for school, self- destructive actions) as "freestanding" actions without any link to antecedents, current situation, others' reactions, or your own thoughts or feelings. (pp. 109)

Accordingly, the idea of self-awareness within individuals with ASD is not consistent, and it varies from one person to another as there are differences in the level of self-awareness. Williams (2010) discusses an experiment whereby children with ASD are presented with a box of Smarties and asked what it contains. The box does not, in fact, contain Smarties, but a pencil. After checking the contents of the box, children with ASD are now aware of what the box contains. They are then asked what they thought was in the box before they checked its contents. Their

answers were not consistent with those of children who answered questions relating to the false-belief task, i.e., they answered correctly. Williams suggests that this could be due to the children's memory of what they said rather than what they thought. This is supported by Baron-Cohen (1991).

Using a different methodology, an investigation by Williams and Happé (2010) sought to test whether children with ASD represent their own intentions. A knee-jerk task was given to participants with ASD and others without ASD in which they were asked if they planned to move their leg after a knee reflex was stimulated. Perner (1991) hypothesized that a meta-representational ToM is required to appropriately recognize one's own reflexes as unintended. Lang and Perner (2002) observed that performance on a false-belief test among early normally developing children was substantially linked with performance on the knee-jerk task. Williams and Happé (2010) discovered that children with ASD performed much worse on this knee-jerk test than age and ability-matched comparison participants, asserting falsely more frequently that their reflex response was under their conscious control. Furthermore, regardless of age or language ability, their performance on this test was significantly linked to their performance on false-belief tasks. These findings corroborate the theory that people with ASD have a reduced awareness of their own goals. Since many children with ASD also fail basic false-belief tasks, I believe this is grounds to draw a connection between self-awareness and recognizing others' mental states. In other words, the former leads to the latter.

Hurlburt, Happé, and Frith (1994) used Hurlburt's (1990) introspective sampling method to evaluate reports of interior experiences among three high-functioning, intellectual persons with ASD. Each participant wore a gadget that beeped at random intervals throughout the day. The respondent paused what they

were doing each time the device beeped and wrote down exactly what they were thinking just before the beep. The initial discovery of Hurlburt et al. (1994) was the link between ToM and the quality of the introspective report provided. While the lone person who consistently completed advanced ToM tasks found describing their internal sensations to be reasonably simple, the participant with the worst ToM skills was utterly unable to report what he was thinking when the gadget beeped. This is further evidence of the relation between ToM and self-awareness. It is possible for one to be a result of the other. However, we cannot be certain that ASD is the main cause of reduced ToM capabilities. There is another condition called Alexithymia that frequently co-occurs with ASD that might be the culprit.

3.3 Alexithymia

Alexithymia is a personality construct whereby the alexithymic individual is unable to identify emotions in themselves (Sifneos, 1973). More recently, it has come to describe a range of emotion processing deficits (Parker, Keefer, Taylor, & Bagby, 2008). It has a comorbidity with other disorders, most notably ASD. Although the prevalence of ASD within the general population is somewhat low at around %10 (Salminen, Saarijärvi, Äärelä, Toikka, & Kauhanen, 1999), its co-occurrence with ASD is at around %50 (Hill et al., 2004). For the purposes of this thesis, alexithymia is usually associated with impaired interoception (Herbert, Herbert, & Pollatos, 2011). Fitzgerlad and Bellgrove (2006) state the following:

> James Grotstein (in Disorders of Affect Regulation by Taylor et al., 1997) describes Alexithymia as ‘‘an affect processing disorder that interrupts or seriously interferes with the organisms self-organising and reorganising processes’’ (page 12). This means that they have a diffuse sense of self. (pp. 573)

Alexithymia is, therefore, the mental disorder, while impaired interoception is the tangible result of that disorder when analyzed conceptually and practically. Krystal (1998) mentions that ASD patients have a cognitive style in which there is an "… absence of the human quality [which] contributes to making these patients' thoughts 'operational' or 'thing oriented'" (p. 247). They can acquire roughly socially suitable responses, but they are frequently overly formal. These utterances "… lack nuance, lack proper prosody, and are often very formal, and pedantic" (Fitzgerald & Bellgrove, 2006, p. 574). They come off as strange or unduly formal to the listener. Autism spectrum illnesses, such as Asperger's disorder, are known to include speech and prosodic irregularities (Shriberg et al., 2001). Alexithymic patients, according to McDougall (1978), utilize speaking as "an act rather than a symbolic means of communicating ideas or affect" (p. 45). Carruthers (2011) proposes that emotional responses of non-alexithymic individuals to perceiving emotions in others pass through an intermediate step of introspection. One would become aware of a similar emotion in themselves before completely comprehending the emotion in the other. The lack of this intermediate step in alexithymic individuals could be the reason for their lack of understanding of others' emotions. An interoceptive capacity is required to be able to understand one's own emotions clearly, something that is lacking in people with alexithymia.

Carruthers (2011) concludes that alexithymic people's lack of understanding of others' emotions, the example he gives being pain response, could be the result of a lack of empathic responses in themselves. As explained by Feldmenhall, Dalgleish, and Mobbs (2013), alexithymia is defined by a marked impairment in emotional awareness (Sifneos, 1973), social attachment (Vanheule, Desmet, Meganck, & Bogaerts, 2007), and interpersonal connecting (Berthoz et al., 2002). As a result,

alexithymics struggle to recognize and appreciate the feelings of others (Taylor, Bagby, & Parker, 1997), which is thought to lead to unempathic and inefficient emotional responses (Bernhardt & Singer, 2012; Taylor et al., 1997).

According to Valdespino, Antezana, Ghane, and Richey (2017), behavioral data demonstrate that alexithymia is linked to and causes empathetic impairments. Silani et al. (2008) discovered links between alexithymia and empathy in both healthy and sick people. “Links between alexithymia and empathy have also been found in the context of empathizing based on facial expressions” (p. 2). In a study by Moriguchi et al. (2007), participants were asked to rate how much pain they thought people were experiencing based on their facial expressions. Those with high alexithymia rated the pain of the observed person lower. This indicates that perceiving one’s own internal state and the state of others could be directly correlated. There is evidence which suggests that children with ASD have more difficulties with cognitive empathy, i.e., understanding and adopting others' emotional perspectives, than they do with affective empathy, i.e., feeling others' emotional states (Deschamps, Been, & Matthys, 2014). In relation to this, ASD has been linked to a lack of empathy by Baron-Cohen and Wheelwright, (2004). Bird et al. (2010) tried to establish a link between alexithymia and empathy deficits in ASD. They tested for insula activation in those watching others experiencing pain. They relied on this to detect or measure the degree of empathy. During the empathy condition, alexithymia was inversely linked to insula activation. The authors concluded that alexithymia, not the severity of ASD, is what causes empathic deficiencies. Since, according to Hill et al. (2004), alexithymia rates are disproportionately higher in people with ASD, alexithymia could be a key factor in explaining empathy deficiencies in this population (Valdespino et al., 2017).

The insula was similarly linked to alexithymia and reduced empathy in ASD, according to Silani et al. (2008). The anterior insula was also activated when rating the pleasantness or unpleasantness of negative emotional images. Overall, these data show that alexithymia with its related insula processing may be linked to lower empathy in people with ASD. Furthermore, the findings imply that the cognitive subtype may be used to distinguish a subgroup of ASD patients with significant alexithymia.

I will now move on to a discussion of the link between empathy and ToM. After discussing some different conceptions of empathy, I will present my argument for the relation between empathy and simulation mentioned above. A counter argument based on Scheler's views will be presented, which will be addressed in the final section of the next chapter.

CHAPTER 4

THE RELATIONSHIP BETWEEN EMPATHY AND SIMULATION THEORY

In order to draw a relationship between simulation theory and empathy, I will first present a few definitions of empathy and try to consolidate them into what I believe represents an account for the necessary conditions of empathy. The cases that we usually ascribe to empathy share three aspects, namely similarity of emotion, other-directedness, and activation after perception in the other. Later on, I will state how empathy is related to simulation theory. I believe that simulation is a cognitive mechanism which exhibits tangible or perceptible results that are often described as empathy.

4.1 Some definitions of empathy

I will start by presenting a few different definitions of empathy from various researchers in order to have a general view of how the concept of empathy has been viewed previously and today. Not all definitions of empathy from the literature will be presented, but the ones that will be discussed serve as an overview of how definitions of empathy have changed over time. Starting with Guiora (1965), he defined empathy as follows:

> Empathy is a process of comprehending in which a temporary fusion of self/object boundaries, as in the earliest pattern of object relations, permits an immediate emotional apprehension of the affective experience of the other, this sensing being used by cognitive functions to gain understanding of the other. (pp. 782)

Also, according to Guiora, Taylor, and Baldwin (1968), empathy necessitates a reaction to little emotional cues. The responses to such stimuli are frequently swift. They don't demand the receiver's imagination and are likely to require less

background information to support them. It is possible that more empathic people react to a wider range of events and cues from people from various cultures who would act differently. From this definition, we can assume that people who have a better understanding of others' emotions and try to effectively understand those they interact with the most – whether voluntarily or involuntarily – may be more motivated to continuously improve their communication skills with others than most people. To put it another way, empathy motivates people to want to understand others, which may lead to them taking actions to do so.

Heidi Maibom listed some definitions of empathy in her book Empathy (2020), one of which is by Preston and de Waal (2002):

> A perception-action model of empathy specifically states that attended perception of the object's state automatically activates the subject's representations of the state, situation, and object, and that activation of these representations automatically primes or generates the associated autonomic and somatic responses, unless inhibited. (pp. 4)

A key word here is "automatically". This puts us in a situation where the only possible type of empathy is that which is triggered without any prior planning. This is not useful for the purposes of this thesis since it ignores possible scenarios where empathic responses may be preconceived. If a link with simulation is to be accounted for, high-level simulation would be inconsistent with this account because of its conscious and non-automatic nature.

Outside of Maibom's list, Baron-Cohen (2003) states that there appear to be three separate senses, dimensions, or types of empathy: (1) Empathy can relate to reading another' mental states through E-imagination or simulation. In persons with ASD, this appears to be lacking. (2) Empathy can refer to a strong desire to understand the mental processes of another. Individuals with ASD are also missing this component. They appear uninterested in other people's mental lives. (3) People

who are empathic are interested in learning about other people's feelings and have compassionate reactions to those feelings. Autistics, on the other hand, are unconcerned about other people's sentiments.

It is important to note that I cannot base my argument on the first sense explained by Baron-Cohen since it relates empathy to simulation through its definition. My goal from this thesis is to establish a link between simulation and empathy. Therefore, relying on this definition would create a problem whereby I would be begging the question. Baron-Cohen (2003) gives the example of Richard Borcherds, a Cambridge University mathematician with Asperger's syndrome who won the Fields Medal, the mathematical equivalent of the Nobel Prize. Borcherds was perplexed by his sense of isolation from humans despite his mathematical prowess. People were intricate, mysterious beings who were difficult to fathom for him. He wasn't completely mindless, so he understood they had feelings and thoughts, but it was difficult for him to tell which ones they were having at any one time. As this example shows, the mathematician finds it very difficult to socialize and understand others. I think his experience indicates a possible lack of simulation capabilities. It is possible that he was having difficulties understanding others due to his lack of interoceptive capacities. The second sense of empathy relates to a desire to understand others. I believe that the desire to understand others is a precursor to simulation. Within a society, this desire is beneficial since it paves the way for mutual understanding between its members. Without this understanding, conflicts may arise. Therefore, it may be that this desire in most people has developed out of necessity to avoid conflict and establish more prosperous communities. Someone with ASD doesn't seem to give forming a community much importance. They tend to isolate themselves and focus on matters of personal interest rather than indulge in

societal or group affairs. This is further evidence of the lack of simulation capabilities for people with ASD. Finally, and in relation to the third sense of empathy provided by Baron-Cohen, interest in other's feelings and compassion towards them is also a precursor to simulation. This may be both a high-level and low-level simulation issue. For the former, one would usually try to consciously understand others' desires as a means of offering help. For the latter, mirror neurons are triggered in the simulator when a certain emotion is communicated from the other side. As for those with ASD, this reaction for low-level simulation and conscious interest for high-level simulation is lacking.

The following three definitions are all listed in Maibom's *Empathy* (2020). Eisenberg (2005) defines empathy as:

> … an affective response that stems from the apprehension or comprehension of another's emotional state or condition and is similar to what the other person is feeling or would be expected to feel in a given situation. (pp. 75)

This definition is more general than those of Guiora (1965) and Preston and de Waal (2002) in that it encompasses two types of responses due to either apprehension or comprehension. When one apprehends, they don't necessarily fully understand what is being conveyed. It may be a case of low-level simulation whereby certain emotions are perceived by the simulator and a response similar to said emotions is shown. As for comprehension, it is a more conscious process where the receiver understands the situation as much as possible before reacting or responding. In such cases, high-level simulation may be at play. The reaction or response would, therefore, be thought of beforehand as opposed to apprehension where it might be automatic.

Another definition comes from Coplan (2011). She posits three features of empathy. These features are "…affective matching, other-oriented perspective

taking, and self-other differentiation" (p. 6). I think this suggests a more conscious approach. Even though affective matching may be unconscious, there is a higher chance that taking the perspective of someone else and differentiating oneself from the other would require slightly complex processes. When another person's perspective is taken, a process of high-level simulation is utilized. One would have to consciously think what they would do in a similar situation given the current premises relating to the subject being simulated. At the same time, the simulator needs to keep in mind that even though there is a simulation process in progress, the simulated is still another person. This differentiation is necessary for minor adjustments to the simulation when new information is learned.

Batson (2014) offers a conception of empathy that seems closer to folk psychological notions than the definitions mentioned above. He says that "Empathic concern refers to other-oriented emotion elicited by and congruent with the perceived welfare of a person in need" (p. 41). This definition focuses on the moral aspect of empathy by mentioning that the welfare of a person in need is the prime concern of someone experiencing empathic feelings. This moral aspect could have characteristics that are either on the conscious or subconscious level. Similar to Coplan's (2011) definition, the aspect of this emotion being other-oriented is mentioned as well. It is unclear from this definition alone whether this could be related to low-level or high-level simulation. It is possible that according to this definition, different situations incite different reactions, whether conscious or unconscious.

Blanchet (2020) demonstrates the direct perception theory of empathy. This theory claims that there is no simulation present within empathy. Rather, we directly perceive others' mental states as a result of perceiving their behavior. There is no

middle process between perceiving an expression or behavior and recognizing a mental state, emotion, or desire. Blanchet goes on to present his own theory of empathy that, in his view, is a strong contender to the direct perception theory called the third-person theory of empathy. Briefly, it forgoes the requirement of directly perceiving another's behaviors and expressions. It becomes sufficient but not necessary in that one could experience empathy indirectly through, for instance, hearing about another's experience.

Harrelson (2020) employs a different approach whereby he starts with a basic theory of empathy and continuously modifies it to account for what he calls surrogate and out-group cases. Surrogate cases are when empathy is "directed at the object when the recipient lacks relevant knowledge or perspective" (p. 2). Out-group cases are "directed at the recipient or object when the recipient has relevant differences from the empathizer" (p. 2). In order to account for these cases, he attempts to replace matching theories with an indexical model. This leads to employing the idea of virtual self-reference and adding an epistemic condition in place of self-other distinction. The final form of his theory is as follows:

> E empathizes with R in regard to p→o if and only if:
>
> (xxx) it would be appropriate for R to have p→o
> (xxxi) E is having p→o
> (xxxii) E indexes p→o to R, or E*R(p→o), and
> (xxxiii) E makes the relevant distinctions between facts about E and facts about R (pp. 16).

The E above is the empathizer, R is the recipient, p→o is a representation p towards an object o, and E* is a virtual self-reference of E. I believe this doesn't differ from how simulation is understood on a fundamental level. E would put themselves in R's shoes as a form of simulation while maintaining the distinction between R and themselves. E believes that R should have a particular representation of an object as

a result of some knowledge mostly about themselves. As a result, E ascribes that representation to R, or in Harrelson's words, indexes p to o while making a kind of virtual self-reference to themselves.

4.2 Accounting for empathy through simulation theory

The definitions presented in the previous section have some common aspects which can be taken as starting points to further understand how empathy and simulation theory can be linked. When one experiences empathic feelings towards someone, feelings experienced by both parties are similar in nature (Eisenberg, 2005; Coplan, 2011). Moreover, the emotional state in the receiver is activated after the corresponding similar emotion is perceived in the person which empathy is directed towards (Preston & de Waal, 2002). This also means that the receiver's emotion is directed towards the other party rather than internalized (Batson, 2014).

We can observe that each definition takes an aspect of human-human emotional interaction as the locus of empathy. Therefore, empathy is sometimes looked at in a fashion similar to that of many emotional states. Take, for example, "love". When an attempt is made to define or describe love, more often than not, the symptoms of this emotional state are given instead of a description of the concept of love itself. Likewise, understanding of empathy by the average person usually pertains to what they know about its effects. For instance, an individual might feel care towards someone and consequently explain it as empathy. What they are observing here is the effect of empathy, which in this case is care. In other words, the representation of empathic feelings towards the other is instantiated by care. However, by doing this, they do not offer an explanation of the concept, rather they point out the effect they experience. Similar explanations through listing effects are offered for different circumstances or situations where empathy is at play. Such

effects could be happiness, sorrow, anger, etc. Due to the wide array of emotions one can experience as a result of having empathy towards someone, we have come to ascribe empathy to many different situations without completely grasping its causes. Empathy has, therefore, become an umbrella term that describes a plethora of cases which mainly share the three aspects stated above, namely similarity of emotion, other-directedness, and activation after perception in the other.

In the case of an infant, and especially in terms of its relationship with its mother, simulation could be the precursor to developing capacities needed for survival. Consequently, an infant would be a prime case to utilize simulation since a more theory-theory oriented method would require some background knowledge and a more cognitively aware approach. As I have explained above, low-level simulation is used here as opposed to high-level simulation since the infant does not need to be aware of their participation in a simulation of someone else's mind. It is the work of mirror neurons that sets this in motion.

It is possible that mirror neurons in a mother are similar to those in her infant which means they would trigger whenever the infant needs care, i.e., when it is hungry, in pain, or ill. Olivares-Cuhat (2012) found in a study that females are generally more empathic than males. This may be evidence in favor of the claim that mothers utilize a form of mindreading whereby unconscious responses are relied upon to attend to her infant's needs. Baron-Cohen (2003) describes autism as "an empathy disorder: those with autism have major difficulties in 'mindreading' or putting themselves into someone else's shoes, imagining the world through someone else's eyes and responding appropriately to someone else's feelings" (p. 137).

Baron-Cohen (2003) also states:

> Empathy involves a leap of imagination into someone else's head. While you can try to figure out another person's thoughts and feelings by reading their face, their voice and their posture, ultimately their internal world is not transparent, and in order to climb inside someone's head one must imagine what it is like to be them. (pp. 24)

Going back to the example of Borcherds the mathematician in section 4.1, understanding and dealing with mathematics requires one to be acquainted with how numbers, symbols, and formulas relate to one another. As opposed to social interaction, there is no process of simulation. Had Borcherds needed to employ a ToM within his mathematical work, I think he would have found it difficult to arrange his ideas in an efficient manner. Explaining and predicting others' thoughts, desires, and intentions through simulation is preceded by taking into account one's own background knowledge, as well as factors pertaining to the life of the individual simulated. In addition, facial expressions, cultural nuances, and innuendos come into play when deciphering speech. As opposed to issues pertaining to mathematics, one would need to take all of this into account before they can start or continue a conversation, let alone predict intentions and desires. A robust ToM needs to have been developed before this can be achieved. Without an adequate ToM, the individual trying to understand and predict others would be confused. In other words, a ToM is not needed to solve and understand mathematical problems since there is only one person involved tackling problems that are deductive in nature. Inducing other people's ideas, intentions, and desires requires different kinds of mental gymnastics at which those with autism are not adept.

The link with simulation theory is now more apparent. To reiterate, simulation is a cognitive process utilized to understand, explain, and predict others' intentions and feelings. Throughout this process, we can observe mechanisms that

are also present in instantiations of empathy. One may be prompted to simulate as a result of a similarity of emotions experienced between them and the receiver. This is further a result of the perception of that emotion. Afterwards, the simulation is played out while taking the simulated person's perspective. This is not strictly directed towards the other person. However, the simulator taking the perspective of the simulated individual might reinforce the idea that the emotions experienced are shared, something which is present in empathy. I believe this is grounds to view simulation theory as the cognitive process through which empathy is realized. This is similar to the relationship between alexithymia and interoception, where the former is the mental disorder, and the latter is our conception of its consequences. Simulation is a process that has tangible or perceptible results. These results are often categorized within the wide-ranging features of empathy.

4.3 Empathy and simulation theory: a counter argument

For Scheler, one is empathetic when the focus is on the other's thoughts and emotions (Scheler, 1954, p. 39). On the other hand, Goldman views empathy as a projective process in which one's own thoughts and feelings are experienced first. They are then projected onto the other with the goal of understanding them and predicting their behavior. This is done by assuming the role of the other and drawing conclusions, whether consciously or unconsciously, as to how or what they would think or feel (Goldman, 2006).

Goldman isn't the only simulationist who makes this distinction. Zahavi (2008) explains how Stueber (2006) differentiates between basic empathy and reenactive empathy. The former is "… a mechanism of inner imitation that underlies our theoretically unmediated quasi-perceptual ability to recognize other creatures

directly as minded creatures" (Zahavi, 2008, p. 515). The latter is identified as "… involving the use of our cognitive and deliberative capacities to reenact or imitate the thought process of other" (Zahavi, 2008, p. 515). As such, Goldman proposes that reflecting another's emotional experience automatically initiates a similar emotion in the one reflecting, and that this first-personal experience then serves as the basis for my third-person ascription of the emotion to the other (Zahavi, 2014, p. 110). According to Goldman, mindreading is an "extended form of empathy" (Goldman, 2006, p. 4).

Scheler discusses different types of emotional interchanges which add up to form empathy, and in some cases, they are distinct from empathy. This is important to note as some scholars have lumped all these interchanges into empathy. Accordingly, having a distinct definition for each gives a broader spectrum of the relationship between empathy and simulation theory. Scheler differentiates between emotional contagion, empathy, sympathy, and emotional sharing (Zahavi, 2014, p. 120). "… whereas empathy has to do with a basic understanding of expressive others, sympathy adds care or concern for the other" (Zahavi, 2010, p. 286). Consider a second set of scenarios. You can walk into a bar and be completely taken away by the festive atmosphere. Emotional contagion is distinguished by the fact that an emotion is caught similar to how one catches a virus, hence the use of the word contagion (Scheler, 1954). It has been passed on to you. It takes on a life of its own. Indeed, you might be infected by other people's happy or sad moods without even realizing they are separate people. The special status of emotional contagion is clear. Empathizing or sympathizing with someone entails that the experience remain with the other. The focus is on the other in both of these circumstances, and the space between self and other is maintained. An emotional contagion is further

distinguished by the fact that it affects the emotional quality and not the emotion's object. Something similar to an infection happens where the receiving end feels happy or laughs without knowing why they are feeling this way or doing so. For Gatyas (2022), an emotional contagion is distinguished from emotional sharing since sharing an emotion is similar to sharing an object or sharing a life. He proposes four conditions for sharing: mutual awareness, things shared are tokens, sharing changes some features of those tokens, and sense of ownership from all sharing parties. The opposite of emotional contagions is emotional sharing as defined by Scheler. Zahavi (2010) presents a situation in which a father and mother stand beside the body of a loving child. This circumstance, according to Scheler, provides the potential of sharing both an emotion, namely sadness or misery, and the emotion's object. Emotional sharing, on the other hand, must be differentiated from empathy and sympathy. Consider the scenario in which a common acquaintance contacts the parents who are in distress. He can relate or, more likely, sympathize with their grief without having experienced the same misery, and that is why his mental state is qualitatively different from any of theirs. Their grief and his empathy or sympathy are plainly two different emotions. His empathy or sympathy is directed toward their sorrow (Scheler, 1954, p. 12–13). Gatyas (2022) describes how sharing does not involve taking others' perspective in a strictly self-oriented manner. It is essential that one would be aware that the emotion is being shared. Holding this knowledge has the corollary of knowing that the two emotions are not identical. Gallagher (2017) corroborates this by saying that only a sufficient amount of patterns should be recognized in order for the emotion to be shared. A few details may be missing, but the overall sense of the emotion would still be present in both parties. Schwan (2019) mentions what he calls the wide matching account of affective empathy. It is similar

to the example described above in that the emotional state of the target is not completely reflected in or shared by the empathizer. While there is a sense of matching between both parties' emotional states, the side trying to reflect the original emotion could only have an estimation of what the other side is experiencing. Schwan (2019) gives an example of someone empathizing with his friend after learning that he has only six weeks left to live. This demonstrates the difference between a wide matching account of affective empathy and an actual affective matching account where the emotion in question is shared rather than estimated. In brief, empathy is not about how people can transmit what they experience across each other. Rather, it is about the process of being able to connect with them. Contrary to what Goldman appears to be implying, it is a result of our comprehension of the other's feeling, not a precondition or prerequisite for it. Accordingly, if I understood Scheler's description of empathy, it is more of a conscious behavior or consideration to the other person's feelings than an unconscious state of mind. Empathy for Scheler is, therefore, akin to what I described in section 4.2 as the name given to the group of phenomena observed as a result of simulation. In this case, it is a form of high-level simulation as it is done consciously. In order to connect with others on an emotional level, one consciously simulates their mental or emotional states, resulting in a state of caring or consideration for that particular situation. The phenomenon analogous to low-level simulation is the emotional contagion Scheler speaks of. Therefore, I believe Scheler is presenting two phenomena that are results of cognitive mechanisms within a single category, namely simulation, albeit empathy as related to high-level simulation, and emotional contagion as related to low-level simulation.

Also, a noteworthy study that might be considered another counter argument is that of Chapple, Davis, Billington, Williams, and Corcoran (2022). It describes that adults with ASD experience empathy similar to and sometimes more than those without ASD when reading literature. Their reactions towards events and characters in the readings point towards an empathic capacity usually detected in non-ASD adults.

4.4 Counter to the counter argument

Children with ASD often do not react differently to their caregivers' faces than they do to strangers' faces (Powell, 2004, p. 1055). Evolutionarily speaking, it is in the infant's best interests to develop a distinct reaction to the faces of their caretakers. The child may be able to tell their caregivers when they are hungry, which could help them survive (Sullivan, Perry, Sloan, Kleinhaus, & Burtchen, 2011). I believe that the lack of a distinct reaction towards the caregiver reflects indifference rather than an inability to distinguish between faces. Wire (2005) also states that children with ASD tend to keep to themselves and do not engage in a lot of verbal conversation. A child's inability to perceive cues and signals from others, including their caregiver might point to a lack of development of overall empathy in the future. This means that children with such a condition should have a substantially harder time projecting empathic feelings than their non-ASD counterparts, as I propose in this thesis. However, other studies suggest that people with ASD may have increased sensitivity which would make it harder to filter out unimportant distractions and process the needed information to properly understand others or predict their intentions (Favre et al., 2015). In either case, I think there would be a shortage in simulation skills. If we study this from a low-level simulation perspective, the person with ASD could have

a shortfall in their mirror neuron activity on the assumption that ASD causes a lack of empathic feelings. Experiments have shown that such a shortage exists (Dapretto et al., 2005). Being unable to put oneself in another's shoes might result in such apathy in high-level simulation. In addition, since people with ASD have heightened concentration in comparison to non-ASD individuals, i.e., they might pick up on things that would otherwise be overlooked by those without ASD, the receiver might be overwhelmed with stimuli where they would not be able to properly analyze the mental states of people around them on both a low and high level. For the former, superfluous stimuli in specific situations might cause their mirror neurons to react, while for the latter, not knowing what is crucial in visualizing another's condition might lead to incorrect predictions. I believe this demonstrates that there is a strong possibility that those with higher simulation capacities have a higher-than-average ability to express empathic feelings more easily than others.

Going back to Blanchet (2020) and Chapple et al. (2022), I believe their accounts do not pose a major difficulty towards the idea of empathy deficits in individuals with ASD. This is because the tests conducted by Chapple et al. suppose that when one interacts with literature, they convey similar results and utilize the same cognitive processes had they been interacting with another human. When exposed to a piece of fiction, especially written material, one is not interacting with another individual who possesses ideas, emotions, and desires. There is some room for interpreting the materials presented, and one has time to go back and forth to improve upon a ToM. In human-human interactions, there is limited time to think and act. Although time is a variable that might be disregarded in the context of empathy, I think the speed with which one creates a ToM or reacts with empathy or another kind of cognitive or emotional response could be one factor in determining

the degree of one's empathic capacity. This applies to Blanchet's (2020) theory as well. Indirectly learning about one's experiences may not have the same effect as directly perceiving those same experiences. Many variables may be missing in the process. Moreover, Gallagher (2017) states that Stein (2010) differentiates between learning about someone's experience through reading about it versus directly perceiving that experience through an in-person interaction. I believe it is counter-intuitive to equate the two experiences as they happen under different circumstances and with different triggers and stimulants. Further, Ratcliffe (2007) claims that situational understanding is crucial as a corequisite for a phenomenological account of perception. He means that one understands and predicts others not only through theory utilization or simulation, but rather perceiving what a specific situation presents plays an equally important role. I agree with this claim, and I think that it supports my claim above pertaining to the importance of face-to-face interactions.

Therefore, there is still grounds to claim that ASD individuals lack some empathic as well as simulation qualities. This is the claim I have supported throughout this thesis. As mentioned previously, this points towards a possible relationship between simulation theory and empathy whereby simulation is the cognitive process, and empathy is what we name many of simulation's results observable through human-human interaction.

CHAPTER 5

CONCLUSION

Throughout this thesis, I have attempted to link empathy to simulation theory by way of an analysis of the deficiencies in ToM within the population suffering from ASD. Firstly, I went over what a theory is as a precursor to an explanation of ToM. Within ToM, there are four main sub-types or approaches that mind theorists usually adopt to explain social cognition. These four sub-types are theory theory, simulation theory, modular theories, and hybrid models.

As the main purpose of this thesis is the link between empathy and simulation theory, I discussed the two forms of simulation, namely high-level simulation and low-level simulation. In brief, low-level simulation is an involuntary action mainly triggered by the firing of mirror neurons. As for high-level simulation, it is a conscious process whereby the simulator voluntarily contemplates what they would think or do. Since individuals with ASD suffer from a variety of deficiencies pertaining to social awareness and interaction (American Psychiatric Association, 2013), it is prudent to ask whether empathy is affected as well. Through an explanation of mindreading and mind blindness, I discussed three experiments that are indicators of whether one could understand that others' mental states could be false or different than their own. Through these experiments, it was shown that individuals with ASD have a difficult time answering questions relating to false beliefs (Perner & Leekam, 2008).

In chapter 3, I presented what is currently known about ASD. Since those with ASD are less aware of their own thoughts and desires than those without ASD

(Williams, 2010), I conjectured that this might correlate with a lack of simulation capabilities. After all, one has to start from themselves before being able to simulate.

Finally, in chapter 4, I presented how empathy is linked to simulation theory since there is a demonstrable lack of both capacities in people with ASD. Even though Scheler (1954) separates empathy from emotional contagion, I believe it is merely high-level vs. low-level simulation. With that in mind, and taking into account the lack of both empathy and simulation capabilities in individuals with ASD, we can conclude that empathy, as an umbrella term covering a variety of feelings within social cognition and interaction, is directly related to simulation. This relation is that simulation is the cognitive process by which we attempt to understand others, and empathy is what we call the many observable effects of this process through interaction between the simulator and the simulated. Although this may not be a one-to-one relation, I believe it is a steppingstone from which further research can be conducted.

REFERENCES

American Psychiatric Association, D. S., & American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5 (Vol. 5)*. Washington, DC: American psychiatric association.

Andrews, K. (2012). *Do apes read minds?: Toward a new folk psychology*. MIT Press.

Baird, G., Charman, T., Baron-Cohen, S., Cox, A., Swettenham, J., Wheelwright, S., & Drew, A. (2000). A screening instrument for autism at 18 months of age: a 6-year follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*(6), 694-702.

Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., & Tidball, G. (1996). Infants' reliance on a social criterion for establishing word- object relations. *Child Development, 67*(6), 3135-3153.

Baron-Cohen, S. (1991). The development of a theory of mind in autism: deviance and delay. *Psychiatric Clinics of North America, 14*(1), 33-51.

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. MIT press.

Baron-Cohen, S. (2003). *The essential difference: the truth about the male and female brain*. New York: Basic Books.

Baron-Cohen, S., Cox, A., Baird, G., Swettenham, J., Nightingale, N., Morgan, K, Drew, A., & Charman, T. (1996). Psychological markers in the detection of autism in infancy in a large population. *The British Journal of Psychiatry, 168*(2), 158-163.

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

Baron-Cohen, S., Lesli, A. M., & Frith, U. (1986). Mechanical, behavioral, and intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology, 4*, 113-125.

Baron-Cohen, S., Tager-Flusber, H., & Cohen, D. (1999). *Understanding other minds II: perspectives from autism and cognitive neuroscience*. Oxford: Oxford University Press.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders,* 34(2), 163-175.

Batson, C.D. (2014). Empathy-induced altruism and morality: No necessary connection. In Maibom, H. L. (Ed.), *Empathy and morality* (pp. 41-58). New York: Oxford University Press.

Bauman M. L., & Kemper T. L. (2005). Neuroanatomic observations of the brain in autism: a review and future directions. *International Journal of Developmental Neuroscience, 23*(2-3), 183–87.

Bernhardt, B. C., & Singer, T. (2012). The neural basis of empathy. *Annual Review of Neuroscience, 35*, 1-23.

Berthoz, S., Artiges, E., Van de Moortele, P. F., Poline, J. B., Rouquette, S., Consoli, S. M., & Martinot, J. L. (2002). Effect of impaired recognition and expression of emotions on frontocingulate cortices: An fMRI study of men with alexithymia. *American Journal of Psychiatry, 159*(6), 961-967.

Bird, G., Silani, G., Brindley, R., White, S., Frith, U., & Singer, T. (2010). Empathic brain responses in insula are modulated by levels of alexithymia but not autism. *Brain, 133*(5), 1515-1525.

Black, I. B. (1998). Genes, brain, and mind: The evolution of cognition. *Neuron, 20*(6), 1073-1080.

Blanchet, R. (2020). Empathy as the opposite of egocentrism: Why the simulation theory and the direct perception theory of empathy fail. *Topoi, 39*(4), 751-759.

Bloom, P. & German, T.P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition, 77*, 25-31.

Carruthers, P. (2011). *The opacity of mind: An integrative theory of self-knowledge.* OUP Oxford.

Chapple, M., Davis, P., Billington, J., Williams, S., & Corcoran, R. (2022). Challenging empathic deficit models of autism through responses to serious literature. *Frontiers in Psychology, 13*, 828603-828603.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*(4), 377-395.

Craig, A. D. (2003). Interoception: the sense of the physiological condition of the body. *Current Opinion in Neurobiology, 13*(4), 500-505.

Coplan, A. (2011). Understanding empathy: Its features and effects. In Coplan, A. & Goldie, P. (Eds.), *Empathy: Philosophical and psychological perspectives* (pp. 3-18). New York: Oxford University Press.

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., & Iacoboni, M. (2005). Understanding emotions in others: minor neuron dysfunction in children with autism spectrum disorder. *Nature Neuroscience, 9*(1), 28-30.

Deschamps, P. K., Been, M., & Matthys, W. (2014). Empathy and empathy induced prosocial behavior in 6-and 7-year-olds with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 44*(7), 1749-1758.

Doherty, M. (2009). *Theory of mind. How children understand others' thoughts and feelings*. Hove (East Sussex).

Eisenberg, N. (2005). The development of empathy-related responding. In Carlo G. & Edwards C. P. (Eds.), *Nebraska symposium on motivation. Volume 51. Moral motivation through the life span* (pp. 73-117). Lincoln, NE: University of Nebraska Press.

Elmose, M. (2016). A conceptual framework for understanding characteristics of self-awareness associated with autism spectrum disorder. *Scandinavian Journal of Child and Adolescent Psychiatry and Psychology, 4*(3), 109-114.

Favre, M. R., La Mondola, D., Meystre, J., Christodoulou, D., Cochrane, M. J., Markram, H., & Markram, K. (2015). Predictable enriched environment prevents development of hyperemotionality in the VPA rat model of autism. *Frontiers in Neuroscience, 9*(127).

FeldmanHall, O., Dalgleish, T., & Mobbs, D. (2013). Alexithymia decreases altruism in real social decisions. *Cortex, 49*(3), 899-904.

Fitzgerald, M., & Bellgrove, M. A. (2006). The overlap between alexithymia and Asperger's syndrome. *Journal of Autism and Developmental Disorders, 36*(4), 573-576.

Fodor, J. A. (1983). *The modularity of mind*. MIT press.

Frith, U. (2001). Mind blindness and the brain in autism. *Neuron, 32*(6), 969-979.

Frith, U., & Happé, F. (1998). Why specific developmental disorders are not specific: On- line and developmental effects in autism and dyslexia. *Developmental Science, 1*(2), 267-272.

Gallagher, S. (2007). Logical and phenomenological arguments against simulation theory. In Hutto, D. & Ratcliffe M. (Eds.), *Folk psychology re-assessed* (pp. 63-78). Springer.

Gallagher, S. (2017). Empathy and theories of direct perception. In Maibom H. L. (Ed.), *The Routledge handbook of philosophy of empathy* (pp. 158-168). Routledge.

Gallese, V. (2001). The 'shared manifold' hypothesis: from mirror neurons to empathy. *Journal of Consciousness Studies, 8*(5-6), 33–50.

Gatyas, M. (2022). Emotion sharing as empathic. *Philosophical Psychology*, 1-24.

Goldman, A. I. (2005). Imitation, mind reading, and simulation. In Hurley, S. & Chater, N. (Eds.), *Perspectives on imitation: From neuroscience to social science-volume 2: Imitation, human development, and culture* (pp. 79-93). MIT Press.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford University Press on Demand.

Goldman, A. I. (2012). Theory of mind. In Margolis, E., Samuels, R., & Stich, S. P. (Eds.), *The Oxford handbook of philosophy of cognitive science* (pp. 402-424). New York: Oxford University Press,

Goldman, A. I. & Jordan, L. C. (2013). Mindreading by simulation: The roles of imagination and mirroring. In Baron-Cohen, S., Tager-Flusberg, H., & Lombardo, M. V. (Eds.), *Understanding other minds: perspectives from developmental social neuroscience (3rd ed)* (pp. 448-466). New York: Oxford University Press.

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language, 7*(1-2), 145-171.

Gordon, R. (2004). Folk psychology as mental simulation. In Zalta, N. (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved from http://plato.stanford.edu/archives/fall2004/entries/folkpsych-simulation/

Gordon, R. & Cruz, J. (2004). *Simulation theory. Encyclopedia of cognitive science*. London: Nature Publishing Group.

Gopnik, A. & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. MIT Press.

Gopnik, A., Meltzoff, A. N., & Kuhl, P. K. (1999). *The scientist in the crib: Minds, brains, and how children learn*. William Morrow & Co.

Gordon, R.M. (1986). Folk psychology as simulation. *Mind and Language, 1*(2), 158-171.

Gottschling, V. (2019). Modularity. In Robins, S., Symons, J., & Calvo, P. (Eds.), *The Routledge companion to philosophy of psychology* (pp. 296-321). Routledge.

Grotstein, J. (1997). Alexithymia: the exception that proves the rule – of the unusual significance of affects. In Taylor, G., Bagby, R., & Parker, J. (Eds.), *Disorders of affect regulation: Alexithymia in medical and psychiatric illness* (pp. XI-XVIII). Cambridge University Press, Cambridge.

Guiora, A.Z. (1965). On clinical diagnosis and prediction. *Psychological Reports, 17*, 779-784.

Guiora, A.Z., Taylor, L.L., & Brandwin, M.A. (1968). The role of empathy in second language behavior. *Center for Research on Language and Language Behavior*. Retrieved from http://files.eric.ed.gov/fulltext/ED024952.pdf

Happé, F. & Frith, U. (1996). The neuropsychology of autism. *Brain, 119*(4), 1377-1400.

Harrelson, K. (2020). Intention and empathy. *Philosophical Psychology, 33*(8), 1162-1184.

Herbert, B. M., Herbert, C., & Pollatos, O. (2011). On the relationship between interoceptive awareness and alexithymia: Is interoceptive awareness related to emotional awareness?. *Journal of Personality*, *79*(5), 1149-1175.

Heyes, C. M. & Frith, C. D. (2014). The cultural evolution of mind reading. *Science, 344*(6190), 1243091.

Hill, E., Berthoz, S., & Frith, U. (2004). Brief report: Cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives. *Journal of Autism and Developmental Disorders*, *34*(2), 229-235.

Hurlburt, R.T. (1990) *Sampling normal and schizophrenic inner experience*. New York: Plenum Press.

Hurlburt, R.T., Happé, F. & Frith, U. (1994) Sampling the form of inner experience in three adults with Asperger syndrome. *Psychological Medicine, 24*, 385-395.

Khan, N. Z., Gallo, L. A., Arghir, A., Budisteanu, B., Budisteanu, M., Dobrescu, I., Donald, K., El-Tabari, S., Hoogenhout, M., Kalambayi, F., Kawa, R., Espinoza, I. L., Lowenthal, R., Malcolm-Smith, S., Montiel-Nava, C., Odeh, J., de Paula, C. S., Rad, F., Tarpan, A. K., Thomas, A. G. F., Wang, C., Patel, V., Baron-Cohen, S., & Elsabbagh, M. (2012). Autism and the grand challenges in global mental health. *Autism Research, 5*(3), 156-159.

Krystal, H. (1998). Affect regulation and narcissism: Trauma, alexithymia, and psychosomatic illness in narcissistic patients. In Ronningstam E. F. (Ed.), *Disorder of narcissism: diagnostic, clinical and empirical implications* (pp. 299-326). American Psychiatric Association.

Lang, B. & Perner, J. (2002). Understanding of intention and false belief and the development of self-control. *British Journal of Developmental Psychology*, *20*, 67-76.

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind". *Psychological Review*, *94*(4), 412.

Maibom, H. L. (2020). *Empathy*. Routledge.

McDougall, J. (1978). Primitive communication and the use of countertransference. *Contemporary Psychoanalysis, 14*, 173-209.

Meltzoff, A. N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*(5), 838.

Moriguchi, Y., Decety, J., Ohnishi, T., Maeda, M., Mori, T., Nemoto, K., Matsuda, H., & Komaki, G. (2007). Empathy and judging other's pain: an fMRI study of alexithymia. *Cerebral Cortex*, *17*(9), 2223-2234.

Morin, A. (2004). A neurocognitive and socioecological model of self-awareness. *Genetic, Social, and General Psychology Monographs, 130*(3), 197-224.

Morin, A. (2006). Levels of consciousness and self-awareness: A comparison and integration of various neurocognitive views. *Consciousness and Cognition, 15*(2), 358-371.

Morton, A. (1980). *Frames of mind: Constraints on the common-sense conception of the mental*. Oxford University Press.

Nichols, S. & Stich, S. P. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Clarendon Press/Oxford University Press.

Olivares-Cuhat, G. (2012). Does empathy make a difference in the foreign language classroom?. *Alicanto, 5*, 62-72.

O'Reilly, C., Lewis, J. D., & Elsabbagh, M. (2017). Is functional brain connectivity atypical in autism? A systematic review of EEG and MEG studies. *PLOS One, 12*(5), e0175870.

Parker, J. D., Keefer, K. V., Taylor, G. J., & Bagby, R. M. (2008). Latent structure of the alexithymia construct: a taxometric investigation. *Psychological Assessment, 20*(4), 385.

Perner, J. (1991) *Understanding the Representational Mind*. Cambridge, MA: MIT Press.

Perner, J. & Leekam, S. (2008). The curious incident of the photo that was accused of being false: Issues of domain specificity in development, autism, and brain imaging. *Quarterly Journal of Experimental Psychology, 61*(1), 76-89.

Popper, K. (1934). *The logic of scientific discovery*. New York.

Powell, K. (2004). Opening a window to the autistic brain. *PLOS Biology, 2*(8), e267.

Preston, S. D. & De Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 25*(1), 1-20.

Ratcliffe, M. (2006). *Rethinking commonsense psychology: A critique of folk psychology, theory of mind and simulation*. Springer.

Repacholi, B. M. (1998). Infants' use of attentional cues to identify the referent of another person's emotional expression. *Developmental Psychology, 34*(5), 1017.

Salminen, J. K., Saarijärvi, S., Äärelä, E., Toikka, T., & Kauhanen, J. (1999). Prevalence of alexithymia and its association with sociodemographic variables in the general population of Finland. *Journal of Psychosomatic Research, 46*(1), 75-82.

Scheler, M. (1954). *The Nature of Sympathy* (Peter Heath, Trans.). London: Routledge & Kegan Paul. (Original work published 1913)

Scheuffgen, K., Happé, F., Anderson, M., & Frith, U. (2000). High "intelligence," low "IQ"? Speed of processing and measured IQ in children with autism. *Development and Psychopathology, 12*(1), 83-90.

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*(2), 410-417.

Schneider, D., Slaughter, V. P., & Dux, P. E. (2015). What do we know about implicit false-belief tracking?. *Psychonomic Bulletin & Review, 22*(1), 1-12.

Schwan, D. (2019). Does affective empathy require perspective-taking or affective matching?. *American Philosophical Quarterly*, *56*(3), 277-288.

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.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research, 44*(5), 1097-1115.

Sifneos, P. E. (1973). The prevalence of 'alexithymic' characteristics in psychosomatic patients. *Psychotherapy and Psychosomatics, 22*(2-6), 255-262.

Silani, G., Bird, G., Brindley, R., Singer, T., Frith, C., & Frith, U. (2008). Levels of emotional awareness and autism: an fMRI study. *Social Neuroscience, 3*(2), 97-112.

Stein, E. (2010). *Zum Problem der Einfühlung*. Freiburg: Herder.

Stueber, K. R. (2006). *Rediscovering empathy: Agency, folk psychology, and the human sciences.* MIT Press.

Sullivan, R., Perry, B.S., Sloan, A., Kleinhaus, K., & Burtchen, N. (2011). Infant bonding and attachment to the caregiver: Insights from basic and clinical science. *Clinics in Perinatology, 38*(4), 645-655.

Taylor, G., Bagby, R., & Parker J. (1997). *Disorders of affect regulation: Alexithymia in medical and psychiatric illness*. Cambridge University Press.

Thomas, K. (1962). The structure of scientific revolutions. *International Encyclopedia of Unified Science, 2*(2).

Tooby, J. & Cosmides, l. (1992) The psychological foundations of culture. In Barkow, J., Cosmides, L. and Tooby, J. (Eds.), *The adapted mind* (pp. 19-136). Oxford: Oxford University Press.

Valdespino, A., Antezana, L., Ghane, M., & Richey, J. A. (2017). Alexithymia as a transdiagnostic precursor to empathy abnormalities: The functional role of the insula. *Frontiers in Psychology, 8*, 2234.

Vanheule, S., Desmet, M., Meganck, R., & Bogaerts, S. (2007). Alexithymia and interpersonal problems. *Journal of Clinical Psychology, 63*(1), 109-117.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust. *Neuron, 40*(3), 655-664.

Williams, D. (2010). Theory of own mind in autism: Evidence of a specific deficit in self-awareness?. *Autism, 14*(5), 474-494.

Williams, D.M. & Happé, F. (2010). Representing intentions in self and others: Studies of autism and typical development. *Developmental Science, 13*(2), 307-319.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103-128.

Wire, V. (2005). Autistic spectrum disorders and learning foreign languages. *Support for Learning, 20*(3), 123-128.

Zahavi, D. (2008). Simulation, projection and empathy. *Consciousness and Cognition, 17*(2), 514-522.

Zahavi, D. (2010). Empathy, embodiment, and interpersonal understanding: From Lipps to Schutz. *Inquiry, 53*(3), 285-306.

Zahavi, D. (2014). *Self and other: Exploring subjectivity, empathy, and shame*. Oxford University Press, USA.

Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. *Cognition, 35*, 41-68.