Facial Deception in Humans and ECAs

Isabella Poggi[1], Radoslaw Niewiadomski[2], and Catherine Pelachaud[2]

[1] Dipartimento di Scienze dell'Educazione,
Università Roma Tre, Italy
poggi@uniroma3.it
[2] IUT de Montreuil, Université Paris 8, France,
{niewiadomski,pelachaud}@iut.univ-paris8.fr

Abstract. Deception is a relevant issue for the theories of cognition and social interaction. When we deceive, we influence others through manipulating their beliefs. This paper presents a definition of deception and of its functions in terms of a model of cognition and social action. We define as deceptive any act or omission aimed at making others believe something false or not believe something true about the invironment, our identity or our mental states. A typology of ways to deceive is outlined (omission, concealment, falsification, masking, negation, and false confirmation), and applied to deception in facial and bodily expression. An ECA is presented that can simulate, mask, or suppress facial expressions of emotions. The relationship of deception and politeness is investigated theoretically and through analysis of a video corpus. The results of the analysis are used to determine when an ECA masks, suppresses or simulates emotional expression.

Key words: deception, social influence, ECA, politeness

1 Introduction

The notion of Deception, traditionally studied since antiquity by moral philosophers ([1, 6, 37, 40]) has been investigated in the fields of social psychology (see for example, [2, 10]), nonverbal communication ([25–27, 29, 30, 33]), ethology and evolution studies ([47, 63]), mainly in connection with Machiavellanism and Theory of Mind ([11, 12, 21, 38, 39, 64]).

Deception is an intriguing and relevant issue for the theory both of cognition and of social interaction, since by deceiving, we influence other people through manipulating their beliefs. Thus, studying deception is a good chance to investigate the mechanisms of knowledge and sociability, and just as observing Animals' deception can shed light on those device in Humans, their implementation in ECAs - Artificial Agents that exhibit intelligent motion and communicative multimodal behaviour driven by internal representations - is a way to test the cognitive and social mechanisms hypothesised.

An ECA is a socially aware autonomous entity with a humanoid aspect [14]. She is able to communicate verbally and nonverbally. She can manifest her emotional and intentional states through facial expression, voice, gaze direction, head

movement, gesture, body movement, etc. Her model is based on studies from human communication, emotion and social behaviours (e.g. [7, 20, 29, 28, 41, 52]). Thus, an ECA's behaviours are not drawn manually by animators but are automatically driven through algorithms [3, 4, 42, 58]. Most ECAs embed models of emotions, communication skills, or personality. They are used mainly as interface in human-machine interaction. An ECA is viewed also as a *social* entity able to follow social norms when interacting with a user. She can be used on a web site to provide information (i.e. for short interaction only) but she can also be a companion, a friend, or even a virtual salesman.

In this paper we present a definition of deception and of its functions in terms of a model of cognition and social action [14], a typology of cases of deception, and we apply them to the construction of an Embodied Agent that exhibits deceptive facial expressions.

2 A General Model of Mind and Social Action

To define the notion of deception and to understand its functions in social interaction we present a model of mind, social action and communication in terms of goals and beliefs.

According to Conte and Castelfranchi [20], the life of an agent (be it a human, an animal or a machine, an individual or collective entity) is regulated by goals. The notion of **goal** in this model is a very abstract one: a goal is a state that regulates the behaviour or the morphological features of the agent; when the perceived state of the world is discrepant with the goal, the agent performs actions or exhibits morphological features until the goal is achieved. So the notion of goal does not necessarily refer to an individual's conscious and deliberate intention, but applies as well to natural and artificial, individual and collective agents (persons, animals, machines, social organisations); and more specific notions such as need, motive, instinct, intention, social end, biological function can all be seen as sub-cases of goals. Actually, many goals are generally represented in an agent's mind, but some goals (e.g., biological functions) regulate the agent from outside; all the internal goals of the agent are in fact subgoals of these functions.

The structure of action is hierarchical; often a single action is not sufficient to achieve a goal, so the agent has to plan and perform more or less complex plans: sets of actions ordered according to a hierarchy of goals. An action is a means for Goal 1, and Goal 1 can in turn be a means for a further goal (Supergoal) G2. To pursue her goals the agent makes plans [46], that is, more or less complex hierarchies of goals, sub-goals, and super-goals, to be achieved by using external and internal **resources**; internal resources encompass actions of the agent's repertoire, and beliefs about the world and the agent itself, while external resources include the objects present in the environment and, more generally, world conditions; but they also include the "social resource", that is, other agents.

A **belief** is some assumption represented in the agent, whether consciously or not (that is, one that may be meta-believed or not). Beliefs are necessary to an agent since without them the agent could not achieve her goals. They are necessary at all the stages of goal pursuit: first, from the very decision whether to pursue some goal or not; then, to check if the conditions of pursuit are met; but also to assess what are the right actions to do, and to evaluate if they have been effective. The human animal, weaker and more clumsy than other animals, depends on the acquisition, elaboration and use of beliefs more than other animals do. This is why beliefs are a very important resource for Humans. Beliefs are relevant because (and inasmuch as) they are useful to goals, and so, among the most important goals of an agent's life are epistemic goals: the goals of knowing, of acquiring, checking, processing, elaborating beliefs. Moreover, beliefs are generally useful to the extent to which they are "true", that is, respondent to the world, an adequate and reliable map of reality.

Often an agent does not have all the resources or cannot perform all the actions she needs to achieve her goals. This gives rise to the need for cooperation among different agents. In fact, as an agent A has the internal or external resources needed to achieve his goals, he has power, but when he does not have those resources, he is in a situation of lack of power. And if another Agent B happens to have the resources that A needs, then A is dependent on B. According to the model presented [20], social exchange and interaction multiply the resources of people and their potential to reach more goals than they could by themselves. This is done through the device of goal **adoption** - the fact that people put their own resources to the service of other people's goals. An agent B adopts the goal of an agent A [17] when B comes to be regulated by a goal of A's (she pursues A's goal as her own goal): when B "helps" A to reach A's goals. Different kinds of adoption are possible: instrumental adoption (I lend you my car so I can avoid accompanying you), exploitation (I host you in the hut so you raise and care my cotton fields), social exchange (I lend you my car so you lend me your pied-à-terre), cooperative (I maintain the ladder while you pick the apples for both of us), altruistic (I dive into the sea to save you), affective (I take you to a movie I do not like just because I love you), and normative adoption (I let you cross the street because the light is red).

The device of adoption multiplies the agents' power, since it allows them to achieve many goals they could not achieve by themselves. But sometimes, in order to have other agents adopt our goals, we may have the sub-goal of **social influence**. An agent A has the goal to influence another agent B when she has the goal of raising or lowering the likeliness for B to pursue some goal: A influences B if A causes B to pursue some goal that B would have not pursued before, or not to pursue a goal B would have pursued. So influencing others is to induce goals in them.

Also social influence, like goal adoption, may be either selfish or altruistic. Selfish influence occurs when A influences B to pursue a goal that is an interest of A, for example if I order you to fetch me the newspaper, which is a goal of mine; altruistic influence occurs when A influences B to pursue a goal that is an

interest of B; for example, if I am your doctor and I advice you not to eat fried food for the sake of your liver, which is an interest of yours.

3 The Two Faces of Communication

The notions of belief and social influence are a key to understand both communication and deception. In order to influence people you need to generate new goals in other people, or to activate goals they do not presently pursue. But since in the human animal goal generation and activation are generally cognitively determined, that is, they are triggered by beliefs [16], to influence a human's beliefs is a means to influence his or her goals. Both communication and, as we shall see, deception, are means to influence other people, either in our or in their own interest.

Let us first take communication. Communication is a two-face social behaviour: on the one side it is a gift that an agent gives to another, because by communicating we give beliefs to others, and beliefs are useful to achieve goals. But on the other side, communication is an act aimed at influencing other people. From the former point of view, communication is a case of reciprocal adoption of goals; in fact, it is ruled by a norm that Grice [36] calls the "Cooperation Principle". Castelfranchi and Poggi [18] agree with Grice [36] about the existence of such a principle and of the consequent Maxim of Quality "tell the truth", but they more generally propose the existence of a norm they call the "reciprocal altruism of knowledge", which sounds as follows:

If A beliefs belief X,
and X is relevant to a goal of B,
A should communicate X to B.

This is but an instantiation, in the domain of knowledge and communication, of the general norm of "reciprocal altruism" stated by Trivers [61]. This is the bright face of communication, tending to rule out deception. But on the latter side, every act of communication is a request to my Addressee to do something for me; more precisely, with every communicative act we provide the other with information about our goals, relying on the principle that the very fact that we display our goals to the other is a good reason for him or her to adopt them [17]. So with requestive speech acts I tell my goal that the other perform some action, with interrogatives that he provide some belief, with informative acts, that he believe some belief.

To sum up, communication is on the one side an act of adoption of the other's goal to have beliefs, and on the other side a request for adoption, an act of influence, a way to have the other do what I want. And from this point of view, even the positive aspect we saw before - the fact that giving others beliefs is adopting their need for knowledge - can be turned into a means for influence. In fact, since, as we said, in humans the decision to pursue goals is triggered by beliefs, to influence a human's beliefs is a way to influence his or her goals.

The function of communication for social influence is based on the following device:

If A has the goal for B to pursue goal G1,
and if believing belief X causes B to pursue goal G1,
then A will have the goal for B to come to believe X.

And conversely,

If believing Y causes B to pursue goal G2,
and A has the goal for B not to pursue goal G2,
then A will have the goal for B not to believe Y.

This is the reason, for example, for the existence of persuasion. A Persuader has to convey some beliefs, and to convince the Persuadee of them, in order to have B decide to pursue some goals [52]. But if A is really concerned that B should pursue G1, and if letting B believe X is the best or the only way to do so, what matters that X be true or not? In principle, A's goal would be to have B believe X, whether true or false. So, here it comes the goal of deceiving.

To sum up, this is the relation between deception and the goal directedness of behaviour: to achieve one's goals an agent often needs to influence others, that is to induce them to pursue some goals; but when the beliefs apt to trigger those goals are different from true, the agent may have the goal to deceive.

4 The Default Rule of Telling the Truth

If the function of communicating is to influence others, and to influence we have to provide the information apt to trigger the goals we want, no matter the information is true or not, in principle either of the following rules could apply as default rule:

1. Deceive, except for when you have good reasons to tell the truth
2. Tell the truth, except for when you have good reasons to deceive.

Yet, in the human mind the latter seems to be implemented, as is affirmed not only by the Grician Maxim "tell the truth", but also by the very fact that we feel deception as a morally reproachable act. Actually, deceiving is by definition an aggressive act, since it thwarts a relevant goal of the other; more, a natural right: the right to knowledge. Since Grotius [37] on, the right to know the truth can be seen as a natural right of humans. So, the default rule is 2: to tell the truth. And it is possible to account for why this is so. The former rule would not be adaptive, mainly for two reasons. One is that if everybody always deceived, no one could trust the other [45, 60], and so the great adaptive advantage of the reciprocal exchange of information would be lost. The second reason is that, just because communicating is a means to influence others, would we never tell the truth, the other would never know what we want of him; thus, again, the adaptive advantage of reciprocal influence and goal adoption would be lost [18].

Nonetheless, the rule has its exceptions: we generally feel we must tell the truth but sometimes we do not. This depends on our relationship to the Addressee and on the specific beliefs at issue. The former criterion here is that "you lie to your enemy". If the Addressee of my potential communication is someone whose goals I want to thwart, or not to adopt, then I will be likely to deceive him: if true beliefs are a gift, I will not give them to one I hate, one whose goals are in competition with mine. This is why, for example, war is the most typical habitat of lies and deception; as well as politics, commerce, criminal acts, love betrayal and so on. The latter criterion, that cuts across with the former, concerns the valence of the beliefs for the goals of the Addressee. So, "you withhold useful beliefs from your enemy, and harmful beliefs from your friend". Beliefs are usually good, because they are useful to plan for achieving goals; but some beliefs are very unpleasant to know and very hard to bear, they can cause you anxiety, confusion, pain; thus, it may be an altruistic act to withhold them from our friend, and a cruel action to reveal them.

5 What Is Deception?

According to Castelfranchi and Poggi [18], an agent A is deceiving an agent B when

- A has the **goal** for B to either have false beliefs or not to have true beliefs (hence, in any case, B should **not know the truth**) about some topic that is **relevant** for B
- The goal of A is a **non-communicative goal**, in that A also has the goal for B not to know of it
- in order to this goal A uses one of a number of means (e.g., A displays a morphological feature, or performs some behaviour, or omits to perform one).

Let us illustrate the elements of this definition.

- **Goal:** necessary condition for deception is that A has the goal of not letting B have a true belief. But we can distinguish two types of it, finalistic versus functional (biological) deception. In finalistic deception, the goal of not letting B have true beliefs is a conscious goal of the agent, as it is in human deception. In functional deception, typical for other animals, the goal of not letting B have a true belief is a biological function: for example, in the aggressive mimicry of a chameleon, that changes its colour to hide and predate, or in the defensive mimicry of the shelduck, that pretends walking with a limp to distract a predator from its chicks.
- **Non–communicative goal:** the goal of not letting B believe true beliefs is noncommunicative, that is, for A not to communicate it is a necessary condition to achieve it. Obviously, if I tell you X and also tell you I want you not to believe something true, you will not believe X. This is where, for instance, deception differs from irony. In irony I tell you something I do not

believe is true, but I want you to know that I don't believe it true. In fact irony can be defined [18] as a case of "recitation", that is, a case in which one communicates something different from what one believes is true, but also communicates one does not believe it true. "Recitation", and therefore irony, are thus cases of "revealed" deception, hence not really deception.

- **Non-truth:** in deception A has the goal for B either to ignore relevant true beliefs or to have false beliefs: in both cases B does not really know the truth.
- **Truth:** but what is truth? In general, a belief is an agent's representation of the external or internal world, and it is a "true" belief when the agent believes, with a high level of certainty, that it is a reliable "model" of that world. But here again we can distinguish. For functional deception, the definition of truth is an objective one: true as a belief that an abstract omniscient agent sees as actually corresponding to the external world; truth as a responding, adequate representation of the world. For finalistic deception, instead, where the goal of not having the other believe something true is a conscious goal, a goal the individual does know he has, the definition of truth is a subjective one: true as *what the agent A believes is true* [18], after [1]. So, if one tells something that he believes is true, even if it is not true objectively, this does count as deception. For example, Ptolomaeus in saying that the earth was the centre of the universe was not deceiving, even if this was not objectively true. And conversely, if one tells something he believes is false, even if it is true in fact, he is deceiving.
- **Relevant:** beliefs are means for goal pursuit and planning, so much that not to give beliefs is an aggressive act that obstructs the other's achievement of goals. But beliefs are relevant because - and inasmuch as - they are useful to our goals. So if I am paying at McDonald's, I ate a hot dog but I say a hamburger, if hamburger and hotdog cost the same, it is not deception. If I know I have AIDS and I do not tell my boyfriend, it is deception; but if I knew it about a friend of mine, and I do not tell my girlfriend, it is not.
- **Means:** deception may be performed by a number of means: a speech act, like a sentence; a communicative nonverbal action, like a hypocrite handshake; a non-communicative action, like hiding your lover in the wardrobe; an omission, if I simply do not tell you anything; an object produced by the deceiving act, like a false banknote; an object used during an act of deception, like the mask used in a robbery; and for animal deception, generally for deception governed by a biological function, even a morphological feature, like the chameleon's colour [18].

6 Why Do We Deceive?

As mentioned above, the general reason why people deceive is that they want to influence others. But in everyday life, what are the specific reasons why we deceive others? According to Boffa et al. [5], we deceive either to cause or to prevent that an event occur, that someone assume a *belief* or an *evaluation*, feel an *emotion*, do some individual *action* or some *social action* toward someone else. Let us give some examples.

1. Deceiving to prevent an event:
   A's dad B suffers from heart disease. A does not tell him she failed her exam to prevent a heart attack to B.
2. To have someone assume a *belief*:
   In Stanley Milgram's experiment (as well as in innumerable studies in psychology) the subject believes that "pupil" are really striken by electrical shocks, while they are actors, confederates of the experimenter, that pretend suffering from the shocks. The whole experiment is carried on to obtain very important knowledge about obedience to authority.
3. To have someone assume an *evaluation*:
   The teacher asks a pupil if his dad is an alcoholic, and while he is in fact, the pupil denies to protect his father's image. A clerk tells his boss a false gossip about a colleague in order to induce a bad evaluation of him.
4. To have someone feel an *emotion*:
   Jago lets Othello believe that Desdemona is betraying him in order to arouse his jealousy.
5. To have someone do some *action*:
   B is a vegetarian, and he goes to dinner to his friend C. A, C's mother, is convinced that not to eat meat is unhealthy, so she makes vegetables stuffed with meat, but she tells him they are stuffed only with egg and bread, in order for him to eat them.
6. To have someone do some *social action*:
   The child lies to his mother to avoid being punished. As one can see, the goals aimed at by acts of deception can be either good or bad, either selfish or altruistic.

A case in which deception is not generally aimed at selfish goals is politeness. In politeness we care the goals of image and self-image of others [7]. This in some cases is at the service of selfish goals: for example in hypocrisy or adulation, where our true goal (the superordinate goal of our being polite) is not really complying with the other's goal of image, but influencing him to do something for us. Yet, in other cases politeness can simply be seen as a way not to hurt the other's feelings, not to wound his image or self-image. In some cases politeness could hardly considered a case of deception, but it is rather one of "recitation". Yet, even through what I literally communicate is false, and both you and I know it is false, what I meta-communicate in doing so - that I respect your goal of image - is true.

7 What Do We Deceive About?

In communication we can provide beliefs about three domains: the world around us - we talk about objects, persons, events; our identity - our socio-cultural roots and personality; and finally about our own mind - our beliefs, goals and emotions. Also deception may concern the world, our identity and our mind. Take the pirate who hides his treasure; or a case of "tactical deception" like the rooster's call: the hen comes, thinking there is food here, while the rooster in fact only wants

to copulate: in both cases deception is about the world. The chameleon's colour, as well as the false pretense of a cheater, are cases of deception about one's identity. And finally, an hypocritical smile to someone we hate, to mask our real feelings, is deception bout our mind.

Within states of our mind, in fact, sometimes we deceive about our emotions. An emotion is a complex subjective state that encompasses cognitive elements as beliefs, evaluations, causal attributions, images, as well as subjective feelings, physiological states, readiness to action, but also, typically, expressive traits of behaviour. So, emotions are mental states that naturally tend to expression: a goal of displaying them is embedded in their very neurophysiological program. Nonetheless, sometimes other goals suggest us to deceive about our emotions, and we may do so through all possible types of deception. Let us see some examples of the ways to deceive, about our emotions or other topics.

8 In How Many Ways Do We Deceive?

There are multiple ways to deceive, as witnessed by the huge number of words in a language that name acts of deception: to lie, to pretend, to cheat, to mislead... But all of these ways to deceive can be grouped in some basic ways in which a Deceiver manipulates the beliefs of the Dupe. The first important distinction is that between withholding true beliefs and providing false beliefs. But this main action over the Dupe's mind must be done in different ways depending on the previous condition of her mind before I give her that belief or not. Since any act of deception performs a manipulation of the Other's mind, the deceiver must be sophisticated in the Theory of Mind, and form an articulated representation of the other's beliefs. For example, it is not the same if I want you not to believe something that you already know vs. something you don't know anything about. So we can distinguish at least six different ways to deceive:

- **Omission.** A has the goal for B not to believe a belief K, but since B does not really know anything of K, for A it is sufficient not to do anything. *A is B's boyfriend, and he just discovered he has AIDS, but he is afraid of being forsaken by B. So he does not tell B.* As for emotions, we have a case of deceptive omission if one feels an emotion but does not display it at all. For example, *A is angry at B but he does not move any muscle of his face, nor does he become more tense in his movements.*
- **Concealment**. A does an action which is not communicative to prevent B from believing a true belief. For example, *A closes the door of the wardrobe to hide her lover from her husband.* A feels Emotion E1 but performs some non-communicative action to prevent B from perceiving his display. *B is ridiculous for A, so A is going to laugh, but not to offend B, he bites his lips to conceal his laughter.*
- **Falsification**. A provides B with a false belief, that is a belief different from what A believes is true. A pays B with false money. A does not feel Emotion E1 but displays a body expression of E1: *while meeting a colleague that she does not love particularly, she smiles at her very cordially.*

- **Masking**: a case of concealment brought about through falsification: providing a false belief in order to conceal a true belief. A mask is a prototypical example: *I conceal my true face with a fake one.* A feels Emotion E1, but pretends (falsifies) to be feeling an Emotion E2: so A conceals the expression of a true emotion by displaying the expression of a false emotion. *A is angry at B but conceals his angry expression with a friendly smile.*
- **Negation**: denying a true belief. *Seeing your dirty face, I ask you if it is you who ate the jam, and you say: "No". Or: I ask you if you are disappointed, you are but you say "Not at all!".*
- **False confirmation**: the opposite of negation. You make a hypothesis that in fact is not true, but I let you believe it is.

These theoretical distinctions among the ways we deceive about our emotions partially correspond to the categories proposed by Ekman [29]. While analysing human facial expressions of emotions, Ekman distinguished between: modulating, falsifying, and qualifying an expression. In particular, falsifying a facial expression means to simulate it (to show a fake emotion), neutralize it (to show neutral face) or mask it (to hide an expression as much as possible by simulating another one). In our terms, simulating corresponds to what we call "falsifying": you display an emotion you really do not feel, but this is not aimed to hide a true one; masking as meant by Ekman is, like in our sense, producing a fake expression to conceal a true one; neutralizing may correspond either to concealing or to omitting; it depends how "active" you are in avoiding leakage of your spontaneous expression. If, say, you are quite an inexpressive person, you have not so much to do to have a neutral expression, this is your natural face; but if you are spontaneously expressive, having a neutral face requires some active inhibition of an expression that is ready to display.

9 Detecting Deception

Classical works by Ekman [25, 26] and Ekman and Friesen [30, 29] first revealed the work of facial, gestural and bodily expression in lying and deceiving, and hence also determined how one can detect deception from them. Cues to deception are typically borne by:

1. Concomitant signals of emotions caused by deception. When you deceive, you may feel specific positive or negative emotions [25]: embarrassment, shame or guilt of being deceiving (one more evidence of the default rule of telling the truth); fear of being caught; joy or amusement for being able to dupe the other. So, if the expressions of these emotions are superimposed onto those of the message to convey, this can be a cue that the Sender is lying: if for instance A is touching her own body often, B may understand that A is embarrassed and from this infer that A might be deceiving.
2. Detection from contradiction. In deceiving, our mind keeps both the deceptive and the true belief in memory, and thus, in communicating the false belief the expressions concerning the true one can leak along with the fake

ones. These expressions can leak either subsequently (immediately before or after), or in parallel, concomitant to the deceptive ones. In both cases, this is the route to deception detection:

A is displaying belief X
A is displaying belief Y
the two displays are contradictory
since
between two contrasting beliefs one must not be true
then
B infers that A must be deceiving

Let us see some examples of contradiction between subsequent expressions.

1. In a trial of high political import, the Italian politician Paolo Cirino Pomicino is being examined by the public accuser, Antonio Di Pietro, about his having taken money for his party by Italian industries. During the whole examination, Cirino Pomicino has a defying stare, sometimes an ironic smile, as if telling Di Pietro: *"I am not afraid of you, I have more power than you".* But just for a moment, for a very short time, he stares off into space, and his ironic smile disappears from his mouth. His whole expression means: *"I am dismayed, I am very tired of still opposing him: I do not know how long I will still be able to resist him".*

2. Upon hearing the Teacher's question, a Student pulls her head between her shoulders, as if trying to skip a hard blow. Immediately after, she takes on the typical posture and gaze of someone reflecting to find the right answer: she directs her eyes up in the sky, a type of gaze that communicates: *"I am trying to make inferences".*

And now two examples of contradiction between concomitant expressions.

3. During an interview about racism, a subject tells he is very tolerant, but at the same time he *steps back*, with crossed arms.

4. A female politician, in commenting electoral defeat, displays sadness through her oblique eyebrows, but happiness through her smile.

10 Building an Embodied Conversational Agent that Deceives

In the previous sections we saw that many different skills and capacities from high level cognitive reasoning [13, 19, 23] to low level control of facial muscles can be involved in an act of deception. In following sections we describe an embodied conversational agent (ECA) that is able to deceive about her emotional state by using *deceptive facial expressions* instead of the "spontaneous" ones.

That is the agent uses different types of strategies to display emotions: her true expressions of emotions are *inhibited*, *masked* by other expressions, or *fake expressions* are displayed instead. In particular our agent is able to modify her own facial expressions to take into account her relationship with her interlocutor. Thus, in certain situations, she will hide her true emotions covering them by socially appropriate facial expressions with the goal of being consistent with "the community", and thus will *deceive* interlocutors according to the definition of deception presented in section 5.

Now, one could wonder: is it useful, and possible, to build a Virtual Agent that can have, express, and deceive about emotions? As far as mere expression is concerned, the question finds a definitely positive answer in all research about Embodied Agents and Affective Computing [4, 31, 49, 51, 57]. This is both possible and useful for practical reasons, because the interaction with an ECA that expresses emotions may be more satisfying and motivating for the User than it is with a metallic voice or a motionless cold avatar [55]. But of course, that an Agent can even "feel" an emotion is a much more tricky issue ([18, 32]). Here, one should first distinguish the question "is it possible?" from that "is it useful?", but second, independent on whether it is possible or not, also distinguish "useful for what?" As to the latter question, one should conclude that, even if to have Agents feel emotions is no good for practical purposes, yet, research aimed at this goal could still have relevant results for the advancement of human knowledge. In line with a traditional trend in AI, we think that simulation is an important test-bed for theories of human cognition. The same, in our view, could be said for deception. Simulating deception in a Virtual Human is in itself a valuable aim for theoretical purposes; moreover, doing so is also possible, as witnessed by previous research [23].

Moreover, we believe that ECAs should take different social roles related to the activities they carry out. For instance, when being a virtual salesman an ECA ought to be polite (e.g. she should respect the social distance with a new customer). When interacting with the user the ECA may, for example, display less negative emotions than it results from her virtual appraisal state. Thus an ECA ought to be aware of social rules and be able to control her behaviours. For example, an agent with nonverbal politeness knowledge will know which expressions to display and when; a deceiving agent can be used whenever displaying various social rules is desired (for example in pedagogical agents, virtual salesmen, virtual companions). The agent can cover emotional states (though the expressions of her true emotions can leak over the mask ones) and deceive about them by showing some expressions that are deceptive but adequate to the situation.

In this paper we imagine a Virtual Agent that does not really "feel" emotions but, 1. can express emotions, in that she can exhibit the facial expression that corresponds to a given emotion; 2. can "have" emotions in that, if not the internal feelings, she can trigger an internal representation of the cognitive structure of some emotions [50] or of their "mental ingredients" [53]; 3. can express emotions different from those she has - that is, she can deceive about her emotions. Since,

according to the definition above, A deceives when it has the goal to let B assume something different from what A itself believes, a simple comparison between the structure of the emotion actually triggered in the agent with the structure of the emotion expressed in her face can tell if she is "deceiving" or not: if my internal representation tells me *I am now having the emotion X*, and my face expression, deliberately, either displays no emotion or an expression corresponding to emotion Y, I am deceiving.

In the following sections we introduce an ECA, that is able to display, in addition to spontaneous expressions of her emotional states, various deceptive expressions. Deceiving about emotions does not mean simply to "replace" one expression by another. People are able to recognize the real emotion from the face even if it is hidden or masked [24]. Similarly they distinguish between fake and felt expressions [33, 35]. Thus a deceptive act of our agent also needs to be perceived by human interlocutors as such. Let us consider two different expressions: an expression of a true emotional state and a fake expression, that is one in which the displayer only pretends to be in an emotional state. A spontaneous smile informs the addressee that something positive happened to the displayer. A fake smile can have another (maybe even contrary) meaning. Thus, should the agent have the capacity to display these two facial expressions instead of only one, this would obviously increase her communicative skills. It is important to notice that the distinction between different types of facial expressions in the case of embodied agents makes sense only if these two facial expressions are distinguishable. Some recent experiments have indicated that using deceptive facial expressions in ECA systems is relevant. In a study on a deceiving agent, Rehm and André found that users were able to differentiate, even unconsciously, between the agent displaying an expression of felt emotion versus an expression of fake emotion [57].

In the next sections we present an ECA that is able to:

- express deceptive facial expressions,
- know which factors influence the display of facial behaviour,
- know how they influence facial behaviour.

Out of the deceptive strategies distinguished by Ekman [29], we first modelled cases of simulation (fake expression), inhibition (expression of emotion covered by a neutral face), and masking (superposition of a fake over a true expression). Our model embeds these varieties of deceptive facial expressions. To compute deceptive expressions (e.g. inhibited anger) we use the "spontaneous" expressions that is the expressions of the true emotion (e.g. expression of anger) and modify them. After presenting our computational model of deceptive expressions we propose a set of rules for facial behaviour management (see section 13). This set is based on the annotation of a video corpus, the results of experiments about facial expression management [8, 43, 44], and the theory of politeness [7]. Consequently our agent is not only able to display deceptive expressions but she "knows" in which situations she should use which facial expression management. At the moment we focused on interpersonal relations in which spontaneous facial behaviour should be altered. Although previous solutions of this problem exist

in ECA's domain (see section 11), in our approach a variety of facial expressions (i.e. true expression, masking, inhibition, and fake expressions) is used for the first time. Knowing her own affective state and the type of relations between interlocutors our ECA automatically adapts the facial behaviour to the social context.

11 State of Art

Few animated agents implement the regulation of facial expressions. Prendinger et al. modelled "social role awareness" in animated agents [54]. They introduced a set of procedures called "social filter programs". These procedures are a kind of rules for facial expression management. Defining social filter programs Prendinger et al. considered both social conventions (politeness) and personalities of the interlocutors. The social filter program defines the intensity of an expression as the function of a social threat (power and distance), user personality (agreeableness, extroversion), and the intensity of emotion. As a result, it can either increase, decrease the intensity of facial expression, or even totally inhibit it.

The Reflexive Agent by De Carolis et al. [22] is also able to adapt her expressions of emotions according to the situational context. This agent analyses various factors in order to decide about whether displaying her emotional state or not. These factors are of two types: emotional nature factors (i.e. valence, social acceptance, emotion of the addressee) and scenario factors (i.e. personality, goals, type of relationship, type of interaction). In particular the Reflexive Agent uses *regulation rules* that define for which values of these factors the felt emotion should (or should not) be displayed [22]. Although many factors related to the management of facial displays are considered in this model, it uses only one type of deception, i.e. inhibition.

Different types of facial expressions were considered by Rehm and André [57]. For the purpose of the study on deceptive agents, they manually defined facial expressions of an agent according to Ekman's description (see section 9). They found that users were able to differentiate when the agent displays an expression of felt emotion versus a fake expression of emotion [57]. Moreover, a non-deceiving agent was perceived as more reliable, trustworthy, convincing, and credible [57].

Out of the models presented in this section, only the last one distinguishes between expression of a true and of a fake emotion. But the expressions of fake emotions were manually defined. In our work, we aim at building an agent that will modify her facial expressions depending on the relations she has with her interlocutors: our agent will deceive the interlocutor about her emotional state, but she will also (inadvertently) allow the human to perceive and interpret her behaviour. Our general aim is dual: build an agent able to know when and how to adapt her facial expression depending on the social context she is placed in. Moreover, the display of the modified expression should be interpreted perceptually as being different from the expression of felt emotion. Thus we introduce

a diversification among facial expressions. Consequently our agent will be able to generate and use different types of deceptive displays.

12 Deceiving About Emotions in ECA

In this section we present an algorithm to compute the facial expression of the agent when she masks, fakes and inhibits her emotional expressions: e.g., expressing joy, masking anger, inhibiting sadness, expressing fake disappointment (see also [27, 29, 34, 41]). Our model of facial expressions [4, 48] is based on Paul Ekman's studies [26–29]. We define facial expressions using a face partitioning approach. The face is divided in eight facial areas F_i, i= 1,..,8 (i.e., F_1 - brows, F_2 - upper eyelids, F_3 - eyes, F_4 - lower eyelids, F_5 - cheeks, F_6 - nose, F_7 - lips movement, F_8 - lips tension). Each facial expression is a composition of these facial areas, each of which could display signs of emotion. In the case of deceptive facial expressions (as in an expression masking another one) different emotions can be expressed on different areas of the face (e.g. anger is shown on the eyebrows area while sadness is displayed on the mouth area). Recapitulating, the main idea of our algorithm is to assign expressions of the emotions to the different parts of the face. For this purpose we defined a set of rules that describe the composition of facial areas in deceptive expressions.

12.1 Rules for Generating Deceptive Facial Expressions

Researchers have proposed a list of *deception cues* i.e. the features of expressions that are useful in distinguishing between fake and felt expressions [29, 25, 26]. At the moment, two of them are implemented in our model: *reliable features* and the *inhibition hypothesis*.

First of all humans are not able to control all their facial muscles. Consequently expressions of felt emotions may be associated with specific facial features like sadness brows [29] or orbicularis oculi activity in the case of joy [26]. Such *reliable features* lack in fake expressions as they are difficult to do voluntarily. Fake expressions are similar to the corresponding sincere - spontaneous - expressions, but slightly differ from them for some subtle differences that are not easy to detect. For example, a false smile is distinguished from a true one as it lacks crow's feet near the eyes that appear by the contraction of the *orbicularis oculi* [30].

On the other hand, people are not able to fully inhibit felt emotions. According to the *inhibition hypothesis* [26], the same elements of facial expressions, which are difficult to show voluntarily in the case of unfelt emotions, are also difficult to inhibit in the case of felt emotions. Ekman enumerated all facial areas that leak over the mask during the emotional displays management [29].

Deceptive facial expressions for six emotions - anger, disgust, fear, joy, sadness, and surprise - are described in the literature [29, 27]. Based on these studies, we have defined a set of fuzzy rules that describe the characteristic features in terms of facial areas for each type of deceptive expression. To each emotion

corresponds a rule. Thus we have defined six rules for each type of deceptive expression.

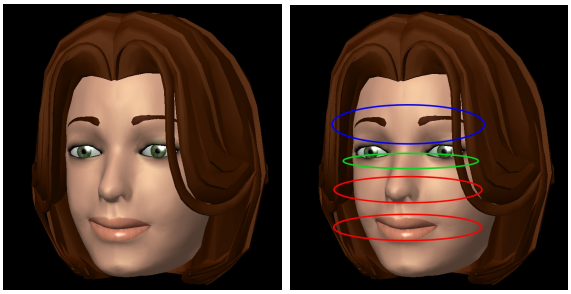
12.2 Algorithm

In our approach any facial expression is described by a set of fuzzy sets. The main advantage of this approach is that slightly different expressions can be described by one label (like "joy" or "contempt"). For each of them we can find a deceptive expression (masking, inhibition, fake expression) by applying suitable rules. Let the *input expression* be an expression of emotion for which a deceptive facial expression needs to be established. In case an input expression for which the deceptive facial expression is not defined explicitly by our rules (e.g. expressions of contempt or disappointment) a fuzzy similarity based algorithm is used in order to establish the degree of similarity between the input expression and expressions whose deceptive expressions are described by our rules (see [48] for details). Our fuzzy similarity algorithm compares any two facial expressions and outputs a value of similarity in the interval [0,1] (0 meaning "not similar at all" while 1 means identical expressions). Once the most similar expression (chosen among the 6 ones) is known we can apply corresponding rules to our input expression. For example, when we want to compute the deception face of contempt or of disappointment, we look to which expression of six-elements set mentioned above it is the most similar and we use the associated rule. For example, our fuzzy similarity algorithm outputs that the expression of disappointment is similar to the one of sadness. Thus masked, inhibited or fake facial expressions of two *similar* facial expressions are created using the same rules.

12.3 Example of Deceptive Expression

Figure 1 presents the agent displaying an expression of disappointment, that is masked by fake happiness. In Figure 1b the parts of expression copied from the expression of disappointment are marked with blue and those of happiness with red circles. We can notice that *orbicularis oculi* activity (characteristics of felt happiness) is not activated. This absence is an indicator of fake expression of happiness. Also the position of the brows can be observed (inner raised brows), which is characteristics of disappointment as the expression of disappointment is very similar (according to our fuzzy similarity based algorithm) to the expression of sadness. According to Ekman [27, 29] the features of felt sadness that leak over the masking expression are: forehead, brows, and upper eyelids. In our model these elements of expression are represented by the facial areas F_1(forehead and brows) and F_2 (upper eyelids). As a consequence, facial areas F_1 and F_2 can be observed in inhibited sadness, and thus, they can be also observed in masked disappointment (figure 1a). On the other hand the mouth area displays a smile (sign of happiness).

Similarly we can generate any cases of inhibited or fake facial expressions. Facial expressions generated with our model were evaluated in a study based on the "copy-synthesis" method (see [9] for details).



a b

Fig. 1. a) Example of disappointment masked by happiness; b) decomposition of the expression in areas: blue areas correspond to the fake expression (happiness), while the red one to the felt one (disappointment).

13 Facial Expression Management

Deceptive facial expressions are often used in given social context. Due to cultural, professional, and social rules we are often required not to show our felt emotions. With time we have learned to inhibit them, to mask them, to put on a fake one. An ECA being setup in an interaction i.e. a social context, ought to know when, where, and to whom she can show the expression of her emotion. That is she needs to know to manage the display of her facial expressions.

In this section we present an ECA that uses a variety of deceptive facial expressions. Depending on her interpersonal relations with the interlocutor and her emotional state, the agent modifies her own "spontaneous" facial behaviour. That is in certain social contexts our agent will use some deceptive facial expressions instead of spontaneous ones. For that purpose we need to find rules that associate the factors (variables), which influence the display of facial behaviour with the occurrence of particular types of facial expressions. Our rules of facial behaviour management are mostly based on the results of the annotation of a video-corpus we have made for this purpose and the theory of politeness [7]. In the next subsection we briefly introduce the latter.

13.1 Politeness Strategies

Brown and Levinson proposed a computational model of politeness in language [7]. According to this theory, any linguistic act like request or promise can threaten the "face" of the speaker and/or hearer. Politeness consists in taking remedial actions to counterbalance the negative consequences of these face threatening acts.

Brown and Levinson proposed the classification of all actions that prevent face threatening. They defined five different strategies of politeness: bald, positive and negative politeness, off-record, and "don't do the action". The more antagonistic a given act is, the more likely a higher strategy is to be chosen.

The decision about the strategy to be used depends on the level of threat of an action (FTA). Brown and Levinson proposed to estimate the FTA of an action by using three variables: social distance, power relation, and absolute ranking of imposition of an action. Finally, the FTA value is computed as a sum of these three values [7].

13.2 Annotation Scheme and Results

In order to define rules of facial behaviour management we decided to re-use the approach proposed by Rehm and André [56]. They analysed the relationship between different types of gestures and politeness strategies in verbal acts. They built a video-corpus called SEMMEL that contains various examples of verbal and nonverbal behaviour during face threatening interactions. Inspired by the encouraging results of Rehm and André's experiment, we decided to analyse the same video-corpus in order to find relations between politeness strategies and facial behaviour.

We used 21 videos of the SEMMEL video-corpus [56]. The overall duration of the analysed clips is 6 minutes and 28 seconds. In this study we used the original annotation of politeness strategies proposed by Rehm and André. They considered four of the politeness strategies from Brown and Levinson's model [7]: bald, positive politeness, negative politeness, and off-record strategy [56]. In our study the facial expressions (and corresponding emotional states) were annotated by a native speaker annotator. In our annotation scheme we considered four types of facial displays: expression of the true emotional state, inhibited, masked, and fake expression. Because of a relatively small number of examples analysed so far we decided to distinguish only between positive, negative emotions, and neutral state. Consequently, we did not consider distinctive emotional states and their corresponding expressions, but, instead, we consider *patterns* of facial behaviour. For example, a pattern called "positive masked" describes any facial expression that occurs in a situation in which a positive emotion is masked by another one. The following patterns of facial expression were considered in the annotation process: negative masked, negative inhibited, negative expressed, fake negative, neutral expression, fake positive, positive masked, positive inhibited, positive expressed.

We analysed the frequency of occurrence for each of them and we found that different types of facial expressions were not evenly distributed along the different strategies of politeness (see [48] for details). For example the "neutral expression" pattern was the most often observed (52% of all cases) while the "fake positive" pattern was observed in 26.5% cases. Some patterns were not observed at all. None of the "positive masked" expressions or "fake negative" expressions was annotated. We use this information to build our model of facial deceptive behaviour management in interpersonal relations.

13.3 Variables of Facial Expression Management

Different sources [7, 8, 43, 62, 65] show that two factors, social distance (SD) and social power (SP), are important to describe interpersonal relations. Social distance (SD) refers to the degree of intimacy and the strength of relation, while social power (SP) expresses the difference in status and the ability to influence others [59]. According to Wiggins *et al.*[65] all personality traits relevant to social interaction can be located in a two dimensional space defined by the orthogonal axes of dominance and affiliation. So two variables, dominance (corresponding to SP) and affiliation (corresponding to SD), are sufficient to describe interpersonal relations. Furthermore, Brown and Levinson include SD and SP in their theory of politeness (see section 13.1). Finally, power (SP) and social distance (SD) are two factors that influence human expressions according to various studies about facial behaviour ([8, 43, 62]).

Facial behaviour management is also conditioned by emotional factors. In particular, facial behaviour depends on the valence (Val) of an emotion [8, 44]. Negative emotions are more often masked or inhibited, while positive emotions are often pretended.

Thus, in our model, we consider three variables to encompass the characteristics of the interaction and the emotional state of the displayer, namely: social distance (SD), social power (SP), and emotion valence (Val).

13.4 Facial Expression Management Model

In our model we consider three different emotional states: negative, positive, and neutral. For each of them we looked for the pattern of facial behaviour that best corresponds to each politeness strategy. The choice is based on the frequency of the co-occurrence of a strategy j and a pattern i in the annotated video clips (see section 13.2). When the data gathered in the annotation study were insufficient to make a choice, we also used the conclusions from other experiments [8, 43, 44]. Table 1 shows which pattern of facial expression i will be used for each type of emotion (positive, neutral, negative) and strategy of politeness.

face threat	bald	positive	negative	off-record
positive emotion	positive expressed	positive expressed	positive inhibited	positive expressed
neutral state	neutral expressed	fake positive	neutral expressed	fake positive
negative emotion	negative expressed	negative expressed	negative inhibited	negative masked

Table 1. Facial behaviour and strategies of politeness.

In our model all three variables *SD*, *SP*, and *Val* take values in the interval [0,1]. In Table 1 different strategies of facial behaviour are ordered in columns

according to the value of threat. We establish this value as the difference: $w = SD - SP$ which takes values in the interval [-1,1]. Then we split the interval of all possible values of w into four equal parts: $w \in$ [-1,-0.5] (very low) is associated with the bald strategy, $w \in$ (-0.5,0] with positive politeness, $w \in$ (0,0.5] with negative politeness, while $w \in$ (0.5,1] (very high) with the off-record strategy. Finally our facial management rules are of the type:

if Val(E_i) is {positive | negative | zero } and
w is {very low | low | high | very high}
then the expression of E_i is {expressed | fake | inhibited | masked}.

Using our set of rules we decide on the facial expression pattern of an emotion E_i. In the case of any "negative masked" expression or "fake positive" expression we use the expression of fake joy or masked joy. Finally, for any emotional state E_i, values of social distance *SD* and of social power *SP*, by using our rules, we can generate the corresponding facial expression (see [4] for details). Our model was evaluated in a study in which we analysed the perception of interpersonal relations from the facial behaviour of our agent [48].

The examples in Figure 2 illustrate the variety of facial reactions displayed by the agent at the same instant in different interpersonal relation settings. In particular, the first row includes the reactions of the agent when she discovers the dishonesty of her interlocutor, while in the second row we can see the agent's reactions for an unpleasant event.

Fig. 2. Examples of different facial expressions displayed by the agent (in consecutive columns, interaction with: a friend, a stranger, a superior). Top row: the agent discovers the dishonesty of her interlocutor; lower row: the agent displays a reaction to an unpleasant event.

14 Conclusion

In the first part of this paper we presented a theoretical model of deception, its function, performance and detection, showing its relevance in terms of its cognitive mechanisms, its verbal and multimodal means, its specific goals and its impact over social interaction. As we deceive, we manipulate the other's mind in order to influence him, that is, to have him pursue the goals we want, be they in the other's or in our own interest. Therefore deception, as well as communication, may be either a selfish or an altruistic behaviour. In the second part of the paper, we described a model to generate facial expressions of an agent deceiving about her emotional state. Our model introduces the diversification of facial expressions for masked, inhibited and fake expressions. As a consequence, these different types of deceptive facial expressions can be distinguished by the user, because their appearance is different. Then we also presented how an ECA can deceive about her emotional state by adapting her facial behaviour to the social context. In particular we showed that different types of facial expressions used in the same situation can express interpersonal relations.

Acknowledgement
This work has been partially done while two of the authors were part of the "Embodied Communication" research group at ZIF. Part of the research is supported by the EU FP6 Network of Excellence HUMAINE and by the EU FP6 Integrated Project Callas.

References

1. Augustinus, A.: De Mendacio. In: Migne, J.P. (ed.): Patrologica Latina, tom. 40 S. Aurelii Augustini Opera Omnia 6. Turnolti: Brepols (1968) 488–518
2. Bavelas, J.B., Black, A., Chovil, N., Mullett, J.: Truths, lies, and equivocations: The effects of conflicting goals on discourse. Journal of Language and Social Psychology 9 (1990) 129–155
3. Becker, C., Kopp, S., Wachsmuth, I.: Simulating the Emotion Dynamics of a Multimodal Conversational Agent. In: André, E., Dybkjaer, L., Minker, W., Heisterkamp, P., (eds.): Affective Dialogue Systems, Springer Verlag (2004) 154–165
4. Bevacqua, E., Mancini, M., Niewiadomski, R., Pelachaud, C.: An expressive ECA showing complex emotions. In: Proceedings of the AISB Annual Convention. Newcastle, UK (2007) 208–216
5. Boffa, S., Pascucci, F., Poggi, I.: La banca delle bugie. In: Bonfantini, M., Castelfranchi, C., Martone, A., Poggi, I., Vincent, J. (eds.): Menzogna e simulazione. Edizioni Scientifiche Italiane, Napoli (1997) 227–240
6. Bok, S.: Lying. Moral Choice in Public and Private Life. New York, Pantheon (1978)
7. Brown, P., Levinson, S.C.: Politeness: some universals on language usage. Cambridge University Press (1987)
8. Buck, R., Losow, J., Murphy, M., Costanzo, P.: Social facilitation and inhibition of emotional expression and communication. Journal of Personality and Social Psychology 63 (6) (1992) 962–968

9. Buisine, S., Abrilian, S., Niewiadomski, R., Martin, J.-C., Devillers, L., Pelachaud, C.: Perception of blended emotions: From video corpus to expressive agent. The 6th International Conference on Intelligent Virtual Agents. Marina del Rey, USA (2006)
10. Buller, D.B., Burgoon, J.K.: Interpersonal Deception Theory. Communication Theory 6 (3) (1996) 203–242
11. Byrne, R.W.: The Ape Legacy: The Evolution of Machiavellian Intelligence and Anticipatory Planning. In: Goody, E. (ed.): Social Intelligence and Interaction. Expressions and Implications of the Social Bias in Human Intelligence. Cambridge University Press (1995)
12. Byrne, R.W., Whiten, A.: Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans. Oxford, Clarendon Press (1997)
13. Carofiglio, V., de Rosis, F., Grassano, R.: An Interactive System for Generating Arguments in Deceptive Communication. Springer LNAI 2175 (2001)
14. Cassell, J.: Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents. In: Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (eds.): Embodied Conversational Characters. MITpress, Cambridge, MA (2000)
15. Castelfranchi, C.: To believe and to feel: The case of "needs". In: Canamero, D. (ed.): Emotional and Intelligent: The Tangled Knot of Cognition. California: AAAI Press (1998) 55–60
16. Castelfranchi, C.: Reasons: Belief Support and Goal Dynamics. Mathware & Soft Computing 3 (1996) 233–247
17. Castelfranchi, C., Parisi, D.: Linguaggio, conoscenze e scopi. Il Mulino, Bologna (1980)
18. Castelfranchi, C., Poggi, I.: Bugie, finzioni, sotterfugi. Per una scienza dell'inganno. Carocci, Roma (1998)
19. Castelfranchi, C., de Rosis, F.: Which User Model do we need, to relax the hypothesis of "Sincere Assertion" in HCI? UM99 Workshop on Attitudes, Personality and Emotions in User-Adapted Interaction. Banff, Canada (1999)
20. Conte, R., Castelfranchi, C.: Cognitive and social action. London, University College (1995)
21. Christie, R., Geis, F.L.: Studies in Machiavellianism. New York, Academic Press (1970)
22. De Carolis, B., Pelachaud, C., Poggi, I., De Rosis, F.: Behavior Planning for a Reflexive Agent. In: Proceedings of the Conference IJCAI 2001. Oporto, Portugal (2001)
23. de Rosis, F., Castelfranchi, C., Carofiglio, V.: On various sources of uncertainty in modeling suspicion and how to treat them. Workshop on Deception, Fraud and Trust in Agent Societies, Autonomous Agents 2000 Barcelona (2000)
24. Devillers, L., Abrilian, S., Martin, J.-C.: Representing real life emotions in audiovisual data with non basic emotional patterns and context features. In: Proceedings of First International Conference on Affective Computing & Intelligent Interaction. Pekin, China (2005) 519–526
25. Ekman, P.: Telling lies: Clues to Deceit in the Marketplace, Marriage, and Politics. New York, Norton (1985)
26. Ekman, P.: Darwin, deception, and facial expression. Annals of the New York Academy of Sciences 1000 (2003a) 205–221
27. Ekman, P.: The Face Revealed. Weidenfeld & Nicolson, London (2003b)
28. Ekman, P., Friesen, W.V.: The repertoire of nonverbal behavior's: Categories, origins, usage and coding. Semiotica 1 (1969) 49–98
29. Ekman, P., Friesen, W.V.: Unmasking the Face. A guide to recognizing emotions from facial clues. Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1975)

30. Ekman, P., Friesen, W.V.: Felt, false, and miserable smiles. Journal of Nonverbal Behavior 6 (1986) 238–252
31. Elliott, C.: The Affective Reasoner: A Process Model of Emotions in a Multi-agent System. PhD thesis, Institute for the Learning Sciences, Northwestern University (1992)
32. Fellous, J.M., Arbib, M.A. (eds.): Who needs Emotions? The brain Meets the Robot. New York, Oxford University Press (2005)
33. Frank, M.G., Ekman, P., Friesen, W.V.: Behavioral Markers and Recognizability of the Smile of Enjoyment. In: Ekman, P., Rosenberg, E.L. (eds.): What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). Oxford University Press (1995)
34. Gonzaga, G.C., Keltner, D., Londahl, E.A., Smith, M.D.: Love and commitment problem in romantic relation and friendship. Journal of Personality and Social Psychology, 81 (2) (2001) 247–262
35. Gosselin P., Kirouac, G., Doré, F.Y.: Components and Recognition of Facial Expression in the Communication of Emotion by Actors. In: Ekman, P., Rosenberg, E.L. (eds.): What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). Oxford University Press (1995) 243–267
36. Grice, H.P.: Logic and Conversation. In: Cole, P., Morgan, J.L. (eds.): Syntax and Semantics, Vol III, Speech Acts. New York, Academic Press (1975)
37. Grotius, H.: De iure belli ac pacis libri tres. Paris (1625)
38. Ivkovic, V., Grammer, K.: Communication, Deception, Indoctrination, Dogmatism and Memes - Neuroethology of Mind Control. Collegium Antropologicum 26 Suppl. (2002) 91
39. Janovic, T., Ivkovic, V., Nazor, D., Grammer, K., Jovanovic, V.: Empathy, communication, deception. Collegium Antropologicum 27 (2) (2003) 809–822
40. Kant, I.: Grundlegung sur Metaphysik der Sitten, Leipzig, 1785. Engl. Tr. Groundwor of the Metaphysics of Morals, ed. by Gregor, M., Introduction by Korsgaard, C. Cambridge University Press (1997)
41. Keltner, D.: Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. Journal of Personality and Social Psychology 68 (1992) 441–454
42. Kopp, S., Wachsmuth, I.: Synthesizing Multimodal Utterances for Conversational Agents. The Journal Computer Animation and Virtual Worlds, 15 (1) (2004) 39–52
43. La France, M., Hecht M.: Option or Obligation to Smile: The Effects of Power and Gender and Facial Expression. In: Philippot P., Feldman R.S., Coats E.J. (eds.): The Social Context of Nonverbal Behavior (Studies in Emotion and Social Interaction). Cambridge University Press (2005) 45–70
44. Manstead, A.S.R., Fischer, A.H., Jakobs, E.B.: The Social and Emotional Functions of Facial Displays. In: Philippot, P., Feldman, R.S., Coats, E.J. (eds.): The Social Context of Nonverbal Behavior (Studies in Emotion and Social Interaction). Cambridge University Press (2005) 287–316
45. Maynard Smith, J.: Do animals convey Information about Their Intentions. Journal of Theoretical Biology 97 (1982) 1–5
46. Miller, M., Galanter, E., Pribram, K.H.: Plans and the structure of behavior. New York, Holt, Rinehart and Winston (1960)
47. Mitchell, R.W., Thompson, N.S. (eds.): Deception: Perspectives on Human and Nonhuman Deceit. Albany (NY), Suny (1986)
48. Niewiadomski, R.: A model of complex facial expressions in interpersonal relations for animated agents. Ph.D. Thesis. University of Perugia (2007)

49. Ochs, M., Niewiadomski, R., Pelachaud, C., Sadek, D.: Intelligent Expressions of Emotions. In: Proceedings of First International Conference on Affective Computing and Intelligent Interaction (ACII'2005) Beijing, China (2005)
50. Ortony, A., Clore, G.L., Collins, A.: The cognitive structure of emotions. Cambridge: Cambridge University Press (1998)
51. Picard R.: Affective Computing. MIT Press (1997)
52. Poggi, I.: The goals of persuasion, Pragmatics and Cognition 13 (2) (2005) 297–336
53. Poggi, I.: Emozioni, credenze e scopi. Incognito (2007)
54. Prendinger, H., Ishizuka M.: Social role awareness in animated agents. Proceedings of the fifth international conference on autonomous agents, Montreal, Quebec, Canada (2001) 270–277
55. Reeves, B., Nass, C.: The media equation: how people treat computers, television, and new media like real people and places. Cambridge University Press (1996)
56. Rehm M., André E.: Informing the Design of Embodied Conversational Agents by Analysing Multimodal Politeness Behaviors in Human-Human Communication. Workshop on Conversational Informatics for Supporting Social Intelligence and Interaction (2005)
57. Rehm, M., André, E.: Catch Me If You Can - Exploring Lying Agents in Social Settings. Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS'2005) (2005) 937–944
58. Rickel, J., Marsella, S., Gratch, J., Hill, R., Traum, D., Swartout, B.: Towards a New Generation of Virtual Humans for Interactive Experiences. IEEE Intelligent Systems (2002) 32–38
59. Spencer-Oatey, H.: Reconsidering power and distance. Journal of pragmatics 26 (1996) 1–24
60. Swift, J.: Gulliver's travels. Dublin, Faulkner (1726)
61. Trivers, R.L.: The evolution of reciprocal altruism. Quarterly Review of Biology 4 (1971) 35–37
62. Wagner, H.L., Smith, J.: Facial expression in the presence of friends and strangers. Journal of Nonverbal Behavior 15 (4) (1991) 201–214
63. Whiten, A., Byrne, R.: Tactical Deception in Primates. Behavioral and Brain Sciences 11 (1988) 233–244
64. Whiten, A., Perner, J.: Fundamental Issues in the Multidimensional Study of Mindreading. In: Whiten (ed.): Natural Theories of Mind. Oxford, Basil Blackwell (1991)
65. Wiggins, J.S., Trapnell, P., Phillips, N.: Psychometric and geometric characteristics of the Revised Interpersonal Adjective Scales (IAS-R). Multivariate Behavioral Research 23 (3) (1988) 517–530