This is a pre-print of an article published in: **Artificial Intelligence Review**
The final authenticated version is available online at:
https://doi.org/10.1007/s10462-019-09753-0

Modeling Empathy: Building a Link Between Affective and Cognitive Processes

Özge Nilay Yalçın · Steve DiPaola

Published online: 21 August 2019

Abstract Computational modeling of empathy has recently become an increasingly popular way of studying human relations. It provides a way to increase our understanding of the link between affective and cognitive processes and enhance our interaction with artificial agents. However, the variety of fields contributing to empathy research has resulted in isolated approaches to modeling empathy, and this has led to various definitions of empathy and an absence of common ground regarding underlying empathic processes. Although this diversity may be useful in that it allows for an in-depth examination of various processes linked to empathy, it also may not yet provide a coherent theoretical picture of empathy. We argue that a clear theoretical positioning is required for collective progress. The aim of this article is, therefore, to call for a holistic and multilayered view of a model of empathy, taken from the rich background research from various disciplines. To achieve this, we present a comprehensive background on the theoretical foundations, followed by the working definitions, components, and models of empathy that are proposed by various fields. Following this introduction, we provide a detailed review of the existing techniques used in AI research to model empathy in interactive agents, focusing on the strengths and weaknesses of each approach. We conclude with a discussion of future directions in this emerging field.

Keywords Empathy · Affective Computing · Cognitive Modeling · Artificial Agents

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Social Sciences and Humanities Research Council of Canada (SSHRC).

Ö. N. Yalçın · S. DiPaola
School of Interactive Arts and Technology
Simon Fraser University,
250 -13450 102 Ave. Surrey
BC, CANADA
E-mail: oyalcin@sfu.ca
E-mail: dipaola@sfu.ca

1 Introduction

Recent developments in technology and artificial intelligence (AI) research has created an environment that allowed virtual agents to become a part of our daily lives, not only as mere tools, but as assistants and even companions. The increasing capabilities of computational systems for sensing and processing have made it possible to process social and affective cues and naturally interact with people. Computational empathy is a novel paradigm that emerges from this environment. It refers to the ability of computational systems to understand and respond to the thoughts, feelings, and emotions of humans.

Empathy is a complex socio-emotional phenomenon that can be defined as the capacity to understand and react towards the feelings, thoughts, and experiences of others. This capacity allows us to perceive another's point of view by resonating with their emotions. It has been argued that empathy plays an essential role in forming and maintaining social interactions; it helps to coordinate actions, understand the intentions of others, and facilitate prosocial behavior between individuals (Omdahl 1995). Since empathy is so important to social interactions, the integration of empathic capability for computational systems would also be useful. It could enhance interactive systems such as educational applications, medical assistants, companions, psychotherapy, and gaming applications where social capabilities are of great importance. Moreover, as an amalgamation of the affective and cognitive processes, computational empathy can provide us grounds to examine the link between emotions and cognition.

Since the introduction of the term "empathy" in the 19th century, a diversity of disciplines have contributed to its study. These sources of knowledge are invaluable for the computational models of empathy. However, the rich variety of fields contributing to the empathy research has resulted in a number of approaches used in computational modeling of empathy. This in turn has led to vague definitions and the absence of common ground in conceptualizing the processes underlying empathy. This paper presents a holistic approach to empathy modeling that integrates the theoretical and empirical background of empathy research from various fields. Our goal is to provide a systematic summary of the field to aid theoretical positioning and further research in the emerging field of computational empathy.

In order to achieve this, the following section provides a comprehensive theoretical background of empathy to capture the variety of behaviors attributed to empathy from this broad spectrum of fields (Section 2). This section is followed by computational models based on three main components that can be used to methodologically compare these approaches: emotional communication competence, emotion regulation and cognitive mechanisms (Section 2.3). The subsequent section presents the methods used in implementing the computational models of empathy in interactive agents as theory-driven and data-driven approaches (Section 3). We conclude with a discussion of these approaches and future directions in this promising field.

2 Theoretical Background of Empathy

Empathy has been an influential concept in ethics and moral theory (Hoffman 2000; Slote 2007; Smith 1959), aesthetics theory (Smith 2011), social/developmental psychology (Hoffman 2000; Batson 2012), clinical psychology (Clark 2014), and neuroscience (Goldman 2011; De Waal 2010), which all have followed the foundational work in philosophy. Contributions from this rich variety of fields has resulted in a number of definitions, functions and proposed components of empathy, which pose a challenge to studying empathy (Coplan 2011).

However, research efforts are starting to converge towards certain aspects of empathic behavior that merge emotional and cognitive phenomena. A definition of empathy as an umbrella term for any type of process triggered by observing the others' emotional states and activating one's own (de Waal and Preston 2017), has been used as an attempt to cover this broad range of behaviors. Behaviors that have been used to define empathy (Batson 2009; Coplan and Goldie 2011; de Waal and Preston 2017; Leiberg and Anders 2006) can be listed as: motor mirroring (mimicry), affective matching (emotional contagion), concern about another's state (sympathy, empathic concern), consolidation behavior (altruistic helping), understanding another's state and thoughts (theory theory), imagining another's thoughts (perspective-taking), and projection of self into another's situation (projective empathy, simulation theory). Empathy research often focuses on each of these phenomena individually. However, empathy research in psychology, neuroscience, and ethology suggests these behaviors represent the levels of empathic behavior that are connected through evolutionary processes, where each layer is built on top of the other, without replacing the layer before (de Waal and Preston 2017).

Most AI research on empathy has depicted a binary categorization and evaluation of empathy as empathic and non-empathic behavior (Prendinger and Ishizuka 2005; Brave et al. 2005; Rodrigues et al. 2014). More recently, researchers (Ochs et al. 2012; Boukricha et al. 2013) have represented levels of empathy by modulating emotions using the personality, mood, and social link parameters together with an appraisal of the situation. Another line of research in robotics focuses on developmental aspects of empathy (Asada 2015; Lim and Okuno 2015), which suggests that empathic behavior are learned to some extent and need to be based on neurological foundations.

Despite much progress, these research efforts are still in their infancy, and they highlight the need for developing and testing the models of empathy to challenge and solidify existing approaches. A computational model of empathy should support the theoretical background and the empirical research gathered from related fields. In the later sections, we provide an overview of definitions and models of empathy, including recent findings from psychology, neuroscience and ethology research.

2.1 Definitions of Empathy

One of the earliest mentions of empathic behavior can be seen in the work of Hume (1739), who used the term "sympathy" as a notion that is related to what we now refer to as low-level or affective empathy (see (Wisp 1987) and (Nowak 2011) for a complete history of the term). Hume conceptualized sympathy as an automatic process that allows for emotion contagion, morality and aesthetic pleasure while focusing on the communicability of affect for further cognitive processing. Similarly, Adam Smith (1959) distinguished sympathy from both pity and compassion by assigning it a communicative function. He refers to sympathy as a higher-level process and cognitive capacity that is related to perspective taking and simulation of other minds. In these early definitions of sympathy Hume and Smith focus on different levels of empathy that continue to be widely adopted today.

The first mention of the term "empathy" is made in the late 19th century (as cited in (Coplan and Goldie 2011; Wisp 1987)) as "Einfühlung" by Theodor Lipps, which soon became a fundamental concept for aesthetics and understanding other minds. The fundamental work of Lipps inspired a generation of thinkers to study "Einfühlung" (which means "feeling into") in several domains, including psychology, aesthetics, and philosophy; Titchener later translated this term to English as "empathy" (Titchener 1909).

The term "empathy" later came to largely replace "sympathy" in a number of fields, and the concept acquired additional behavioral functions. In addition to being an involuntary and affective feedback mechanism, it was assigned new functions as observations about others' affective, behavioral and mental states (Kohut 2011). It came to refer to an affective response that reflects another's situation rather than one's own, (Hoffman 2000), and an imagination of another's thought processes (Stueber 2006).

As empathy has continued to be studied by psychology, neuroscience and artificial intelligence, the number of related processes and cognitive capacities involved in empathic behavior have increased. Coplan (2011) lists these empathic capabilities as emotional contagion, caring behavior, perspective taking, being affected by emotions of others, theory of mind, as well as a combination of these. Some researchers focus on a narrow definition of empathy, specifying that empathy only occurs when the observer is in an affective state as a result of having imagined or observed the target's affective state with a clear self-other distinction (De Vignemont and Singer 2006). This definition excludes from empathy concepts such as sympathy, emotional contagion, and personal distress. Consistent with this narrow definition, Coplan (2011) argues that self-other differentiation is at the core of empathy and it can only be said to exist when three features of empathy are present: affective matching, other-oriented perspective taking, and self-oriented perspective taking.

In contrast to these narrow views of empathy, new research by several researchers has focused on a broader definition of empathy that aims to generate empathy models that can explain how the interaction of various processes may give rise to the broad spectrum of concepts. In the following section, we will examine these models of empathy and demonstrate how they are supported by psychology, neuroscience and ethology research.

2.2 Models of Empathy

There are two main approaches to theoretical models of empathy: categorical and dimensional. Categorical models focus on the two distinct levels of empathy mechanisms that are referred to as high/cognitive and low/affective empathy. In contrast, dimensional models propose a more multidimensional system where the levels of empathy are functionally integrated.

The former categorical approaches identify two levels of the term 'empathy' by distinguishing affective from cognitive levels of empathy. These different levels are also referred to as low/high empathy (Goldman 2006), basic/re-enactive empathy (Stueber 2006) and mirroring/constructive empathy (Goldman 2011). Omdahl (1995) attempts to connect these levels by categorizing empathy as affective, cognitive and a mixture of the two. Similarly, Hoffman (2000) argues for five modes of empathic arousal that include low-level processes of mimicry and afferent feedback, classical conditioning, and association to one's own experience; as well as high level mediated association and perspective-taking, where a mixture of these modes can be observed in an individual.

Affective (or low-level) empathy is considered to be the automatic mimicking of another's emotional response as one's own. This level is suggested to arise from the Perception-Action mechanism (PAM), which has its biological roots in mirror neurons (Preston and De Waal 2002). Some researchers distinguish the mirroring of bodily states in a category other than cognitive and emotional empathy, calling it kinesthetic empathy (Wood 2016; I. 2004). Cognitive (or high-level) empathy is defined as the ability to understand the target's mental state by imagining how they feel. This ability is related to perspective taking and theory of mind (ToM) (Batson 2009).

Evidence for both affective (low-level) and cognitive (high-level) empathy has been found in various mammalian species, suggesting a Darwinian assumption on underlying processes (de Waal and Preston 2017; Preston and De Waal 2002). Recent findings in neuroscience, especially the discovery of mirror neurons (Rizzolatti and Fabbri-Destro 2010), also support an evolutionary basis of empathy. Research on mirror neurons inspired empathy research in neuroscience and established a foundation for the functional architecture of empathy mechanisms in humans (Iacoboni 2011; Goldman 2011). From mimicry of motor actions to sharing affective experiences, the neural activation of self-experience of emotions while being exposed to another's emotional experience has been suggested to be the core mechanism for empathy (Singer and Lamm 2009).

Evidence for cognitive (higher-level) empathic behavior comes from pathological studies, autism studies, and developmental psychology (Baron-Cohen et al. 2003). Research has found that a malfunction in high- or low-level empathy mechanisms can lead to a spectrum of social behavior disorders such as autism and sociopathy (Preston and De Waal 2002). For example, cognitive empathy is found to be impaired in individuals with autism spectrum disorder, even though they may have an intact lower-level empathic function (Baron-Cohen et al. 2013). People with autism are therefore able to understand the type of emotions a person expresses, without being able to understand the reasons for those emotions. Psychopathy, on the other hand, is related to an intact ability for perspective-taking but an inability to share the resulting emotions (Preston and De Waal 2002). Following these observations and evidence from neuroscience, distinct neural routes were hypothesized as mirroring and reconstructive routes of empathy (Goldman 2011).

On the other hand, dimensional approaches to modeling empathy consider it to result from a set of interrelated constructs. Researchers that advocate this approach point to the evidence from evolutionary and neurological mechanisms that affective and cognitive levels of empathy are interconnected (de Waal and Preston 2017; Goldman 2011). For example, De Waal (2010) argues for an evolutionary foundation for the emergence of empathy based on the evidence of different levels of empathy mechanisms in social animals other than humans.

One example of a dimensional model is from Davis (1994), who also stresses the importance of the evolutionary roots of empathy mechanisms. In this model, Davis proposes to separate the processes taken place within the observer and the outcomes of these processes as affective and cognitive outputs. The model focuses on studying an empathic "episode" where the observer is exposed to emotional stimuli from a target. This exposure will result in an empathic reaction, which can have different properties according to four related constructs: antecedents, processes, intrapersonal outcomes and interpersonal outcomes. Antecedents refer to the characteristics of the person (biological capacities, learning history) and the situation (strength of the observed emotion, the degree of similarity). Three mechanisms primarily constitute the processes that produce empathic behavior: non-cognitive mechanisms (mimicry, primary reaction); simple cognitive mechanisms (classical conditioning, labelling, direct association); and advanced cognitive mechanisms (linguistic associations, perspective taking). The intrapersonal (affective and non-affective) and interpersonal (relationship-related outcomes) outcomes refer to the empathic behaviors arising from these mechanisms. Most computational models (McQuiggan and Lester 2007; Ochs et al. 2012) refer to a sub-set of intrapersonal outcomes called parallel and reactive outcomes. These consist of the affective responses that differ from the target's emotional behavior (reactive empathy) and resonate to the target's emotion (parallel empathy).

Following this view and the evidence of empathic behavior in social animals, De Waal and Preston proposed the Russian-doll model of empathy (Preston and De Waal 2002; De Waal 2007). This model, Preston and De Waal (2002) proposes that the capacity of empathy arises from the simple emotional contagion mechanisms that are based in mirroring and Perception-Action Mechanisms in mammals (De Waal 2007; de Waal and Preston 2017). Starting from this basic involuntary mechanism as the foundational layer, the model consists of three hierarchical layers where each layer depends on the previous layers. The second layer requires emotional self-regulation mechanisms that give rise to empathic concern and consolidation behavior. The last layer consists of cognitive functions such as perspective taking and theory of mind. This model allows for a wide range of empathic behaviors and emotional patterns starting from simple reflex-like copying mechanisms to higher-level cognitive functions. This theoretical model has been widely used in computational empathy research (Yalçın and DiPaola 2018; Asada 2015).

Another significant model by De Vignemont and Singer (2006) proposes a contextual approach to modeling empathy where empathy is modulated by appraisal processes. In their model, the authors distinguish between low-level affective behaviors, such as emotional contagion, sympathy and personal distress (which they refer as "narrow empathy"), to the empathic behavior that involves higher-level appraisal processes. The appraisal processes are then further categorized into early and late appraisals, where the former involves the direct appraisal of the emotional stimuli and the latter includes modulation factors after the appraisal having taken place. This extra processing in the late-appraisals results in a slower response. According to this model, the modulation factors can be related to the features of emotions (valence, intensity, saliency), to the relationship between target and observer (affective link, familiarity, similarity), to context (appraisal of the situation), and to features of the observer (mood, personality, gender, age, emotional regulation capacities). Although this approach seems to disregard lower-level/affective empathy behaviors, the authors argue that these behaviors can still arise when the appraisal processes modulate the reflex-like processes such as mimicry.

Table 1 shows an overview of the most prominent theoretical models of empathy mentioned in this section. Although each model seems to have unique details, there is significant overlap between these models that can be systematically analyzed. In the following section, we will examine the similarities between these models to be used in the computational modeling of empathy.

2.3 A Systematic Categorization of Empathy Models

The various theoretical approaches to modeling empathy can be united as a set of cognitive and behavioral capacities, which we call components. In an attempt to arrive at a comprehensive computational model of empathy, we propose to classify these components as emotional communication, emotion regulation, and cognitive mechanisms. Others have suggested the similar categories of empathic responses (Paiva et al. 2017; Boukricha et al. 2013), empathy modulation and empathy mechanisms, in which low-level perceptual mechanisms are categorized together with high-level cognitive functions in the empathy mechanisms category. Our approach differs by separating low- and high- level functions and this is intended to reflect the distinct –but functionally integrated– routes of empathy responsible for various empathic behaviors.

The proposed components have been selected to reflect the broad spectrum of behaviors assigned to empathy as a result of different types of mechanisms. Each component has its roots in empathy and emotion research, which we will review in the subsections below. Assigning each empathic behavior to its required mechanisms eases the translation of the theoretical knowledge into the design and implementation of computational models.

Computational models can improve our understanding of empathy mechanisms as well as help enhance interactive agents by equipping those agents with socio-emotional capabilities (Marsella and Gratch 2014). A computational model of empathy is therefore beneficial to the research community as it can provide a means for testing theoretical work. Regardless of the chosen theoretical model, a computational framework of empathy should reflect the broad spectrum of empathic behaviors that arise from the interaction of affective and cognitive processes. For a systematical comparison of the theoretical and computational approaches, we examine the models according to three main components: emotional communication competence, emotion regulation, and cognitive mechanisms.

Emotional Communication Competence Scherer (2010a) argues that any type of emotional behavior requires emotional production and recognition competence. This distinct capacity of perceiving, accessing, and generating emotions in order to influence behavior, reasoning, or understanding has been called as emotional intelligence in psychology (Salovey and Mayer 1990). By definition, all levels and intensities of empathic behavior require the underlying mechanism of perceiving and expressing emotions. This core component is represented as Perception-Action Mechanisms (De Waal 2007) in the Russian Doll Model of empathy, as mentioned previously. Neurological disorders caused by a dysfunction in this mechanism such as autism spectrum disorders and psychopathy suggest that empathic behavior

Table 1 Theoretical Models of Empathy

Author	Emotional Communication	Emotion Regulation	Cognitive Mechanisms	Empathy Levels
(De Waal 2007)	PAM mechanisms; motor mimicry; emotional contagion	Self-regulation	Perspective taking	Connected levels; degrees of empathy
Davis (1983, 1994)	Affective Empathy Parallel Outcomes Noncognitve Processes	Antecedents	Cognitive Empathy Cognitive Processes	Interrelated Categories
Omdahl (1995)	Affective Empathy	Emotion regulation	Cognitive Empathy	Interrelated categories
Hoffman (2000)	Mimicry Afferent feedback Classical conditioning	Empathic bias	Association to self-experience Mediated association Perspective Taking	Distinct categories
Stueber (2006)	Basic Empathy	-	Re-enactive Empathy	Interrelated Categories
Goldman (2006, 2011)	Low-Level Empathy Mirroring	-	High-level Empathy Constructive Empathy	Interrelated Categories
Coplan and Goldie (2011)	Affective Matching	Inhibitory and regulatory mechanisms	Self-other differentiation self and other-oriented perspective taking	All-or-none
De Vignemont and Singer (2006)	Emotional Capacities	Modulatory factors	Appraisal Processes	Intensity Levels

on various levels are linked to this capacity (Preston and De Waal 2002). This communicational competency of the individual is also thought to affect the intensity of empathic responses (De Vignemont and Singer 2006). Therefore, emotional communication competence is an essential component of empathy, which includes emotion recognition, expression, and representation. Low-level empathy, including variations of mirroring behavior such as motor mimicry, yawn contagion, and emotional matching, is a natural consequence of the basic interaction of the subcomponents of emotional communication competence.

Emotion Regulation This component is related to a range of social, psychological, and biological factors. It is argued that the low-level communicational mechanisms are shaped by regulatory processes such as the biological and psychological features of the observer and the relationship between target and observer (Davis 1994; Paiva et al. 2017). The features of the observer (self-related parameters) may include unconscious characteristics that work on different timescales such as attention, arousal, mood, and personality (De Vignemont and Singer 2006). Relationship-related parameters include social links such as familiarity, closeness, relatedness and perceived similarity (de Waal and Preston 2017; De Vignemont and Singer 2006; Hoffman 2000). As these mechanisms are unconscious and automatic, they require inhibitory mechanisms to mediate consolation behavior (de Waal and Preston 2017), which is a mid-level empathic behavior that is present in various mammals. Moreover, Coplan (2011) argues that inhibition of these mechanisms is crucial for higher-level empathic behavior such as other-oriented perspective taking.

Cognitive Mechanisms These capabilities include the appraisal and re-appraisal of the situation, theory of mind and mental simulation. Higher-level empathy mechanisms such as perspective taking and targeted helping require a conscious evaluation of the event and control over the low-level components. Appraisal theories of emotion focus on the connection between emotions and the goals, needs and desires of an individual, and propose that this connection may provide a foundation for high-level empathy mechanisms (Omdahl 1995). There are many appraisal theories, but the underlying common assumption is that emotions are a result of subjective evaluation of events according to the goals and needs of the individual (Roseman and Smith 2001). In stimulus check theory, Scherer (2001) proposes that appraisals can happen at various levels of cognitive processing. This idea also aligns with the Russian Doll model (De Waal 2007), in which the highest most sophisticated component of empathy, the outer components, are fundamentally linked to the perceptual and regulatory inner layers. Moreover, appraisal processes can be followed by a re-appraisal of the situation (Lazarus 1966). Understanding the effect of the event over the observed behavior, and simulation of the possible effects of the same event over the observer requires these appraisal and re-appraisal processes. Even though appraisal processes lay at the center of understanding of the cause and effect relationship between the event and observed emotion, higher-level empathic capabilities, such as self- and other-related perspective taking requires more than appraisal and re-appraisal processes. The ability to differentiate between self and other, and assigning distinct mental attributes to other minds is necessary for perspective-taking behavior. This ability is called the theory of mind (Premack and Woodruff 1978), which is related to simulation theory and theory

theory (Leiberg and Anders 2006). However, theory of mind alone is not sufficient for perspective-taking without the ability to evaluate the event-action coupling. Hence, theory of mind and appraisal mechanisms should work together.

Table 2 Summary of empathy components, the mechanisms that are responsible and the corresponding behavior.

Empathy Component	**Mechanisms**	**Empathic Behavior**
Communication Competence	Emotion Recognition Emotion Expression Emotion Representation	Mirroring Affective Matching
Emotion Regulation	Features of the observer Features of the relationship	Empathic Concern Consolation
Cognitive Mechanisms	Emotional Appraisal Theory of Mind	Alturistic Helping Perspective-Taking

Table 2 shows an overview of the variety of behaviors linked to these components of empathy. The following section (Section 3) will examine the theoretical and methodological differences in computational models of empathy while focusing on these components: communication competence, emotion regulation, and cognitive mechanisms. Following this section, we will give further detail on the processes used to implement these models based on theory-driven (top-down) or data-driven (bottom-up) approaches (Section 4).

3 Computational Models of Empathy

Computational approaches of empathy in artificial agents can follow a variety of models, even though they might share the same theoretical foundations. These differences are usually due to the differences in the aim and context of the application. A virtual agent for teaching children, a healthcare robot for people with disabilities or a VR environment that is designed for assisting meditation would have a diverse set of behavioral capabilities and goals to show empathy. Additionally, some researchers might focus on creating a computational framework of empathy, whereas others may investigate the effect of empathic behavior on human-computer interaction.

This section examines approaches that are used or can be used for modeling empathy in artificial agents. In line with the theoretical background, we review current computational models of empathy in social agents focusing on the proposed categorization of empathy on three levels: communication competence, emotion regulation and cognitive mechanisms. We also refer to the relevant research on affective computing and social computing communities that address similar problems, which can be integrated into artificial empathy research.

3.1 Communication Competence

Affective computing research has focused on emotion recognition (Zeng et al. 2009; D'mello and Graesser 2012; Poria et al. 2017) and expression in artificial agents for many decades (Calvo and D'Mello 2010; Picard 2014; Cambria 2016). These advancements also tackle the problem of understanding multi-modal emotional content (Poria et al. 2016), which is crucial for designing an interactive model that is capable of emotion recognition. However, the systems used in artificial empathy research are usually limited in their emotion recognition and expression capabilities.

An example of implementation (Yalçın in press) and detailed examination of emotional communication competence can be seen in the recent work of Yalçın and DiPaola (2019). In this work, an embodied conversational agent performs empathic listening behaviors by employing three different mechanisms: backchanneling, mimicry and affective matching. The agent perceives the speech signal as well as a combination of face recognition parameters (Ekman and Friesen's Facial Action Units and facial landmark positions) and processes this information according to the selected empathic behavior. The system uses both categorical and dimensional scale to represent emotions in order to choose proper emotional expression feedback with facial expressions and head nods.

One of the most advanced affective communication techniques can be seen in the EMMA framework (Boukricha et al. 2013), which was integrated into an embodied virtual agent platform. The agent, MAX, uses Pleasure-Arousal-Dominance (PAD) space for expressing and representing emotions. Its expressive repertoire consists of Ekman and Friesen's Facial Action Units (AUs), emotional speech (based on prosody changes), and eye blinking and breathing frequencies. EMMA uses AUs to perceive the emotion of the interaction partner.

Prendinger and Ishizuka (2005) present the Empathic Companion in a job interview task where the emotional state of the user is mapped onto a valence-arousal dimensional space. The emotional states of the users are recognized using physiological data (skin conductivity and electromyography). The character can respond via text according to three different scenarios consisting of irritation, boredom and high arousal/high valence state. The agent does not have a high-level empathy model but rather can sympathize according to pre-determined categories.

Rodrigues and colleagues (2014) only used facial expressions to recognize and express emotions. These expressions are coupled with text that either compliments or insults the other agents in the simulation environment. The emotions are presented as a tuple of "type", "valence", "intensity" and "cause" parameters based on OCC theory of emotions (Ortony et al. 1990).

In the CARE framework McQuiggan and colleagues (2008) uses self-reported affective states of the users in ten available emotion categories. Using self-reporting in emotion recognition is controversial (Scherer 2005; Calvo and D'Mello 2010) and may be misleading. Also, the categorical approach in emotion classification disregards the intensity of the emotions (Russell 1980; Scherer 2005) and ignores blended emotions where human emotions are usually not isolated (Jaimes and Sebe 2005). For emotion expression, the authors used predefined emotional sentences that were presented as text.

Brave et al. (2005) use a Blackjack game scenario where the agent can display emotions with picture-based facial expressions and textual expressions. These expressions were only classified in two categories: self-related and other related emotions. The authors did not capture the emotional expression of the users, but the empathic feedback was only given according to the situation of the user in the game. Another categorical approach can be seen in the iCat robot by Leite and colleagues (2014), who used the valence of nonverbal behavior as emotion recognition components. The expressive behaviors of iCat include spoken utterances that are divided into supportive categories of information support, tangible assistance, esteem support and emotional support.

Moridis and Economides (2012) use parallel empathy and reactive empathy based on Davis's (1994) definition of empathy. In their implementation of an empathic tutoring agent, they use six basic emotions to show pedagogical feedback to students' happy, sad and fear emotions extracted from facial expressions. The parallel empathy shows affective state matching, where reactive empathy is an emotional reaction to the user's emotions.

3.2 Emotion Regulation

Empathy regulation involves several factors that influence the extent of empathic behavior such as valence, intensity, and saliency of emotions, social relations, context, as well as mood, personality, gender, age and emotional repertoire of the agent (Paiva et al. 2017). Most of the regulation factors mentioned in the previous sections have been extensively studied in social computing research. Social computing research on personality, social link, and mood provides means of regulating the emotions based on the agent's characteristics. Some researchers (Ochs et al. 2012; Boukricha et al. 2013) used these regulation factors as a way of demonstrating different levels of empathy, where others only used a binary empathic-nonempathic classification (Brave et al. 2005; Prendinger and Ishizuka 2005; Rodrigues et al. 2014).

Boukricha and colleagues (2013) used the distance between the mood of the empathizer and the modulated empathic emotion presented by the empathizer to measure the degree of empathy. They used empathy modulation factors such as mood, desirability for self, liking and familiarity. Liking and familiarity values ranged in [0, 1] scale and predefined modulation factors. It was not clear how the desirability-for-self parameter, which is introduced in the OCC model of emotion, was calculated within the framework. The only dynamic parameter seems to be the mood of the agent, which changes with each interaction.

McQuiggan and colleagues (2008) created the CARE framework, where the agent learns to show empathy in parallel to the target's emotions or reactively to a specific set of attributes collected from the target. The authors prefer to use the self-reported emotion categorization of the user and train their model by using age, gender, context, empathic index and goal directedness features of the user. They used Naive Bayes, decision trees and SVM approaches to model empathic reactions.

Rodrigues and colleagues (2014) use affective link, similarity, mood and personality as modulation factors of the potential empathic emotions. The first two factors are parameters of the social relationship between the target and the empathizer, where the latter two are psychological factors that only concern the empathizer. Similarity is calculated by the distance between the intensities and valences of the emotions of empathizer and the target. In this sense, similarity is only calculated according to the emotional responses towards the event. Affective link and personality of the virtual agents are predetermined parameters.

Asada (2015) proposes a different approach for emotion regulation that does not mention any of the psychological, social or cultural parameters previously mentioned in this section. He proposes that emotion regulation is a part of the cognitive regulatory mechanisms that are intentional. His framework does not include the effect of mood, personality and similar factors.

3.3 Cognitive Mechanisms

Empathy mechanisms are capabilities to successfully understand the affective states of others. This also includes understanding the context, which is the situation the agent is in, that requires reasoning capabilities. De Vignemont and Singer (2006) proposed that the empathy process is tightly linked with appraisal dynamics and dependent on the observer's situation. Cognitive appraisal theories (Roseman and Smith 2001) state that the subjective assessment of an event triggers emotions. The lack of cognitive appraisal and high-level reasoning capacities in empathy models can only account for lower-level or affective empathy. However, the higher level empathic processes such as other-oriented perspective taking require a clear distinction of self-other and theory of mind components.

Affective computing and emotion research, and especially research on appraisal theory is closely related to empathy theories (Omdahl 1995). Affect is the result of the cognitive assessment of an individual where the situation and events are appraised. Computational models of empathy often use the appraisal models in emotion and affective computing research such as the OCC model (Ortony et al. 1990) or Scherer's appraisal model (2010b). Alternatively, other models of appraisal processes such as EMA (Marsella and Gratch 2009), which is commonly used in embodied conversational agents, can also be used in these frameworks. However, appraisal theories alone cannot account for achieving the empathic perspective taking and self-other distinction.

Bates and colleagues (1994) suggests the role of emotion and appraisal theories in modeling believable agents. Omdahl (1995) states that most appraisals are communicated through verbal information, and therefore linguistic information is more critical in providing context than nonverbal behaviors. Speech act theory (Searle 1969) can be used to translate appraisal theories into a dialogue context (Ochs et al. 2012). Ochs and colleagues (2012) propose a formal model of empathic emotions based on Scherer's appraisal model of emotion (2010b). Here, empathy is formulated as an emotion towards the user as a result of the successful or unsuccessful completion of the user goals.

In the work of Rodrigues and colleagues (2014), the appraisal system consists of the recognition of emotion according to the agent's own appraisals as well as the selection of the response and the modulation of that response using affective links, similarity, mood and personality. An important distinction here is that the agents use their own belief systems and goals to determine the appraisal, not the target's. This disregards the link between the Theory of Mind and empathy research, where the other's world view is needed to infer their beliefs and intentions. However, the potential response is selected in a way that allows mimicking of the same emotional response where it does not have a direct mapping of appraisal situations.

In the EMMA framework, Boukricha and colleagues (2013) use the Belief-Desire-Intention (BDI) model (Lefimann et al. 2006) in their version of the late

Table 3 Theory-Driven Approaches to Empathy and Related Empathy Components

		Empathy Components		
Author	Model	Communication Competence	Regulation	Cognitive Mechanisms
Brave et. al. (2005)	Self and other oriented emotions	Emotional pictures text	gender	-
Prendinger & Ishizuka (2005)	Sympathy	Physiological responses text	-	-
McQuiggan et al. (2008)	Parallel and Reactive Empathy (Davis 1994)	Self-reported affect recognition Emotional sentence construction	Age, Gender	Goal directedness
Moridis & Economides (2012)	Parallel and Reactive Empathy (Davis 1994)	Facial emotion recognition Facial expression and text	-	-
Ochs et al. (2012)	Formal model of empathic emotions (Scherer 2010b)	Dialogue with facial expressions	-	Appraisal model (Scherer 2010b)
Lisetti et al. (2013)	Affective Empathy	Facial emotion recognition Verbal & nonverbal expression	User model	OCC Appraisal Theory (Ortony et al. 1990)
Boukricha et. al. (2013)	Late Appraisal Model De Vignemont and Singer (2006)	Facial emotion recognition Facial expression Speech Prosody Blinking, Breathing	Mood desirability liking familiarity	Belief-Desire- Intention (Bratman 1987) (Rao and Georgeff 1991)
Rodrigues et. al. (2014)	Narrow Empathy De Vignemont and Singer (2006)	Facial expressions text	Affective link Similarity, Mood Personality	OCC Appraisal Theory (Ortony et al. 1990)
Leite et al. (2014)	-	Facial expressions gaze, speech expressive behaviors	Mood	Appraisal model (Scherer 2001)
Asada (2015)	Russian Doll Model (De Waal 2007)	Facial expressions	-	Perspective- Taking Theory of Mind
Yalcin and DiPaola (2018)	Russian Doll Model (De Waal 2007)	Emotion Recognition Emotion Expression Emotion Representation	Self and Other Related Modulation	Appraisal Reappraisal Perspective-taking

appraisal model of De Vignemont and Singer (2006). Beliefs represent the agent's knowledge about the world, where desires represent the goals of the agent. Intentions are the action plans to achieve the goals. The BDI Framework (Lefimann et al. 2006) was initially used for action planning in embodied conversational agents and is suitable for the appraisal mechanisms of an empathy model.

User modeling and adaptation research also indirectly focuses on empathy mechanisms in a way where the agent adapts to the user's behavior over time. These mechanisms relate to the Theory of Mind (TOM), where a model of the interaction partner is made that allows the self and other distinction. Paiva and colleagues (2017) highlight the importance of context, history and user modeling in achieving high-level empathic behavior. Lisetti and colleagues (2013) adopt user modeling techniques for their embodied conversational agent to reason with the outcomes of empathic actions it takes. Their approach shows an interesting direction towards combining the recent techniques in social computing and use it to provide reasoning capacities that are required in higher-level empathic behavior.

4 Modeling Empathy in Artificial Agents

This section is a continuation of the previous section, which outlined the current approaches to computational models of empathy in terms of empathy categories. This section continues the survey on these theoretical approaches of modeling computational empathy and further focuses on how theoretical models are implemented in artificial agents in detail. We will adopt two main methods for implementing a computational model of empathy: theory-driven approach and data-driven approach.

Theory-driven approaches (also called top-down or analytical approaches) are models that are formulated based on theories and concepts of empathy. This approach is intended to provide an overview of the mechanisms related to empathy and then use these mechanisms to model the smaller components of the system hierarchically. This approach has a strong explanatory power where it allows for a basis to test the theoretical components systematically.

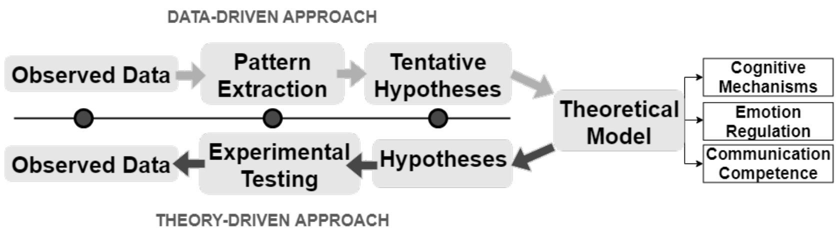

Fig. 1 Top-down and bottom-up approaches for implementing empathy models in artificial agents.

On the other hand, data-driven approaches (also called bottom-up or empirical approaches) builds up from the empathic behavior data to train computational models and algorithms in an automated fashion. These models can be used for predicting or re-generating empathic behavior based on the desired application.

Hybrid approaches can also be used where theoretical components of empathy mechanisms are modeled separately to be used by the data-driven training methods. Given the success of emotion recognition algorithms (Cambria 2016) and increasing computational power, these hybrid methods show great potential for the future of empathy research.

Both approaches can be used to reach or confirm a theoretical model, and have their strength and weaknesses. In the following sub-sections, we will review the computational approaches to empathy modeling while focusing on these methodological categorizations. Following this section, we will talk about the challenges of each method in the discussion (Section 5).

4.1 Theory-Driven Approaches

Theory-driven approach to modeling empathy has been increasingly popular within the artificial intelligence community due to the richness of theoretical models from a variety of disciplines, which we covered in Section 2.2. Researchers tend to use the components of these models as a starting point for their computational implementation. Depending on the selected model and details of each component, these models and implementations may differ in terms of the types of empathic behavior they cover.

A recent example of these approaches is seen in the component model of Yalçın and DiPaola (2018). This proposed model of empathy for interactive agents is highly inspired by the Russian Doll Model of Empathy (De Waal 2007). In this model, the authors follow an evolutionary approach that connects behavioral patterns with related mechanisms. This model is composed of distinct but interconnected components that are responsible for the various degrees of empathic behavior: emotional communication competence, emotion regulation and cognitive mechanisms. According to this model, the input from the event/context and the emotion are processed through the three layers of components. Each layer has information about the previous layer, as well as the processed information on these inputs. This approach allows for the possible implementation of low-level empathic behavior in isolation while providing a framework for simulating high-level empathic behaviors.

In his work, Asada (2015) also adopts the Russian Doll Model of empathy by De Waal (2007) from a developmental perspective. Although de Waal's model focuses on the evolutionary roots of the empathy mechanism rather than developmental changes, Asada's model draws parallels between the developmental theories of self-other distinction by proposing that developmental empathy should be a part of Cognitive Developmental Robotics (CDR). A similar approach that follows CDR is seen in the work of Lim and Okuno (2015), where empathic behaviors are learned through interaction of brain regions associated with specific mechanisms of empathy. However, their model only focuses on the affective/low-level empathic mirroring mechanisms.

Rodrigues and colleagues (2014) follow another approach that is linked with the theoretical approach of De Vignemont and Singer (2006) that states that humans do not feel empathy with every emotion and situation they encounter but instead, they select the response according to their appraisals. According to this view, empathy is never passive and is modulated according to similarity, affective link, mood and personality. Their proposed model is aimed to achieve more engaging and believable interactions with virtual humans.

Similarly, the virtual human EMMA (Boukricha et al. 2013) follows the late appraisal model of De Vignemont and Singer (2006). The EMMA framework consists of three modules: Empathy Mechanisms, Empathy Modulation and Expression of Empathy. A different approach by Ochs and colleagues (2012) provides a formal model of empathic emotions based on the theoretical formulations of Scherer (2010b) which were considered by Omdahl (1995) as the best candidate to model empathic emotions.

In their work, Ochs and colleagues (2012) represent emotions based on type (satisfaction, frustration, irritation, sadness and anger), intensity, emotion target, trigger event and the intention affected by the event. The empathic emotions use an additional variable for the target emotions in order to provide a logical formulation of them. One drawback of this approach is that it provides a very narrow view of empathy that fails to account for low-level empathic processes such as affective matching and high-level empathic processes such as perspective taking.

In their robotic empathic companion iCat, Leite and colleagues (2014) use predetermined empathic strategies that are inspired by the research on empathic behavior in teaching (Cooper et al. 2000) combined with appraisal theories by Scherer (2010b) rather than following a clear theoretical framework. Similarly, some researchers focused only on mimicking (Gonsior et al. 2011; Riek et al. 2010) and affective matching (Lisetti et al. 2013; Hegel et al. 2006) capabilities, rather than presenting the entire spectrum of empathic behavior. Affective matching and mimicking behaviors have also been studied outside of empathy research as an indication of emotional intelligence (Burleson and Picard 2007).

Additionally, some empathic behavior can be explained with current models of emotion that do not specifically address empathy. Scherer's sequential check model (2001) defines the appraisal processes as a sequence of evaluations done at different layers of processing: sensory-motor level, schematic level and the conceptual level. The sensory-motor level involves the innate assessment of the situation based on biological needs, such as hunger or pain. The schematic level is based on a learned assessment of cause-effect relationships. The conceptual level is based on symbolic cortical mechanisms that require consciousness. This framework can allow for different helping behavior which requires understanding causal relations of a stress-inducing event of another. Consolidation behavior, learned helping and targeted helping may result from the appraisal of the observed situation in different levels of processing. However, this model does not account for low-level mirroring and state matching behaviors, as they do not require an assessment of the situation. Mechanisms for theory of mind are also missing from this model, as it only accounts for appraisal of the situation in relation to self.

Broekens and colleagues (2008) suggest a formalism of the structure of appraisal. They provide a set of functions to perception, appraisal and mediation processes, as the core mechanisms of emotional behavior. Perception processes provide a link between the external world and internal representations. Appraisal processes use these current representations from working memory and assign appraisal values based on the evaluation. Lastly, mediating processes map these appraisals to emotion-component intensities, which is a set of emotional behaviors. This formalized model has a very similar underlying structure to the component model of Yalçın and DiPaola (2018). However, they differ in some significant ways.

Firstly, similar to Scherer's theory, this is a serial model that does not allow any behavioral output before it follows the process line. Being a formalization of appraisal processes, this approach does not include the direct mapping between perception-action mechanisms which is responsible for mirroring and affective matching behavior. Moreover, this model does not allow for other-oriented perspective taking, which requires mechanisms for the theory of mind.

4.2 Data-Driven Approaches

Data-driven approaches to modeling of empathy are used to recognize, predict or generate empathic behavior from several types of behavioral data. The research on data-driven empathy can be said to be in its infancy, where there are only a few attempts to empirically model empathy using this approach in the last decade. The lack of datasets and differences in the labeling methods poses a challenge in building data-driven models of empathy (see Section 5).

Earlier work focuses on classifying empathy levels using linear models with various behavioral features. Xiao and colleagues (2014) build a computational model to classify therapist empathy levels by using prosodic features from the speech signal. They extracted pitch, energy, jitter, shimmer and utterance duration from the audio recordings of the psychotherapy interaction. These audio recordings were evaluated by three people using Motivational Interviewing Treatment Integrity (MITI) system (Moyers et al. 2016) to classify the empathy levels in seven categories. Similarly, Gibson and colleagues (2015) use motivational interviews to predict therapist empathy which was evaluated using the MITI Scale. They extract 13 psycholinguistic norm features from the text such as affective norms (Valence, Arousal, Dominance, Pleasantness), familiarity, gender ladenness, context availability in addition to n-grams.

Recent advances in machine learning allow vast amounts of data to be trained efficiently, where more complex models using neural networks can be used to link data to predict and even generate empathic behavior. Rashkin and colleagues (2018) provide a dataset of approximately 25k dialogues with empathic listener responses, called EmpatheticDialogues dataset. The authors used this dataset to fine-tune their pre-trained dialogue model in order to generate empathic responses to the utterances. The produced responses were evaluated by its performance on showing an "understanding of the feelings of the other person" using crowd-sourcing. Authors showed that their model produces more empathic responses both in retrieval and generation tasks.

Kumano and colleagues (2015) use video recordings of group interactions to provide a computational model to predict empathy. They use a Bayesian approach that connects the pairwise synchronization of gaze and facial expression information to the perceived empathy of the users. This is the first study that uses group dynamics in order to evaluate the empathy of the individuals as the input of the empathy model. Authors suggest that the collective evaluation of the individuals would remove the individual bias in scoring by allowing inter-group comparison. This idea, while being similar to the third-person evaluation of crowd-sourcing, provides a collective second-person evaluation. However, the evaluation is based on selecting the empathy levels on a 5-level empathy scale between "Strong Empathy" and "Strong Counter-Empathy", where the definition of empathy was not provided.

Most of these data-driven models of empathy use a single modality, such as speech, text or video in order to train their models. However, the current advances in multi-modal emotion recognition techniques are most likely to change this trend soon. A recent distinct example is the OMG-Empathy Challenge (Barros forthcoming), which provided a dataset of interaction videos between a "speaker" and "listener", where the speaker provides scripted stories. The interaction is recorded, and the felt empathy of the listener is then rated by the listener while watching this recording. One model from this challenge uses multi-modal inputs as well as the idea of synchronization in order to predict the empathy level of from the interaction video (Tan et al. 2018).

Another notable approach can be seen in the hybrid model by McQuiggan and colleagues (2008), who developed a framework, CARE, where the empathy model trained on human-agent social interaction data that is capable of extracting intentions, actions, age, gender as well as affective states and biofeedback response. Learning from the user behavior in a simulation environment called Crystal Island, the authors aim to model artificial agents that can generate empathic responses. Their framework uses the theoretical model from Davis (1994) that is trained on user interaction data. Authors use Interpersonal Reactivity Index (Davis 1983) to measure the empathic nature of the users, and their goal orientation is used to train the model using Naïve Bayes. This approach applies both the top-down and bottom-up methodology, which shows great promise for the future of empathy modeling research.

5 Discussion

Table 3 shows a summary of some of the empathy research regarding the methodology, theoretical background, and its coverage of three empathy components: emotional communication competence, emotion regulation, and cognitive mechanisms. Regarding the empathy components, only a few researchers used a complete spectrum of empathy mechanisms. However, research efforts are advancing in terms of applying a complete model of empathy and including a broader range of behaviors.

As the research on empathy in artificial agents is an emergent field, current models and techniques often fail to capture the broad spectrum of empathic behavior. There seems to be a tendency in artificial empathy research to refer to any system that can respond to affective signals as empathic, which ignores the cognitive and high-level processes involved in empathy mechanisms. Although it is not necessary for every attempt of modeling empathic behavior to adopt all the components of empathy, it is important that the theory, models, and implementation researchers provide are congruent with each other.

Research on social agents and affective computing has examined concepts that are similar to empathy without necessarily mentioning empathy itself. For example, communicational competence of agents has been studied in affective computing, especially in the areas of emotion recognition (Soleymani et al. 2017) and expression (Calvo and D'Mello 2010). Also, user modeling and personality research in the field of social computing has investigated emotion regulation mechanisms as well as modeling self/other distinction, which are both key to high-level empathy.

As they are not directly intended to model empathy, these research efforts only cover some aspects of empathy rather than providing a complete picture. However, they nonetheless provide valuable insights that are useful for modeling empathy in artificial agents.

5.1 Methodological Issues

Some of the challenges in computational modeling of empathy are specific to the chosen methodology. Theory-driven approaches allow for the translation of the observed relationships to a framework. These approaches are easier to test and have more explanatory power. However, theories are often more difficult to translate into computational implementation. This mismatch can result in ambiguity, which in turn can generate different implementations of the same theoretical model.

Additionally, even though theory-driven approaches are the only way to test the theories of empathy, a complete implementation and evaluation of each component is required before evaluating the models as a whole. Moreover, the implementation from vague theoretical definitions can be challenging, especially in higher-level cognitive mechanisms. It is not clear, for example, how to achieve theory-of-mind in a computational setting.

On the other hand, one of the main problems of the data-driven models is data collection. Most of the data-driven approaches focus on the behavioral cues of empathizer. However, being a complex socio-emotional phenomenon, directly linking visual and textual information to the evaluation of empathy can be misleading. The behaviors can only be considered as “empathic” when they are a response to another’s emotional stimuli. Also, the annotation of the data, types of available modalities, and variety in context are all factors that affect the empathic behavior during an interaction.

5.2 Evaluation of the Model

Developing a reliable, sensitive, and valid measure of empathy for artificial agents is not an easy task. Evaluating agents on their empathic competence mostly relies on subjective user perception of a spectrum of characteristics, rather than on the application of objective measurement. When such evaluation tools are used, they tend to show many differences in the preferences of research subjects; subject preferences are very dependent on the capabilities of the computational model. There might also be differences between users in their definition of empathy due to its varying use in daily life. Considering the variations in the definitions and behavioral attributes, this vagueness in the term “empathy” poses an extra challenge to the research community.

To overcome this challenge, some researchers have chosen to measure perceived empathy by referring to empathy as “feeling with” (Boukricha et al. 2013), “felt sorry” (Rodrigues et al. 2014), “matching emotion” (McQuiggan et al. 2008), and “caring” (Brave et al. 2005). It is crucial to note which of empathic behavior these terms relate to during the evaluation of the system. Moreover, the assessment of the empathic behavior is often treated as a binary classification and coded as either an empathic or a non-empathic response. These approaches disregard the different levels of empathic responses as well as other components of empathy mentioned in the previous sections. In order to clarify the aims of the particular type of research, it is crucial to indicate which sense the term “empathy” is being used and to provide a precise positioning in the theoretical framework.

In psychology research, empathy of individuals is generally measured with the Empathy Quotient (Baron-Cohen and Wheelwright 2004) which is a self-report scale that has been validated (Lawrence et al. 2004). Empathy Quotient and Systemizing Quotient are used to determine patients with Autism Spectrum Disorders, where the average scores of both autistic men and women have been found to be lower than their healthy counterparts (Baron-Cohen et al. 2003). Additional tests of empathy focus on distinct features of empathic capacity, such as perception of emotions (Tavassoli et al. 2014) in pictures (Baron-Cohen et al. 2001a), voice (Golan et al. 2007) and movies (Golan et al. 2006), where similar tests can be used to evaluate emotional communication competence. Understanding, initiating and maintaining social relationships such as friendship (Baron-Cohen and Wheelwright 2003), can be used as a guideline to prepare evaluations for the emotion regulation component. Understanding appraisals (Baron-Cohen et al. 1986; Lawson et al. 2004) and intuitive physics (Baron-Cohen et al. 2001b) can also be used to determine whether or not the system is capable of understanding cause and effect relationships, which is based on higher-level cognitive mechanisms.

Given the component model of empathy, it is beneficial to evaluate the model in both the component level and the system level. Component-level evaluation can provide an incremental assessment of the hierarchical nature of empathic behaviors. A poorly performing appraisal mechanism can be due to the malfunction of perception mechanisms or any other mediator, as well as a system-level misrepresentation of the components. Thus, evaluating the features separately before the system-level evaluation would help to provide useful insights into the nature of a problem. Nevertheless, this does not invalidate the necessity of providing a system-wide evaluation.

Moreover, studies have shown (Riek et al. 2009) that “human-likeness” and “believability” of the agents have a dramatic effect on feeling empathy towards the agent, especially when the situation evokes negative emotions (Rosenthal-Von Der Pütten et al. 2014). Considering the influential research of Reeves and Nass (1996) that demonstrated humans treat artificial agents as social actors, it is safe to assume that aesthetic decisions as well as perceived social traits of the agent would impact the perception of empathy during an interaction. Similarly, data-driven approaches might collect contradicting behavioral data from an empathic participant with regards to their interaction partner and the context. These social and aesthetic variables should be considered in the evaluation of empathic behavior.

6 Future Directions

Computational empathy research is still in its infancy. The challenges mentioned earlier need to be overcome in order to achieve competent empathic agents that can interact with humans in real time. Agreeing on better evaluation methods that are most suitable for measuring computational empathy is one of the most crucial issues that need to be solved. Novel evaluation metrics and questionnaires that are validated specifically for empathic agent research should be one of the central topics that the computational empathy community focuses on.

Another critical aspect for the future of computational empathy is to make use of the state-of-the-art research in affective computing and user modeling research. Recent advances in interactive systems, as well as the best practices of evaluation in both fields, can be translated into the implementation of the theoretical empathy models. Moreover, although human-level empathy does not often include the input from modalities such as skin conductivity, breathing, heartbeat, and the electrical activity from the brain, these modalities have been shown to be useful in gathering affective information. Greater use of these recent innovations in computational empathy research would surely contribute to the progress in the field.

Finally, the field of computational empathy is growing very rapidly and showing great promise for the future of AI. Our daily interactions are increasingly including artificial agents, which are evolving from mere tools to interaction partners and even companions. This rapid transition prompts important questions regarding the necessity of creating moral agents. Although the relationship between empathy and morality is a controversial topic (Decety and Cowell 2014; Prinz 2011), examining the effect of empathic capabilities on creating moral agents may provide significant insights.

7 Conclusion

Computational modeling of emotions has been useful for understanding, testing, and developing the theoretical framework of emotions in affective computing research. Computational modeling of empathy is a further development to this field. Research from philosophy, psychology, ethology and neuroscience provides a broad theoretical and empirical foundation for empathy modeling. A computational model of empathy should be grounded in this background research so that it can incorporate the full range of empathic behaviors. This paper is aimed to provide a holistic approach to empathy modeling that can account for the diversity of behaviors observed in various disciplines. We believe that a successful model and implementation of the empathic capacity not only would enhance our interaction with technological systems but also with each other as a society.

Acknowledgements The authors would like to thank the reviewers for their valuable comments and suggestions to improve this article. This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2019-06767] and the Social Sciences and Humanities Research Council of Canada (SSHRC) [435-2017-0625].

References

Asada M (2015) Towards artificial empathy. International Journal of Social Robotics 7(1):19–33

Baron-Cohen S, Wheelwright S (2003) The friendship questionnaire: An investigation of adults with asperger syndrome or high-functioning autism, and normal sex differences. Journal of autism and developmental disorders 33(5):509–517

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

Baron-Cohen S, Leslie AM, Frith U (1986) Mechanical, behavioural and intentional understanding of picture stories in autistic children. British Journal of developmental psychology 4(2):113–125
Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I (2001a) The "reading the mind in the eyes" test revised version: a study with normal adults, and adults with asperger syndrome or high-functioning autism. The Journal of Child Psychology and Psychiatry and Allied Disciplines 42(2):241–251
Baron-Cohen S, Wheelwright S, Spong A, Scahill V, Lawson J, et al. (2001b) Are intuitive physics and intuitive psychology independent? a test with children with asperger syndrome. Journal of Developmental and Learning Disorders 5(1):47–78
Baron-Cohen S, Richler J, Bisarya D, Gurunathan N, Wheelwright S (2003) The systemizing quotient: an investigation of adults with asperger syndrome or high–functioning autism, and normal sex differences. Philosophical Transactions of the Royal Society B: Biological Sciences 358(1430):361–374
Baron-Cohen S, Tager-Flusberg H, Lombardo M (2013) Understanding other minds: Perspectives from developmental social neuroscience. Oxford University Press
Barros P (forthcoming) Omg empathy challenge. IEEE international conference on automatic face and gesture recognition
Bates J, et al. (1994) The role of emotion in believable agents. Communications of the ACM 37(7):122–125
Batson CD (2009) These things called empathy: eight related but distinct phenomena.
Batson CD (2012) The empathy-altruism hypothesis: Issues and implications. Empathy: From bench to bedside pp 41–54
Boukricha H, Wachsmuth I, Carminati MN, Knoeferle P (2013) A computational model of empathy: Empirical evaluation. In: Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on, IEEE, pp 1–6
Bratman M (1987) Intention, plans, and practical reason
Brave S, Nass C, Hutchinson K (2005) Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. International journal of human-computer studies 62(2):161–178
Broekens J, Degroot D, Kosters WA (2008) Formal models of appraisal: Theory, specification, and computational model. Cognitive Systems Research 9(3):173–197
Burleson W, Picard RW (2007) Gender-specific approaches to developing emotionally intelligent learning companions. IEEE Intelligent Systems 22(4)
Calvo RA, D'Mello S (2010) Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Transactions on affective computing 1(1):18–37
Cambria E (2016) Affective computing and sentiment analysis. IEEE Intelligent Systems 31(2):102–107
Clark AJ (2014) Empathy in counseling and psychotherapy: Perspectives and practices. Routledge
Cooper B, Brna P, Martins A (2000) Effective affective in intelligent systems–building on evidence of empathy in teaching and learning. In: Affective interactions, Springer, pp 21–34
Coplan A (2011) Understanding empathy: Its features and effects. In: Coplan A, Goldie P (eds) Empathy: Philosophical and psychological perspectives, Oxford University Press, pp 3–18
Coplan A, Goldie P (2011) Empathy: Philosophical and psychological perspectives. Oxford University Press
Davis MH (1983) Measuring individual differences in empathy: Evidence for a multidimensional approach. Journal of personality and social psychology 44(1):113
Davis MH (1994) Empathy: A social psychological approach. Brown and Benchmark Publishers
De Vignemont F, Singer T (2006) The empathic brain: how, when and why? Trends in cognitive sciences 10(10):435–441
De Waal F (2010) The age of empathy: Nature's lessons for a kinder society. Broadway Books
De Waal FB (2007) The 'russian doll'model of empathy and imitation. On being moved: From mirror neurons to empathy pp 35–48
Decety J, Cowell JM (2014) The complex relation between morality and empathy. Trends in cognitive sciences 18(7):337–339
D'mello S, Graesser A (2012) Autotutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. ACM Transactions on

Interactive Intelligent Systems (TiiS) 2(4):23
Gibson J, Malandrakis N, Romero F, Atkins DC, Narayanan SS (2015) Predicting therapist empathy in motivational interviews using language features inspired by psycholinguistic norms. In: Sixteenth Annual Conference of the International Speech Communication Association
Golan O, Baron-Cohen S, Hill JJ, Golan Y (2006) The “reading the mind in films” task: complex emotion recognition in adults with and without autism spectrum conditions. Social Neuroscience, 1(2):111–123
Golan O, Baron-Cohen S, Hill JJ, Rutherford M (2007) The ‘reading the mind in the voice’test-revised: a study of complex emotion recognition in adults with and without autism spectrum conditions. Journal of autism and developmental disorders 37(6):1096–1106
Goldman AI (2006) Simulating minds: The philosophy, psychology, and neuroscience of mindreading. Oxford University Press
Goldman AI (2011) Two routes to empathy. In: Coplan A, Goldie P (eds) Empathy: Philosophical and psychological perspectives, Oxford University Press, pp 31–44
Gonsior B, Sosnowski S, Mayer C, Blume J, Radig B, Wollherr D, Kühnlenz K (2011) Improving aspects of empathy and subjective performance for hri through mirroring facial expressions. In: RO-MAN, 2011 IEEE, IEEE, pp 350–356
Hegel F, Spexard T, Wrede B, Horstmann G, Vogt T (2006) Playing a different imitation game: Interaction with an empathic android robot. IEEE, pp 56–61
Hoffman ML (2000) Empathy and Moral Development: Implications for Caring and Justice. Cambridge University Press, DOI 10.1017/CBO9780511805851
Hume D (1739) A treatise of human nature: being an attempt to introduce the experimental method of reasoning into moral subjects. printed for John Noon, London
I H (2004) Some speculative hypotheses about the nature and perception of dance and choreography. Journal of Consciousness Studies 11(F0020003):79–110
Iacoboni M (2011) Within each other: Neural mechanisms for empathy in the primate brain. In: Coplan A, Goldie P (eds) Empathy: Philosophical and psychological perspectives, Oxford University Press, p 45
Jaimes A, Sebe N (2005) Multimodal human computer interaction: A survey. In: Sebe N, Lew M, Huang TS (eds) Computer Vision in Human-Computer Interaction, Springer Berlin Heidelberg, pp 1–15
Kohut H (2011) The Search for the Self: Volume 1: Selected Writings of Heinz Kohut 1950-1978. Karnac Books
Kumano S, Otsuka K, Mikami D, Matsuda M, Yamato J (2015) Analyzing interpersonal empathy via collective impressions. IEEE Transactions on Affective Computing 6(4):324–336
Lawrence EJ, Shaw P, Baker D, Baron-Cohen S, David AS (2004) Measuring empathy: reliability and validity of the empathy quotient. Psychological medicine 34(5):911–920
Lawson J, Baron-Cohen S, Wheelwright S (2004) Empathising and systemising in adults with and without asperger syndrome. Journal of autism and developmental disorders 34(3):301–310
Lazarus RS (1966) Psychological stress and the coping process / [by] Richard S. Lazarus. McGraw-Hill series in psychology
Lefimann N, Kopp S, Wachsmuth I (2006) Situated interaction with a virtual human perception, action, and cognition. Situated communication 166:287
Leiberg S, Anders S (2006) The multiple facets of empathy: a survey of theory and evidence. In: Anders S, Ende G, Junghofer M, Kissler J, Wildgruber D (eds) Understanding Emotions, Progress in Brain Research, vol 156, Elsevier, pp 419 – 440, DOI 10.1016/S0079-6123(06)56023-6, URL http://www.sciencedirect.com/science/article/pii/S0079612306560236
Leite I, Castellano G, Pereira A, Martinho C, Paiva A (2014) Empathic robots for long-term interaction. International Journal of Social Robotics 6(3):329–341
Lim A, Okuno HG (2015) A recipe for empathy. International Journal of Social Robotics 7(1):35–49
Lisetti C, Amini R, Yasavur U, Rishe N (2013) I can help you change! an empathic virtual agent delivers behavior change health interventions. ACM Transactions on Management Information Systems (TMIS) 4(4):19
Marsella S, Gratch J (2014) Computationally modeling human emotion. Commun ACM 57(12):56–67, DOI 10.1145/2631912, URL http://doi.acm.org.proxy.lib.sfu.ca/10.1145/2631912

Marsella SC, Gratch J (2009) Ema: A process model of appraisal dynamics. Cognitive Systems Research 10(1):70–90
McQuiggan SW, Lester JC (2007) Modeling and evaluating empathy in embodied companion agents. International Journal of Human-Computer Studies 65(4):348–360
McQuiggan SW, Robison JL, Phillips R, Lester JC (2008) Modeling parallel and reactive empathy in virtual agents: An inductive approach. In: Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1, International Foundation for Autonomous Agents and Multiagent Systems, pp 167–174
Moridis CN, Economides AA (2012) Affective learning: Empathetic agents with emotional facial and tone of voice expressions. IEEE Transactions on Affective Computing 3(3):260–272
Moyers TB, Rowell LN, Manuel JK, Ernst D, Houck JM (2016) The motivational interviewing treatment integrity code (miti 4): Rationale, preliminary reliability and validity. Journal of Substance Abuse Treatment 65:36–42
Nowak M (2011) The complicated history of einfühlung. Argument : Biannual Philosophical Journal 1(2):301–326, URL https://doaj.org/article/9ae0e55e3b624e85b6ba40405d45219f
Ochs M, Sadek D, Pelachaud C (2012) A formal model of emotions for an empathic rational dialog agent. Autonomous Agents and Multi-Agent Systems 24(3):410–440
Omdahl BL (1995) Cognitive appraisal, emotion, and empathy.
Ortony A, Clore GL, Collins A (1990) The cognitive structure of emotions. Cambridge university press
Paiva A, Leite I, Boukricha H, Wachsmuth I (2017) Empathy in virtual agents and robots: A survey. ACM Transactions on Interactive Intelligent Systems (TiiS) 7(3):11
Picard RW (2014) The promise of affective computing. The Oxford handbook of affective computing 11
Poria S, Cambria E, Howard N, Huang GB, Hussain A (2016) Fusing audio, visual and textual clues for sentiment analysis from multimodal content. Neurocomputing 174:50–59
Poria S, Cambria E, Bajpai R, Hussain A (2017) A review of affective computing: From unimodal analysis to multimodal fusion. Information Fusion 37:98–125
Premack D, Woodruff G (1978) Does the chimpanzee have a theory of mind? Behavioral and brain sciences 1(4):515–526
Prendinger H, Ishizuka M (2005) The empathic companion: A character-based interface that addresses users'affective states. Applied Artificial Intelligence 19(3-4):267–285
Preston SD, De Waal FB (2002) Empathy: Its ultimate and proximate bases. Behavioral and brain sciences 25(1):1–20
Prinz J (2011) Is empathy necessary for morality. Empathy: Philosophical and psychological perspectives 1:211–229
Rao AS, Georgeff MP (1991) Modeling rational agents within a bdi-architecture. KR 91:473–484
Rashkin H, Smith EM, Li M, Boureau YL (2018) I know the feeling: Learning to converse with empathy. arXiv preprint arXiv:181100207
Reeves B, Nass CI (1996) The media equation: How people treat computers, television, and new media like real people and places. Cambridge university press
Riek LD, Rabinowitch TC, Chakrabarti B, Robinson P (2009) How anthropomorphism affects empathy toward robots. In: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction, ACM, pp 245–246
Riek LD, Paul PC, Robinson P (2010) When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. Journal on Multimodal User Interfaces 3(1-2):99–108
Rizzolatti G, Fabbri-Destro M (2010) Mirror neurons: from discovery to autism. Experimental brain research 200(3-4):223–237
Rodrigues SH, Mascarenhas S, Dias J, Paiva A (2014) A process model of empathy for virtual agents. Interacting with Computers 27(4):371–391
Roseman IJ, Smith CA (2001) Appraisal theory. Appraisal processes in emotion: Theory, methods, research pp 3–19
Rosenthal-Von Der Pütten AM, Schulte FP, Eimler SC, Sobieraj S, Hoffmann L, Maderwald S, Brand M, Krämer NC (2014) Investigations on empathy towards humans and robots using fmri. Computers in Human Behavior 33:201–212
Russell JA (1980) A circumplex model of affect. Journal of personality and social psychology 39(6):1161

Salovey P, Mayer JD (1990) Emotional intelligence. Imagination, cognition and personality 9(3):185–211
Scherer KR (2001) Appraisal Considered as a Process of Multilevel Sequential Checking. Oxford University Press
Scherer KR (2005) What are emotions? and how can they be measured? Social science information 44(4):695–729
Scherer KR (2010a) Emotion and emotional competence: conceptual and theoretical issues for modelling agents. Blueprint for affective computing pp 3–20
Scherer KR (2010b) The component process model: Architecture for a comprehensive computational model of emergent emotion. Blueprint for affective computing: A sourcebook pp 47–70
Searle JR (1969) Speech acts: An essay in the philosophy of language, vol 626. Cambridge university press
Singer T, Lamm C (2009) The social neuroscience of empathy. Annals of the New York Academy of Sciences 11561(1):81–96
Slote M (2007) The ethics of care and empathy. Routledge
Smith A (1959) The theory of moral sentiments. Printed for A. Millar and A. Kincaid and J. Bell, London
Smith M (2011) Empathy, expansionism, and the extended mind. Empathy: Philosophical and psychological perspectives pp 99–117
Soleymani M, Garcia D, Jou B, Schuller B, Chang SF, Pantic M (2017) A survey of multimodal sentiment analysis. Image and Vision Computing 65:3–14
Stueber KR (2006) Rediscovering empathy : agency, folk psychology, and the human sciences
Tan ZX, Goel A, Nguyen TS, Ong DC (2018) A multimodal lstm for predicting listener empathic responses over time. arXiv e-prints arXiv:1812.04891, 1812.04891
Tavassoli T, Hoekstra RA, Baron-Cohen S (2014) The sensory perception quotient (spq): development and validation of a new sensory questionnaire for adults with and without autism. Molecular autism 5(1):29
Titchener EB (1909) Lectures on the experimental psychology of the thought-processes. Macmillan
de Waal FB, Preston SD (2017) Mammalian empathy: behavioural manifestations and neural basis. Nature Reviews Neuroscience 18(8):498
Wisp L (1987) History of the concept of empathy. Empathy and its development pp 17–37
Wood K (2016) Kinesthetic empathy: Conditions for viewing. Oxford University Press
Xiao B, Bone D, Segbroeck MV, Imel ZE, Atkins DC, Georgiou PG, Narayanan SS (2014) Modeling therapist empathy through prosody in drug addiction counseling. In: Fifteenth Annual Conference of the International Speech Communication Association
Yalçın ÖN (in press) Empathy framework for embodied conversational agents. Cognitive Systems Research
Yalçın ÖN, DiPaola S (2018) A computational model of empathy for interactive agents. Biologically Inspired Cognitive Architectures 26:20–25
Yalçın ÖN, DiPaola S (2019) Evaluating levels of emotional contagion with an embodied conversational agent. In: Proceedings of the 41st Annual Conference of the Cognitive Science Society
Zeng Z, Pantic M, Roisman GI, Huang TS (2009) A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE transactions on pattern analysis and machine intelligence 31(1):39–58