Empathy in Virtual Agents and Robots: A Survey

ANA PAIVA, INESC-ID and Instituto Superior Técnico, University of Lisbon, Portugal
IOLANDA LEITE, KTH Royal Institute Technology
HANA BOUKRICHA and IPKE WACHSMUTH, Faculty of Technology, Bielefeld University

This article surveys the area of computational empathy, analysing different ways by which artificial agents can simulate and trigger empathy in their interactions with humans. Empathic agents can be seen as agents that have the capacity to place themselves into the position of a user's or another agent's emotional situation and respond appropriately. We also survey artificial agents that, by their design and behaviour, can lead users to respond emotionally as if they were experiencing the agent's situation. In the course of this survey, we present the research conducted to date on empathic agents in light of the principles and mechanisms of empathy found in humans. We end by discussing some of the main challenges that this exciting area will be facing in the future.

CCS Concepts: • **General and reference → Surveys and overviews**; • **Computing methodologies →** *Intelligent agents*; Cognitive robotics;

Additional Key Words and Phrases: empathy, virtual agents, social robots, affective computing, human-computer interaction, human-robot interaction

ACM Reference format:
Ana Paiva, Iolanda Leite, Hana Boukricha, and Ipke Wachsmuth. 2017. Empathy in Virtual Agents and Robots: A Survey. *ACM Trans. Interact. Intell. Syst.* 7, 3, Article 11 (September 2017), 40 pages.
https://doi.org/10.1145/2912150

1 INTRODUCTION

We live in the company of others. As such, understanding those around us and their intentions, motivations, and feelings is key to the well-being and survival in our complex social environment. During the past few decades, our social environment has been populated by technology that is dramatically revolutionizing the way we live. The presence of technological artefacts that act autonomously in social environments is now becoming a reality. Autonomous cars, vacuum cleaners, smartphones, virtual agents, and robots are all part of a reality that embraces technological autonomy to support all types of activities. Yet, for these new autonomous systems to inhabit our social environment, they must be able to interact with us humans. Such interaction has to be as natural

The reviewing of this article was managed by associate editor Catherine Pelachaud.
This work is partially funded by national funds through the Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013, and by the project AMIGOS, also funded by the FCT with reference PTDC/EEI-SII/7174/2014.
Authors' addresses: I. Leite, KTH Royal Institute of Technology, School of Computer Science and Communication, Department of Robotics, Perception and Learning (RPL), Teknikringen 14, 6th floor, SE-100 44 Stockholm, Sweden; email: iolanda@kth.se; A. Paiva, INESC-ID & Instituto Superior Técnico, University of Lisbon, TagusPark , Avenida Prof. Cavaco Silva, Taguspark, 2744-016 Porto Salvo, Portugal; email: ana.paiva@inesc-id.pt; H. Boukricha and I. Wachsmuth, Universität Bielefeld, Technische Fakultät, P.O. Box 10 01 31, 33501 Bielefeld.

ACM 2160-6455/2017/09-ART11 $15.00
https://doi.org/10.1145/2912150

as possible and very much inspired by the way we interact with each other. In this regard, these systems must be able to act in a *social* way; that is, they must not only be reactive and proactive but also have social capabilities, particularly when interacting with humans (Wooldridge 2002) (cf. Boukricha (2013)). As such, a social agent—that is, an agent with social ability—must be endowed with the capability to understand others and their intentions, motivations, and feelings. Such capability entails perceiving and understanding others' affective states and acting accordingly, a capability often referred to as *empathy*.

Empathy is central in the study of human social relationships and thus one of the major elements in human social interaction. Nowadays, the term empathy is currently used (and perhaps even abused) by the media. We can find newspaper headlines associated with the manifestation of empathy in rats, the lack of empathy in business relationships, or the right empathic edge for winning or losing presidential elections. In a world where many voices raise concern about the dehumanization of our modern way of living, scientists are investigating the factors that lead to such dehumanization, and empathy seems to be at the centre of this concern. Indeed, according to Coplan (2011), empathy plays a major role in our society because of its relation to a wide range of important issues, such as the "nature and conditions of morality," "how do we understand each other," or even "what makes certain political candidates appealing, what characterizes psychopaths and bullies, and how medical workers should interact with their patients." It has been shown that empathy promotes stronger relationships and collaboration and that it is one of the pillars of morality (Hoffman 2001; Davis 1994; de Waal 2010). Furthermore, empathy is not only specific to humans but is also present in other species and even across species. Plutchik (1987), who has studied affective communication in different species, claims that such behaviour patterns have implications for survival and shows that some affective signals, particularly *alarm* signals, have a genetic component. From an evolutionary perspective, empathy is considered to bond individuals to one another (Plutchik 1987).

Given the strong role of empathy in shaping the way that communication and social relationships are established in different species, it is obvious that empathy is a major element in human–machine interaction. It has been shown that humans feel empathy not only for their peers but also for fictional characters, game characters, and even robots (see Section 4).

However, machines are seldom able to reciprocate such behaviour. Although they are to some extent able to portray expressions of emotion (e.g., *sad* or *happy*), many of the processes found in empathic behaviours between humans were, until recently, poorly captured in machines. As a consequence, during the past few years, some researchers are working to address this problem. Studies from social psychology show that humans tend to unconsciously mimic the behaviors of others (Chartrand and Bargh 1999), follow others' expressions, and respond empathically, leading to smoother interactions and increased bonding. In this article, we will review different models and studies that show that the presence of empathic responses by machines (virtual agents and robots) leads to better, more positive (Ochs et al. 2012) and appropriate (Hegel et al. 2006) interactions. In particular, different studies presented here show that agents and robots endowed with empathic capabilities lead to more trust (Brave et al. 2005; Cramer et al. 2010), increased social presence (Leite et al. 2014), reduced stress (Prendinger and Ishizuka 2005), and more engagement (Leite et al. 2014), and are seen as more friendly (Leite et al. 2013b), likable (Brave et al. 2005), and caring (Brave et al. 2005).

This article provides a state of the art in the recent research field of virtual agents and robots endowed with empathic capabilities. In Section 2, we give an overview on the theoretical background of empathy to clarify some of the central concepts related to empathy in psychology and neuropsychology. Further, in Section 3, we identify the requirements needed for the computational modelling of empathy. In Sections 4 and 5, we provide some examples of the efforts made to develop empathic virtual agents and robots. Finally, we draw some conclusions and identify directions for future research.

2 THEORETICAL BACKGROUND ON EMPATHY

The computational modelling of empathy for virtual agents and robots is an interdisciplinary endeavor that relies on existing psychological and neuropsychological theories. This section gives an overview of the debate over a universal definition of empathy, as well as on the theoretical background crucial for its computational modelling in virtual agents and robots.

2.1 Empathy: Definitions and Processes

Empathy has no universally agreed upon definition, and its multiple definitions can be subdivided into three major categories (cf. Omdahl (1995)): **(1)** empathy as an affective response to others' emotional states (affective empathy), **(2)** empathy as the cognitive understanding of others' emotional states (cognitive empathy), and **(3)** empathy as composed of both an affective and a cognitive component.

However, researchers such as Hoffman (2001), Davis (1994), and Preston and De Waal (2002) have attempted to unify different perspectives on empathy by adopting a multidimensional approach and by providing comprehensive models of empathy. In particular, Hoffman (2001) defines empathy as:

> *A psychological process that makes a person have "feelings that are more congruent with another's situation than with his own situation."*

Hoffman supports that an empathic response does not have to be a close match to the emotional state of the other but can be any emotional reaction compatible with the other's situation. He argues that empathy is multidetermined and that it is defined in terms of the processes required for an empathic response (Hoffman 2001). Similarly, Davis defines empathy as "a set of constructs having to do with the responses of one individual to the experiences of another" (Davis 1994, pp. 12). Furthermore, according to Preston and De Waal, empathy can be seen as a broad umbrella concept that encompasses many of other related phenomena (and processes) such as emotional contagion and sympathy Preston and De Waal (2002).

Such an inclusive view of empathy is not completely shared by other researchers, such as Coplan, who considers that grouping all the low-level processes involved in empathy together with higher level ones "is not the best way to highlight their importance" (Coplan 2011). However, because our focus is on human–machine interaction, we may need to consider both low- and high-level processes as part of a wider interaction framework that machines are a part of. In this regard, models such as those introduced by Hoffman (2001), Davis (1994), and Preston and De Waal (2002) offer comprehensive and detailed views of empathy that unify the different perspectives and that are crucial for the computational modelling of empathy. In fact, De Waal defines empathy as "the capacity to (a) be affected by and share the emotional state of another, (b) assess the reasons for the others state, and (c) identify with the other, adopting his or her perspective." In this article, we adopt this definition of empathy, considering it as a process that makes a person (or agent) have "feelings that are more congruent with another's situation than with his own situation."

Following this definition, Figure 1 illustrates the different elements involved in a empathic situation, or "empathic encounter":

- The *observer* (empathiser) as the person/agent that responds emotionally to the affective state of another one.

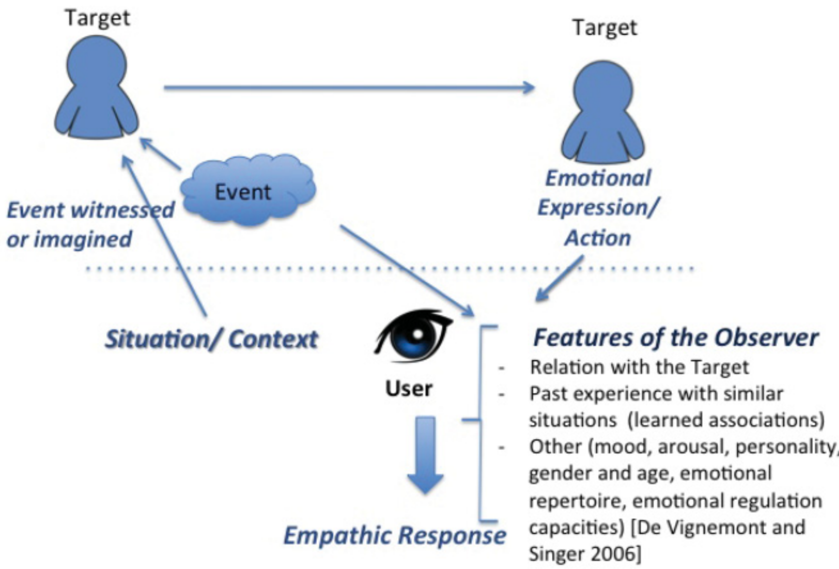


Fig. 1. General elements involved in an empathic encounter.

— The *target* as the person/agent that expresses an emotional state or is in an emotional situation perceived by the observer.
— The *event* that happens and is witnessed (directly or indirectly) by the observer.
— The *emotion* experienced by the target (or even imagined by the observer as felt by the target) and to which the observer responds.
— The *situation/context* that characterises the occurrence of an event leading to the emotion.
— *Mediating factors* and aspects, such as the relation between the observer and the target, the mood of the observer, the presence or absence of other agents, past situations, and the like.

Furthermore, considering the theoretical background on empathy, three different processes that are central in an empathic encounter are identified (Boukricha 2013):

— The **empathy mechanism** as the process by which an empathic emotion arises.
— The **empathy modulation** as the process by which both an empathic emotion is modulated and a degree of empathy is determined.
— The **empathic responses** as the process by which an empathic emotion is expressed/communicated and actions are taken.

These processes of empathy are crucial for the modelling of empathy in social agents (see Section 3) and are introduced and discussed in Sections 2.2, 2.3, and 2.4, respectively.

2.2 Mechanisms of Empathy

Empathy arises as a result of internal mechanisms. Preston and De Waal have proposed that at the core of the empathic responses lies a mechanism that provides the observer with access to the subjective emotional state of the target. This is done through the observer's own perception and internal processes and is mediated by different factors. For example, the more socially related the observer is to the target, the more these processes are activated. At the core of these mechanisms lies the perception/action mechanism (PAM) (Preston and De Waal 2002). This mechanism follows the "Perception-Action Hypothesis" grounded in the notion that perception and action share a common code of representation in the brain, and, as such, perception of a behavior in another person (or agent) will automatically activate one's own representations of that behavior (De Waal 2008). Similarly, Hoffman (2001) also distinguishes internal mechanisms that the observer uses to "get under the skin" of the target. These can be automatic, preverbal, and involuntary, as in the case of mimicry.

Mimicry in particular is defined as the tendency to imitate facial, vocal, or postural expressions of other individuals with whom one interacts. Hoffman (2001) defines mimicry as based on two subsequent processing steps: *imitation* and *afferent feedback*. Imitation is the automatic response of an observer's body to the other's motor expressions by producing similar motor changes. Furthermore, the changes that happen to the observer, be they facial, postural, or vocal, trigger an afferent feedback in the observer, producing feelings that may match or are more in line with those of the other (thus empathy). Emotional mimicry (Hess and Fischer 2014), more specifically, is considered the imitation of the emotional expressions of others based on their emotional signals. This entails congruent emotional displays of the observer (Hess and Fischer 2014), which, combined with an afferent feedback, may produce feelings that match those of the other. Although empathy should be seen as broader than emotional mimicry, it may result from an internal mechanism that involves mimicry. In fact, it is generally accepted that empathic individuals exhibit more mimicry of the postures, mannerisms, and facial expressions of others than do nonempathic individuals.

These core processes are in line with neuropsychological findings that show that a mechanism similar to "mirror neurons" (Rizzolatti 2005) is at the basis of emotional understanding and thus fundamental for empathy (De Vignemont and Singer 2006a). That is, the observation of another's emotional state activates brain areas involved in experiencing that same emotional state. In line with this, the perception-action hypothesis (Preston and De Waal 2002) also suggests that the perception or imagination of someone else's emotional state automatically activates a representation of that state in the observer, thus leading to an empathic response. Furthermore, De Vignemont and Singer (2006a) claim that the observation or imagination of another's emotional state automatically elicits an empathic response based on the activation of a shared neural network.

But empathy should cover all processes by which an individual is affected by the emotional state of another and has feeling more congruent to the other person's. Some of these processes, as just discussed, do not require "conscious and effortful" processing and occur automatically, but others require cognitive processing. Recent work points to two possible systems for empathy: (1) a basic emotional and unconscious one and (2) a more advanced cognitive perspective-taking system. In fact, Hoffman considers that an empathy-arousing mode that demands higher level cognitive processing can be further divided into *mediated association* and *role- or perspective-taking*.

According to Hoffman (2001), role-taking is the ability of an observer to put him- or herself in another's situation and to imagine how the other feels. This mode of empathic arousal requires higher level cognitive processing due to the mental effort involved in role-taking (Hoffman 2001). The work by Stotland (1969) showed that participants' empathic responses in a given situation were higher when they imagined themselves being in this situation rather than imagining how the other feels or simply focusing on the other's movements. In line with Stotland's (1969) findings, Hoffman (2001) defined two types of role-taking: *self-focused role-taking* as an observer's ability to imagine how he would feel in a given situation and *other-focused role-taking* as an observer's ability to imagine how the other feels in a given situation. He further claimed that other-focused role-taking is a more cognitively demanding process than self-focused role-taking because it requires taking the other's mental states into account. A similar distinction has also been made by Higgins (1981) between *situational* and *individual role-taking*. This mechanism of role-taking is often linked to cognitive empathy, which is empathy combined with the appraisal of the situation given the context and the understanding of what caused the other's emotional state (Preston and De Waal 2002).

Given that we will be considering the creation of empathic models for robots and virtual agents in this article, we will look at how some of these mechanisms have been incorporated into them. However, due to the constraints of the embodiments, we will concentrate on the mechanisms that have been addressed or created artificially for these two types of embodied agents and, in particular, *affective mimicry* and *role- or perspective-taking* (see Section 5).

2.3 Empathy Modulation

According to the perception-action hypotheses, the observation or imagination of another's emotional state automatically elicits an empathic response (De Vignemont and Singer 2006a; Preston and De Waal 2002) (cf. Section 2.2). Furthermore, de Waal (2010) claims that "we do not decide to be empathic, we just are." In this regard, De Vignemont and Singer state that the automatic elicitation of empathy would result in being in a "permanent emotional turmoil" ((De Vignemont and Singer 2006a), p. 436). In consequence, they introduce several *modulation factors* that may affect the extent of empathic brain responses and discuss empirical evidence for "the modulation of the empathic brain" (De Vignemont and Singer (2006a), p. 437) (cf. Boukricha (2013)). Hoffman (2001) refers to such modulation factors as *empathy bias,* and Davis (1994) refers to them as *antecedents.* In their cognitive appraisal theory of emotion, Ortony, Clore, and Collins (OCC) (Ortony et al. 1988) define fortunes-of-other emotions as empathic emotions and introduce a number of *local variables* that influence the intensity of an empathic emotion. Following De Vignemont and Singer (2006a), we group these factors into four categories (see Table 1) (cf. (Boukricha 2013)): **(a)** features of observed emotion, **(b)** social relationships between observer and target, **(c)** situation and context, and **(d)** features of the observer.

2.3.1 Features of Observed Emotion. The features of the observed emotion may impact the strength of an empathic response. Hoffman (2001) considers the *salience* and *intensity* of the observed other's emotional state and states that the more salient and intense the other's emotional state is, the stronger the observer's empathic response. Furthermore, Davis (1994) introduces the *strength of situation* as the strength of the observed emotional state. Similarly, the OCC theory (Ortony et al. 1988) defines the local variable *desirability-for-other* as the degree to which an observed event is appraised as desirable or not desirable for the other. Furthermore, De Vignemont and Singer (2006a) propose *valence*, *intensity*, *saliency*, and *primary* versus *secondary* emotions as features of the observed emotion that impact empathic responses. Regarding primary and secondary emotions, they state that it may be easier to empathise with primary emotion, such as *sad* and *happy*, rather than with secondary emotions such as *jealousy*.

2.3.2 Social Relationships. The strength of social relationships between the observer and the target may also impact the strength of an empathic response. Hoffman (2001) considers *in-group*, *friendship*, and *similarity* bias and refers to them as *familiarity* bias. That is, an observer is likely to have a more intense empathic response with family members, group members, friends, and people with similar goals and needs. Similar to Hoffman's (2001) definition of a similarity bias, Davis (1994) introduces *observer-target similarity* as a social relationship factor modulating the degree of an empathic response. Further, De Vignemont and Singer (2006a) propose the notion of *affective link* (liking), *familiarity*, *similarity*, and *communicative intentions* as an observer's intention to communicate the desire for empathy as mediating factors for the resulting emotions. Another important feature that may influence empathic responses is *identification* as an essential mechanism that allows us to adopt the emotions, situations, and behaviours of those we are empathising with. In fact, it is difficult to identify with people whom we see as different or who belong to a different

Table 1. Empathy Modulation Factors Adapted from Boukricha (2013)

Authors	Features of the observed emotion	Relationship between the observer and the target	Situation/ context	Features of the observer
Hoffman (2000)	Salience and intensity of observed emotion	Familiarity bias: in-group bias friendship bias similarity bias	Here-and-know bias	–
Davis (1994)	Strength of situation	Observer-target similarity	–	Biological capacities, individual differences, learning history
OCC (1988)	Desirability-for-other	Liking	Deservingness	Desirability-for-self
De Vignemont & Singer (2006)	Valence, intensity, saliency, primary vs. secondary emotions	Affective link, familiarity and similarity, communicative intentions	Appraisal of the situation, display of multiple emotions	Mood, arousal, personality, gender and age, emotional repertoire, emotional regulation capacities

group. Identification is stronger in moments of high emotion and stronger with familiar people, in particular our close relatives (Hall et al. 2005).

2.3.3 Situation and Context. In addition, the features of the situation and the context may impact the strength of an empathic response. Hoffman (2001) considers a *"here-and-now"* bias associated with the presence or absence of the target. That is, observers empathise to a higher extent with those who are present rather with than those who are absent. In OCC, Ortony et al. (1988) define *deservingness* as the degree to which the target deserves or does not deserve his current desirable or undesirable situation. Further, De Vignemont and Singer (2006a) define *appraisal of the situation* and *display of multiple emotions* as features of a situation and context that may impact the strength of an empathic response. Regarding the display of multiple emotions, they claim that it becomes difficult to empathise with different people displaying different emotions. An aspect that needs to be taken into account is the target's *behaviour transparency* in a given situation. According to Malle (2005), expressive behaviours and other movements that are performed by the target are "embedded functionally in the physical context." This fact makes them somehow more transparent and easier to understand without too much mental inference. Another aspect that impacts the empathic response is related to the *presence of others* in the situation. Darley and Latane (1968) state that the higher the number of people present in a situation, the less compelled they are to help a victim. Furthermore, Garcia et al. (2002) have shown that this process can even be implicit and that just imagining the presence of others leads to a similar effect of responsibility diffusion.

2.3.4 Features of the Observer. The features of the empathiser also have an impact on the strength of an empathic response. Davis (1994) considers *biological capacities* as the observer's capability for higher level cognitive processing such as role-taking, *individual differences* as differences in the dispositional tendency to experience particular empathic outcomes and to engage in particular empathic processes, and *learning history* as an observer's acquired values of behaviours regarding empathy. De Vignemont and Singer (2006a) propose *mood*, *level of arousal*, *personality*, *gender*, *age*, *emotional repertoire*, and *emotional regulation capacities*. In terms of gender differences in empathy, the view that females are more empathic than males is a stereotype widely held. According to a survey conducted by Lennon and Eisenberg (1987), gender differences in empathy are not as strong as the stereotype leads us to believe, and the differences found seem to be dependent on the manner in which empathy is operationalised and measured. With regard to observers' *personality*, it has been shown that people who score high on social responsibility also score high on dispositional empathy. Social responsibility is considered to be a moral obligation to fulfill expectations that others have on one's behaviour and is often considered to be one important aspect of a person's personality.

2.4 Empathic Responses

Based on the definition of empathy as an affective response more appropriate to another's situation than to one's own, the result of the empathic process is an emotion triggered in the observer and possibly some actions taken. This view is also supported by De Vignemont and Singer (2006b) who argued that empathy exists only if the empathizer (observer) is in an affective state. Thus, the outcome of the empathic process is an empathic emotion felt by the observer, which can then potentially trigger specific behaviours.

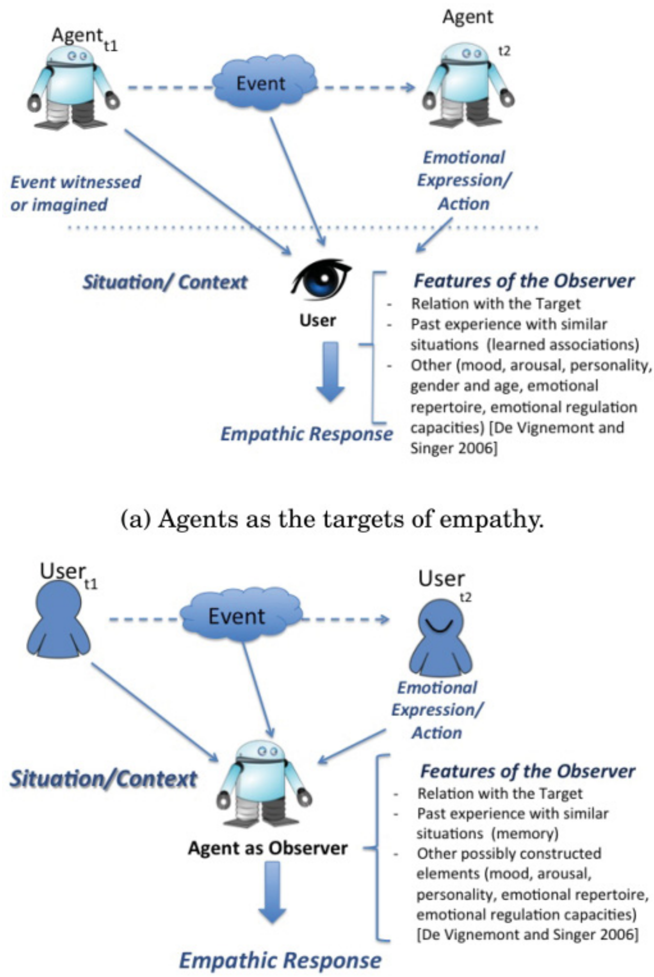
Considering the emotion itself, as discussed previously, it will be mediated by different factors. In fact, we may have an empathic emotion in a certain situation but not in another very similar situation. De Vignemont and Singer consider that empathy is modulated by appraisal processes and not merely the consequence of the passive observation of emotional cues of the target. Thus, the empathic emotion does not need to be the same as the target one. Furthermore, the intensity of the empathic emotion felt by the observer is modulated by different factors such as the mood of the observer, his or her personality, and the like.

The empathic emotion can be expressed through facial expressions, body expression, physiological responses, and action tendencies. One interesting aspect of the study of empathic outcomes has been associated with the notion of empathic concern, which is reliably associated with caring and helping behaviours. In terms of behavior, empathy per se is not enough to induce prosocial behavior. Yet empathic responses might result in personal distress and thereby motivate not only self-related responses, such as avoidance and withdrawal, but also other-related prosocial behavior.

Accordingly, Hoffman (2001) underlines the role of empathy in motivating prosocial actions and moral principle and refers to caring and justice as prosocial moral behaviours. Thus, action tendencies and behaviours resulting from the empathic processes may lead to helping and social behaviours as interpersonal empathic outcomes (Davis 1994). According to Gaertner and Dovidio (1977), the degree of empathy of an observer is systematically related to the speed of her subsequent helping action. Furthermore, Davis (1983a) claims that prosocial acts are mediated by individual differences that exert a significant influence on the helping actions.

3 EMPATHY IN SOCIAL AGENTS

In line with the theoretical background introduced in the previous section, this section focuses on the main aspects that need to be considered in the creation of computational models of empathy for virtual agents and robots.



(a) Agents as the targets of empathy.

(b) Empathic agents as observers.

Fig. 2. Two different perspectives of studying empathy in social agents.

There are two distinct ways to look at empathy in social agents (Paiva 2011). The first is to consider the human interaction partner as the observer (empathiser) and the agent as the target that triggers empathy in the human partner (see Figure 2, (a)). In this case, the agent does not necessarily need to be endowed with empathic behaviour, but is designed to evoke empathy in the human observer. Such agents have been employed, for example, in scenarios where it is important to persuade the human partner to select the "right" action (Paiva et al. 2004). The second way to look at empathy in social agents is to consider agents as observers that empathize with other agents, particularly with human partners as targets of empathy (see Figure 2, (b)). Accordingly, we follow Paiva's definition of empathic agents, which encompasses these two perspectives (Paiva 2011):

Empathic agents are (1) agents that respond emotionally to situations that are more congruent with the user's or another agent's emotional situation or (2) are agents that, by their design and behaviours, lead users to respond in a way that is more congruent with the agent's emotional situation.

3.1 Agents as Targets of Empathy

We are walking down a corridor and in front of us a cute robot is moving purposefully toward a door. Halfway down the corridor, the robot loses its balance and falls. We rush toward the robot and try to help it. In such situations, we know that we evoke empathic concern toward artificial or fictional characters, such as narrative literary characters, film characters, and even game characters. But which processes of human empathy are relevant for artificial agents to evoke empathy in human observers? And what characteristics must the agents have for that to happen?

In the area of social agents, there has been research toward developing agents that evoke empathic responses in a human observer, in particular with the aim of promoting the human observer's attitudes and behaviour change.

But, for behavior change to occur, people need to be immersed in engaging situations, connect what happens in those situations to their interests and preferences, and, most of all, care. Recent such examples include social agents for intercultural training (focusing on intercultural empathy) (Mascarenhas et al. 2013), to help children deal with bullying (Paiva et al. 2004), or to train young doctors on how to interview patients with depression (Marsella et al. 2000).

Yet, to design these agents, particular care needs to be placed in the features of such agents so that the human observer can identify him- or herself with the character and, as a result, become immersed in the observed situation. The existing works in this area (see Section 4) involve the creation of situations where various processes and characteristics for social agents were considered for them to act as *targets of empathy*. These processes include:

— *Situation/context*: The situation that the agent is in is shared with the observer in such a way that empathic responses may arise, including aspects that characterise the context (e.g., social and cultural aspects);
— *Features of the observer*: The user's characteristics, such as age, gender, culture, and so on. These factors modulate the empathic response by the observer;
— *Empathy mechanisms*: One may consider two types of mechanisms for empathy-triggering in the users. Based on the theory automatic emotional responses, empathy can be triggered due to an emotional cue or expression of the agent. Furthermore, by creating a story and manipulating the relationship between the user and the agent , a more advanced cognitive perspective-taking mechanism can be triggered.
— *Characteristics of the agent (target)*: The agent's characteristics, such as degree of agency, its physical appearance, and even internal mental capabilities like the agent's "personal history" must be carefully designed and shared with the observer. Furthermore, the degree of expressivity of the target's affective state is also important because it constitutes a fundamental cue for the empathy process to occur (see Mechanisms for Empathy). This aspect is crucial, especially if the agent has a human-like embodiment. Nonverbal behaviours may also contribute to the elicitation of mimicry or the activation of shared neural networks in the human observer (cf. Section 2.2) and need to be designed/generated accordingly.
— *Empathy modulation*: The agent can evoke different degrees of empathy in the human observer based on several empathy modulation factors, such as the ability to establish and maintain social relationships (see Table 1). Other modulation factors include the agent's mechanisms for social relations. For example, rapport (Tickle-Degnen and Rosenthal 1990) may influence the perception of the agent as a close entity and thus modulate the empathic response of the user.

It is important to stress that these characteristics are expected to change the likelihood that the human observer feels empathy towards the agent.

3.2 Empathic Agents as Observers

The perspective of computational empathy on the development and evaluation of agents with empathic behaviour is one that has received great attention so far, especially in the virtual agents community. Research on empathic virtual agents has shown, among other things, that these agents are perceived as more likeable, trustworthy, and caring (Brave et al. 2005) and that they can build and sustain long-term socioemotional relationships with human partners (Bickmore and Picard 2005). In the field of human–robot interaction, there is a growing interest in studying how human interaction partners react to empathic robots (Cramer et al. 2010; Riek et al. 2010). Research on empathic robots has also shown that they are able to build and sustain long-term social relationships with human partners (Leite et al. 2014). In this regard, a crucial question is: *Which computational processes are needed for agents to be perceived as empathic?* In this section, we address this question by discussing the main computational processes relevant for simulating empathy in artificial agents.

For social agents to exhibit empathic behaviour autonomously, they need to simulate the empathic processes computationally.

Supported by the theoretical background reported in Section 2, we identified the following processes that agents should simulate to exhibit an empathic behaviour:

— *Empathy mechanisms*: The recognition of others' affective states is unanimously acknowledged as the first step of any empathic behaviour (cf. Section 2.2). Additionally, understanding others' actions is equally relevant because there are situations where the target deliberately inhibits the expression of certain emotions, which means that empathy can only be achieved by perceiving the overall situation (e.g., the target's actions and context). When the target is represented in a virtual environment as a virtual human, recognizing its actions and emotions is less complex than when it is a real human. In the latter case, the state-of-the-art systems for the perception of human targets still have some limitations, especially regarding the perception of affective states that can be extremely subjective and user -dependent (e.g., some people are more expressive than others). Acceptable performance rates for affect recognition can be achieved by combining several modalities such as vision, audio, physiological signals, context interpretation, and so on (for a complete survey on this area, please see Zeng et al. (2009)). But, combined with recognising the emotional cues of the target, empathic agents should be able to appraise the user's situation by using the empathy mechanisms described in Section 2.2. As such, the movements of the user or other agents can be perceived and even mimicked, including low-level behaviours by the motor systems such as eye contact and shared attention. Furthermore, the process of gradually adapting or changing the agent's affective state to match the affective state of the target agent/person based on the observed cues can also be implemented. Finally, mechanisms associated with cognitive empathy involve the perception of the context in which the event has occurred or even imagining the situation and the possibility of an event occurring, as well as reasoning about aspects of the situation at hand.

— *Empathy modulation*: As discussed in Section 2.3, empathy occurs to different degrees depending on different modulation factors (e.g., mood, personality, and social relationships) (see Table 1). While this aspect is crucial in enhancing agents' social behaviour, it has received little attention in the simulation of agents' empathic behaviour, as will be discussed in Section 5.

— *Empathic responses*: To be perceived as empathic, in addition to the internal empathic processes that trigger an empathic emotion, empathic agents should be able to display and express their internal states in a way that is understood by others (cf. Section 2.4) (e.g., in human–agent interaction). As will be discussed in Section 5, most existing empathic agents express their affective states and actions through verbal and nonverbal behaviours. Finally, agents must also have the capability to act in the environment by performing actions and/or displaying reactions that may change the other's state (e.g., emotion regulation or prosocial behaviour).

To successfully build empathy in social agents, researchers have to consider these three processes. However, depending on various factors such as the application domain, the context where the agent will be placed, or the agent's interactive capabilities, the degree of complexity of these processes may vary. Furthermore, we also need to consider that empathic responses may play a more or less important role depending on the context in which they are going to operate. If the agent or robot is not supposed to interact with humans or be exposed to a social setting, empathy does not need to be constructed as part as its competencies. For contexts in which task performance is the most relevant feature and social competencies are minimal, empathic mechanisms may not need to be engineered into agents. However, in scenarios where humans, agents, or robots are collaborating and executing tasks together—where emotional responses are important and social interaction becomes a relevant feature for the task—empathy will play an important role in guaranteeing a natural interaction flow. Yet, in some situations (e.g., competitive situations), the presence of the empathic capability may not show a clear impact in the interaction and may even lead to negative effects. For example, the study reported in Becker et al. (2005) and described later in this article, featuring the virtual agent MAX, has found that the display of empathic emotions in a competitive card game scenario was actually arousing and stress-inducing (by comparison to a nonempathic one).

3.3 Methodological Approaches

McQuiggan and Lester (2007) argue for two different approaches for developing computational models of empathy: analytical and empirical. The analytical approach is based on theoretical models of empathy established in psychological and neuropsychological research. This seems to be the most popular approach among researchers in the field (see Section 5). In analytical approaches, researchers take on board theories that describe human empathic processes generalised from behavioural or physiological observations and use them to computationally model these processes. This approach has also been the backbone of the most significant work in emotional modelling in artificial intelligence (AI) (Marsella et al. 2010).

The empirical approach relies on data collected in studies that examine peoples' social and emotional interactions in empathic situations. Here, empirical data can be used to train computational models using algorithms that generalise patterns of empathic behaviour. The obtained models are used to generate empathic behaviours in similar situations to those existing in the training data. However, to accurately train such models, large amounts of data are necessary, which means conducting several experiments and analysing many hours of data. Some researchers are also following this approach, as we will see in Section 5.

There are conceptual and methodological difficulties with both these approaches. While theoretically based approaches seem to be the most adequate to pursue, we should not forget that these theories have emerged mainly from behavioural data that were obtained by different researchers and represent accumulated knowledge abstracted into theories and models. This means that what we transfer to computational models is already an abstraction. As such, we need to find the computational mechanisms that can replicate those abstracted models and processes. This process can be quite challenging because most theoretical models do not provide enough detail to become concrete implementations with instantiated parameters and clearly defined rules that replicate cognitive processes. As a result, in the adaptation from theoretical to computational models, not only information can be lost in translation, but also the end result can deviate significantly from the original model due to misinterpretation. When many "engineering" operations are performed to fit a theoretical model into computational processes, the main argument that such a model is theoretically based becomes harder to prove. Furthermore, analytical approaches are often more difficult to evaluate, not to mention validate.

By contrast, empirically based approaches do not suffer from the "lost in translation" problem in the same way theoretical models do, but one should keep in mind that the developed models are also an abstraction of the data used to generate such models. There are further problems with this approach. First, to collect data, we need to know what type of data to collect and at which level of granularity and detail. The decisions on the type and granularity of data to collect can significantly influence the generated model, and they can compromise the main assumptions of this approach. Second, empirical approaches tend to be very domain dependent and thus more difficult to generalize to other application scenarios.

We additionally identified a third approach which takes a more user-centred view of the problem. Instead of focusing on the internal mechanisms necessary to generate empathy in an automated way (either based on data or theory), this approach encompasses simple models, carefully scripted and controlled in Wizard-of-Oz (WoZ) scenarios (Dahlbäck et al. 1993). These WoZ scenarios are developed mainly to evaluate and identify the impact that empathic virtual agents and robots have in human–agent interactions. In fact, this approach is adopted widely for many studies in social agents. Finally, more recently, some researchers are adopting a developmental approach (see the work by Asada (2015) and Lim and Okuno (2015)) where empathy is not directly built into the agents (although one can say that it is theoretically inspired), but emerges and evolves from other internal processes.

3.4 Different Communities Addressing the Same Challenges

The community of embodied virtual agents has embraced the problem of creating empathic agents. This problem has been addressed not only to try to give empathic abilities to agents, but also to try to make human partners feel empathy toward agents. The same research challenges are being addressed in the social robotics community.

However, empathy is such a broad field that some aspects of it are also investigated in other fields. For instance, the user modelling and user adaptation communities identify empathy as the capability of a system to adapt to the user and to the many different aspects and characteristics of that user. Still, we should distinguish between user adaptation and empathy in an interactive system. User modelling and adaptation is related to the capability of the system to model different states of the user: for example, how well he or she is progressing in a concrete task, the workload perceived, the experience that the user has on the topic, or even the plan that the user is following while performing a certain task. Empathy in social agents is related to user modelling and adaptation when we consider that the agent perceives the user and tries to identify her emotions, leading to the system's own affective response in a way that affects how the system will perform. As such, adaptation to the user does not necessarily mean empathy by the system. Additionally, the system may be empathic and respond in a very fast and reactive manner to the user relying on lower level empathy mechanisms such as mimicry (cf. Section 2.2) without explicitly storing any user model or changing to that user as the interaction occurs. However, for higher level processes, such as perspective/role-taking (cf. Section 2.2), storing and reasoning about the user becomes a main requirement.

3.5 Summary

In this section, we started by introducing our definition of empathic agents, which comprises two different perspectives: agents as *targets* of empathy and agents as *observers*. Subsequently, we identified the main requirements for creating social agents with empathy by considering the theoretical work presented earlier in Section 2. For agents as targets of empathy, the main elements that need to be considered and computationally manipulated are **situation and context**, **characteristics and behaviour of the agent (target)**, and **empathy modulation**. For agents as observers empathising with a target, the main elements are **an empathy mechanism**, **empathy modulation**, and **empathic responses**.

4 EVOKING EMPATHY WITH SOCIAL AGENTS AND ROBOTS

In this section, we discuss "empathic agents" as *agents that, by their design and behaviours, lead users to emotionally respond to a situation that is more congruent with the agent's situation than the user's*. The main questions we investigate here concern those elements in the agents and the created situations that evoke empathic responses in humans. Which situations are most conducive to empathic responses? How do particular characteristics of the agents (personality, anthropomorphism, etc.) lead to different emotional responses? For example, it is acknowledged that anthropomorphic agent characteristics are a factor that may enhance the social effects (and emotions) evoked in humans. And what about the users themselves? What responses are usually triggered, and how can they be captured and measured? For the past decade, systems using different types of agents have addressed some of these questions.

The literature review presented in this section provides concrete examples of systems where relevant aspects of these research questions are addressed. We review them in both the areas of virtual agents and social robots. We divide the review of agents evoking empathy into two section because the type of the agent's embodiment can affect users' perception and experience (Pereira et al. 2008; Bainbridge et al. 2011; Leyzberg et al. 2012).

4.1 Evoking Empathy with Virtual Agents

One of the first systems in this area was developed by de Rosis et al. (2005) who created an Embodied Conversational Agent (ECA) (Cassell et al. 2000) that acted as a therapist on the topic of healthy eating. The ECA was a head-only character driven by a model of dialogue inspired by research on behaviour change (Prochaska et al. 1992). The goal was to achieve some level of "involvement" between the agent and the user, leading users to respond to the "message" given by the agent to change their eating habits. An initial WoZ study was conducted with 30 subjects who interacted with the character controlled remotely by a human wizard. The interaction featured the character advising the user about what to eat and what not to eat. In the study, the users' responses were analysed, and the results have shown that the amount of social responses given by users was positively correlated with subjects' involvement in the scenario. However, it was not possible to reach any conclusion regarding the impact that the ECA had on evoking empathy. However, in spite of this, the study opened avenues for studying in a principled manner how ECAs can impact the user through empathy. The main questions one may ask are: (1) what types of behaviours should agents exhibit to trigger empathy, (2) what appearance and features are relevant for triggering empathy (given the empathy mechanisms discussed), (3) how important is the context and situation for triggering empathy, (4) what types of mediating factors should be considered and in what situations, and (5) how can we measure the effects that agents have on users in terms of empathic responses?

One system that uses empathy and identification to foster change was developed at the University of Southern California and aimed at improving the problem-solving skills of mothers of paediatric cancer patients. The system, Carmens Bright IDEAS (Marsella et al. 2000) (see Figure 3), targeted mothers of very sick children. The interaction featured a session between the main virtual character (Carmen, the mother) and a virtual counsellor (Gina). The goal was to help the user

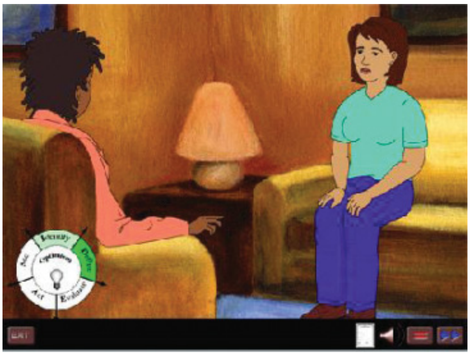


Fig. 3. Scenario of Carmen's bright IDEAS (Marsella et al. 2000).

Fig. 4. Agents in FearNot! (Paiva et al. 2004).

address her own stress by interacting with the characters and the story. In a 2D virtual world, Carmen and Gina discuss strategies for how to deal with different situations following a pedagogical approach called IDEAS (Marsella et al. 2000). The system and the agents in it were designed to ensure both dramatic and pedagogical goals without compromising the believability of or empathy towards the characters. The system's drama management is driven by the agent (Gina), which not only controls the dialogue that unfolds but also acts as director of the system, ensuring that critical goals for the drama are achieved. Thus, empathy is evoked through the relation the user establishes with Carmen (the character that the user partially controls) and Gina. Gina's actions rely on dialog and emotional expressions to motivate the user, whose interventions and responses influence Carmen's thoughts and the entire flow of the system. In fact, empathy plays an important role and was achieved not only by the creation of a strong context with very well-crafted situations inspired by real cases, but also by the presence of a director agent that manages the interactive agent-based drama.

Following a similar approach, FearNot! (Paiva et al. 2004, 2005) (see Figure 4) explored the empathic responses to virtual characters to achieve attitude change in children. The game featured embodied virtual characters in a 3D learning environment and used role-play to help children address bullying situations. By providing the child a "safe" environment for bullying situations to occur, FearNot! allows children to explore different coping strategies and techniques for dealing with bullying in an individual and personal manner. The main idea was that a child/user would play the role of a friend of a victim (a virtual agent) and would help that agent to address the bullying it suffers. This system explores empathy by embedding various constructs that lead the user/child to feel sorry for the victim character and thus act out in a proactive manner. By stimulating empathy towards the character, the system aimed at helping children who are victims themselves. The system works as follows: a child sees a bullying situation unfolding in a 3D virtual environment, which is populated with other virtual agents. One of them is a victim of bullying and one is a bully. This scenario is generated as a result of the autonomous actions of the autonomous characters that are present in the scene. Once a bullying episode occurs, the child/user is allowed to interact with the victim through a natural language interface and advise the victim character on what to do next.

Some modulating factors were considered in order to intensify the emotional responses of the user. In particular, FearNot! takes advantage of the similarity between the user and the main character (a modulating factor) in order for the empathic responses to become stronger. To do that, the development was centred around a concrete user group: 9- to 11-year-old children. Aspects such as the physical appearance and the character's clothes were studied and varied according to the target user's group (e.g., the characters for the UK schools had uniforms, whereas the characters for the German and Portuguese schools had normal clothes). The gender issue was also considered, and two types of episodes were explored: bullying between females, which involved relational bullying, such as exclusion from parties or games, name calling, and the like, and bullying between boys, which is more direct and physical, involving punching and hitting. Finally, to clearly address the issue of empathy evocation, the systems and the virtual characters must portray emotions that are easily recognizable by the users. To do that, the facial expressions of the characters were exaggerated and easy to identify (also because the characters were cartoon characters). Furthermore, the story and narrative explored (the contextual situation) in the system was based on concrete, real situations involving children of the target group age, which were obtained by user-centred methods.

FearNot! was evaluated in a study conducted in the United Kingdom and Germany where 1129 children (with a mean age of 8.9) (see Sapouna et al. (2010) for more details of the study) interacted with FearNot! over a few weeks. The main goal of the study was to assess whether the victims of bullying (who had been identified a priori) would escape bullying after being exposed to interaction with FearNot! The results obtained after a three-week intervention showed that children who were initially identified as victims were more likely to stop being victims than were the children in a parallel control group. This was a good result, yet it was only significant on the first reassessment and only for the UK children. Nevertheless, these results suggest that evoking empathy with autonomous virtual characters has the potential for social change on sensitive topics. However, there are many mediating factors that need to be considered.

As discussed in the opening sections, empathy occurs to different degrees depending on different modulation factors, such as the mood of the empathizer, the situation, the social relationships established, or even the perception of in-group and out-group. In a recent study conducted with virtual characters in an immersive virtual environment featuring a bar (Slater et al. 2013), users are immersed in a simulated reality where they perceive the world through wide field-of-view stereo-vision and sound, giving the strong illusion of being there (see Figure 5). The featured situation includes a 3D character (the perpetrator) who starts an argument with another virtual character (the victim), and the argument escalates to a level of violence. The user (a bystander) witnesses these unfolding events happening to the character (victim) and may, at some point, intervene. The study explores how the social identity of the user and the victim affect the speed of the intervention by testing the situation with fans of the Arsenal Football Club and creating the social identity of the victim character by giving him an Arsenal t-shirt. The results of the evaluation show that users (who were Arsenal fans) responded more quickly to the violent situation if the agent had a t-shirt of the Arsenal team. That is, by manipulating the social identity of the agent (through a simple T-shirt), different empathic and prosocial responses were triggered in the user.

Recently, virtual agents are being used to simulate patients (Gratch et al. 2014; Stevens et al. 2006), doctors (Robb et al. 2014), and nurses (Bickmore et al. 2009), not only for training situations but also for interacting with patients. One example is the SimSensei Kiosk (DeVault et al. 2014) that features a virtual human who acts as an interviewer designed specifically to create interactions that make users feel comfortable and willing to share information. The SimSensei Kiosk is aimed at addressing problems of depression, anxiety, or post-traumatic stress disorder in patients. In that context, empathy is seen as an essential element of the relationships between physicians, therapists, and patients. In fact, medical students are often trained to exhibit empathy in their communication with patients. The work by Cordar et al. (2014) features an application to train young doctors using a virtual patient (a 21-year-old student with depression) how to interview people with depression. The study evaluating this system explored how the presence of a backstory

Fig. 5. Responses to a Violent Incident in an Immersive Virtual Environment (Slater et al. 2013).

Fig. 6. Empathic Agents for Intercultural Training in Traveller (Mascarenhas et al. 2013).

for the virtual agent had an impact on the empathic responses by those medical trainees who interacted with that virtual patient. To do that, two conditions were created: one where the virtual patient was given a backstory (created as a set of four cut scenes assembled with The Sims 3 game) and a second where the virtual agent simply interacts with the student without providing details about its life. A total of 35 first-year medical students interacted with one of the two conditions of the virtual agent. Immediately after the interaction, they had to talk with a human standardized patient actor who was also portrayed as suffering from depression. The interactions were recorded and adequately coded. The results showed significant differences in two items associated with communication between the participants and the human standard patient: Participants offered more encouraging, supportive, and empathetic statements in the backstory condition and appeared more warm and caring. These results suggest that the presence of backstories in the empathy-evoking agents (thus providing context and grounding the interaction in past events) is an efficient way by which virtual agents can facilitate greater empathic responses from users.

Some recent work in the area of video games has shown that playing prosocial video games may have an impact on altruistic responses from the players in subsequent situations. Embodying a character, or even experiencing flying in a virtual reality setting (like a superhero), has been shown to lead to more prosocial behaviour (see the work by the Bailenson team (Rosenberg et al. 2013)). In the area of cultural sensitivity, the work by Mascarenhas et al. (2013) has focused on the triggering of intercultural empathy in humans as a way to train intercultural awareness. Traveller (see Figure 6) is a serious game that follows an interactive storytelling approach, where the user plays an active role in a story that takes place in different fictional countries. The user travels to countries with different cultures and solves practical problems, such as finding directions to a hotel. Solving these problems requires the user to engage in social interaction with small groups of autonomous characters that will behave in a culturally distinct manner, particularly in the way they treat users and respond to their actions. By exploring different cultures with in-group and out-group, the aim is to raise intercultural awareness and empathy. To do that, agents that can be parameterized with different cultural characteristics were built to test the impact that these agents would have on users from different cultures. The initial results showed that users' responses indeed depend on both the culture portrayed by the agents and the culture of the user as two mediating factors.

Table 2 provides a summary of the analysis and classification of works on empathy evocation with virtual agents with respect to their approach in terms of *situation/ context*, *features of the*

Table 2. Summary of Related Work on Virtual Agents that Evoke Empathy

Authors	**Situation/ Context**	**Observer's Features**	**Agent's Characteristics and Emotion Expression**	**Empathy Modulation**
F. di Rosis (2005)	Dialogue with an ECA providing advice on eating habits	Teenagers (seeking for advice), passive role	ECA (head only), WoZ, use of emotions in facial expression and voice	Through the situation
Marsella et al. (2000)	Interactive drama featuring a mother of a sick child and a therapist	Adults, active role (controlling actions of a virtual agent)	ECA in a scripted virtual environment; use of facial expression, voice, actions	Identification with the character and the situation
Aylett et al. (2005)	Interactive drama featuring a virtual agent suffering from bullying	Children (aged 9–12), active role (influencing the decisions taken by agent)	ECA in a virtual environment, autonomous generation of behaviour; use of facial expression, voice, actions	Identification with the character and the situation
Slater et al. (2013)	Immersive virtual environment featuring a bar with agents	Adults, active role (possibility to intervene or not when witnessing an aggression)	ECA in a scripted virtual environment; use of facial expression, voice, actions	Through the identification with the victim's character
Mascarenhas et al. (2013)	Game with virtual characters featuring different cultures	Young adults, active role (interaction with characters from different cultures)	ECA in a virtual environment, autonomous generation of behaviour; use of facial expression, voice, actions (culture-dependent)	Through the identification with the culture of the agents
Cordar et al. (2014)	Dialogue with an ECA featuring a virtual patient with depression	Adults (training doctors), active role (interacting with the agent through text	ECA scripted behaviour selected from a large set of utterances; mainly natural language	Through the presence (or not) of a back story

observer (user), *characteristics of the agent and its emotion expressiveness*, and *modulation factors used*.

4.2 Evoking Empathy with Robots

Another area that has seen significant investigation into how to evoke empathy in users is the area of social robotics. Without a doubt, this area has now become an exciting and challenging field of research (Breazeal 2004; Fong et al. 2003) as robots are increasingly endowed with capabilities to act in a social way with others, including humans. Because of their physical presence, and with interaction capabilities, social robots have the potential to evoke emotional responses in humans.

Such potential for empathy is shown in the work by Matsumoto and colleagues (Matsumoto et al. 2004), who conducted a study based on "letters" sent by fans of a Japanese interactive robotic doll (called Primopuel) as the material for understanding the relations fans have to fictional characters and robots. In that study, a link was found between the "attachment" that the fans had to the interactive robotic doll and the empathy evoked by it. Furthermore, in a study conducted by L. Riek and colleagues (Riek et al. 2009), the impact that the "human-likeness" of a robot has on its potential for evoking empathy was explored. To do so, the researchers designed an experiment where five "robots" (protagonists) with different human-likeness features were tested with a web-based questionnaire. At one extreme, robots such as Roomba, a robotic vacuum cleaner, were featured, representing robots that do not exhibit human-like features. At the other, scenes with a human child, extracted from the film "The 400 Blows," were portrayed representing a "very human" type of character. For each protagonist in the study, two sets of videos were selected. In one set, the protagonist was treated in a neutral manner by others. In the second set, the protagonist was treated in an unfair and emotionally negative way by another human. For example, in one video, the protagonist was mocked or even hit. Two main hypotheses were considered in the experiment: The first one concerned the idea that people with a higher Empathy Quotient (EQ) (Baron-Cohen and Wheelwright 2004) score will tend to feel more empathic towards robots than will people with a low EQ score. This hypothesis was not proved by the experiment conducted. The second research question concerned the difference in terms of empathic reaction. The hypothesis was that people would be more empathic towards more human-like protagonists. Thus, participants were asked to rate video clips with the protagonists stating how sorry they felt for the protagonist, with each video clip lasting approximately 30 seconds. The results showed that people felt more sorry for the human and the more humanoid robots than for the nonhumanoid ones (the Roomba and AUR robot). These results are interesting and raise important questions concerning the appearance of robots and its relation to the robot. However, one should note that because the experiment relied on scenes from films with five quite different protagonists, the situations portrayed for emotional and neutral scenes were different for each protagonist.

Investigating a similar effect but placing users in a situation with real robots, Kwak et al. (2013) conducted a study with the Mung robot (Kim et al. 2009) to evaluate if the perceived "agency" of the robot impacts the empathy felt towards it. Inspired by the Milgram's experiment (1963), participants were given the role of teaching a robot. Then, the robot had to answer a test to assess what it had learned. At this stage, if the robot failed to answer the questions correctly, participants could give electric shocks to the robot. The robot used, Mung (see Figure 7), was designed to be able to recognize emotional speech and express emotions through light-emitting diodes, thus simulating some type of "bruises" to express a negative emotional states. Two conditions were created: one where the robot acted as a mediator (a type of proxy of a human) and a second one where the robot is perceived as an autonomous entity acting in an independent manner. These two conditions allowed for a manipulation of the "agency" level of the robot. Thirty children participated in this experiment, and the results showed that participants empathized more with the robot that acted

Fig. 7. Scenario Mung interacting with humans (Kwak et al. 2013).

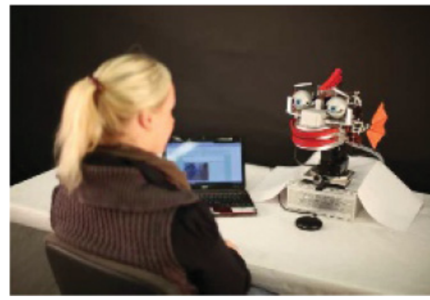


Fig. 8. Scenario of EDDIE playing Akinator (Gonsior et al. 2011).

as a mediator than with the "autonomous" one. These results, in spite of being limited to the particularities of the population and the concrete embodiment chosen, show that empathy towards robots is influenced not only by their human-like features but also by their perceived agency, suggesting that the more "autonomous" robots are, the less empathy people may feel towards them. In a similar study D. Jo et al. (2013) investigated the impact that the presence of a laughing robot had on the expression of emotions of people watching a funny film. The study compared two conditions: one where the participants watch the film with a robot laughing at certain parts of the films and another condition where participants watched the same film but with a "laughing" human instead. The results of the study carried out with college students seem to indicate that the robot discouraged positive responses from the participants (in contrast to the human, who seemed to increase positive responses). With regard to negative emotions, both conditions seem to decrease the expression of negative feelings.

Gonsior et al. (2012) created a system where a robotic head, EDDIE (see Figure 8), had perceptive capabilities to detect the facial expressions of the user and to understand specific keywords. The robot was built with a social component that allowed it to react to various situations and change its internal state accordingly (including emotional state). This robotic head was used with the bonding game Akinator.[1] Users and the robot interacted in spoken dialogue during guessing sessions where EDDIE tried to guess the "person" the users were thinking off. A study was performed with this scenario, and the results showed that people felt more empathy towards the robot in the condition where the robot expressed emotions and mirrored the user's expressions, in contrast to the neutral condition, where the robot did not express emotions. Further, the same team (Gonsior et al. 2012) explored which mechanisms are more likley to trigger prosocial human reactions towards a robot. Again, two types of mediating factors were used: explicit emotion expression by the robot and similarity through emotional adaptation. These were achieved by using the PAD model in the same way as previous work by Boukricha and Wachsmuth (2011). The model was tested in a controlled scenario where users interacted with EDDIE. Two conditions were tested: a control condition where the robot did not exhibit emotional responses nor was able to generate emotional adaptation and a condition where the empathic characteristics were present. After interacting with the robot through small talk and playing Akinator, participants were faced with the possibility of helping the robot in a task of classifying objects in a picture. The results showed that people were more likely to help the robot in the condition where empathy (and generally emotional behaviour) was used. Additionally, in the condition where empathy was used, people rated

[1]See http://en.akinator.com/.

the robot higher in the dimensions of anthropomorphism and animacy in the Godspeed (Bartneck et al. 2009) questionnaire.

Rosenthal-von der Pütten et al. (2013) and colleagues addressed the similar problem of "how robots affect human's emotions and how humans treat robots based on the sentiment they evoke." Using both psychophysiology and self-reported measures, a study was conducted with two sets of videos of Pleo, a robot-like dinosaur, to assess the emotional responses of users to emotionally charged situations with this robot. In one set of videos, Pleo is portrayed being tortured (choked, beaten up, dropped, etc.) by a person wearing a black jumper (the videos feature the arm of the person), while in the second set of videos Pleo is treated nicely. In both cases, participants could also hear the sounds made by Pleo according to the situation shown. The measures used aimed to obtain the emotional responses and attitudes of people toward the robot and included both physiological arousal (captured through electrodermal activity) and self-reported measures, including the Positive and Negative Affect Schedule (PANAS) questionnaire (Watson et al. 1988). Furthermore, an additional scale with 12 items was constructed to assess how the participants felt about Pleo and the situation. These items included, for example, "I felt pity for the robot," "I felt for the robot," or "I did not mind what happened to the robot at all." Finally, participants were also assessed regarding various traits, including "empathy" in particular (by using a German version of the IRI index (Davis 1983b)). The results showed that people reported more pity and empathic concern when Pleo was tortured than when Pleo was treated in a natural manner. Furthermore, this situation was shown to influence people's subjective feelings because they reported more negative and less positive emotion after watching the video in which the robot was tortured compared to the ones where Pleo was treated nicely. The authors also investigated whether the effects of prior exposure to the robot (by interacting for approximately 10 minutes with it) would cause a difference in the results (thus seeing if interacting for a short period was a mediating factor in the situation). Surprisingly, the results did not show any effect on the empathic responses. However, the results achieved are very important because they reveal that people indeed experience empathic concern for artificial creatures (agents and robots) in spite of the nonhuman form studied in this case.

In a more recent study, Hayes et al. (2014) reported a study where a Keepon robot competed with a human participant in a counting task. The study had three different conditions: neutral (the robot's comments during the game were neutral), self-directed (the robot appealed to the participant, urging him or her to let the robot catch up for the robot's own benefit), and externally directed (the robot would make distressed pleas to the participant to let the robot win for its programmer's benefit). The authors found that, if a social robot is capable of evoking empathic responses from participants, they will be more willing to help the robot in both empathy-related study conditions. Additionally, the externally directed condition was the one in which participants were more likely to help the robot.

An interesting study was also conducted exploring the difference between a physical robot and its virtual representation with regard to the empathic responses of users towards a character (Seo et al. 2015). Inspired partially by Batson et al.'s experiments (Batson et al. 1997), which were performed to assess peoples responses towards a member of a stigmatized group, and using a similar instrument to assess situational empathy, participants in this study witnessed a scripted "sad" situation about the NAO robot (or its virtual counterpart) exhibiting a faulty behaviour. Using a WoZ-controlled scenario, participants interacted with the NAO robot performing a collaborative task, and, at some point, the robot started having "problems" due to a virus, which made it work badly. Eventually, that fault led to its memory being deleted. The same scripted situation was mirrored with a virtual representation of the same robot. The study results indicate that it is possible to induce empathy in a human–robot interaction and that the empathic responses were higher with the physical robot than with its virtual representation. The embodiment of a "physical" entity appears to have an impact on the empathic responses of users. Although some of the previous studies concerning social responses to robotic versus virtual agents would indeed suggest that there is a difference in the social interaction, this study shows that empathic responses are affected by the physicality of the situation.

Empathy, and cognitive empathy in particular, is usually associated with placing oneself into the shoes of another. Using this idea, an interface was designed in the work of Marti et al. (2013) to allow the user to look through the eyes of a robot, thus taking its perspective. Through a tablet-based interface, users are able to control the robot and put themselves in the perspective of the robot, thus seeing what the robot sees and even understanding the intentions of the robot. Furthermore, the remote-controlled interface "masks" the interaction by adding "filters" that aim to promote perspective-taking. For example, one mask may signify "surprise" or even "disgust." The "mask" idea was evaluated through a set of videos presented to participants who judged the interface with and without perspective-taking. The results obtained showed that the interface had some impact on the understanding of the intentions of the robot. Furthermore, in a limited way, the results showed that participants found that the scenarios containing the empathic masks were more engaging, resulting in more empathic concern and perspective-taking.

There have been some advances in the creation of robots that trigger empathy, but there is still a lot to be studied. Table 3 provides a summary of the analysis and classification of works on empathy evocation with robots with respect to their approach in terms of *situation/context*, *features of the observer*, *characteristics of the agent and its emotion expressiveness*, and *modulation factors used.*

4.3 Discussion

Evoking empathy using artificial social agents, either robotic or virtual, is an area that deserves further attention from the research community in the near future. Significant research has been conducted on virtual agents that lead users to change their behaviour (persuasive technology) or, in the area of robotics, to motivate users to collaborate with robots. Tables 2 and 3 summarize the work reported here.

To determine whether a virtual agent or a robot evokes empathic responses in the users, a number of measures need to be adopted to assess the levels of empathy in users. There is no consensus yet on how to assess the empathic responses of robots. For example, in the study performed by Kwak et al. (2013) empathy was measured as the difference between the reported perceived emotion in the robot and the participants' reported emotional response. Other researchers relied on behavioural responses (e.g., how users act, their emotional responses, their prosocial behaviour, etc.) (Slater et al. 2013), or on self-reported measures based on questionnaires. Furthermore, the adoption of well-established questionnaires, such as the friendship questionnaire (Mendelson and Aboud 1999) or social presence (Gunawardena and Zittle 1997), are promising candidate measures given the link between empathy and social relationships. However, in a study designed by Chaminade et al. (2010), rather than looking at high-level behaviours or attitudes, they investigated if various neural substrates involved in the way humans perceive and respond to emotions in other humans are also involved in their responses to robots. The results showed that those neural responses are also found in the responses to emotions in robots. This analysis was performed using functional magnetic resonance imaging (fMRI) to assess how the brain responds to emotional stimuli consisting of human facial expression, on one hand, and robot facial expressions, on the other hand. This study, and subsequent studies following this approach, bring to light the fact that humans respond to these robots in a very deep and human way.

However, in more interactive settings, and particularly with robots, it is difficult to judge the implications of empathy evocation in a broader sense and to foresee the ramifications that may arise

Table 3. Summary of Related Work on Robots that Evoke Empathy

Authors	Situation/ Context	Observer's Features	Agent's Characteristics and Emotion Expressiveness	Empathy Modulation
Gonsior et al. (2011)	Akinator game with a robotic head which guesses what the person is thinking	Adults, active role (speech interaction)	Perceives and mimics user facial expressions, recognizes key words. Use of facial expressions, voice	Mimics user facial expressions, situation (playing a game together)
Gonsior et al. (2012)	Akinator game with a robotic head which guesses what the person is thinking; user helps the robot (or not, in another task)	Adults, active role (speech interaction and decision of whether or not to help the robot); perceives and mimics the user facial expressions, recognizes key words	Facial expressions, voice, prosocial behavior	Mimics user facial expressions, situation (playing a game together)
Doori Jo et al. (2013)	A humanoid robot and a user watch a film together	Adults, passive role (watch a film)	Wizard controlled; use of laughter (pre-recorded)	Situation given by the film and the presence of the robot
Kwak et al. (2013)	Participants teach a robot and then are allowed to punish the robot if it answers incorrectly	Children, active role (teaching and punishing the robot)	Robot responds to the "punishments" with bruises; "bruises" appear in the robot's embodiment	Through the degree of agency of the robot (manipulated in the experiment)
Marti et al. (2013)	Controlling a robot in a home scenario	Elderly, active role (controlling the robot by putting themselves in the shoes of the robot)	Robot partially controlled by the users; use of an "emotional mask" as the robot interface that allows the user to perceive the robot's emotional state	Perception of the world as the robot perceives it

(Continued)

Table 3. Continued

Authors	Situation/ Context	Observer's Features	Agent's Characteristics and Emotion Expressiveness	Empathy Modulation
Hayes et al. (2014)	Competitive counting game between a user and a robot	Adults, active role	Robot asks participant to let it win by evoking self or externally directed reasons; use of expressive voice (pre-recorded).	Situation (based on the state of the game)
Seo et al. (2015)	Collaborative game in which a robot faces a sad situation	Adults, active role	Robot controlled by WoZ exhibits faulty behaviour; expressive voice and gestures used	Situation (robot experiencing memory loss)

when robots and virtual agents interact with us on a daily basis. Furthermore, the ethics behind the use of purposeful empathy-evoking situations with agents and robots need to be addressed, and the subsequent results have to be taken into consideration.

As advocated by Reeves and Nass (1996), computers and thus agents can trigger responses that are more appropriate to other humans, and empathy can be—and is—one such response. People feel distress because of an agent's or a robot's distress, sad because an agent is sad, and so on. The work developed so far is only the starting point of a challenging area of research that places machines together with humans as prosocial actors.

5 EMPATHY IN VIRTUAL AND ROBOTIC AGENTS

Research on empathy in virtual agents and robots has been increasing. Some of the work is primarily focused on investigating the impact of empathy on human–machine interaction, with the findings highlighting empathy's major role in enhancing human–machine interaction. Motivated by these findings, further work has mainly focused on providing computational models of empathy shaped by the theory. In this section, we discuss seminal work on allowing virtual agents and robots to empathize, and we conclude by classifying this work with respect to the requirements listed in Section 3: *empathy mechanism*, *empathy modulation*, and *empathic outcomes*. As in the previous section, we separate our review by embodiment type (virtual vs. robotic).

5.1 Empathy in Virtual Agents

Prendinger and Ishizuka (2005) proposed the Empathic Companion, an embodied virtual character that supports users in the setting of a virtual job interview task (see Figure 9). By means of biopotential signals and context-based information, the Empathic Companion recognizes users' affective states and responds with empathic feedback during a simulation of a job interview. There are three alternatives for empathic feedback. First, the character shows empathy for a user whose emotional state is recognized as indicating high arousal and negative valence. Second, the character encourages a user whose emotional state is recognized as not aroused. Third, the character congratulates a user whose emotional state is recognized as indicating high arousal and positive

Fig. 9. The empathic companion (Prendinger and Ishizuka 2005).

valence. The purpose of developing the Empathic Companion was to investigate the impact of empathic feedback on reducing users' stress levels in frustrating conditions. Biopotential signals are used to measure users' perception's of the Empathic Companion and to thus evaluate the impact of its empathic behaviour on users. Two versions of the Empathic Companion were distinguished: a *nonempathic* version where the character does not provide empathic support and an *empathic* version where it does. The findings show that, when receiving an interview question, users' arousal and stress levels are significantly reduced in the empathic condition of the study.

Becker et al. (2005) designed a Skip-bo card game scenario where the virtual human MAX (Kopp et al. 2005) provides game partners with empathic feedback (see Figure 10). The virtual human recognizes the partner's emotional state based on the previous work by Prendinger and Ishizuka (2005). As empathic feedback, MAX displays facial expressions mirroring the partner's recognized emotional state. As in Prendinger and Ishizuka (2005), biopotential signals are used to evaluate the impact of MAX's empathic behaviour on game partners. In this regard, four different conditions were distinguished: a *nonemotional* condition where MAX does not display emotions, a *self-centred* condition where MAX displays his own emotions with respect to his own game moves, a *negative empathic* condition where MAX displays his own emotions and non-congruent emotions to those of his game partner, and a *positive empathic* condition where MAX displays his own emotions congruent to those of his game partner. The results show that the display of empathic emotions in a competitive card game scenario is arousing and stress-inducing.

Brave et al. (2005) evaluated the psychological impact of an empathic agent on human users in a scenario of a casino-style blackjack. The agent was represented by a picture displaying a human face and was playing with the human user against a disembodied dealer. Based on a very simple emotion model, the agent communicated emotions by facial and textual expressions. Positive emotions were expressed if the agent or the user won, and negative ones were expressed if the agent or the user lost. To evaluate the impact of the agent's empathic behaviour on users, Brave et al. asked participants to complete a questionnaire after the game. In this regard, they distinguished two different conditions, a *self-oriented emotion* condition and an *other-oriented empathic emotion* condition. Their findings show that in the *other-oriented empathic emotion* condition, the agent was perceived as more likable, trustworthy, and caring than in the *self-oriented emotion* condition.

Bickmore and Picard (2005) introduced the relational agent Laura as a computer agent able to build and sustain long-term socioemotional relationships with human users. As the advisor in a fitness computer program (see Figure 11), Laura interacts with the user via speech and nonverbal

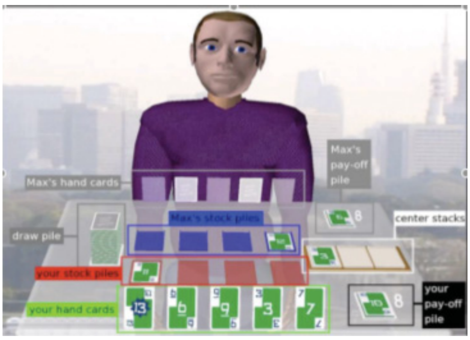


Fig. 10. The empathic agent MAX playing Skip-Bo (Becker et al. 2005).

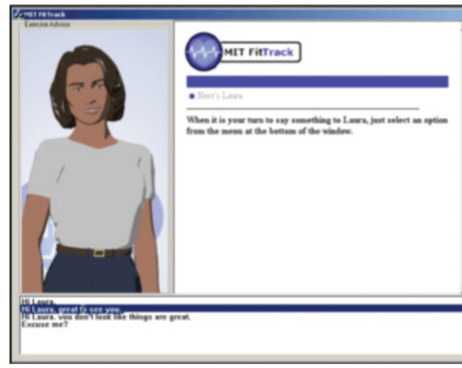


Fig. 11. The relational agent Laura (Bickmore and Picard 2005).

behaviour. In this scenario, Laura exhibits empathic behaviour, realized through scripting, as a crucial socioemotional aspect that contributes to the introduced relational agent approach.

McQuiggan et al. (2008) proposed an inductive framework called *CARE* (Companion Assisted Reactive Empathizer) to model reactive and parallel empathy as defined by Davis (1994). Their framework is based on learning empirically grounded models of empathy during human–agent interaction. In a scenario of a virtual training environment (Crystal Island), user data, such as their actions, intentions, affective states, age, and gender, are collected during their interactions with virtual characters. Accordingly, the virtual characters respond with either parallel or reactive empathy, and the users are asked to evaluate the characters' empathic reactions using a 4-point Likert scale. In a learning phase, the data collected during interaction are used to learn models of empathy from "good examples". In a test phase, the induced models of empathy are used to drive the virtual characters' empathic behaviour. An evaluation of the proposed framework with respect to training and test data sets shows that it generates appropriate empathic behaviour.

Ochs et al. (2012) provided a computational model of empathic emotions based on the results of an empirical analysis and a theoretical model of emotions. The empirical analysis was conducted to determine the conditions of elicitation of users' emotions during human–agent dialog interaction. The theoretical model of emotions is based on Scherer's appraisal theory of emotion (Scherer 2010). Accordingly, based on a number of appraisal variables, the type and intensity of a user's elicited emotion are determined, and the agent's empathic emotion from the same type is triggered towards the user. The agent is represented as a 3D talking head (see Figure 12). Its empathic emotion is conveyed by facial expressions, and its base intensity is affected by the degree to which the agent likes the user and by the degree to which the user deserves his current situation (cf. (Ortony et al. 1988)). To empirically evaluate the agent's empathic behaviour, three conditions were defined similarly to those defined by Becker et al. (2005): a *nonemotional condition*, an *empathic* condition, and a *noncongruent emotional* condition. The results show that, in the *empathic* condition, the agent was perceived more positively while in the *noncongruent emotional* condition it was perceived more negatively.

Rodrigues et al. (2009, 2014) introduced a generic computational model of empathy based on neuropsychological findings on empathy. Their model considers that empathy is a process that involves two stages: the empathic appraisal (combining emotion cues recognition and role-taking) and the empathic responses. The model was integrated into an existing affective agent architecture (Dias and Paiva 2005).

Fig. 12. Empathic 3D talking head (Ochs et al. 2012).

Fig. 13. A scenario of four virtual characters with different roles and values of affective link (Rodrigues et al. 2009).

A perceived emotional cue in another agent is input to the emotion recognition module of the empathic appraisal component, which provides a number of candidate emotions where the one with the highest likelihood is set as the default. A perceived event that elicits emotions in another agent is input to a self-projection appraisal module of the empathic appraisal component where, by means of self-projection, the agent assumes the other's situation and appraises it with its own appraisal rules. The results of both modules are combined to provide an empathic emotion as the output of the empathic appraisal component. Within the empathic appraisal module, the intensity of the empathic emotion is affected by several modulation factors, such as the affective link between the agents and the empathizing agent's mood. The empathic emotion output by the empathic appraisal component triggers a situation-appropriate action within the empathic response component. The proposed model was applied in a scenario of four virtual characters that were provided different roles and values of affective link (see Figure 13). To evaluate their model, the authors defined two versions of this scenario: a *nonempathic* and an *empathic* version. The results show that the perceived values of empathy and affective link between the empathizing agent and a target agent were significantly higher in the *empathic* version.

Boukricha and Wachsmuth (2011) proposed a computational model of empathy for virtual humans. Their model is based on three processing steps (cf. Section 2): the empathy mechanism as the process by which an empathic emotion is produced; the empathy modulation as the process by which the empathic emotion is modulated by factors like the virtual human's mood and its social relationships (such as liking and familiarity) and a degree of empathy is calculated; and the expression of empathy as the process by which the empathic emotion is communicated based on a repertoire of modalities. As empathy mechanisms, they considered facial mimicry and situational role-taking as defined, respectively, by Hoffman (2001) and Higgins (1981). The proposed model was realized for the virtual humans MAX Kopp et al. (2005) and EMMA (Boukricha 2013) in two different interaction scenarios. In the spatial interaction task scenario, the virtual human MAX engages in a cooperative tower-building task with a human partner (Boukricha et al. 2011). Within this scenario MAX's empathic emotion is elicited by means of situational role-taking, and his degree of empathy impacts his spatial helping action towards the human partner. In the conversational agent scenario, both MAX and EMMA can engage in a small-talk dialog with a human partner who can trigger their emotions positively or negatively by being either kind or rude to

Fig. 14. The virtual human EMMA empathizing with MAX in a conversational agent scenario (Boukricha 2013).

them (see Figure 14). In this scenario, EMMA empathizes with MAX's triggered emotions to different degrees depending on her mood and her defined value of relationship with MAX. EMMA's empathic emotion is triggered by means of facial mimicry. An empirical evaluation of the model (Boukricha et al. 2013) was conducted in the context of the conversational agent scenario. In this regard, three conditions were defined that distinguished three different degrees of empathy: neutral, medium, and maximum empathy. The results show that EMMA's expression of empathy, her degree of empathy, and her value of relationship to MAX were significantly recognized in all three conditions.

One of the most successful applications for empathic virtual agents is in the area of health behaviour change. A recent example is the work by Lisetti and colleagues [2013] who developed a virtual counsellor that delivers empathetic advice with the goal of reducing alcohol consumption. This embodied conversational agent uses decision trees that take into account the user's facial expression, head movement, smiles, valence of the previously asked question, and consequent user's answer to decide the verbal and nonverbal behaviour of the agent. These behaviours include, for instance, mimicking the user's facial expression or reflective listening. The acceptance of the virtual counsellor was evaluated in a controlled study with three different conditions: the empathic agent, an agent with less empathic capabilities (with neutral facial expressions), and a text-based interface without an embodied agent. Several acceptance metrics (e.g., attitude towards the agent, intention to use, perceived enjoyment) were combined with agent-oriented perceptions (e.g., animacy and perceived intelligence) in a postinteraction questionnaire. Although the differences between the two embodied agents were not always clear, participants showed clear preference towards the embodied agent in metrics such as reported intention of use.

5.2 Empathy in Robotic Agents

In the field of social robotics, researchers have only more recently started to assess the effects of endowing robots with empathic behaviour (compared to the research in the virtual agents community). One possible reason for this is that it is relatively easier to infer the user's affective states in interaction scenarios with virtual agents than it is in human–robot scenarios. Usually, while interacting with virtual agents, the user is sitting in front of a computer screen and there is often the possibility of selecting predefined dialogues to inform the agent of the user's emotional state, as in Bickmore and Picard (2005). The interaction with robots tends to be more open-ended, and thus perceiving the user's affective state is more challenging. Nevertheless, along with the significant

Fig. 15. Participant interacting with the chimpanzee head robot (Riek et al. 2010).

improvements in automatic affect recognition using different modalities such as vision, speech, or physiological signals (Zeng et al. 2009), the study of empathy in social robots has witnessed significant progress in the past few years in diverse application domains.

Health care, entertainment, and education are some of the application areas where robots need to interact with humans in a natural way. As we move towards truly engaging robots, empathy needs to be addressed and modelled. Tapus and Mataric (2007) proposed a model of empathy to be used in the context of Socially Assistive Robotics. Several capabilities were identified, for example, recognizing and interpreting another's emotional state, the capability of processing and expressing emotions, and the capability of communicating and perspective-taking. These processes were modelled based on the work of Davis (1983a), where empathy is not seen only through its processes but also by the outcomes it leads to.

Hegel and colleagues [2006] conducted a study with an anthropomorphic robot that recognizes the user's emotional state through speech intonation and then mirrors the inferred state using a corresponding facial expression. The results suggest that users who interacted with this version of the robot found its responses adequate in terms of appropriateness to both the social situation and timing when compared to subjects who interacted with the robot without affective expressions. In another study (Riek et al. 2010), a robot with the form of a chimpanzee head mimics the user's mouth and head movements (see Figure 15). When interacting with this robot, most subjects considered the interaction more satisfactory than did participants who interacted with a version of the robot without mimicking capabilities.

Cramer et al. (2010) studied how empathy affects people's attitudes towards robots. In a between-subjects design, two groups of participants viewed a four-minute video with an actor playing a cooperative game with an iCat robot (see Figure 16). The experimental manipulation consisted of causing the robot to express empathic behavior towards the actor in an accurate or inaccurate manner (i.e., incongruent to the situation) depending on the control group. In this study, there was a significant negative effect on the user's trust in the inaccurate empathic behaviour condition. Conversely, participants who observed the robot displaying accurate empathic behaviours perceived their relationship with the robot as closer.

More recently, a scenario where an iCat robot is able to play chess with children while reacting emotionally to the moves played on the chessboard was developed to study empathy in human–robot interaction (see Figure 17). After every child's move on the chessboard, the robot provides empathic feedback on that move by conveying facial expressions influenced by the child's affective state and the state of the game. To infer children's affective states, an SVM-based Affect Recognition System returns the probabilities of children's valence (positive, negative, and neutral). In addition to the empathetic facial expressions, if the child's affective state is negative and below a certain threshold, the robot also displays social supportive behaviours. The developed empathic

Fig. 16. Screenshot of the video and task used in Cramer et al. study (2010).

Fig. 17. Child playing chess with the iCat robot (Leite et al. 2014).

model had a positive impact in long-term interaction between children and the robot (Leite et al. 2014). In a repeated interaction study where the same group of children interacted with the robot five times, children's ratings of social presence, engagement, help, and self-validation remained similar after five weeks, contrasting with the results obtained in an initial long-term study where the robot was not endowed with the empathic model. Video observation and interviews were also employed to further analyse these results, suggesting that, over time, users were even more aware that their actions influenced iCat's behaviour. The results also showed that children felt supported by the empathic robot in a similar manner to how, in general, children feel supported by their peers.

The same authors developed a variation of this scenario in which the iCat robot acts as a social companion to two players in a chess game (Leite et al. 2013b). However, in this case, the emotions of the human participants were captured using perspective-taking. In other words, instead of using the multimodal affect detection system, the robot used the context of the game to "put itself in the shoes of the players." After assessing the game from the perspective of one of the players, the empathic responses of iCat were modulated by its relationship with the players (companion or opponent). Towards the companion, iCat reacted in a very empathic way, providing encouraging comments and being enthusiastic towards his or her moves. In contrast, to the other player, iCat's relation was mostly neutral, and, as such, the comments were mainly factual and less emotive. The results of this study indicate that users towards whom the robot behaved empathically perceived the robot as friendlier, as compared to participants to whom the robot displayed neutral behaviours only.

5.3 Discussion

The computational modelling of empathy in virtual agents and robots provided valuable findings on the role of empathy in enhancing virtual agents' and robots' social behaviour as well as support for its underlying theories and hypothesis.

Tables 4 and 5 show the classification of seminal works on empathy in virtual agents and social robots with respect to their contribution to the requirements listed in Section 3: **empathy mechanism**, **empathy modulation**, **empathic outcomes**, **adequacy**, and **universality**.

Regarding the empathy mechanism, facial mimicry as defined by Hoffman (2001) (imitation and feedback; cf. Section 2) has not received much attention, with the exception of the work by Boukricha et al. (2013). However, in the empathy model by Boukricha et al. (2013), the empathy mechanisms of facial mimicry and situational role-taking are considered separately in two different context scenarios. Most of the computational models of empathy have focused on the generation (empathy mechanism) and expression (empathic outcomes) of an empathic emotion, while the calculation of different degrees of empathy (empathy modulation) has received little attention. For instance, this issue has been addressed in the empathy model by Ochs et al. (2012). However, the perception of the virtual agent's degrees of empathy was not evaluated. In contrast, Rodrigues et al.

Table 4. Related Work on the Computational Modeling of Empathy in Virtual Agents Classified with Respect to Its Contribution to the Requirements Formulated in Section 3 (Adapted from Boukricha (2013), p. 68)

Authors	Empathy Mechanism	Empathy Modulation	Empathy Expression
Prendinger & Ishizuka (2005)	Multimodal emotional Recognition	–	Language
Becker et al. (2005)	Multimodal emotional Recognition	–	Facial expression
Brave et al. (2005)	Emotional appraisal	–	Multimodal emotional expression
Bickmore & Picard (2005)	Emotional appraisal	–	Multimodal emotional expression
McQuiggan et al. (2008)	Multimodal emotional recognition	Type of empathic emotion	Multimodal emotional expression
Ochs et al. (2012)	Emotional appraisal	Intensity of empathic emotion	Multimodal emotional expression
Rodrigues et al. (2014)	Self-projection; emotional recognition	Intensity of empathic emotion	Multimodal emotional expression
Boukricha (2013)	Facial mimicry; situational role-taking	Type and intensity of empathic emotion	Multimodal emotional expression
Lisetti et al. (2013)	Multimodal emotional recognition	–	mimicking, reflective listening

(2009) evaluated the perception of different values of the empathy modulation factors considered in their model in two different conditions, *nonempathic* and *empathic,* with the results supporting their model. Furthermore, Boukricha et al. (2013) successfully evaluated their empathy model based on three different conditions that distinguished three different degrees of empathy, thus allowing for a fine-grained evaluation of their model and of its underlying parameters.

While significant advances have been made in enabling virtual agents and robots to empathize, most of the work in this domain relies on *ad hoc* domain-specific behaviours. One possible reason for this is that empathy as a psychological phenomenon is not fully explained in the areas of cognitive and social sciences, and thus the empathic behaviours and prosocial strategies used by empathic virtual agents and robots are tied to the considered context and are difficult to generalize to other domains. Furthermore, the works presented in this section show some limitations. First, the methods used to assess interaction partners' emotional states are not always reliable, which impacts the appropriateness of agents' empathic behaviour. Second, as described in the literature, empathic responses can extend beyond simple emotional reactions, and they are very often realized through more prosocial actions or coping behaviours. However, little exploration has been made, for example, on the effects of prosocial actions that can be taken to reduce interaction partners' levels of distress.

6 CONCLUSION, CHALLENGES, AND OPEN ISSUES

In this article, we presented a survey on the topic of computational empathy for virtual agents and robots. We started with an overview of the main theories and central concepts related to empathy,

Table 5. Related Work on the Computational Modelling of Empathy in Social Robots with Respect to Its Contribution to the Requirements Formulated in Section 3

Authors	Empathy Mechanism	Empathy Modulation	Empathy Expression
Hegel et al. (2006)	Speech intonation	–	Facial expression
Riek et al. (2010)	Facial expression	–	Mouth and head movements
Cramer et al. (2010)	Situational appraisal	–	Verbal responses and facial expression
Leite et al. (2012)	Self-projection	Relationship with the user (companion or opponent)	Verbal responses and facial expression
Leite et al. (2013)	Self-projection and automatic affect recognition	–	Verbal responses and facial expression

and these were used to identify a taxonomy of elements to be considered for building computational empathy. We then presented the state of the art of agents and robots that evoke empathic responses from users, as well as agents and robots that model and/or exhibit empathic behaviour.

6.1 Challenges and Open Issues

There is still a long way to go before fully empathic agents are out there in the real world. Many challenges still need to be addressed, and new technologies need to be developed. We believe that some of the challenges presented here will shape future research in this area.

— *What about context?*
For empathic agents to extend beyond the simple scenarios developed so far and be able to grasp some aspects of empathic encounters in a deeper way, *context* needs to be not only taken into account but also captured and modeled. For that, rich sensory apparatuses combined with mechanisms for memory formation are needed. Information coming from sensors in the environment will allow for a more concrete perception of contextual features—of what is going on—allowing for a better characterization of what leads to empathic encounters. Furthermore, information about users' physiological responses will provide data to the agents, allowing them to form a better picture of users' concrete responses. For example, Leite et al. (2013a) developed a system in which responses from children were captured using physiological data, but the adaptation to those responses remains unsolved. Finally, for context to be concretely used, the sensory information needs to be interpreted, filtered, stored, and used to create a history of events that are salient and important for our empathic agents. This is perhaps one of the great challenges of this area.

— *More autonomy*
Creating agents that exhibit empathy characteristics is extremely complex. Thus, significant work has been conducted to evaluate specific elements of empathic interactions with agents by using videos or WoZ studies (de Rosis et al. 2005; Gonsior et al. 2012) or predefined behaviours (Cordar et al. 2014). Theses approaches are, without a doubt, important for advancing the understanding of the underlying processes in humans' reactions, but we should

not forget that they also should represent a step towards building fully autonomous empathic agents.

From an engineering perspective, autonomy is the ultimate aim. As such, more work needs to be carried out towards achieving higher degrees of autonomy. By relying on the lessons learned from controlled WoZ studies, one can learn what works and what does not. However, current technology for building the different empathic mechanisms identified in this article is still limited. Emotion recognition in real time is still hard, specially in natural settings, and often not reliable enough for many types of users. Computational models of empathy using either symbolic rule-based approaches or machine-learning based ones need to represent, maintain, and update information not only about the user, but also about past events and context. This is an area of research on its own. Plus, algorithms for the identified processes capturing different empathy mechanisms or empathy modulation are difficult to achieve because symbolic approaches are limited and very scenario-dependent, and decision theoretical approaches often simplify the problem in order to be implementable. As such, most approaches presented here are constrained to specific scenarios, and the systems hardly generalize. Developmental approaches such as Lim and Okuno (2015) are also limited in their capability to capture high-level processes and need intensive data and training to be effective. Nevertheless, we need to advance the building of automatic mechanisms for perception (behavioural and emotional), reasoning, and action that, in an integrated manner, will constitute our fully operational believable empathic agents.

— *What shall we measure?*
One question that we quite often need to address when we create "empathic" agents concerns the measures used to evaluate the systems (particularly involving user studies). The evaluation usually has a tremendous impact on many future developments and on whether we should give more emphasis to specific algorithms or certain constructs and not others. Such evaluations also provide data that will influence the creation of new models for our agents' behaviours, which in turn will impact the many different new applications we can pursue. However, to clearly assess "empathy" in social agents, we need to not only rely on instruments that currently exist but also often develop our own measures. To study empathy in humans, several questionnaires are available, in particular the Davis' questionnaire (Davis 1983b) which is, without a doubt, a benchmark in measuring individual differences in empathy. Using four subscales, particularly a Perspective-Taking Scale (PT) and an Empathic Concern Scale (EC), this questionnaire would be interesting to consider when looking at empathy in agents and robots. However, given that it is a subjective questionnaire, it cannot be used as such, and there is not yet a parallel for examining the empathy of artificial entities. To evaluate empathy in social agents, we have to rely on users' perceptions of empathic characteristics, which means we need to measure "perceived empathy." In fact, researchers from an EU-funded project[2] recently developed (but did not validate) a new questionnaire inspired by the PT and EC dimensions of the Davis questionnaire to assess the perceived empathy of a social robot in learning scenarios. However, thus far in social agents and robots, the majority of researchers use related measures such as the quality and adequacy of emotions expressed by the agents, believability (Bartneck et al. 2009), social presence (Harms and Biocca 2004), friendship, anthropomorphism and perceived intelligence (Bartneck et al. 2009), or intention to use the agent/system (the Almere questionnaire, for example (Heerink et al. 2010)). Yet the link between those measures and a measure of empathy still needs to be be uncovered. For example, following our theoretical analysis, it

[2]http://gaips.inesc-id.pt/emote/.

seems natural that the dimension of "Likeability" in the Godspeed questionnaire should be related to a measure of perceived empathy, but the degree of this dependence needs to be properly examined and validated.

We believe that there is space for research into new measures and questionnaires adequate to this field, as well as research into validation studies that will enable a better view of how empathic robots and virtual agents are. The development of validated metrics for evaluating social robots and agents in their diverse competencies is beyond the scope of this survey, but it is something that the community should discuss and work towards.

— *From theories and back*
To create a computational model of empathy, one must operationalize the different processes involved in empathic encounters. That is, we rely on the different psychological theories (as we presented in the previous sections) and operationalize processes, making them sufficiently specific, quantifying the internal relationships and their structures, and thus making them concrete enough to map them into computational software entities (Aylett and Paiva 2012). To achieve this, we, researchers and engineers, make choices about representations, structures, processes, and even parameters that impact the artificial reality built. However, the "implemented" models can not only be used for the purpose they were designed for but also serve as a new source of data for those studying social and emotional relationships among humans. Social computational modeling constitutes an amazingly new opportunity in how we study interactions in humans. Not only can these computational models be used for data gathering, but their features, processes, and elements, which are made concrete in the implementation, can be studied extensively through many simulations and cases and used to rethink the theoretical underpinnings of these areas. Research and development of artificial empathic agents needs to walk hand in hand with psychology, and the results we engineers may achieve can be used to lead to the development of new research methods and even theories in psychology.

— *Towards more "authentic" empathy*
One can discuss what constitutes true empathy or even true emotions in a virtual agent or a robot. Can a set of rules programmed to generate prototypical emotional expressions be seen as a model of true emotions? Aren't they just fake emotions? If the agent does not have the physiological artefacts for affective experiences to emerge, can such artificial agents ever be viewed as having empathy or emotions? Are those emotions ever authentic? We believe that such authenticity depends on how we look at the problem. Our view is to consider the creation of synthetic creatures with different embodiments, and, as such, emotions and empathy have to be engineered in the context of those embodiments. That is, synthesizing empathic processes and responses must be constrained by the limited bodies of the agents and thus may have to be considered always artificial (thus, not authentic). In spite of this, building synthetic empathy mechanisms should be inspired by what we know about empathy and the different processes that lead to it. In that respect, one can also question the depth that such models present and the processes by which they are created. Some researchers, in particular Asada (2015), consider that deeper models for empathy can be achieved if they arise from a developmental approach. The area of Cognitive Developmental Robotics (CDR) (Asada et al. 2009) aims at providing "new understandings of how humans' higher cognitive functions develop by means of a synthetic approach that developmentally constructs cognitive functions." In particular, in what concerns emotions, a conceptual model was proposed where empathy develops from internal processes built into the agents, such as the self–other discrimination. Similarly, Lim and Okuno (2015) proposed a model where empathy is an emergent phenomenon resulting from three main neurophysiological subsystems: a

mirror neuron system, somatosensory cortices, and an insula. Empathy thus emerges from interaction with users through infant-directed "baby talk." These deeper and developmental models for empathy can be seen as more natural because they can emerge from interactions with humans in particular parenting scenarios. Furthermore, they address the issue of authenticity by trying to mimic as much as possible in the agent architecture some of the neurological processes involved in empathy, and processes, such as empathy, become emergent or learned from interactions with others. By contrast, some systems presented here follow a more superficial and shallow approach where the processes are completely engineered or scripted. However, in both approaches, empathic processes and responses will still be artificial and human engineered and certainly conditioned by the "artificial" embodiment chosen.

— *Long-term interaction*
Social relations between humans impact their empathic responses in the long term. For example, we tend to be more empathic to friends than to strangers. With a few exceptions (Bickmore and Picard 2005; Leite et al. 2014), the studies evaluating empathic virtual agents or robots involve interaction partners being exposed to such agents for a short period of time or in a single interaction. Likewise, most of the proposed computational models of empathy are not designed to handle multiple interactions with the same user. The number of candidate empathic responses given a particular situation is usually limited, which can lead to repetitive behavior by the agent. Moreover, as pointed out by Leite and colleagues [2014], individual differences can influence user preferences for certain types of empathic responses (e.g., emotional support vs. concrete task assistance), so user adaptation is important once the agent can reason about these preferences. For these reasons, more long-term interaction studies are necessary not only to verify whether the results of empathic agents remain after extended periods of time, but also to explore the issues of user adaptation and variability of behavior in this particular domain.

6.2 Final Remarks

This is an exciting new field of research that has witnessed significant advances in the past few years. The notion that virtual agents or robots should be able to recognize and understand people's emotions, put themselves into our shoes, and act in an empathic manner is necessary if we want to achieve believable and natural interactions with users in a variety of contexts. In addition, if we want to build agents that impact people's behaviours, we need to build them in a way that touches users' inner feelings (i.e., by evoking empathy). The development of computational empathy and fostering empathy through machines may help to increase the prosocial responses of both humans and machines.

Indeed, we live in a world where technology is not only changing the environment but also us humans. Serious questions are currently being raised about the dehumanizing effect of technology. Yet such challenges can lead to a very exciting time as technology starts embracing moral, social, and prosocial concerns. We believe that addressing empathy in the relation between humans and machines, fostering the type of research we just surveyed, may contribute to the creation of a kinder and more empathic future society.

ANNEX: EMPATHY QUESTIONNAIRE FOR SOCIAL ROBOTS

In the EMOTE project (see http://gaips.inesc-id.pt/emote/) and as reported in one of the project deliverables (Barendregt et al. 2016), in order to assess the perceived empathy of a social robot in two learning scenarios, a questionnaire was developed inspired in some of the dimensions of the Davis questionnaire. In the questionnaire, instead of asking about the user's perceptions of their own empathic capacities, the items aim at appraising the robot's empathy. The questionnaire, shown below, has been used in short- and long-term studies in the EMOTE project, both for individual interactions (one person and one robot) and multiuser interactions.

(1) <ROBOT-NAME> can have tender and concerned feelings for people less fortunate than himself.
(2) Sometimes <ROBOT-NAME> found it difficult to see things from my point of view.
(3) Sometimes <ROBOT-NAME> did not feel sorry for me when I was having problems.
(4) <ROBOT-NAME> tries to look at all sides of an issue before he makes a decision.
(5) If <ROBOT-NAME> would see someone being bothered or hurt, he would probably feel protective towards them.
(6) <ROBOT-NAME> sometimes tried to understand me better by imagining how things look from my perspective.
(7) <ROBOT-NAME> is not disturbed when I or someone else am upset.
(8) When <ROBOT-NAME> is sure he is right about something, he doesn't waste much time listening to my arguments or point of view.
(9) If <ROBOT-NAME> would see someone being treated unfairly, he wouldn't feel much pity for them.
(10) <ROBOT-NAME> is often quite touched by things that he sees happening.
(11) <ROBOT-NAME> believes that there are two sides to every question and tries to look at them both.
(12) I would describe <ROBOT-NAME> as a pretty soft-hearted robot.
(13) If <ROBOT-NAME> is upset with someone, he would try to put himself in my shoes for a while to understand the situation.
(14) Before telling me I have done something wrong, <ROBOT-NAME> tries to imagine how he would feel if he was in my place.

ACKNOWLEDGMENTS

We would like to thank our anonymous reviewers for their valuable comments and suggestions to improve this article. We also thank our colleagues who gave their permission to reprint some of the images of empathic agents and robots included in this article.

REFERENCES

Minoru Asada. 2015. Towards artificial empathy. *International Journal of Social Robotics* 7, 1 (2015), 19–33.

Minoru Asada, Koh Hosoda, Yasuo Kuniyoshi, Hiroshi Ishiguro, Toshio Inui, Yuichiro Yoshikawa, Masaki Ogino, and Chisato Yoshida. 2009. Cognitive developmental robotics: A survey. *IEEE Transactions on Autonomous Mental Development* 1, 1 (2009), 12–34.

Ruth Aylett and Ana Paiva. 2012. Computational modelling of culture and affect. *Emotion Review* 4, 3 (2012), 253–263.

Wilma A. Bainbridge, Justin W. Hart, Elizabeth S. Kim, and Brian Scassellati. 2011. The benefits of interactions with physically present robots over video-displayed agents. *Advanced Robotics* 3, 1 (1 Jan. 2011), 41–52.

Simon Baron-Cohen and Sally Wheelwright. 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 (2004), 163–175.

Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics* 1, 1 (2009), 71–81.

C. Daniel Batson, Marina P. Polycarpou, Eddie Harmon-Jones, Heidi J. Imhoff, Erin C. Mitchener, Lori L. Bednar, Tricia R. Klein, and Lori Highberger. 1997. Empathy and attitudes: Can feeling for a member of a stigmatized group improve feelings toward the group? *Journal of Personality and Social Psychology* 72, 1 (1997), 105.

Christian Becker, Helmut Prendinger, Mitsuru Ishizuka, and Ipke Wachsmuth. 2005. Evaluating affective feedback of the 3D agent max in a competitive cards game. In *Affective Computing and Intelligent Interaction.* Springer, Berlin, 466–473.

Timothy W. Bickmore, Laura M. Pfeifer, and Brian W. Jack. 2009. Taking the time to care: Empowering low health literacy hospital patients with virtual nurse agents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1265–1274.

Timothy W. Bickmore and Rosalind W. Picard. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, 2 (2005), 293–327.

Hana Boukricha. 2013. *Simulating Empathy in Virtual Humans*. Ph.D. Dissertation. Bielefeld University.

Hana Boukricha, Nhung Nguyen, and Ipke Wachsmuth. 2011. Sharing emotions and space: Empathy as a basis for cooperative spatial interaction. In *Proceedings of the 11th International Conference on Intelligent Virtual Agents (IVA 2011), LNAI 6895*, S. Kopp, S. Marsella, K. Thorisson, and H. Vilhjalmsson (Eds.). Berlin: Springer, Reykjavik, Iceland, 350–362.

Hana Boukricha and Ipke Wachsmuth. 2011. Empathy-based emotional alignment for a virtual human: A three-step approach. *Künstl Intell* 25, 3 (2011), 195–204.

Hana Boukricha, Ipke Wachsmuth, Maria Nella Carminati, and Pia Knoeferle. 2013. A computational model of empathy: Empirical evaluation. In *Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, Geneva, Swizerland, 1–6.

Scott Brave, Clifford Nass, and Kevin Hutchinson. 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 (2005), 161–178. DOI: http://dx.doi.org/10.1016/j.ijhcs.2004.11.002

Cynthia L. Breazeal. 2004. *Designing Sociable Robots*. MIT University Press.

Justine Cassell, Joseph Sullivan, Scott Prevost, and Elizabeth F. Churchill. 2000. *Embodied Conversational Agents*. MIT University Press.

Thierry Chaminade, Massimiliano Zecca, Sarah-Jayne Blakemore, Atsuo Takanishi, Chris D. Frith, Silvestro Micera, Paolo Dario, Giacomo Rizzolatti, Vittorio Gallese, and Maria Alessandra Umiltà. 2010. Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures. *PloS One* 5, 7 (2010), e11577.

Tanya L. Chartrand and John A. Bargh. 1999. The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology* 76, 6 (1999), 893.

Amy Coplan. 2011. Will the real empathy please stand up? A case for a narrow conceptualization. *Southern Journal of Philosophy* 49, s1 (2011), 40–65.

Andrew Cordar, Michael Borish, Adriana Foster, and Benjamin Lok. 2014. Building virtual humans with back stories: Training interpersonal communication skills in medical students. In *Intelligent Virtual Agents*. Springer, 144–153.

Henriette Cramer, Jorrit Goddijn, Bob Wielinga, and Vanessa Evers. 2010. Effects of (in)accurate empathy and situational valence on attitudes towards robots. In *HRI '10: Proceeding of the 5th ACM/IEEE International Conference on Human-Robot Interaction*. ACM, New York, 141–142.

Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz studies: Why and how. In *Proceedings of the 1st International Conference on Intelligent User Interfaces*. ACM, 193–200.

John M. Darley and Bibb Latane. 1968. Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology* 8, 4p1 (1968), 377.

Mark H. Davis. 1983a. The effects of dispositional empathy on emotional reactions and helping: A multidimensional approach. *Journal of Personality* 51, 2 (1983), 167–184.

Mark H. Davis. 1983b. Measuring individual differences in empathy: Evidence for a multidimensional approach.*Journal of Personality and Social Psychology* 44, 1 (1983), 113.

Mark H. Davis. 1994. *Empathy: A Social Psychological Approach*. Westview Press.

Fiorella de Rosis, Addolorata Cavalluzzi, Irene Mazzotta, and Nicole Novielli. 2005. Can embodied conversational agents induce empathy in users?*Virtual Social Agents* (2005), 65.

Frederique De Vignemont and Tania Singer. 2006a. The empathic brain: How, when and why?*Trends in Cognitive Sciences* 10, 10 (2006), 435–441. DOI: http://dx.doi.org/10.1016/j.tics.2006.08.008

Frederique De Vignemont and Tania Singer. 2006b. The empathic brain: How, when and why?*Trends in Cognitive Sciences* 10, 10 (2006), 435–441.

Frans B. M. De Waal. 2010. *The Age of Empathy: Nature's Lessons for a Kinder Society*. Three Rivers Press.

Frans B. M. De Waal. 2008. Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology* 59 (2008), 279–300.

David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, and others. 2014. SimSensei kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1061–1068.

João Dias and Ana Paiva. 2005. Feeling and reasoning: A computational model for emotional characters. In *EPIA'05 Proceedings of the 12th Portuguese Conference on Progress in Artificial Intelligence*, C. Bento, A. Cardoso, and G. Dias (Eds.), Vol. LNAI 3808. Springer-Verlag Berlin Heidelberg, 127–140.

Wolmet Barendregt et al. 2016. EMOTE Project Deliverable 7.2 Summative and formative evaluations results. (2016).

Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. A survey of socially interactive robots. *Robotics and Autonomous Systems* 42, 3 (2003), 143–166.

Samuel L. Gaertner and John F. Dovidio. 1977. The subtlety of white racism, arousal, and helping behavior. *Journal of Personality and Social Psychology* 35, 10 (1977), 691.

Stephen M. Garcia, Kim Weaver, Gordon B. Moskowitz, and John M. Darley. 2002. Crowded minds: The implicit bystander effect. *Journal of Personality and Social Psychology* 83, 4 (2002), 843.

Barbara Gonsior, Stefan Sosnowski, Malte Buß, Dirk Wollherr, and K. Kuhnlenz. 2012. An emotional adaption approach to increase helpfulness towards a robot. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on.* IEEE, 2429–2436.

Barbara Gonsior, Stefan Sosnowski, Christoph Mayer, Jürgen Blume, Bernd Radig, Dirk Wollherr, and K. Kuhnlenz. 2011. Improving aspects of empathy and subjective performance for HRI through mirroring facial expressions. In *RO-MAN, 2011 IEEE.* IEEE, 350–356.

Jonathan Gratch, Gale M Lucas, Aisha Aisha King, and Louis-Philippe Morency. 2014. It's only a computer: The impact of human-agent interaction in clinical interviews. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems.* International Foundation for Autonomous Agents and Multiagent Systems, 85–92.

Charlotte N. Gunawardena and Frank J. Zittle. 1997. Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American Journal of Distance Education* 11, 3 (1997), 8–26.

Lynne Hall, Sarah Woods, Ruth Aylett, Lynne Newall, and Ana Paiva. 2005. Empathic interaction with synthetic characters: The importance of similarity. *Encyclopedia of Human-Computer Interaction* (2005).

Chad Harms and Frank Biocca. 2004. Internal consistency and reliability of the networked minds measure of social presence. In *The 7th Annual International Workshop: Presence 2004.*

Bradley Hayes, Daniel Ullman, Emma Alexander, Caroline Bank, and Brian Scassellati. 2014. People help robots who help others, not robots who help themselves. In *RO-MAN*. IEEE.

Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. 2010. Assessing acceptance of assistive social agent technology by older adults: The Almere model. *International Journal of Social Robotics* 2, 4 (2010), 361–375.

Frank Hegel, Torsten Spexard, Britta Wrede, Gernot Horstmann, and Thurid Vogt. 2006. Playing a different imitation game: Interaction with an empathic android robot. In *Proceedings of the 2006 IEEE-RAS International Conference on Humanoid Robots (Humanoids06).* 56–61.

Ursula Hess and Agneta Fischer. 2014. Emotional mimicry: Why and when we mimic emotions. *Social and Personality Psychology Compass* 8, 2 (2014), 45–57.

E. Tory Higgins. 1981. Role-taking and social judgment: Alternative developmental perspectives and processes. In *Social Cognitive Development: Frontiers and Possible Futures*, J. H. Flavelland L. Ross (Eds.). Cambridge University Press, 119–153.

Martin L. Hoffman. 2001. *Empathy and Moral Development: Implications for Caring and Justice.* Cambridge University Press.

Doori Jo, Jooyun Han, Kyungmi Chung, and Sukhan Lee. 2013. Empathy between human and robot? In *Proceedings of the 8th ACM/IEEE International Conference on Human Robot Interaction.* IEEE Press, 151–152.

Eun Ho Kim, Sonya S. Kwak, Kyung Hak Hyun, Soo Hyun Kim, and Yoon Keun Kwak. 2009. Design and development of an emotional interaction robot, Mung. *Advanced Robotics* 23, 6 (2009), 767–784.

Stefan Kopp, Lars Gesellensetter, Nicole Kraemer, and Ipke Wachsmuth. 2005. A conversational agent as museum guide – Design and evaluation of a real-world application. In *Intelligent Virtual Agents.* Springer, 329–343.

Sonya S. Kwak, Yunkyung Kim, Eunho Kim, Christine Shin, and Kwangsu Cho. 2013. What makes people empathize with an emotional robot?: The impact of agency and physical embodiment on human empathy for a robot. In *RO-MAN, 2013 IEEE.* IEEE, 180–185.

Iolanda Leite, Ginevra Castellano, André Pereira, Carlos Martinho, and Ana Paiva. 2014. Empathic robots for long-term interaction: Evaluating social presence, engagement and perceived support in children. *International Journal of Social Robotics* (2014), 1–13. DOI: http://dx.doi.org/10.1007/s12369-014-0227-1

Iolanda Leite, Rui Henriques, Carlos Martinho, and Ana Paiva. 2013a. Sensors in the wild: Exploring electrodermal activity in child-robot interaction. In *Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction.* IEEE Press, 41–48.

Iolanda Leite, André Pereira, Samuel Mascarenhas, Carlos Martinho, Rui Prada, and Ana Paiva. 2013b. The influence of empathy in human robot relations. *International Journal of Human-Computer Studies* 71, 3 (2013), 250–260.

Randy Lennon and Nancy Eisenberg. 1987. Gender and age differences in empathy and sympathy. In *Empathy and Its Development*, N. Eisenbergand J. Strayer (Eds.). Cambridge University Press, 195–217.

Daniel Leyzberg, Samuel Spaulding, Mariya Toneva, and Brian Scassellati. 2012. The physical presence of a robot tutor increases cognitive learning gains. *Annual Conference of the Cognitive Science Society* (2012).

Angelica Lim and Hiroshi G. Okuno. 2015. A recipe for empathy. *International Journal of Social Robotics* 7, 1 (2015), 35–49.

Christine Lisetti, Reza Amini, Ugan Yasavur, and Naphtali Rishe. 2013. I can help you change! An empathic virtual agent delivers behavior change health interventions. *ACM Transactions on Management of Information Systems* 4, 4, Article 19 (Dec. 2013), 28 pages. DOI: http://dx.doi.org/10.1145/2544103

Bertram F. Malle. 2005. Three puzzles of mindreading. *Other Minds: An Interdisciplinary Examination*. The Guilford Press (2005), 26–43.

Stacy Marsella, Jonathan Gratch, and Paolo Petta. 2010. Computational models of emotion. *A Blueprint for Affective Computing-A Sourcebook and Manual*. ACM (2010), 21–46.

Stacy C. Marsella, W. Lewis Johnson, and Catherine LaBore. 2000. Interactive pedagogical drama. In *Proceedings of the Fourth International Conference on Autonomous Agents*. ACM, 301–308.

Patrizia Marti, Iolanda Iacono, Michele Tittarelli, and Jelle Stienstra. 2013. Shaping empathy through perspective taking. In *RO-MAN, 2013 IEEE*. IEEE, 751–756.

Samuel Mascarenhas, Rui Prada, Ana Paiva, and Gert Jan Hofstede. 2013. Social importance dynamics: A model for culturally-adaptive agents. In *Intelligent Virtual Agents*. Springer Berlin, 325–338.

Naoko Matsumoto, Akifumi Tokosumi, and Yoko Hirai. 2004. Affection for cohabitant toy dolls. *Computer Animation and Virtual Worlds* 15, 3–4 (2004), 339–346.

Scott W. McQuiggan and James C. Lester. 2007. Modeling and evaluating empathy in embodied companion agents. *International Journal of Human-Computer Studies* 65, 4 (2007), 348–360. Evaluating affective interactions.

Scott W. McQuiggan, Jennifer L. Robison, Robert Phillips, and James C. Lester. 2008. Modeling parallel and reactive empathy in virtual agents: An inductive approach. In *Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Lin Padgham, David C. Parkes, Joerg Mueller, and Simon Parsons (Eds.). Estoril, Portugal, 167–174.

Morton J. Mendelson and Frances E. Aboud. 1999. Measuring friendship quality in late adolescents and young adults: McGill friendship questionnaires.*Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement* 31, 2 (1999), 130.

Stanley Milgram. 1963. Behavioral study of obedience.*Journal of Abnormal and Social Psychology* 67, 4 (1963), 371.

Magalie Ochs, David Sadek, and Catherine Pelachaud. 2012. A formal model of emotions for an empathic rational dialog agent. *Autonomous Agents and Multi-Agent Systems* 24, 3 (2012), 410–440.

Becky Lynn Omdahl. 1995. *Cognitive Appraisal, Emotion, and Empathy*. Lawrence Erlbaum Associates.

Andrew Ortony, Gerald L Clore, and Allan Collins. 1988. *The Cognitive Structure of Emotions*. Cambridge University Press.

Ana Paiva. 2011. Empathy in social agents. *International Journal of Virtual Reality* 10, 1 (2011), 1.

Ana Paiva, Joao Dias, Daniel Sobral, Ruth Aylett, Polly Sobreperez, Sarah Woods, Carsten Zoll, and Lynne Hall. 2004. Caring for agents and agents that care: Building empathic relations with synthetic agents. In *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1*. IEEE Computer Society, 194–201.

Ana Paiva, Joao Dias, Daniel Sobral, Ruth Aylett, Sarah Woods, Lynne Hall, and Carsten Zoll. 2005. Learning by feeling: Evoking empathy with synthetic characters. *Applied Artificial Intelligence* 19, 3–4 (2005), 235–266.

André Pereira, Carlos Martinho, Iolanda Leite, and Ana Paiva. 2008. iCat, the chess player: The influence of embodiment in the enjoyment of a game. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 3*. International Foundation for Autonomous Agents and Multiagent Systems, 1253–1256.

Robert Plutchik. 1987. Evolutionary bases of empathy. In *Empathy and Its Development*, N. Eisenberg and J. Strayer (Eds.). Cambridge University Press, 38.

Helmut Prendinger and Mitsuru Ishizuka. 2005. The empathic companion: A character-based interface that addresses users' affective states. *Applied Artificial Intelligence* 19, 3–4 (2005), 267–285.

Stephanie Preston and Frans B. M. De Waal. 2002. Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences* 25, 1 (2002), 1–20.

James O. Prochaska, Carlo C. DiClemente, and John C. Norcross. 1992. In search of how people change: Applications to addictive behaviors. *American Psychologist* 47, 9 (1992), 1102.

Byron Reeves and Clifford Nass. 1996. *How People Treat Computers, Television, and New Media Like Real People and Places*. CSLI Publications and Cambridge University Press.

Laurel D. Riek, Philip C. Paul, and Peter Robinson. 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 (2010), 99–108.

Laurel D. Riek, T.-C. Rabinowitch, Bhismadev Chakrabarti, and Peter Robinson. 2009. Empathizing with robots: Fellow feeling along the anthropomorphic spectrum. In *3rd International Conference on Affective Computing and Intelligent Interaction, 2009. ACII'2009*. IEEE, 1–6.

Giacomo Rizzolatti. 2005. The mirror neuron system and its function in humans. *Anatomy and Embryology* 210, 5 (2005), 419–421.

Andrew Robb, Casey White, Andrew Cordar, Adam Wendling, Samsun Lampotang, and Benjamin Lok. 2014. A qualitative evaluation of behavior during conflict with an authoritative virtual human. In *Intelligent Virtual Agents*. Springer, 397–409.

Sérgio Hortas Rodrigues, Samuel Mascarenhas, João Dias, and Ana Paiva. 2014. A process model of empathy for virtual agents. *Interacting with Computers* (2014), iwu001.

Sérgio H. Rodrigues, Samuel F. Mascarenhas, João Dias, and Ana Paiva. 2009. I can feel it too! : Emergent empathic reactions between synthetic characters. In *3rd International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, Amsterdam, Netherland.

Robin S. Rosenberg, Shawnee L. Baughman, and Jeremy N. Bailenson. 2013. Virtual superheroes: Using superpowers in virtual reality to encourage prosocial behavior. *PloS One* 8, 1 (2013), e55003.

Astrid M. Rosenthal-von der Pütten, Nicole C. Krämer, Laura Hoffmann, Sabrina Sobieraj, and Sabrina C. Eimler. 2013. An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics* 5, 1 (2013), 17–34.

Maria Sapouna, Dieter Wolke, Natalie Vannini, Scott Watson, Sarah Woods, Wolfgang Schneider, Sibylle Enz, Lynne Hall, Ana Paiva, Elisabeth André, and others. 2010. Virtual learning intervention to reduce bullying victimization in primary school: A controlled trial. *Journal of Child Psychology and Psychiatry* 51, 1 (2010), 104–112.

Klaus R. Scherer. 2010. The component process model: Architecture for a comprehensive computational model of emergent emotion. In *Blueprint for Affective Computing*, K. Scherer, T. Bänziger, and E. Roesch (Eds.). Oxford University Press, 47–84.

Stela H. Seo, Denise Geiskkovitch, Masayuki Nakane, Corey King, and James E. Young. 2015. Poor thing! Would you feel sorry for a simulated robot? A comparison of empathy toward a physical and a simulated robot. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI'15)*. ACM, New York, NY, USA, 125–132.

Mel Slater, Aitor Rovira, Richard Southern, David Swapp, Jian J. Zhang, Claire Campbell, and Mark Levine. 2013. Bystander responses to a violent incident in an immersive virtual environment. *PloS One* 8, 1 (2013), e52766.

Amy Stevens, Jonathan Hernandez, Kyle Johnsen, Robert Dickerson, Andrew Raij, Cyrus Harrison, Meredith DiPietro, Bryan Allen, Richard Ferdig, Sebastian Foti, and others. 2006. The use of virtual patients to teach medical students history taking and communication skills. *American Journal of Surgery* 191, 6 (2006), 806–811.

Ezra Stotland. 1969. Exploratory investigations of empathy. In *Advances in Experimental Social Psychology*, L. Berkowitz (Ed.), Vol. 4. Academic Press, 271–314.

Adriana Tapus and Maja J Mataric. 2007. Emulating empathy in socially assistive robotics. In *AAAI Spring Symposium on Multidisciplinary Collaboration for Socially Assistive Robotics, Stanford, USA*.

Linda Tickle-Degnen and Robert Rosenthal. 1990. The nature of rapport and its nonverbal correlates. *Psychological Inquiry* 1, 4 (1990), 285–293.

David Watson, Lee A . Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology* 54, 6 (1988), 1063.

Michael Wooldridge. 2002. *An Introduction to Multiagent Systems*. John Wiley & Sons.

Zhihong Zeng, Maja Pantic, Glenn I Roisman, and Thomas S. Huang. 2009. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 1 (2009), 39–58. DOI: http://dx.doi.org/10.1109/TPAMI.2008.52

Received October 2015; revised December 2016; accepted December 2016