Cogn Tech Work (2017) 19:211–231
DOI 10.1007/s10111-017-0417-3

ORIGINAL ARTICLE

On human–machine relations

Asaf Degani[1] · **Claudia V. Goldman**[1] · **Omer Deutsch**[1] · **Omer Tsimhoni**[1]

Received: 8 June 2016 / Accepted: 22 June 2017 / Published online: 12 August 2017

Abstract This paper presents an approach to human–machine interactions based on the concept of teamwork and the psychological theory of object relations. We envision the human and the machine in a close relationship that has many aspects of human-to-human relations. Not only does the machine have to relate to and accommodate human wants and needs, but also, to some extent, the human is called to reciprocate. We propose a framework consisting of eleven attributes that describe generic processes in teamwork: commitment, goal definition, common ground, belief, planning, transparency, sensitivity, caring, responsibility, trust, and reflection. Using an automotive climate control system as an example, we show how some of these attributes can be used to evaluate user interactions and point to new design opportunities. Based on results from a pilot study of driver interaction with the climate control system, we operationalized sensitivity and caring for other team members, encapsulated them in a computational architecture, and implemented a control interface. The evaluation of the control interface during a driving experiment suggests that it is markedly better than a regular interface and is almost as good as a human expert who interacts with the climate control system in response to the driver's needs and wants.

Keywords Automotive · Automation · Autonomy · Teamwork · Modelling · Field study · Artificial intelligence

1 Introduction

In their studies of user interactions with computers, television, and other media, Byron Reeves and the late Clifford Nass concluded that humans interact with machines in the same way as they *relate* to another human being (1996). They argued that such user interactions "are fundamentally *social* and *natural*, just like interactions in real life" (p. 5, italics in text). With an eye towards human interaction with automated systems and autonomy, this article discusses how to best operationalize these concepts for the purpose of design. We first define teamwork, which we see as a primary manifestation of *social* interaction in the context of work. We then examine *natural* interactions between humans and machines as a mirror of human relations in real life. We argue that an understanding of these two fundamental concepts can lead to the development of user interfaces that are simple, intuitive, and minimal in terms of interaction features, but at the same time can handle the complexities of current and future technological systems—especially those with autonomy.

In recent years, there has been growing interest in the design of autonomy (e.g. rovers, drones, self-driving vehicles) since these advances are considered the next frontier in technological development. Semantically, autonomy has three meanings. The first, which comes directly from Greek (*auto* = self, *nomous* = laws), is the authority to create and apply one's own laws (Hansen 1992, pp.18−20). The second denotes *self-sufficiency*, or the capacity of an entity to take care of itself without the help of others (*autarky* is the original Greek term), and the third refers to the quality of *self-directedness*, or freedom from outside control (*Eleuthera*, which is commonly translated as freedom). Although all three meanings evoke a sense of independency which mirrors the political science

---

✉ Asaf Degani
asaf.degani@gm.com

[1] General Motors, Herzliya, Israel

essence of the term, modern applications of autonomy involve heavy interactions with humans. In their insightful analysis of human–autonomy interactions and review of future design directions, Bradshaw et al. called for advances in human–machine teamwork that would enable autonomous systems not only to merely to do things *for* people, but to also work together *with* people. They view this capacity for human–machine teamwork as the key requirement towards effective design of future autonomic systems (2013).

Given the increasing capabilities of machines to conduct work that only a few years ago would seems to have been impossible to carry out (e.g. driving a car), we anticipate that in the not-so-distant future human and the machine will be required to join forces. Such joint human–machine "teams" will not only improve performance on the mundane aspect of the mission (flying, driving, etc.), but also perhaps even benefit overall mission effectiveness, specifically in novel and complex situations (e.g. involving moral decision-making) where machine capabilities are lacking. To this end, we envision this future team-oriented system as cooperative (Goldman and Degani 2012). Current developments have focused primarily on teams of automated agents and even on mixed, human, and automated agent teams, but not on the cooperative aspect (Horvitz 1999; Rich et al. 2001). Cooperation, even in human teams, is difficult to achieve because much of the "glue" that binds and melds people into cooperation is emotional and psychological (Woolley et al. 2010).

In the next sections, we define the notion of "team" and survey some of the key literature on team formation, cohesiveness, and maintenance of social groups and teams comprised of humans. We then gradually present work on teams of machine agents, primarily from an artificial intelligence perspective, and then summarize some attempts from within the human factors literature to analyse mixed teams of humans and machines.

A team is generally defined as "a number of persons associated together in work or activity" (Merriam-Webster 2011), but in the context of work involving for example military, medical, or flight operations, a more representative definition is "a distinguishable set of two or more people who interact dynamically, interdependently and adaptively towards a common and valued goal/objective/mission, who have each been assigned specific roles or functions to perform, and who have a limited life span of membership" (Salas et al. 1992, p. 4; Paris et al. 2000). In the context of this paper, we extend the Salas and colleagues definition to include not only "people" but also machine agents.

A significant amount of human work is performed by teams rather than by individuals (Sasou and Reason 1999). While this is particularly true in complex technological systems, it is also a reality in business, medical care, government, and many social organizations. In terms of productivity, there are many advantages to teamwork such as increased performance, efficiency, and redundancy. Emotionally, being part of a group provides mutual aid, support, and a sense of belonging and cohesiveness that most of us strive for. However, learning how to function in a team, let alone how to forge and lead one, is far from trivial.

Interest in team formation and evaluation began during the 1980s with Belbin's influential work on successful teams (Belbin 1981, 1993). Texts on the topic by business writers and management scientists can be divided into two categories: one deals primarily with leadership, role behaviours, and the way a team impacts performance (Davis et al. 1992), while the other line of inquiry focuses on models of team performance and ways to measure this (McFadzean 2002). In social psychology, there has always been a keen interest in team formation and performance. In aviation, the notion of teamwork attracted considerable attention in the late 1980s after several airline accidents where a lack of teamwork skills among pilots was found to be a contributing factor to the accident (Wiener et al. 1993). Of special concern are situations in which the flight crew did not share critical information and ignored team members' inputs and advice (Foushee 1984). This revelation, coupled with extensive research, resulted in the mandate of Crew Resource Management training for pilots by the Federal Aviation Administration (2004). The concept was quickly adopted in the medical, chemical, nuclear, and maritime fields, as a way to train employees to work more safely and efficiently.

In the early 1990s, researchers began to explore the nature of human action and planning of work and activities. Bratman (1992) devised the shared activity framework that captures the way cooperative actions unfold in human teams. He emphasized that a group of individuals doing work is not necessarily a team. Rather, what makes a team involves the presence of three main factors: (1) commitment to mutual support and the overall activity, (2) responsiveness of the members to each other's needs, and (3) "meshing" of the team members' individual plans into a joint plan.

In artificial intelligence, Grosz and Kraus (1996) extended Bratman's framework to develop a theory of teamwork. From their perspective, teams of intelligent agents are formed by developing and implementing shared plans which specify the capabilities needed by agents to plan and act together. Teamwork is seen as a special type of shared activity where the resulting joint plans constitute more than the composite of the individual agents' plans. Beliefs and intentions about other agents' states and actions also need to be considered for the team to function properly. Follow-up research began to consider computational approaches to embedding human behaviour in interactive systems (e.g. agents that adapt their negotiation strategies to humans and agents which provide strategic information to users; see Azaria et al. 2012; Gal et al. 2011).

Within the human factors community, Christoffersen and Woods (2002) identified two fundamental attributes that enable automated agents to become team players: *observability* (to see what the automated agents are doing) and *directability* (as a means to redirect machine activities). Klein et al. (2004) followed up this line of work by examining other team attributes such as "basic agreement," models of each other, goal negotiation, and the like.

This brief overview has covered the concept of teamwork from the humanistic, management and organizational/social psychology points of view, and the standpoint of technological approaches to work practices in teams. The humanistic view of teamwork provides the necessary foundation to think about teams of humans and agents, and their formation. Studies in management science and social/organizational psychology highlight the importance of human relations in teams. Advances in artificial intelligence point to how machine agents can form teams, how they behave in a team setting, and how they actually act. This literature hints at the ways in which joint human–machine teams could be formed, whereas thinking in the field of human factors has defined a number of attributes of joint human–machine teams.

Hence, the focus of this paper is on the necessary and attributes of teams and how to operationalize them in the context of human interaction with machines. We begin by analysing how humans work in groups and what can be learned from this as regards joint teams of humans and machines. We then apply the concept of a joint team to an automotive system to observe how a team-oriented design approach can be realized. We further examine such close human–human interrelations in the context of a comfort control system in an automobile and discuss how these relations can be operationalized. The last section presents an initial implementation and evaluation of a joint human–machine team.

2 A teamwork framework

This section outlines some of the attributes of successful human teams based on the literature on human teams. Throughout this section and for each attribute, we ponder on the implications for teams of humans and machine agents that can cooperate to conduct effective, rational, and fulfilling work. (Fig. 1)

2.1 Commitment

Partnership, writes Lao Tzu, the sixth century BC Taoist philosopher, is first and foremost about commitment (Amon 1998). The continual success of a team depends on its members and their joint commitment to other group members and the task/mission. One way to establish commitment is through a contract, whether formal or tacit. The general problem of a contract can be formulated as follows: how to compose a "bargain" to enhance the likelihood that the contract will be honoured over time, despite unmet expectations, misconduct, deviations, and gross violations. Contracts and their management were analysed extensively by Schelling (1984, 1985) who demonstrated that the foundation of contractual commitment, and especially its maintenance over time, is enhanced by the creation of a (humanistic) *binding conditions* between the parties.

In the context of human–machine team formation, the commitment is to engage in, maintain, and also terminate the joint activity. When a human user decides to enter into a team relation, he or she explicitly communicates this intention. Similarly, the machine communicates its availability and "willingness" to join and enter into this contract. Once the machine has acknowledged its availability, they are both now bound to a joint activity with all its commitments and implied responsibilities.

In human teams, there are many forms of social pressures that obligate group members to honour their commitments and stick with the team. Shame and perhaps retribution can also be enacted against those who fail to keep their commitments, violate rules, or leave the group. The ways in which a machine agent could plausibly violate or deviate from its commitments are a topic of concern in many human–machine systems (Casner et al. 2016). These includes machine failures during unexpected situations and how this process is negotiated and "takeover" assumed (Casner et al. 2013; Degani 2004 Ch. 15 and 17). In aviation, there have been more than a few accidents where a misunderstanding of the machine's commitments, a lack of graceful deterioration, and improper notifications have been a contributing factor in accidents (National Transportation Safety Board 1996, 2014). In such "contracts" between humans and machines, violations can also be assessed by either side prior to entry into contractual relations. For example, a human may decide not to engage in a joint activity with a machine out of concerns for reliability, or simply because the machine shows some indications of actual or potential failure. By contrast, in the future, a vehicle could notice that its human driver is unfit to drive; the machine may select not to enter into a contract or decide to terminate it by coming to a stop at some safe location.

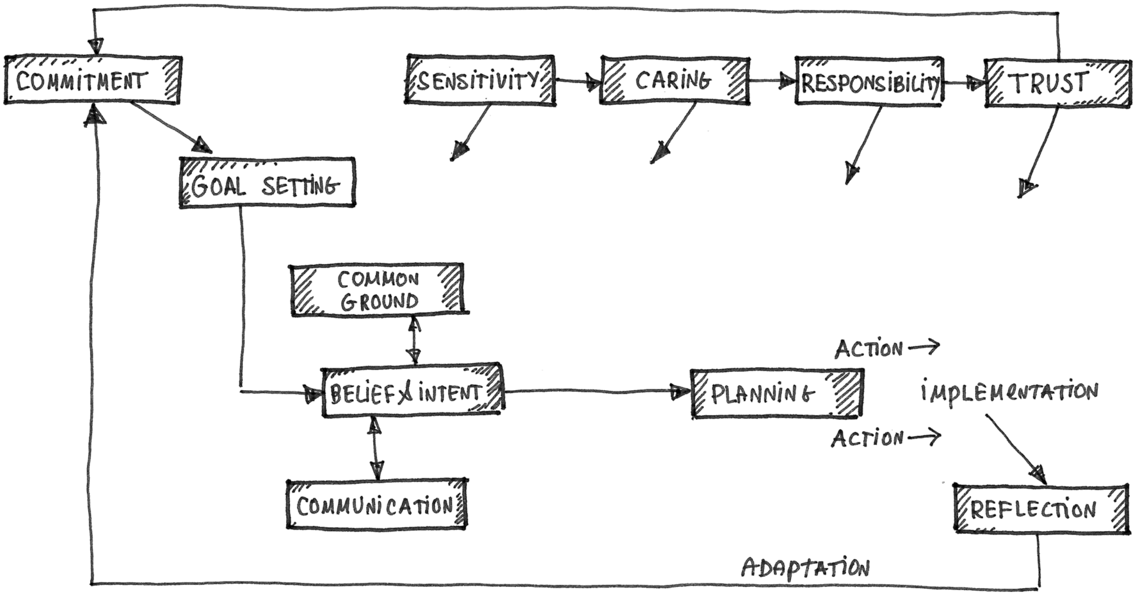

Fig. 1 Eleven-attribute framework of teamwork

2.2 Goals and policies

It is imperative to define the goals and objectives of a given mission. Sometimes the goal and policies are prescribed in the contract. Usually, the overall goals in a team should be shared and not be privy to only a subset of the team members. For example, the clash between the human astronauts and HAL900, the all-powerful computer that "runs" the spacecraft in Stanley Kubrick's *2001: A Space Odyssey* was triggered by the difference between what the astronauts knew as compared to what the computer knew about the purpose of the mission (1968).

The mission objective and goals can be broken down into tasks, with details and assignments. In a team, each member has a task, which is usually associated with his/her specific role that feeds into the overall team goal. Sometimes there is an advantage when team members are not given specific instructions, or a procedure, as to how to do the task, but more of a general policy of action (Barshi et al. 2016). This is one approach to the problem of goal execution, which provides goals and guidelines at the top, and then delegates a great deal of authority and responsibility at the operating level (Bowers 1966, p. 106–107).

2.3 Common ground and shared context

It is almost impossible for a group of individuals to function together as a team when there is no common ground upon which to develop the work. It is only via this common ground that relations among team members in a group can actually be understood and tasks managed. The creation and maintenance of a common ground have been explored in *group analysis*, a psychoanalytical theory that arose from the realization that an individual and his or her close social setting cannot be separated (Brown and Zinkin 1994, p. xii; Foulkes 1964).

In every group, there is an unconscious undercurrent which serves as a common ground that ultimately determines the meaning and significance of all events that take place. All communications, interactions, and interpretations, verbal and non-verbal, in a team emerge from this common ground (Foulkes 1964, p. 292). Foulkes defined two levels of common ground, which he termed the *matrix*: the first is the *foundation matrix* which is based on the common ground preceding the formation of the group; the second is the *dynamic matrix* which undergoes construction as the team is developing and working together: "Even a group of total strangers, being of the same species and more narrowly of the same culture, share a fundamental mental matrix (foundation matrix). To this their closer acquaintances and their intimate exchanges add consistently so that they also form a current, ever moving ever developing dynamic matrix" (Foulkes 1990, p. 228). The *matrix* is the backdrop against which group processes and relations take place as well as their constant re-negotiation. Much of the inner tensions in human teams often arise between the fundamental and dynamical matrix. Wilford Bion, a prominent Neo-Klaninan psychoanalyst, dives deeper into the unconscious foundation of a group with his notion of the "basic assumption" of a group—a set of unconscious fantasies and inner relation that underlie group dynamics—giving rise to behaviours such as "paring," "dependence," "master–slave" relations (Bion 1961, p. 146). These relations impart powerful emotional drives on a group and are the "cement" that keeps the group assembled (López-Corvo 2002, p. 39).

What, then, constitutes the *foundation matrix* of joint human–machine teams? For humans, it is perhaps the expectation that the machine will honour the contract, function as advertised, and will not falter. And as for the *basic assumption*, that perhaps is currently based on inherent fears about machines and robots (leading perhaps to "master–slave" and "fight-flight" images), but may change to "peer-to-peer" relations or others once such robots are fielded and actually experienced. (This issue of the underlying *basic assumption* is currently a topic of much deliberation among designers of autonomous vehicles and robots.) With respect to the dynamic matrix, there needs to be some kind of ongoing negotiation between the human and the machine about expectations (what is permissible now and what is not) also "context understanding"—where the human can supply the machine with the context in which it operates (or the machine can also try to extract this context on its own); the machine, for its part, has to communicate its current level of understanding of the situation.

2.4 Beliefs and intentions

Humans seek to understand the belief system within which the machine operates (its behaviour, operational boundaries, and limitations) as well as its intentions. The machine creates, maintains, and updates a model of its user's belief system (i.e. understanding of the world) as well as its preferences and intents. This model can be extended to include ontologies (i.e. more general information about the world) and the more concrete input/output model of user interactions with the machine (defined as the "user model" in Degani et al. 2013a, b). Humans provide information about their belief system and intent by communication, either directly or indirectly (e.g. gestures, facial expressions, indirect comments), and the machine provides essential information about its behaviour, operational boundaries, limitations, intentions, history, states, and actions.

2.5 Planning and coordination

Many team and group activities centre around planning, coordinating, and then communicating the status, upkeep, and amalgamation of tasks (Stout et al. 1999). In a joint human–machine team, the humans plan and execute actions while considering the machine's plans, whereas the machine computes its plans while taking into account the user's plans. When there are joint planning and synchronization, these plans and activities are said to be "meshed" (Bratman 1992; Grosz and Kraus 1996).

2.6 Transparency and communication

Understanding the other's beliefs and coordinating plans and actions requires a certain degree of transparency and visibility, which can only take place through a shared communication language. The nature of communication between team members is important for any group effort (Sasou and Reason 1999), and this is even more critical for humans and machine where there is little in common in terms of language and communication style. Below we discuss some of the problems between the "language of the user" versus the "language of the machine" and how to overcome this obstacle. Communication theory provides models of what constitutes an acceptable interaction (maxim of information quantity, quality, relation, manner, and clarity) that can be applied to human–machine communication (Grice 1975). Finally, understanding what information is essential and what is not is a rather complicated abstraction process that requires an understanding of both the fundamental and dynamic matrix, beliefs and limitations, as well as goal development (Cannon-Bowers et al. 1990). At times there may be a reality that impedes either or both parties from being transparent even if they want to communicate properly. This brings issues to the forefront such as sensitivity, caring, and the willingness to act in good faith despite impediments, as will be discussed next.

2.7 Sensitivity

The capacity for *sensitivity* is probably one of the key points in establishing a successful group relation. The willingness to be observant and attuned to other team members' innate needs, expressed wants, limitations, and state of mind while trying to achieve a goal is a difficult requirement. Recent research suggests that one factor which differentiates effective and successful teams from others is sensitivity among team members as it relates their teammates' inner needs, limitations, and abilities. Termed the *collective intelligence* factor by Woolley et al. (2010), this "social" sensitivity is correlated with the average social sensitivity of group members, as measured by the ability to assess another's emotional state (Baron-Cohen et al. 2001). What is surprising about the results about collective intelligence is that team effectiveness was shown to be only slightly correlated with the kind of factors that are usually revered in the literature such as maximum individual intelligence of group members and the leader's IQ level. Although the term *collective intelligence* may be a broad brush, it does point to the importance of emotional processes in group dynamics. We assume that being sensitive means being attuned to the foundational matrix of the team, and knowing how to bridge problems by modifying and enhancing the dynamic matrix.

One human capability to achieve sensitivity is *empathy*, which is the capacity to feel and understand the other by distancing oneself from one's singular perspective and placing oneself within the other's frame of reference (in "another person's shoes"). *Sympathy* is somewhat similar, but does not demand to abandon of one's own perspective. Both capabilities emerge from reflective consciousness, or the ability to think about the other (Solomon 2007). From our point of view, empathy and sympathy are stances, or positions, that one has to take in order to achieve sensitivity to the other.

Humans conduct their actions while remaining sensitive to the machine's activities (and limitations) and are responsive to its needs (e.g. information). Humans are aware and sensitive to the subtleties of the machine's state (whether communicated directly or not) and are familiar with its operations. The machine may use knowledge about the user's state of mind and interaction style to create a comfortable atmosphere. The machine behaves in a predictable and proactive manner when it recognizes that the user is having trouble and is respectful of his or her needs.

2.8 Caring

Team building and maintenance over time require social bonding (Salas et al. 2005). In human groups, social bonding is maintained by caring, which is interest and concern for the other's inner and pressing needs. This caring is then transformed into action and support. Caring goes beyond what managers need "to care about others for the job" because it involves actions that are beyond what is expected (Carmeli et al. 2016, p. 49).

Although caring usually entails commitment, the reverse is not always the case. For instance, a hospital orderly may show up for work and do the job as prescribed which that can be defined as committed, but caring for his or her patients is altogether something different. In the context of autonomy, infusing caring in a robot may be somewhat of a stretch, but it is possible to apply some characteristics of caring. The Danish philosopher Søren Kierkegaard argued that unlike erotic love and friendship that cannot be enforced, "caring" can be commanded and may take different forms depending on the needs of the person being cared for. The kind of caring the Kierkegaard was talking about is more about enforcing caring in the sense of "being proper" than an innate passion for others and real caring from the heart. Kierkegaard insisted that caring, of any sort, can only be expressed in concrete action, in the willingness to act, and should not only be judged by the success of its outward accomplishments (Aaron et al. 2008, p. 83). In his formulation, the willingness to act and action itself matters much more than the objective achievement. The ultimate outcome is a concept that can improve trust in human relations. We believe that this will also apply to machine agents.

2.9 Responsibility

There are three basic meanings to "responsibility." One is external and concerns the obligation (being bound) by law or duty, to act. The other is internal and concerns being accountable for one's action and/or inner state. The last concerns the act of being given, or taking, responsibility. Taking responsibility, of any kind, involves establishing ties and sensitivity to what one is responsible for (Saint-Exupéry 1943, Ch. 21). This topic becomes quite relevant when we consider the kind of interrelations to be expected, as well as role gradient, between humans and advanced automated systems in general (Degani 2004) and human–robots relations in particular (e.g. the Three Laws of Robotics devised by the science fiction author Isaac Asimov 1950).

The assumption in groups is that each team member takes on a certain responsibility for his or her actions, the tasks that need to be accomplished, as well as the overall combined group responsibility and perhaps even the overall and long term the implications of the team's output (e.g. moral, social, and environmental impact). Responsibility is derived from the individual's care for oneself, others, and care for the group, as well as his or her role within the social setting of the group (Lewin 1939). Responsibility and caring bring about the requirement of acting in fairness, which is a deeply rooted principle in human relations and reflective of the "warm and fuzzy" social matrix of small bands of early humans (Harari 2015, p. 139).

Reciprocity is one manifestation of responsibility and fairness. The *reciprocity principle* (Chen et al. 2009) has been around since the time of Hammurabi (c. 1792–1750 BC). It is the understanding that one's intentions and actions have implications that will elicit a response from the other (positive in the case of favour, negative as in an "eye for an eye"). For example, we have all been in a situation in a store where the salesperson makes heroic and lengthy efforts to find what we want, thus making it socially awkward to walk out of the store without buying anything. Reciprocity is based on the inner belief of finiteness of one's actions and expectation from others in a group. It can be viewed as a sort of a social "accounting" scheme equivalent of the physical conservation laws (c.f. Social Exchange Theory). Reciprocity and adequacy of social exchange—in the context of business, real estate transactions, communal (e.g. neighbourly), as well as personal relations—have been studied extensively in the Jewish ethical and legal tradition (Tractate Bava Batra, Talmud Bavli).

Respect commonly means esteem and deference, but in a team setting it is somewhat more involved, because it is a fundamental force in human groups, similar to expectations of fairness and justice. In Latin cultures, "respeto," is viewed as a moral value that teaches individuals responsible conduct in a community; a genuine concern and care for others' feelings, dignity, and presence. Respect is the opposite of contempt, humiliation, and disregarding of others which, in team dynamics can lead to conflict. Thus, it can be viewed as a powerful transformational force, in the sense that concessions and successful negotiations within the group (e.g. goals and plans) become easier to attain when the element of respect is present (Farid 2005). Kant's categorical imperative that others should be treated as "ends in themselves" (as opposed to a "means to an end") is the next logical progression of respect (Fromm 2006/1956 p. 14).

In native North American Indian culture, an individual is always seen as part of a whole composed of the environment, other animals, and others (Lame Deer and Erodes 1994, p. 121). Native Americans' notion of the Great Spirit is that of an essence that infuses "unimaginable amounts of force into all things—pebbles, ants, leaves, whirlwinds—whatever you will" (p. 114). This concept is sometime referred to as animism, the view that everything is infused with a soul or spirit. With respect to all others (animate and inanimate) comes the concept of caring and responsibility to everything with which one comes into contact. We argue that in human teams, genuine respect for all members as well as responsibility for one's own and others' actions, is a necessary attribute. The same concepts can be applied to a joint team of human and machine agents where the human is respectful of the machine and its behaviour and is responsible for its upkeep, and the machine is programmed to be responsible for the well-being and welfare of the human team member.

2.10 Trust

The ability to trust is an evolutionary trait in humans that allows us to relate to others and function in groups (Harari 2015). The capacity to trust is either inherent in humans or formed at a very early age (whereas caring requires a more mature understanding of the other—what Donald Winnicott described as the *concern stage*, 1955 p. 264–265). Trust has many advantages in the sense of limiting one's cognitive needs to evaluate multiple options and also in terms of social bonding. In sociology and psychology, trust is a measure of belief in the honesty, fairness, and benevolence of another human being. Trust, however, is not necessarily a belief in the competence of another human being—that is usually defined as "confidence." A failure in the relation with another human being is tolerated to a greater extent if it is perceived as a failure of competence than a lack of benevolence or fairness. In the context of machines, we can only talk about confidence and reliability (a machine cannot be benevolent), and the machine needs to communicate its goals, intention, and limitations (because of our inborn propensity to "over-trust").

2.11 Reflection and adaptation

For a team to develop and improve, it needs the emotional and intellectual capacity to be reflective about its thoughts, feelings, beliefs, plans, and actions and have the space to think and make necessary adjustments and adaptations (Bion 1959). The concept of *reflection* is linked to "self-consciousness," which is the capacity of being aware of oneself and one's own actions not from within but rather from a distance. Reflective consciousness also allows us to be *rational*; i.e. to think critically about what we perceive and to distance ourselves from instinctual responses. Rationality is also the capacity of thinking in the context of a whole and acting properly in the world. Modern philosophers have repudiated the Greek-inspired tradition that rationality is purely logic-based and independent of culture and social standards (Solomon 1993). Rationality of the social kind leads to conformity with the cultural norms and standards, technical, or otherwise of the day.

On a limited scale, the attribute of reflection and adaptation can be applied to a joint team such that humans can acknowledge when the machine performs satisfactorily or not. The machine can use this feedback to update its model and adjust its behaviour with an eye towards the longevity of the relationship (i.e. not necessarily respond immediately). The machine, for its part, can provide feedback to the user about his or her performance and perhaps even provide celebratory moments (Welch and Welch 2011). Nevertheless, we acknowledge that reflection, of the deep thoughtful sense, requires *intentionality* (Searle 1980). Intentionality stresses both understandings of one's "purpose" and also its "meaning." This aspect of reflection, which is derived from the concept of *self-motivation*, is not available to current day machines.

2.12 Towards a framework

The eleven attributes discussed here can be thought of as the necessary ingredients and processes for team formation, conduct, and output. The output may be in the form of actions (work groups), support (social and healing groups), mission completion (in operational domains), generativity and adaptation (in business), as well as creativity and the formation of new ideas (design and research). Note, however, that we do not address such important topics in teams as power distribution, authority, leadership style, and competence. Instead, we focus on those attributes that account for much of the processes in teamwork, with the underlying assumption that they should be present, in some form or another, in mixed human-agent teams.

We propose here a framework that captures some of the interrelations between attributes and hints at a dynamic sequence made up of five stages: *commitment* and the obligation to engage serve as the first stage. Development of a shared *goal* is the second stage, which can only come to fruition and move towards implementation through a shared *common ground*, understanding of the *belief* system, and *communication* (third stage). Intent needs to be communicated such that *planning and coordination* result in actions towards implementation, which is the fourth stage. *Reflection,* for the purpose of adaptation, adjustment, and refinement serve as the final stage of the iteration. Reflection also serves to evaluate the meaning of the entire endeavour, including the processes and outcomes as well as the long-term consequences of the team effort (including renewed commitment). Along with these five stages, we include the four "soft" attributes that, however, impart on all others: *sensitivity* which can be viewed as a fundamental attribute and a necessary step for *caring* and *responsibility*, and the resultant *trust*. Trust is seen as an attribute that develops not only with sensitivity, caring, and responsibility, but also emerges from the group's ability to commit, reflect, and then adapt.

With our first objective of understanding how humans work in groups behind us, we apply the eleven attributes to an actual human–machine system to consider how a team-oriented design approach can actually be realized.

3 Application to a climate control system

We begin by describing the system using *statecharts* (Harel 1987; Harel 2009), a commonly used language to model *reactive* systems (Harel and Pnueli 1985; Harel and Politi 1998). The climate control system is an example of a special class of reactive system, sometimes referred to as an *interactive* system, whose operations involve a significant degree of user interaction. By using statecharts, it is possible to model and describe the behaviour of interactive systems (Parnas 1969), analyse the system and identify problematic design features (Degani et al. 2013a, b; Rushby 2001, as well as develop model-based design solutions (Heymann and Degani 2013). But here, as explained earlier, we wish to take one step further and seek

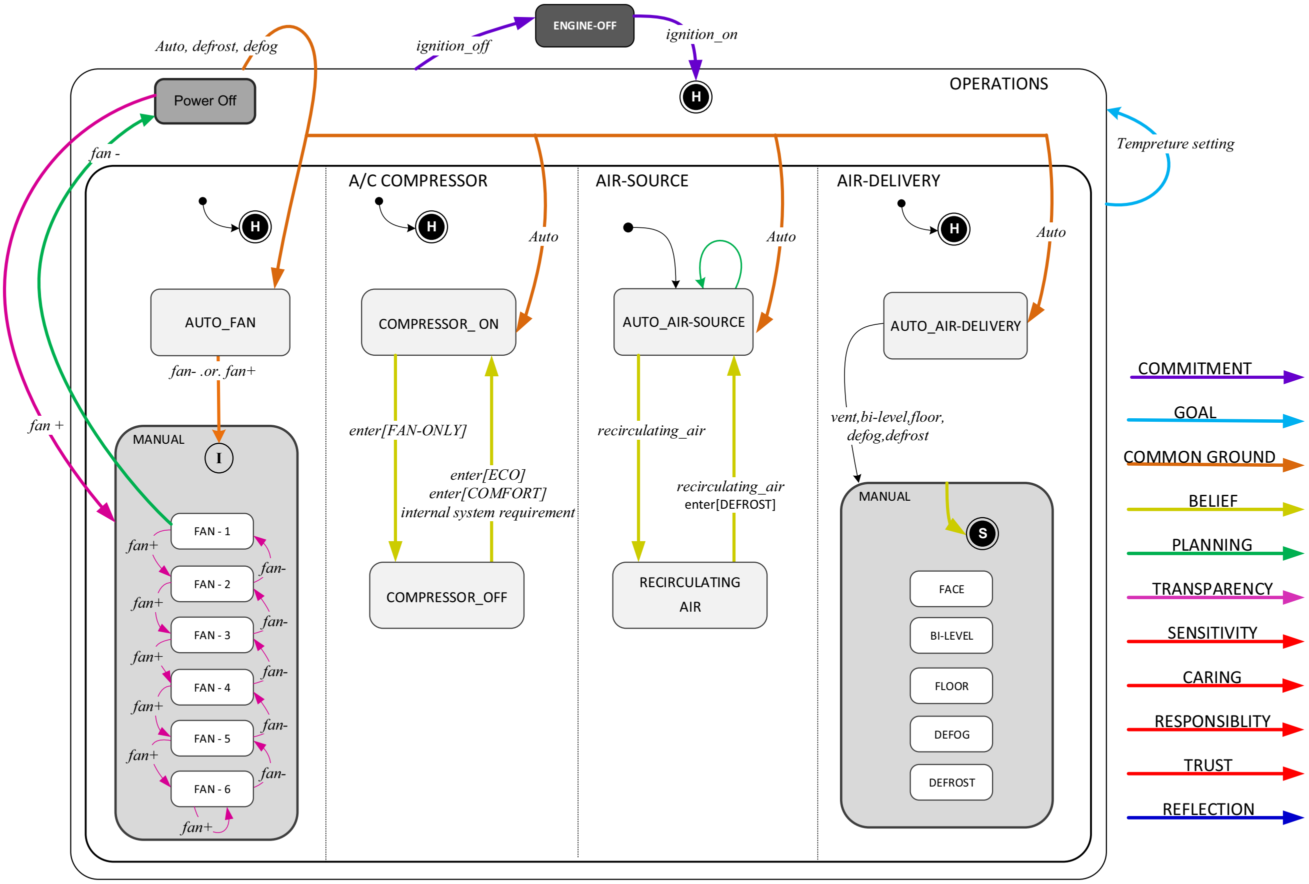

Fig. 2 A statechart model of the climate control system

out opportunities for major design improvement, given a team-oriented approach.

The climate control of an electric car serves as an example here because the system not only has to accommodate the driver's needs, but also be highly efficient (so as to minimize battery consumption and not reduce mileage range), and at times responsible for cooling and heating the batteries. In the following, we describe the main elements that make up the climate control system and their underlying states and transitions (Fig. 2). For almost each state and transition in the system, we discuss the opportunities for team-oriented design improvements (and mark, in colour, the transitions that correspond to each attribute).

3.1 Operations

Once the car is started via the ignition key, the system is operational and ready for work. In most cases, the initial climate control settings (auto, fan speed, air delivery, temperature) are based on the settings from the last ignition cycle (note the H, for history, symbol on the transition *ignition_on*). From a team-oriented approach, turning on the ignition starts the beginning of the partnership between the user and the machine that can be viewed as a *commitment* to engage and maintain joint activity. Given operational constraints (e.g. extreme ambient temperature, limited circulation inside the vehicle, air volumes involved), it may take some time to fulfil the user needs and wants. Thus, the initiation of the "contract" is also an opportunity to explain the *belief set* and limitations to the user under which the machine is working (e.g. a very hot and humid day). At this point, there is also an opportunity for the human to inform the machine about his or her *context* (feeling tired and sweaty) and perhaps preferences.

3.2 Temperature setting

The user sets the temperature and machine attempts to achieve this setting, depending on whether it is in the FAN-ONLY, ECO, or COMFORT mode. In FAN-ONLY mode, the air-conditioning compressor is OFF and only air is used to cool and vent the vehicle. In COMFORT mode, the compressor is on and the fan attempts to deliver maximal power. Since this is an electric car, there is also an ECO mode (for economy), where the compressor is on, but the fan attempts to optimize its setting so as to reduce energy consumption. Although the mode and associated temperature setting define the *goal* of the system (denoted in light blue in Fig. 2), clearly this goal definition is too quantitative and simplistic, because the real goal of the joint human–machine system is to make the user comfortable and content and the temperature, mode, and all other machine setting are simply a means to an end. In a team-oriented design, we can imagine a system where the user communicates his/her wants in terms of human comfort ("I'm hot and sweaty," "it's stifling here," "need air") and not indirect (value) settings in the *language of the machine*. A sensitive system may also attempt to identify the user's underlying needs. Figure 2 depicts the four main components to this system: fan, compressor, air source, and air delivery.

3.3 Fan unit

The fan unit has six speeds (1–6). In manual operations, the system always remembers the last fan speed setting (history) that existed prior to *ignition_off* and returns to this setting when initiated. When the user sets the fan speed to 0, this action, in fact, disables the entire system and no air, conditioned or otherwise, will emerge from the vents. Pressing the "fan increase" button (*fan+*) will reactivate the system. In the ECO mode the fan settings are optimized, which leads to fan speeds that are somewhat different from what the user expects based on his or her actual setting. This is a case of limited *transparency and communication,* denoted via magenta coloured transitions, which can be dealt with by providing better feedback to the user about the state of the fan system (or not providing them at all).

3.4 Compressor unit

The air-conditioning compressor unit is ON when in the ECO and COMFORT modes and OFF in FAN-ONLY. However, because the car is electric there are situations where the batteries need to be cooled and thus the air-conditioning unit will start working on its own, regardless of what the user does. This unique situation concerns the *belief and intentions* of the machine (denoted in yellow-coloured transitions) and can be communicated somehow to the user (*transparency*). How the user and the machine come together to achieve their individual goals (the user is concerned with his/her comfort and the machine is also concerned about minimizing energy consumption) requires a *shared common ground* that needs to be constantly renegotiated (dynamic matrix) depending on the situation.

When the system turns on the air-conditioning or the heating to serve the batteries there is a need to *plan and coordinate activities* with the user who may have other plans. One solution is for the user to shut the vents and set the manual control of the fan to 0 (*fan−*) if he or she does not want cold/hot air entering the cockpit (denoted in the green-coloured transition in the fan unit). The need to plan and coordinate activities requires us to consider the foundational matrix of the user and the kind of foundational matrix that is programmed into the machine (e.g. the user understands that in an electric car he or she may have to sacrifice comfort for range). Similarly, we need to consider the dynamic matrix in the relation and how such negotiations, successful or otherwise, take place so as to minimize frustration (Azaria et al. 2015; Rosenfeld et al. 2012). Such considerations are not part of the design process today.

3.5 Air source

The default setting of the air source is automatic fresh air and the system changes anew to this setting with every engine start (*ignition-on*). This is a unique default condition (represented by an event arrow originating from the small circle in Fig. 2) in comparison to all other initial conditions in this system (fan, compressor, air delivery) where the initial setting is based on the (historical-H) setting from the previous ignition cycle. Here the automatic system tries to maintain the best air quality for the passenger (note the arc transition, in green, indicating *planning* of internal switching between inside air and outside air depending on the outside air quality). Nevertheless, this attempt is not always explained well to the user, which can be considered, from our perspective, as a missed opportunity to better communicate *belief*.

3.6 Air delivery

The air delivery modes include the following airflow possibilities: FACE, BI-LEVEL, FLOOR, DEFOG, and DEFROST. Results from user studies of climate systems indicate that most people are confused about the use of air delivery modes and are willing to settle for a less than optimal configuration. Most people are not aware that it is possible to configure the air delivery to suit almost every want and need. Hence, there is an opportunity to showcase these system capabilities (yellow-coloured transitions) via the interface.

3.7 Auto mode

Although the system provides the user with the ability to take manual control of the climate control system, it is designed to operate in AUTO mode in which the fan speed, air delivery mode, and air source are controlled automatically to achieve maximum effectiveness and efficiency without burdening the driver with mode selection. The general idea is that the user selects auto mode and then sets the temperature, and the rest is done automatically. This design approach has a *common ground* of the underlying matrix (amber-coloured transitions) behind it, but the interface does not afford such understanding. Especially in AUTO mode, there could be ways to indicate to the driver and passenger how the system is trying to fulfil their wants and needs; specifically, that there is ongoing *commitment* as well as all the necessary *planning and coordination* on part of the machine to work hard, and at times minimize efficiency, to keep the user satisfied.

In summary, we identified many opportunities for team-oriented design in terms of the mechanics of team attributes (common ground, belief set, intent, and planning and coordination). However, there was little we could do with the more humanistic features such as sensitivity, caring, responsibility, and trust. In the next section, we focus on some of these humanistic aspects of team attributes.

4 Operationalizing sensitivity and caring

We did not know how to operationalize the humanistic aspects of team attributes let alone implement them. There is no literature in the system design, psychology, or human factors on this topic. However, what we do know from the above literature is that "sensitivity to the Other" is the basis for many humanistic attributes. We decided to approach this problem by conducting a pilot study where we replaced the interface to the climate control system discussed above by an actual human being. The idea was that the user would interact with the system through the human assistant, who was instructed to be sensitive and caring to the driver.

4.1 Pilot study

Eleven participants drove an electric car with an automated climate control system for more than an hour. Figure 3 shows the interface to the driver with all the settings and its three modes—COMFORT, ECO, and FAN-ONLY (top portion).

During the drives, an expert who was well versed with the climate control system occupied the passenger seat next to the driver. The expert was there to manipulate the controls based on the participants' wants and needs. He told the participants that he knew the system very well and that they could communicate whatever they wanted in any way: they could use simplistic phrases such as "I'm hot," "I'm cold," "too windy," or even "I don't feel comfortable," and that based on his close familiarity with their desires and the system capabilities he would find the best configuration to suit their needs.

The pilot study consisted of four consecutive drives, each one about 15–20 min long (depending on traffic) on an intercity highway. The first drive began in a parking garage where the inside temperature was set to a comfortable temperature (25 °C), but the air inside of the car was stale and had a very noticeable smell (produced by an open bag of orange peels that was left in the car overnight). As soon as the participant entered the vehicle, he or she was instructed to drive out of the garage and deal with the climate conditioning while driving. For the second drive, the temperature inside the car was set to hot (30 °C and full fan). Again, the participant was instructed to begin driving

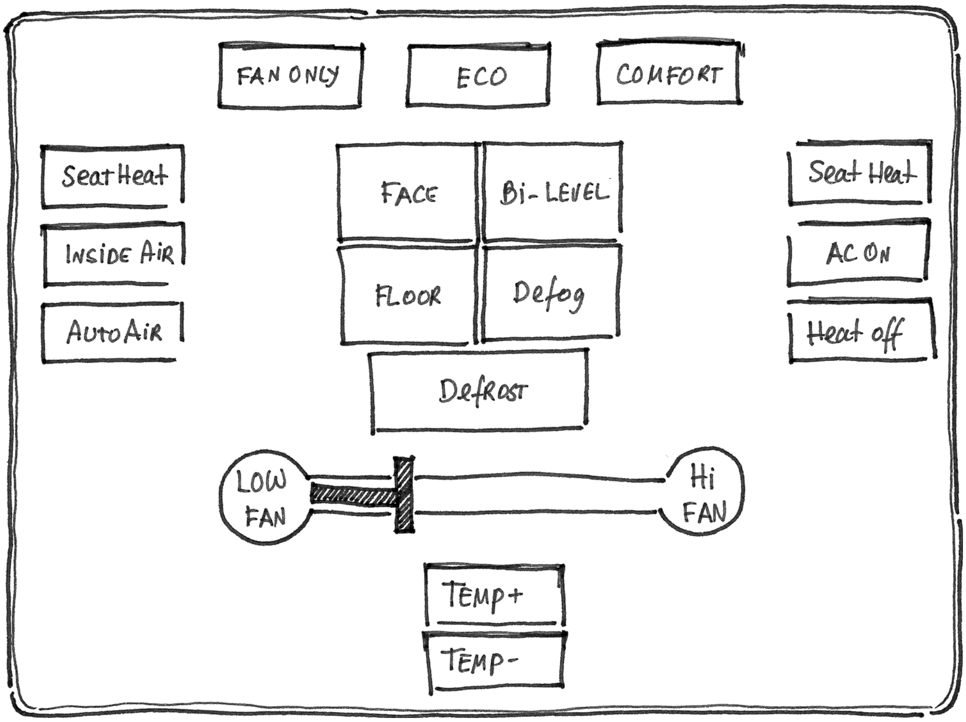

Fig. 3 Schematic of the climate control system

as soon as he or she entered the vehicle and deal with the hot climate while exiting the parking lot and entering the main road. For the third drive, the temperature inside the vehicle was set to cold (16 °C and full fan). For the last drive, the system configuration was reset to the initial setting of the first drive (25°, fan 0, inside air circulation, air delivery to feet) but the seat heating was activated. The general idea was to create some uncomfortable initial state in each drive that would motivate the participant to interact with the climate control system in order to achieve comfort.

Over the course of each trial, as the expert's ability to be sensitive to the user developed, he began using his knowledge of the system to provide the driver with the optimal configuration, which at times was somewhat different from the driver's specific request. For example, see the dataset for participant #11, a rather "specific" participant who made considerable efforts to make sure the system was tailored perfectly to her comfort needs (Fig. 4). As a non-technical person who commonly does not know how to work such systems, she was very satisfied with the (expert operated) climate system because it met her comfort needs fully. For instance, at 14:35 and 14:37 the participant asked to increase the temperature but the expert kept the temperature steady (24 °C) and only changed the fan setting (with satisfaction on the part of the participant at time 14:37—"fine").

What the expert did was to assess the underlying needs of the users and not necessarily only focus on their wants. We also found that as drivers in the study noticed the expert's diligent care for them (using every possible resource in the process including time, effort, knowledge, and courteous affect), they became more comfortable and relaxed. This resulted in their ability to temporarily accept the uncomfortable situation of both cold and heat, knowing that the expert would eventually resolve the situation (*trust*). With time, the expert gained the users' trust, which made them less preoccupied with the climate control system, even when the situation was initially unfavourable. It also made them talk more freely about their even minor dissatisfactions and wants (e.g. 14:14 "same temperature without fan to my hand") and have the expert accommodate them—something that did not occur at the beginning of the drive.

4.2 Object relations

The focus of the pilot study was to try to identify how human's work in a close and intimate team relation. The interaction with the climate control system lends itself to a situation where there was, on one hand, a (needy) user and on the other a (resourceful) provider. The data and the lessons learned by the experimenter shed light on the ways in which a person should be attuned to the needs of the Other, and how to best fulfil these needs. After many debriefings with the experimenter about what he did and how he was able to deal with the stress and discomfort of the driver, we noted that what transpired in the vehicle fit well with the psychoanalytical theory of object relations.

Time	Driver's utterneces	Fan level	Temp. (C^{o})	Air delivery	Circulation
13:43	Drive I: Stuffy cockpit	0	25	1	1
13:46	it's a bit hot	2	23	Floor	inside air
13:49	it's a bit warm you can lower the temperature	3	22	Floor	inside air
13:50	comfortable - yes	--	--	--	--
13:52	air coming from bottom	3	22	Bi-level	inside air
13:53	too much	2	22	Bi-level	inside air
13:54	a bit too strong	1	22	Bi-level	inside air
13:55	temp down, less air not full on me	eco 2	23	Bi-level	inside air
13:56	it's a bit cold	eco 1	24	Bi-level	inside air
14:10	Drive II: Hot cockpit	6	30	Face	inside air
14:12	it's too hot	3	24	Face	inside air
14:12	cooler	4	22	Face	inside air
14:13	keep temp. w/o fan	3	22	Floor	inside air
14:13	it's hot	3	22	Bi-level	inside air
14:14	same temp w/o fan to my hands	4	21	Bi-level	inside air
14:14	that's fine	4	21	Bi-level	inside air
14:25	Drive III: Cold cockpit	6	16	Face	inside air
14:26	it's too cold	3	23	Face	inside air
14:26	too cold	2	23	Face	inside air
14:27	it's better	2	23	Face	inside air
14:27	it's too cold	2	25	Face	inside air
14:28	cold in my arms	2	26	Bi-level	inside air
14:28	more heat	3	27	Bi-level	inside air
14:29	it's warming me up	--	--	--	--
14:30	hot upper body w/o hands	4	27	Floor	inside air
14:30	too much	3	27	Bi-level	inside air
14:30	a bit cooler	3	25	Bi-level	inside air
14:30	cooler	3	24	Bi-level	inside air
14:31	a bit cold, temp fine but less air coming	2	24	Bi-level	inside air
14:32	cold on my arms	2	24	Floor	inside air
14:32	too hot	3	24	Floor	inside air
14:32	a bit more cold, just a little bit	3	23	Floor	inside air
14:33	more	4	22	Floor	inside air
14:33	doesn't come here at all	4	22	Bi-level	inside air
14:33	less cold	3	22	Bi-level	inside air
14:33	less cold	2	23	Bi-level	inside air
14:34	Drive IV	2	23	Bi-level	inside air
14:35	arms cold, face is fine	1	24	Bi-level	inside air
14:36	adjusts vents	--	--	--	--
14:37	less cold, just a bit	eco 3	24	Bi-level	inside air
14:37	fine	--	--	--	--
14:38	-	eco 3	24	Bi-level	inside air
14:40	it's hot now - chair	eco 3	24	Bi-level	inside air

Fig. 4 Dataset for participant #11

This theory, developed by Melanie Klein in the 1930s, is based on the notion of the space between two people—where the subject is dependent on an object who is there to take care of the subject's needs and wants (Klein 1932). The genesis of object relations goes back to the early infant–mother relation, seen from the subject's (infant) viewpoint on the one hand, and the mother's nurturing sensitivity and containing ability on the other. In modern Kleinian psychoanalytic thinking, the understanding and eventual accommodation of the infant's needs are the cornerstone of how we relate and interact with ourselves, other human beings, and the world as a whole (Ogden

2004). In a way, this is the key to *natural* human interaction, which is a move from the one-person view of human behaviour to a perspective that is motivated by a search for another person with whom to relate (Guntrip 1969; Ruszczynski 1993, p. 198). The essence of the relational point of view is that human interaction begins with the relations between two people, involving needs, wants, desires, and expectations, some of which are communicated explicitly, and others which are expressed implicitly.

Can a machine be a viewed as a (psychological) object or at least a stand-in for one? Can we consider human interaction with a machine (such as our climate control system) as reminiscent of an infant–mother relation, where user needs and wants are supplied by a "nurturing" machine? We believe so and discuss two concepts from modern psychoanalytic literature that we think provide meaningful insights into the problem of sensitivity:

The first concept is termed *holding*, which evokes the sensation of a mother embracing her child (Ogden 2004), but in fact is far more inclusive. In her earliest *holding* of the infant (during the first month or two), the mother embeds herself completely into the infant's experience, which results in heightened sensitivity to the infant's needs and well-being (Winnicott 1958/1945). But this heightened sensitivity comes at a hefty price that involves great emotional and physical burdens on the mother's self (e.g. lack of sleep, a dearth of physical replenishment, emotional isolation, and stress). In making this sacrifice, the mother's goal is to foster healthy development by insulating the infant from the existence of "man-made time" and physical reality "and creating in its place the illusion of a world in which time is measured entirely in terms of the infant's physical and psychological rhythms" (Ogden 2004, p. 1350–1351).

It is clear that in order to achieve this maternal capacity of *holding*, there has to be heightened *sensitivity* on part of the mother about the state of the infant. We suggest that Winnicott's formulation of *holding* and its prerequisite, *sensitivity*, can guide the way we think about a team relation. Specifically, a team member needs to be sensitive to the user's needs even when, like an infant, he or she is unable to articulate and communicate them fully. The ability to be sensitive and then willing to "go the extra mile" to help the other, lends itself to the capacity of *holding*.

Winnicott's ontological concept of *holding* maps well to what we discussed above as *caring*. Naturally, a machine cannot encompass what we take as an innate passion for others and caring from the heart. However, it can assume what Kierkegaard described as acting appropriately and the willingness to act, even when it demands some sacrifice. The concept of sensitivity comes close to our initial description in the sense of being attuned to the other's needs, limitations, and state of mind. We thus suggest that some of Klein's, Winnicott' s, and also Bion's formulations (1959) of the object relation can guide, albeit in a limited sense, the design of a capable and adept machine. We operationalize these concepts as follows:

With respect to *Sensitivity*, we suggest that the machine should be attuned to the user's needs, able to cater to the user's needs when he or she is under duress and fatigue, as well as detect the user's needs, even when, like an infant, he or she is unable to articulate and communicate them fully. This sensitivity is achieved by understanding the user's precursors; that is, "knowing" how to read between the lines. Here, of course, we are focusing on what philosophers define as "consciousness as sensitivity" which we share with most other animals and arguably even some plants (Baldwin and Schultz 1983; Solomon 2007, p. 18) and not the more complex aspect of consciousness.

Caring is operationalized here as the use of every possible resource to "suspend time" and change priorities (e.g. sacrificing system efficiency) to care for a user's innate and immediate needs despite physical limitations and external constraints. This, in engineering terms, may mean sacrificing efficiency and perhaps even the utility of other systems to address the user's pressing needs (just as in times of duress, acting with normal constraints is not what we expect from a fellow human being).

4.3 Computational architecture

We began by asking ourselves whether it would be possible to build a machine that could be programmed to have the quality of *sensitivity* to the user's innate needs and the capacity for *caring*. With respect to *caring*, here the objective was to develop mechanisms that could find the optimal solution, and, when necessary, bias the solution by overriding system efficiency, economy, and preservation constraints. We felt satisfied with our ability to formulate and build this facet of *caring* even though we were well aware that the psychological meaning encompasses much more. The more challenging problem was to come up with a formulation for sensitivity that enabled the understanding of precursors; i.e. the person's inner state and underlying needs. To this end, we developed an artificial intelligence-based computational architecture that tries the emulate the human capacity to understand precursors by processing the communications and precursors within the system's functionality, compute a solution, and then deliver it using the system's internal states and possible configurations.

The computational architecture for the system has five modules: *interpretation, analysis, goal recognizer, translation,* and *execution* (see Fig. 5). The *interpretation* module takes the user's inputs (speech, button presses, etc.) and filters out the inputs related to *sensitivity* and the need

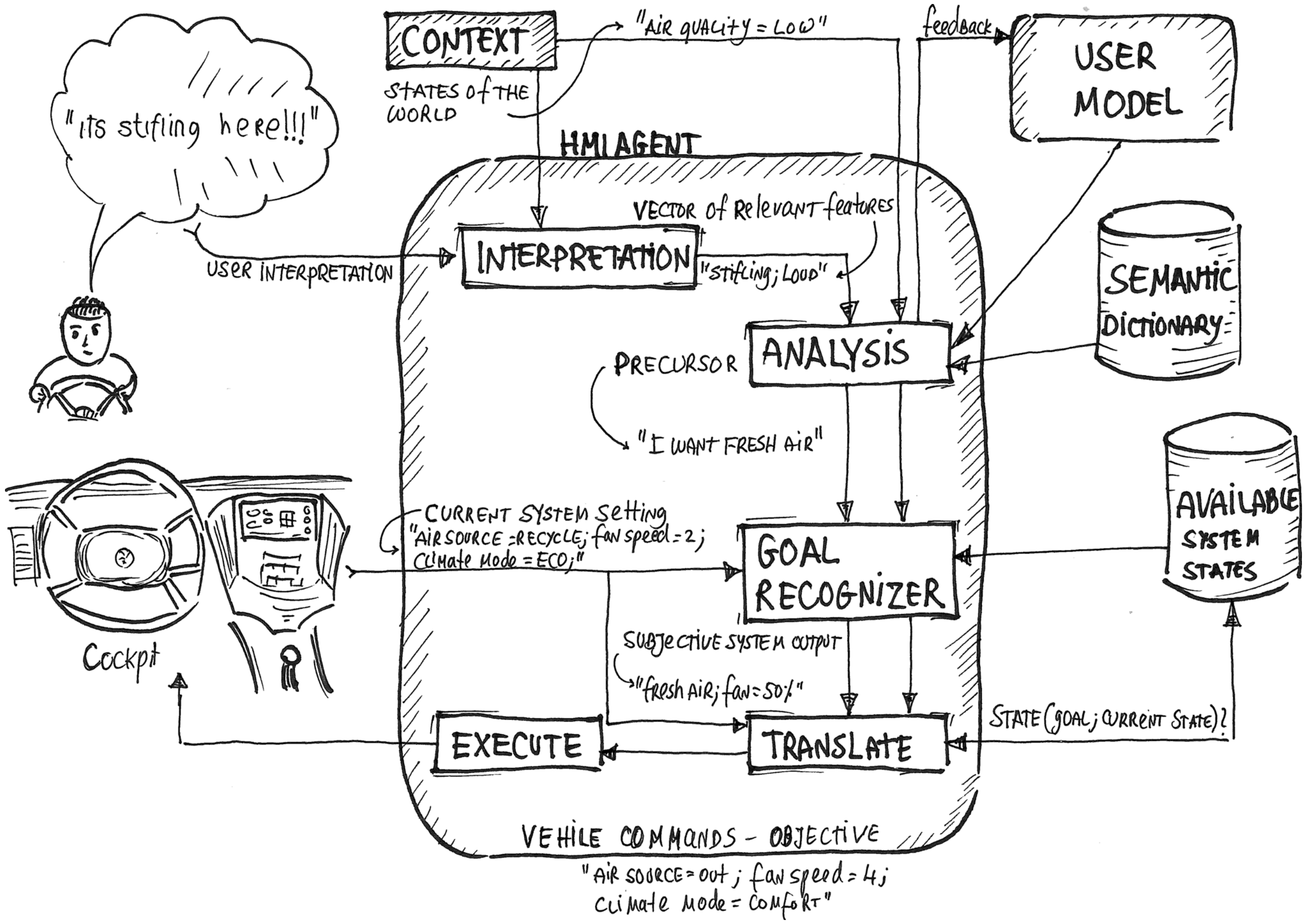

Fig. 5 Computational architecture for sensitivity and holding in a climate control system

to be *held*. The *analysis* component processes the data filtered from the user's input to identify the user's precursors (the "why" and not just the "what"; Bion 1959). The capacity to look for not only what the user wants, but what he or she needs is at the heart of our approach to operationalizing sensitivity. For this process, two databases are used: a *semantic dictionary* that contains the required data for interpreting and understanding user input, and a *user model* that contains a history-based profile of the specific user. The *goal recognizer* is where the solution is computed to accommodate the user's needs and then scheduled over time to accommodate both short-term and long-term outputs; this scheduling can be done by implementing a meta-planner that decomposes inputs into goals with appropriate timings and computes the solution to each one as a function of its dependencies (Koller and Friedman 2009; Russell and Norvig 2009). This is one technical approach for implementing the *planning and coordination* attribute discussed earlier.

To find the most appropriate system configuration to maximize the user's perceived satisfaction rather than system efficiency (*caring*), the *goal recognizer* module uses the *available system outputs* database, which is a dictionary translating all possible system configurations into selected outputs. The *translate* module takes the computed system output (from the goal recognizer) and matches it to a specific heating and cooling system mode and setting (e.g. COMFORT mode, fan level = 4). The final setting is then passed to *execution*. Figure 5 shows how the utterance "it's stifling here!!" is processed until a solution is achieved.

5 Implementation and evaluation

Our first prototype exemplified a simplified version of an intelligent agent. In this version, the precursors interpreted by a rule-based system were mapped to a single goal for each interaction through a search process that took into account the series of last interactions in order to understand the user's intention, and translated it into a score chosen from a preset scale. In the current implementation, the state of the user was simplified to an artificially defined function of subjective feeling and was assumed to be a singleton.

The action made by the system is chosen as to bring the physical system from the current subjective state to the user's desired state of the physical system.

The translation of the current system state onto a numeric scale was done by a human expert, who mapped the system settings to a score scale representing what the user would feel in such state (for example settings that resulted in the same feeling of blower power could be mapped to the same score on the scale even though the fan level was not set differently for these settings). To compute the score of the desired state, we look at the last two states where the user interacted with the system. We evaluate the scores of these states and compute the score of the desired state. This new score depends on the direction of the user's expected change in settings. That is, if the direction was "stronger" and the user keeps asking for "stronger," then the score of the desired state will be the score of the last state increased by some step value (defined in the code). This increase might be larger than the decrease step since the user is further away of his desired state. For example, if the score of an early interaction is 4 and the most recent got a score of 8, and the user keeps asking for "Stronger"—the new state gets a score of 16. Alternatively, if the user asks for a change in the direction then the score of the desired state will decrease by some step value. In the same example, where the score of the less recent interaction is 4 and the most recent got a score of 8, when the user asks for a change in the desired direction ("softer" instead of "stronger")—the desired state gets a score of 6. Once we computed the score of the desired state, we can compute the actual state that will have this score associated with it. This search for the most accurate desired state for the user is performed in order to reach the most accurately comfortable state for the user while requiring least number of interactions as possible. Also, if there is more than one state with the same score, different choice functions can be implemented. Our first prototype chose the more economic option out of all the available states. Another option could be to choose the state that leads to the least number of changes relatively to the current state.

As mentioned earlier, we found that actions based solely on the user's explicit verbal input were not necessarily always the best course of action. In many situations, the user's actual needs were somewhat different from their verbally expressed wants; if the expert followed the verbal inputs verbatim the results could easily get into a spiralling loop of "command and response" that never achieved satisfaction. One successful approach constantly used by the expert to exit the spiral was to first recognize the point of "hopelessness" and then start anew with what he believed was the best configuration to accommodate the user's wants. From our perspective, this can be seen as *sensitivity* to the user with respect to knowing when to stop, and *caring* in the sense of making an extra effort, perhaps also sacrificing efficiency, to find a desirable solution. To mimic a human who would try a new configuration trajectory in reaction to a user being constantly unsatisfied with the settings, the computational search process was adjusted to perform "leaps" in the state space to prevent the user from getting stuck in local minima. Notice that this technique, as well as our entire approach, was aimed at implementing sensitivity and caring and thus differed from traditional approaches that focus improvement efforts on either training the user or the interface.

5.1 Interface design

A team-oriented climate control system demands different types of interfaces than used today. In a traditional human–machine system, users have to go through three steps to fulfil their needs and subsequent wants. They first need to feel the need overtly, then become conscious and cognizant of the want, then somehow translate the want to some new machine configuration in the "language of the machine" (e.g. modes, setting). During this three-step process, there are ample opportunities for divergences from the actual need, with its associated frustration and dissatisfaction. However, a system that is sensitive (understands precursors) and caring (able to remove constraints) provides an opportunity for translating the user's innate needs, expressed in "the language of the user," to the "language of the machine", directly. Such design approach opens the door to a much simpler interface, where many of the modes and settings described in the behavioural model in Fig. 2 can be abstracted out.

In order to identify what should be presented on the display and what can be removed, we analysed the climate control system in terms of possible levels of interface abstraction and refinement (Fig. 6).

The lowest level (TECHNICAL) contains all the system states described earlier in the behavioural model in Fig. 2 and assumes the current interface of Fig. 3. As we go up to the hierarchy, we avoid actual temperature settings by replacing them with relative terms such as "colder" and "hotter" for target temperatures, and "soft" or "strong" for fan setting. At the top of the hierarchy is a level that we call FULL AUTONOMY where the user is not involved in setting climate control at all. The level just below is called FEEDBACK-ONLY, where the user only indicates when he or she is uncomfortable with the existing setting and the machine corrects automatically. The third level from the top, called DIRECTION, is an interface where the user asks the system to cool/heat the vehicle or generate soft/strong air flow, whereas the rest is done automatically (including air-conditioning and air distribution settings).

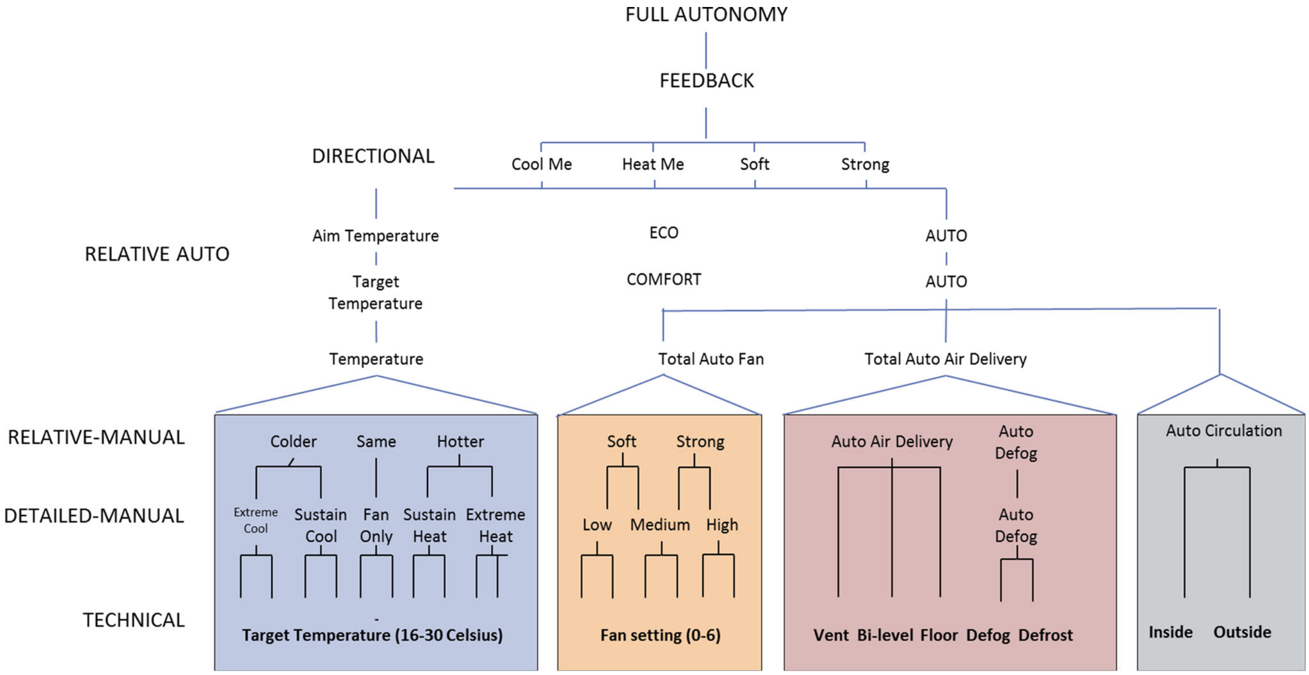

Fig. 6 State-space abstraction of climate control states

Since the FULL AUTONOMY and FEEDBACK-ONLY were currently infeasible due to the lack of an elaborate sensor suite, we settled on implementing the DIRECTION interface. The new interface is shown in Fig. 7. It has six buttons: hotter/colder on temperature, softer/stronger on airflow, and two feedback buttons to the system: comfortable or uncomfortable. Pressing the "comfortable" button provides feedback to the system when the user is content with the achieved climate, and pressing the "uncomfortable" indicates that the user is unhappy with the setting selected by the system. Once the uncomfortable button is pressed after repeated attempts to configure the system to the user's wants, a "leap" is initiated to stop the spiralling loop and begin anew with an altogether new solution.

Finally, note the trade-off between the levels of the abstraction tree: as we go higher in the levels of abstraction the interface is more human natural, but on the other hand gives the system much less information that can be used determine the wanted action. This was compensated in our implementation by an algorithm for translating the user's wants and needs from "the language of the user" to the "language of the machine."

5.2 Experimental driving study

To evaluate the design approach, we conducted yet another driving study. The purpose of the experiment was not to prove or refute the approach, but rather to get an empirical evaluation in terms of its practicality. We assessed whether the system that followed our approach would fare better than a traditional non-team-oriented interface, and, how well it would fare compared with a sensitive and caring human assistant. In this study, eighteen participants were recruited to drive the electric car along the same route as in the pilot study and experience three different experimental conditions: (1) a *human-agent* in which a human expert controlled the climate control system at the request of the driver using any form of language or gesture, just as was done in the pilot study, (2) a *software-agent* in which the driver interacted with the interface in Fig. 7, and (3) a *manual* operation in which the driver interacted with the original interface in Fig. 3. In all conditions, the requests and commands were conducted through voice and the experimenter manipulated the interface based on the participants' commands. We recorded each participant's verbal utterances during all the segments of the drive and summarized them by condition. In addition, we asked the participants to rate their frustration level at the end of each driving segment (1–10 scale). An overall experience rating was conducted at the end of the entire experimental run as well as the administration of the NASA-TLX questionnaire. The ratings were collected as was done on the original NASA-TLX scales (with 21 ticks on a linear line) but later analysed on a 10-point scale.

The experiment was designed as a mixed model with two within-participant factors: interface (three levels) and initial cabin temperature (two levels) and two between-participant factors: order (two levels) and participant gender (two levels). Each participant drove six segments and experienced the three interfaces types (*human-agent,*

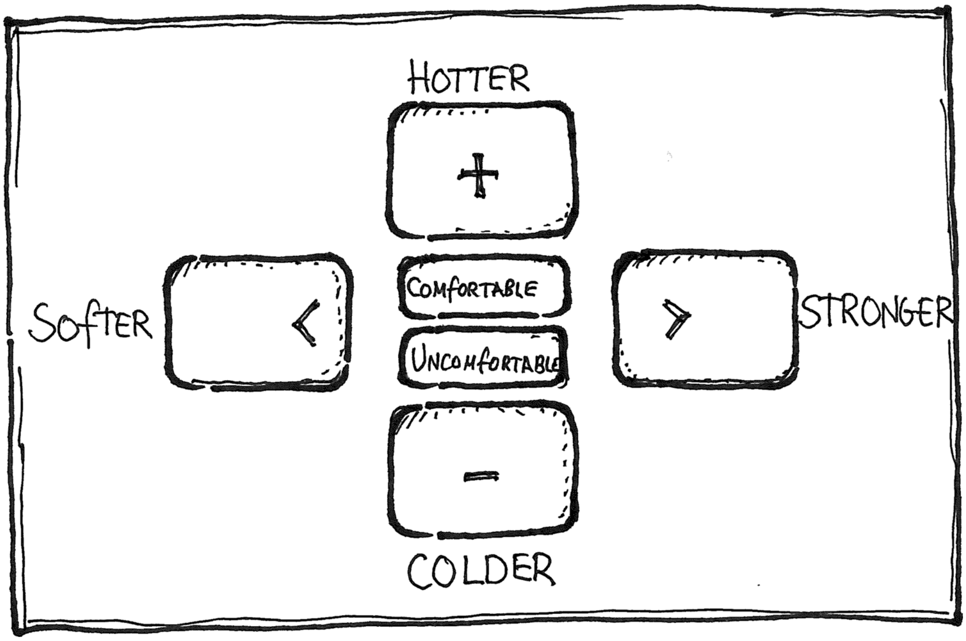

Fig. 7 Simplified interface (*Direction*)

software-agent, manual) twice: once with the initial cabin temperature hot and another time with the initial cabin temperature cold. In both the hot and cold conditions, the initial settings were maximum fan (level 6), inside circulation, air delivery to panel, comfort mode, and vents directed at the torso area. The order of conditions was counterbalanced across participants such that each of the three interfaces types in the two initial cabin temperatures (hot and cold) appeared the same number of times in each of the six legs. For each participant, the initial cabin temperature condition was blocked and counterbalanced across participants such that half of the participants started their first three legs in an initially hot environment and half of the participants started their first three legs in an initially cold environment. Our assumption was that if the *software-agent* was indeed effective, we would expect to see better performance than the manual interface and worse performance than the *human-agent*, which is considered optimal. The data were analysed using the MIXED model analysis of variance in (SAS/STAT, 2008).

The first dependent measure was the number of utterances to achieve a comfortable vehicle temperature and humidity level (from the uncomfortable initial condition). Figure 8a shows that there were significantly more utterances to control the system in the *manual* interface than in the *human*- and *software-agent* interfaces ($F(1, 34) = 67.7$ and 44.9, $p < 0.0001$). In the manual interface, participants made on average 15.2 utterances. In contrast, only 7.3 utterances were made in the *software-agent* interface and 5.5 in the *human-agent* ($F(2, 34) = 38.3$, $p < 0.0001$).

The subjective rating of frustration was significantly affected by interface type $F(2, 34) = 6.04$, $p < 0.01$ (see Fig. 8b). The lowest frustration level, associated with the human-agent (2.4), was significantly below that of the manual (4.0, $p < 0.01$) and moderately lower than the software-agent (3.3, $p < 0.1$). Frustration was not significantly different between the *software-agent* and the *manual* operation interface, and the *software-agent* did not fare much worse than the "optimal" human-agent.

We then tested whether there was a difference in the "loading" required to execute the climate control task in terms of the mental, physical, and temporal demands as well as effort and frustration using the NASA-TLX procedure (NASA 1996; Hart and Staveland 1988). TLX scores were also significantly affected by the type of interface, $F(2, 22) = 4.4$, $p < 0.05$ (Fig. 8c). The score for the *human-agent* (4.2) was significantly lower than both the software agent (6.4) and the manual operation (6.6) $F(1, 22) = 6.0$ and 7.1, $p < 0.05$. In terms of task load, the manual and software interfaces were fairly comparable (note, however, that due to an administrative error, the TLX scores were based on data collected from only twelve participants of the 18 in total in the study). The effect of the interface on overall experience was moderately significant ($F(2, 34) = 2.9$, $p < 0.1$). (Note the strong ceiling/floor effects). Figure 8d shows the results for the overall experience on a 0–10 scale. The rating of the *human-agent* (7.9/10) was 1.1 scale points higher than the *manual* interface (6.8/10), $F(1, 34) = 5.1$, $p < 0.05$. The overall experience of the *software-agent* was 1 scale point higher than the *manual* interface, $F(1, 34) = 3.5$, $p < 0.1$. The difference between the average *software-agent* and the average *human-agent* was very small (0.1/10).

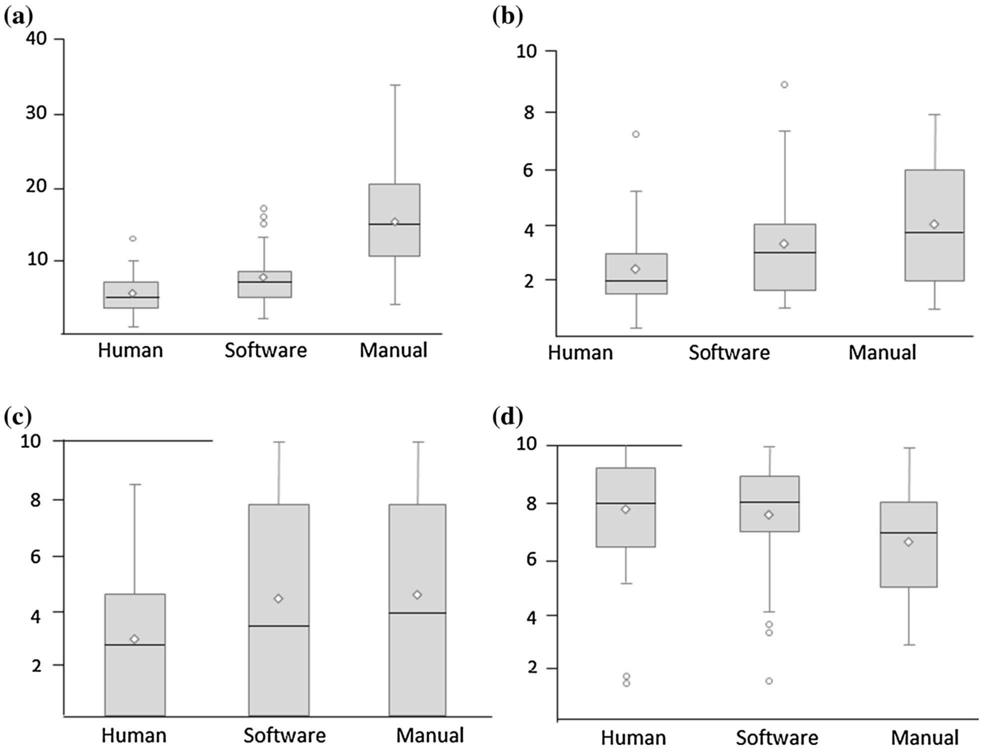

Fig. 8 **a** Number of utterances per interface type, **b** frustration by interface type, **c** TLX (overall) by interface type, **d** experience by interface type

In summary, the software-agent did not fare much worse than the human expert's ability to be sensitive and caring to the user. In some cases, such as overall frustration and overall subjective ratings, there was not much of a difference between the *software*- and the *human-agent*. All in all, the *software-agent's* performance appears to have been only slightly less satisfactory than the *human-agent*. This result is encouraging because the human expert who played the role of the *human-agent* in this study was highly proficient in being sensitive and caring to the users' innate needs (it was the same person who took part in the pilot study). Once he knew what they needed, he used his ability to be sensitive and caring to fulfil and satisfy their all their needs. The fact that the inanimate *software-agent* achieved results that were not that far behind this human expert and almost in all cases well ahead of the voice activated *manual* interface gives us hope that the computation architecture described here and its implementation constitute a step towards more natural human–machine interaction.

6 Conclusion

The starting point of this article was to better understand what is "fundamentally *social* and *natural*" in user interaction. We quickly narrowed the scope to teamwork as a reflection of the social aspects of interaction and searched for the attributes of successful teamwork. Some of these attributes involved management and organizational processes (commitment and contract, goal definition and policies, transparency of actions, reflection), whereas others involved more technical work practices (common ground, beliefs, planning and coordination), and more than a few centred on the humanistic issues that take place in a team. We showed that once a given design is viewed from a team-oriented perspective it calls for a variety of design options that are hardly considered in current design approaches.

Of the eleven team attributes, we focused first on *sensitivity* and then *caring* as the basic glue that binds humans into cohesiveness with another. With an eye towards the future formation of joint human-machine teams, we attempted to operationalize these two attributes using the psychoanalytical theory of object relations. We also found that *natural* interaction, in the sense of interactions where the user feels no barriers to fulfilling his needs and at the same time is cared for, contained, and well understood—occurs in the "theatrical" space between the user (subject) and the other (or parts thereof) which we consider as an object. Klein's object relation theory, and the various extension of her initial work by Winnicott, Bion, and others, provides us with a sound theoretical foundation for understanding this space. Extensive practical work by psychologists offers strategies for dealing with the needs, wants, frustrations, and dissatisfaction manifested in the relations.

The fact that both the *social* and *natural* aspects of interaction draw heavily on the notion of relations between subject and object is surprising and at the same time quite illuminating. Thus, when it comes to designing user interaction that is social and natural as well as building joint teams of humans and machines, considerable emphasis should be placed on the kind of relation formed between the human and the machine to include expectations, demands, unmet requests, frustrations, and perhaps even explanations and attempts at reconciliation.

The psychoanalytical theory of *object relations* was originally developed to understand and heal humans by characterizing the early mother–infant relationship and applying it to the patient–therapist dyad. We suggest that the machine (e.g. autonomic system, robot) can be considered the *object* to which the human (subject) relates, and raise some serious questions about what the relations between subject and object should be. We need to consider and identify the *fundamental* and *dynamic* matrices involved in such human–robot relations, and the social role of the machine in this case. We also need to better understand the Kleinian "space" in which humans and machines interact as well as the specific interface "objects" that the subject relates to (e.g. a voice, visuals, haptics as well as the robotic "being" behind them). The object relational view of human and machines also allowed us to address the problem of interaction language by making a distinction between the (sometimes emotional) "language of the user," on the one hand, and the (logical and rigid) "language of the machine" on the other. We also feel that the psychoanalytical object relation viewpoint opens the door to a better understanding of how humans connect with their nurturing objects and how this relation can be enhanced to develop confidence (the mechanistic equivalent of what we take as *trust*).

The implementation of a climate control interface that had some degree of sensitivity and caring simplified the interface and made user interaction more natural, but we feel that the implications of this study go beyond just improvement in interface and interaction design. They call for a different way of thinking about how we interact with sophisticated machine agents that will be part of our daily lives in the not-so-distant future. Practical applications of this team-oriented and relational design approach are now being considered for autopilots in semi-automatic vehicles (Goldman and Degani 2012) as well as in "fully autonomous" vehicles.

Acknowledgements The authors wish to thank Dr. Hila Degani for the idea to consider the psychoanalytical theory of object relations as a way to view human–autonomy relations; Inbar Sela for transcribing the data from the logbooks and videos and maintaining the dataset and all related experimental forms; Amit Edelstein for helping us to apply, in the technological sense, some of the psychoanalytical concepts; and Esther Singer for her review of the paper and links to related work.

References

Aaron J, Simmons JA, Wood D (2008) Kierkegaard and levinas: ethics, politics, and religion. Indiana University Press, Bloomington

Amon N (1998) The book of Tao. Even Hoshen, Raanana. (Translation from Hebrew is available online in PDF format)

Asimov I (1950) I Robot. Gnome Press, New York

Azaria A, Rabinovich Z, Kraus S, Goldman CV, Gal Y (2012) Strategic advice provision in repeated human-agent interactions. In: Proceedings of the 26th national conference on AI (AAAI 2012), Toronto, Canada

Azaria A, Gal Y, Kraus S, Goldman CV (2015) Strategic advice provision in repeated human-agent interactions. J Auton Agents Multi Agent Syst 30:4–29

Baldwin I, Schultz J (1983) Rapid changes in tree leaf chemistry induced by damage: evidence for communication between plants. Science 15(221):277–279

Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I (2001) the "reading the mind in the eyes" test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. J Child Psychol Psychiatry 42(2):241–251

Barshi I, Mauro R, Degani A, Loukopoulou L (2016) Designing flightdeck procedures. NASA Technical Memorandum #219421. Ames Research Center, Mountain View

Belbin RM (1981) Management teams: why they succeed or fail. Butterworth-Heinemann, Oxford

Belbin RM (1993) Team roles at work. Butterworth-Heinemann, Oxford

Bion WR (1959) Attacks on linking. Int J Psychoanal 40:308–315

Bion WR (1961) Experiences in groups and other papers. Tavistock, London, pp 139–189

Bowers M (1966) The will to manage. Corporate success through programmed management. Mcgraw-Hill, New York

Bratman ME (1992) Shared cooperative activity. Philos Rev 101(2):327–341

Bradshaw JM, Hoffman RR, Johnson M, Woods DD (2013) The seven deadly myths of 'autonomous systems'. IEEE Intell Syst 28(3):54–61

Brown D, Zinkin L (1994) The psyche and the social world. Routledge, London

Cannon-Bowers JA, Salas E, Converse SA (1990) Cognitive psychology and team training: shared mental models in complex systems. Hum Fact Bull 33:1–4

Carmeli A, Jones CD, Binyamin G (2016) The power of caring and generativity in building strategic adaptability. J Occup Organ Psychol 89(1):46–72

Casner SM, Geven RW, Williams KT (2013) The effectiveness of airline pilot training for abnormal events. Hum Fact 55(3):477–485

Casner SM, Hutchins EL, Norman D (2016) The challenges of partially automated driving. Commun ACM 59(5):70–77

Chen Y, Chen X, Portnoy R (2009) Reciprocating positive and negative inequitable offers: culture, emotion, and reciprocity. J Exp Soc Psychol 45:24–34

Christoffersen K, Woods DD (2002) How to make automated systems team players. Adv Hum Perform Cognit Res 2:1–12

Davis J, Millburn P, Murphy T, Woodhouse M (1992) Successful team building: how to create teams that really work. Kogan Page, London

Deer L, Erodes R (1994) Lame deer, seeker of visions. Simon & Schuster, New York

Degani A (2004) Taming HAL: designing interfaces beyond 2001. Palgrave-MacMillan, New York

Degani A, Heymann M, Shafto M (2013a) Modeling and formal analysis of human–machine interaction. In: Kirlik A, Lee J (eds) The Oxford handbook of cognitive engineering. Oxford University Press, New York

Degani A, Heymann M, Shafto M (2013b) Modeling and formal analysis of human-machine interaction. In: Kirlik A, Lee J (eds) The Oxford handbook of cognitive engineering. Oxford University Press, New York

de Saint-Exupéry A (1943/2000). The little prince (Richard Howard, translator). Mariner Books, New York

Farid S (2005) Respect. http://www.beyondintractability.org/essay/respect/?nid=6573. May 3rd 2017

Federal Aviation Administration (2004) Advisory Circular 120-51E: Crew resource management training. Initiated by FAA office AFS-210 on 1/22/04

Foulkes SH (1964) Therapeutic group analysis. George Allen & Unwin, London

Foulkes SH (1990) Selected Papers of S.H. Foulkes: psychoanalysis and group analysis. In: Foulkes E (ed). Karnac Books

Foushee HC (1984) Dyads and triads at 35,000 feet: factors affecting group process and aircrew performance. Am Psychol 39:885–893

Fromm E (1956) The art of loving. Harper & Row Publishers, New York

Gal Y, Kraus S, Gelfand M, Khashan H, Salmon E (2011) Negotiating with people across cultures using an adaptive agent. ACM Trans Intell Syst Technol 3(1):8

Goldman C, Degani A (2012) A team-oriented approach for automation aids in cars. In: Proceedings of the 56th annual meeting of the human factors and ergonomics society, Boston

Grice HP (1975) Logic and conversation. In: Cole P, Morgan JL (eds) Syntax and semantics volume 3: speech acts. Academic Press, New York, pp 41–58

Grosz B, Kraus S (1996) Collaborative plans for complex group action. Artif Intell J 86(2):269–357

Guntrip H (1969) Schizoid phenomena, object relations and the self. International Universities Press, New York

Hansen MH (1992) The ancient Greek city-state. In: Symposium on the occasion of the 250th anniversary of the Royal Danish Academy of Sciences and Letters (July, 1–4)

Harari YN (2015) Sapiens: a brief history of humankind. Vintage, London

Harel D (1987) Statecharts: a visual formalism for complex systems. Sci Comput Progr 8(1):231–274

Harel D (2009) Statecharts in the making: a personal account. Commun ACM 52(3):67–75

Harel D, Pnueli A (1985) On the development of reactive systems. In: Apt KR (ed) Logics and models of concurrent systems, NATO ASI Series, vol F-13. Springer, New York, pp 477–498

Harel D, Politi M (1998) Modeling reactive systems with statecharts: The STATEMATE approach. McGraw-Hill, New York

Hart SG, Staveland LE (1988) Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock PA, Meshkati N (eds) Human mental workload: advances in psychology. North Holland, Amsterdam, pp 139–183

Heymann M, Degani A (2013) Automated driving aids: modeling, analysis, and interface design considerations. In: Proceedings of the 5th international conference on automotive user interfaces and interactive vehicular applications (AU2013), Eindhoven, The Netherlands

Horvitz E (1999) Principles of mixed-initiative user interfaces. Proc ACM CHI 1999:159–166

Klein M (1932) The psycho-analysis of children. Hogarth, UK

Klein G, Woods DD, Bradshaw JM, Hoffman RR, Feltovich PJ (2004) Ten challenges for making automation a "team player" in joint human-agent activity. IEEE Intell Syst 19(6):91–95

Koller D, Friedman N (2009) Probabilistic graphical models. The MIT Press, Cambridge

Lewin K (1939) Field theory and experiment in social psychology: concepts and methods. Am J Sociol 44(6):868–896

López-Corvo R (2002) The dictionary of the work of W.R. Bion. Karnac Press, London

McFadzean E (2002) Developing and supporting creative problem-solving teams: part 1—a conceptual model. Manag Decis 40(5/6):463–476

Merriam-Webster Dictionary (2011)

National Transportation Safety Board (1996) American eagle flight 4184, Roselawn Indiana, October 31, 1994. Report number NTSB/AAR-96/01. Washington, DC

National Transportation Safety Board (2014) Asiana flight 214, San Francisco, California July 6, 2013. NTSB/AAR-14/01. Washington, DC

NASA (1996) NASA task load index (TLX) v.1.0 manual. Ames Research Center, Moffett Field

Ogden TH (2004) On holding and containing, being and dreaming. Int J Psychoanal 85(6):1349–1364

Paris CR, Salas E, Cannon-Bowers JA (2000) Teamwork in multi-person systems: a review and analysis. Ergonomics 43(8):1052–1075

Parnas D (1969) On the use of transition diagrams in the design of a user interface for an interactive computer system. In: Proceeding of the 24th annual ACM conference, pp 379–385

Reeves B, Nass C (1996) The media equation. Cambridge Press, New York

Rich C, Sidner CL, Lesh N (2001) Collagen: applying collaborative discourse theory to human-computer interaction. AI Mag 22(4):15–25

Rosenfeld A, Zuckerman I, Azaria A, Krasu S (2012) Combining psychological models with machine learning to better predict people's decisions. Syntheses 189(Supplement 1):81–93

Rushby J (2001) Analyzing cockpit interfaces using formal methods. Electron Notes Theor Comput Sci 43:1–14

Russell S, Norvig P (2009) Artificial intelligence: a modern approach, 3rd edn. Prentice Hall, Upper Saddle River

Ruszczynski S (1993) Thinking about and working with couples. In: Ruszczynski S (ed) Psychotherapy with couples: theory and practice at the Tavistock institute of marital studies. Karnac Books, London

Salas E, Dickenson TL, Converse SA, Tannenbaum SI (1992) Toward an understanding of team performance and training. In: Swezey RJ, Salas E (eds) Teams: their training and performance. Ablex, Norwood, pp 3–29

Salas E, Sims DE, Burke CS (2005) Is there a "big five" in teamwork? Small Group Res 36:555–599

Sasou K, Reason J (1999) Team errors: definition and taxonomy. Reliab Eng Syst Saf 65(1):1–9

SAS/STAT (2008). User's guide introduction to Statistical Modeling with SAS/STAT Software® 9.2., SAS institute incorporated. NC, Cary

Schelling TC (1984) Choice and consequences. Harvard University Press, Cambridge

Schelling TC (1985) Enforcing rules on oneself. J Law Econ Organ 1(2):357–374

Searle JR (1980) Minds, brains, and programs. Behav Brain Sci 3(3):417–457

Solomon RC (1993) The passions emotions and the meaning of life. Hackett Publishing Company, Cambridge

Solomon RC (2007) The little philosophy book. Oxford University Press, UK

Stout RJ, Cannon-Bowers JA, Salas E, Milanovich DM (1999) Planning, shared mental models, and coordinated performance: an empirical link is established. Hum Fact 41(1):61–71

Welch J, Welch S (2011) How to build a winning team. Newsweek Magazine

Wiener EL, Kanki BG, Helmreich RL (1993) Cockpit resource management. Academic Press, San Diego

Winnicott DW (1955) The depressive position in normal emotional development. Br J Med Psychol 9(28):89–100

Winnicott DW (1958/1945) Primitive emotional development. In: Winnicott DW (ed) Through pediatrics to psychoanalysis. Basic Books, New York, pp 145–156

Woolley AW, Chabris FC, Pentland A, Hashmi N, Malone TW (2010) Evidence for a collective intelligence factor in the performance of human groups. Science 330(6004):686–688