

# Chapter 3

## The Participatory Turn: A Multidimensional Gradual Agency Concept for Human and Non-human Actors

Sabine Thürmel

### 3.1 Introduction to the Participatory Turn in Socio-technical Systems

New varieties of interplay between humans, robots and software agents are on the rise: virtual companions, self-driving cars and the collaboration between humans and virtual agents in emergency response systems exemplify this development. Humans have evolved from “naturally born cyborgs” (Clark 2003) to adaptive, co-dependent, socio-technical agents. Computer-based artefacts are no longer mere tools but may be capable of individual and joint action, too. Turkle characterises this development as follows: “Computational objects do not simply do things *for* us, but they do things *to* us as people, to our ways of seeing ourselves and others. Increasingly, technology puts itself into a position to do things *with* us” (Turkle 2006, 1). This insight was gained when Turkle studied the nascent robotics culture. It is equally valid for software agents. The starting point of the evolution of software agents is constituted by interface agents providing assistance for the user or acting on his or her behalf. As envisioned by (Laurel 1991) and (Maes 1994), they have evolved into increasingly autonomous agents in virtual environments. Moreover software agents may be found in cyber-physical systems (Mainzer 2010, 181). While classical computer systems separate physical and virtual worlds, cyber-physical systems (CPS) observe their physical environment by sensors, process the information, and influence their environment with so-called actuators while being connected by a communication layer. Collaborative software agents can be embedded in cyber-physical systems if the different nodes in the cyber-physical system need to coordinate. Examples include distributed rescue systems (Jennings 2010), smart

---

S. Thürmel (✉)  
Munich Center for Technology in Society, Technische Universität München,  
Munich, Germany  
e-mail: [sabine@thuermel.de](mailto:sabine@thuermel.de)

energy grids (Wedde et al. 2008), and distributed health monitoring systems (Nealon and Moreno 2003). These systems are first simulated and then deployed to control processes in the material world. Humans may be integrated for clarifying and/or deciding non-formalized conflicts in an ad hoc manner. Most agent-based cyber-physical systems aim at enhancing process automation. However, there exist also systems focusing on optimizing the collaboration between humans, robots and software agents. In such environments each participant plays a specific role and contributes in a specific way to the overall problem solution. Examples include diverse areas as self-organizing production systems offering customized products in highly flexible manufacturing environments or managed health care systems, where humans and virtual carers collaborate (Hossain and Ahmed 2012). Thus software agents have been promoted from assistants to interaction partners. The socio-technical fabric of our world has been augmented by these collaborative systems.

Current collaborative constellations between humans and technical agents are asymmetric: their acts are based on different cognitive systems, different degrees of freedom and only partially overlapping spheres of experience. However, new capabilities may emerge over time on the techno level. Self-organization and coalition forming on the group level can occur. New cultural practices come into being. The enactment of joint agency in these heterogeneous constellations is well past its embryonic stage. Therefore it becomes vital to understand agency and inter-agent coordination in purely virtual and cyber-physical systems.

The potential of agent-based virtual or cyber-physical systems becomes actual in testbed environments and real-time deployments. This perspective is elaborated in Sect. 3.2 which is dedicated to the potentiality and actuality of social computing systems. This section is included because many sociologists focus solely on the actuality of socio-technical systems that is on “agency *in medias res*” neglecting the fact that the potential of technical systems is determined by their design. Technical agents are not black boxes just to be observed but may be analysed in detail by computer scientists and engineers.

Agency in socio-technical systems may be attributed in different ways: two of the most relevant ones for attributing agency to both humans and non-humans are presented in Sect. 3.3. These approaches take a technograph’s approach aiming at describing agency as it unfolds. This paper does not intend to evaluate agency in socio-technical systems from an observer’s standpoint. In Sect. 3.4 the agential perspective is characterized as a certain level of abstraction when analysing a system.

In Sect. 3.5 a novel approach to attribute agency in socio-technical systems is presented. A multidimensional gradual agency concept for human and non-human actors is introduced. In this framework individual and joint agency may be taken into view. The framework is applicable to constellations of distributed and collective agency. Scenarios where solely humans act can be compared to testbed simulations.

Finally, an outlook is given on how the framework presented here may support further both the software engineer and the philosopher when modelling and analysing role-based interaction in socio-technical systems.

## 3.2 Potentiality and Actuality of Computing Systems

Computer simulations let us explore the dynamic behaviour of complex systems. Today they are not only used in natural sciences and computational engineering but also in computational sociology. Social computing systems focus on the simulation of complex interactions and relationships between individual human and/or non-human agents. If the simulations are based on scientific abstractions of real-world problem spaces they enable us to gain new insights. For example “crowd simulation” systems are useful if evacuation plans have to be developed. Collaborative efforts may be simulated, too. A case in point is the coordination of emergency response services in a disaster management system based on so-called electronic market mechanisms (Jennings 2010). The humanities, social and political science, behavioral economics, and studies in law have discovered agent-based modeling (ABM) too. Academics have applied ABM to study the evolution of norms (Muldoon et al. 2014), and to explore the impact of different social organization models on settlement locations in ancient civilizations (Chliaoutakis and Chalkiadakis 2014). Since agent-based models may provide a better fit than conventional economic models to model the “herding” among investors, early-warning systems for the next financial crisis could be built based on ABM (Economist 2010). Even criminal behavior, deliberate misinterpretations of norms or negligence can be studied. Therefore it is hardly surprising that the Leibniz Center for Law at the University of Amsterdam had been looking – although in vain – for a specific Ph.D. candidate in legal engineering: He or she should be capable of developing new policies in tax evasion scenarios. These scenarios were planned to be based on ABM (Leibnizcenter for Law 2011). The novel technical options of “social computing” not only offer to explain social behaviour but they may also suggest ways of changing it.

Social simulation systems are similar to numerical simulations but use different conceptual and software models. Both approaches may complement each other. Numerical methods based on non-linear equation systems support the simulation of quantitative aspects of complex discrete systems (Mainzer 2007). In contrast, multi-agent systems (MAS) enable collective behaviour to be modelled based on the local perspectives of individuals, their high-level cognitive processes and their interaction with the environment (Woolridge 2009). Current agent-based software systems range from swarm intelligence systems, based on a bionic metaphor for distributed problem solving, to sophisticated e-negotiation systems (Woolridge 2009).

Simulations owe their attractiveness to the elaborate rhetoric of the virtual (Berthier 2004): “It is a question of representing a future and hypothetical situation as if it were given, neglecting the temporal and factual dimensions separating us from it – i.e. to represent it as actual” (Berthier 2007, 4). Social computing systems are virtual systems modelled, e.g. by MAS, and realized by the corresponding dynamic computer-mediated environments. Computational science and engineering as well as computational sociology systems benefit from these computer-based interaction spaces. The computer is used both as a multipurpose machine and a unique tool used to store and deliver information.

Virtuality in technologically induced contexts is even better explained if Hubig's two-tiered presentation of technology in general as a medium (Hubig 2006) is adopted. He distinguishes between the "potential sphere of the realization of potential ends" and the "actual sphere of realizing possible ends" (Hubig 2010, 4). Applied to social computing systems – or IT systems in general – it can be stated that their specification corresponds to the "potential sphere of the realization of potential ends" and any run-time instantiation to a corresponding actual sphere. In other words: due to their nature as computational artefacts, the potential of social computing systems becomes actual in a concrete instantiation. Their inherent potentiality is actualized during run-time. "A technical system constitutes a potentiality that only becomes a reality if and when the system is identified as relevant for agency and is embedded into concrete contexts of action" (Hubig 2010, 3).

Since purely computational artefacts are intangible, i.e. existing in time but not in space, the situation becomes even more challenging: one and the same social computing program can be executed in experimental environments and in real-world interaction spaces. The demonstrator for the coordination of emergency response services may go live and coordinate human and non-human actors in genuine disaster recovery scenarios. With regard to its impact on the physical environment, it possesses a virtual actuality in the testbed environment and a real actuality when it is employed in real time in order to control processes in the natural world.

In the case of social computing systems, the "actual sphere of realizing possible ends" can either be an experimental environment composed exclusively of software agents or a system deployed to control processes in the material world. Humans may be integrated for clarifying and/or deciding non-formalized conflicts in an ad hoc manner. Automatic collaborative routines or new practices for ad hoc collaboration are established. Novel, purely virtual or hybrid contexts realizing collective and distributed agency materialize.

### 3.3 Attributing Agency in Socio-technical Systems

In order to exemplify the state of the art in attributing collective and distributed agency in socio-technical systems, two schools are briefly summarized: the Actor Network Theory (ANT) and the socio-technical approach of attributing distributed agency of Rammert and colleagues. Both are aimed at analysing constellations of collective inter-agency by attributing agency both to human and non-human actors but they differ in essential aspects.

The ANT approach introduces a flat concept of agency and a symmetrical ontology applicable both to human and non-human actors (e.g. (Latour 2005)) whereas the distributed agency approach of Rammert et al. promotes a levelled and gradual concept of agency based on the "practical fiction of technologies in action" (Rammert and Schulz-Schaeffer 2002; Rammert 2011). Latour focuses on "interobjectivity" (Latour 1996) that is links, alliances, and annexes between all kinds of

objects whereas Rammert takes a more nuanced view on inter-agency based on Anthony Giddens' stratification model of action (1984).

### 3.3.1 *The Actor Network Theory (ANT)*

As a practitioner of science and technology studies and a true technograph, Bruno Latour was the first to attribute agency and action both to humans and non-humans (Latour 1988). Together with colleagues such as Michel Callon, a symmetric vocabulary was developed that they deemed applicable both to humans and non-humans (Callon and Latour 1992, 353). This ontological symmetry led to a flat concept of agency where humans and non-human entities were declared equal. Observations gained in laboratories and field tests were described as so-called actor networks, heterogeneous collectives of human and non-human entities, mediators and intermediaries. The Actor Network Theory regards innovation in technology and sciences as largely depending on whether the involved entities – whether they be material or semiotic – succeed in forming (stable) associations. Such stabilizations can be inscribed in certain devices and thus demonstrate their power to influence further scientific evolution (Latour 1990). All activity emanates from so-called actants (Latour 2005, 54). The activity of forming networks is called “translation” (Latour 2005, 108). Statements made about actants as agents of translation are snapshots in the process of realizing networks (Schulz-Schaeffer 2000, 199). The central empirical goal of the actor network theory consists in reconstructively opening up convergent and (temporarily) irreversible networks (Schulz-Schaeffer 2000, 205). Thus the ANT approach could more aptly be called a “sociology of translation”, an “actant-rhizome ontology” or a “sociology of innovation” (Latour 2005, 9).

It should be noted that Latour has quite a conventional, tool-oriented notion of technology. This may be due to the fact that smart technology and agent systems are nowhere to be found in his studies.

Latour only focuses on actual systems and their modes of existence. However, one may (and should) clearly distinguish between agency (potentiality) and action (actuality) – especially if the investigations are led by techno-ethicists. Moreover, virtual actuality does not equate with real actuality in most circumstances. A plane crash in reality is very different from one in a simulator.

### 3.3.2 *Distributed Agency and Technology in Action*

The conditions under which we can attribute agency and inter-agency to material entities and how to identify such entities as potential agents are important to Werner Rammert and Ingo Schulz-Schaeffer (Rammert and Schulz-Schaeffer 2002, 11). Therefore they developed a gradual concept of agency in order to categorize

potential agents regardless of their ontological status as machines, animals or human beings. Rammert is convinced that “it is not sufficient to only open up the black box of technology; it is also necessary and more informative to observe the different dimensions and levels of its performance” (Rammert 2011, 11). The model is inspired by Anthony Giddens’ stratification model of action (1984). The approach distinguishes between three levels of agency:

- causality ranging from short-time irritation to permanent restructuring,
- contingency, i.e. the additional ability “to do otherwise”, ranging from choosing pre-selected options to self-generated actions, and, in addition, on the highest level
- intentionality as a basis for rational and self-reflective behaviour (Rammert and Schulz-Schaeffer 2002, 49; Rammert 2011, 9).

The “reality of distributed and mediated agency” is demonstrated, e.g. based on an intelligent air traffic system (Rammert 2011, 15). Hybrid constellations of interacting humans, machines and programs are identified.

Moreover, a pragmatic classification scheme of technical objects depending on their activity levels is developed. This enables classification of the different levels of “technology in action”. It starts with passive artefacts, and continues with reactive ones, i.e. systems with feedback loops. Next come active ones, then proactive ones, i.e. systems with self-activating programs. It ranges further up to co-operative systems, i.e. distributed and self-coordinating systems (Rammert 2008, 6). The degrees of freedom in modern technologies are constantly increasing. Therefore the relationship between humans and technical artefacts evolves “from a fixed instrumental relation to a flexible partnership” (Rammert 2011, 13). Rammert identifies three types of inter-agency: “interaction between human actors, intra-activity between technical agents and interactivity between people and objects” (Rammert 2008, 7). These capabilities do not unfold “*ex nihilo*” but “*in medias res*”. “According to [this] concept of mediated and situated agency, agency arises in the context of interaction and can only be observed under conditions of interdependency” (Rammert 2011, 5).

These reflections show how “technology in action” may be classified and how constellations of collective inter-agency can be evaluated using a gradual and multilevel approach. Similar to Latour, these authors are convinced that artefacts are not just effective means but must be constantly activated via practice (enactment) (Rammert 2007, 15).

Since this approach focuses exclusively on “agency *in medias res*”, i.e. on snapshots of distributed agency and action, the evolution of any individual capabilities, be they human or non-human, are not accounted for. Even relatively primitive cognitive activities such as learning via trial and error, which many machines, animals and all humans are capable of, are not taken into account by Rammert’s perspective on agency. A clear distinction between human agency, i.e. intentional agents, and technical agency, a mere pragmatic fiction, remains. In Rammert’s view, technical agency “emerges in real situations and not in written sentences. It is a practical fiction that has real consequences, not only theoretical ones” (Rammert 2011, 6). In

his somewhat vague view, the agency of objects built by engineers “is a practical fiction that allows building, describing and understanding them adequately. It is not just an illusion, a metaphorical talk or a semiotic trick” (Rammert 2011, 8).

### 3.4 Levels of Abstraction

This paper does not intend to analyse agency in socio-technical systems from an observer’s standpoint. The agency of technology is not considered a “pragmatic fiction” as Rammert did (2011). In my view the agential perspective on technology should be characterized as a certain level of abstraction when analysing a socio-technical system.

In the following, the agency of technology is perceived as a (functional) abstraction corresponding to a level of abstraction (LoA) as defined by Floridi. An LoA “is a specific set of typed variables, intuitively representable as an interface, which establishes the scope and type of data that will be available as a resource for the generation of information” (Floridi 2008, 320). For a detailed definition see (Floridi 2011, 46).

An LoA presents an interface where the observed behaviour – either in virtual actuality or real actuality – may be interpreted. Under an LoA, different observations may result due to the fact that social computing software can be executed in different run-time environments, e.g. in a testbed in contrast to a real-time environment. Different LoAs correspond to different abstractions of one and the same behaviour of computing systems in a certain run-time environment. Different observations under one and the same LoA are possible if different versions of a program are run. Such differences may result when software agents are replaced by humans.

Conceptual entities may also be interpreted at a chosen LoA. Note that different levels of abstraction may coexist. Since levels of abstractions correspond to different perspectives, the system designer’s LoA may be different from the sociologist’s LoA or the legal engineer’s LoA of one and the same social computing system. These LoAs are related but not necessarily identical.

The basis of technology in action is not a pragmatic fiction of action but a conceptual model of the desired behaviour. From the designer’s point of view, metaphors often serve as a starting point to develop, e.g. novel heuristics to solve NP-complete (optimization) problems, that is problems for which no fast solution is known. Such metaphors may be borrowed from biology, sociology or economics. Research areas such as neural nets, swarm intelligence approaches and electronic auction procedures are products of such approaches. In the design phase, ideas guiding the modelling phase are often quite vague at first. In due course, their concretization results in a conceptual model (Ruß et al. 2010, 107) which is then specified as a software system. From the user’s or observer’s point of view, during run-time the more that is known about the conceptual model, the better its potential for (distributed) agency can be predicted and the better the hybrid constellations of (collective) action, emerging at run-time, may be analysed. Thus the actuality of

agential behaviour is complemented by a perspective on the system model determining the potential of technical agents. The philosophical value added by this approach not only lies in a reconstructive approach as intended by Latour and Rammert but also in the conceptual modelling and engineering of the activity space. Under an LoA for agency and action, activities may be observed as they unfold. Moreover, the system may be analysed and educated guesses about its future behaviour can be made. Both the specifics of distinct systems and their commonalities may be compiled.

## 3.5 Multidimensional Gradual Agency

### 3.5.1 Introduction

The following proposal for a conceptual framework for agency and action was first introduced in (Thürmel 2012) and expanded in (Thürmel 2013). It is intended to provide a multidimensional gradual classification scheme for the observation and interpretation of scenarios where humans and non-humans interact. It enables appropriate lenses to be defined, i.e. levels of abstraction, under which to observe, interpret, analyse and judge their activities. This does justice to Floridi's dictum that the task of the "philosophy of information" is "conceptual engineering, that is, the art of identifying conceptual problems and of designing, proposing, and evaluating explanatory solutions" (Floridi 2011, 11). In our case, the conceptual problem is how to characterize agency and interagency between humans, robots and software agents such that all current forms of interplay can be analysed. Moreover the framework should be so flexible to allow future technical developments to be included but to be not more complex than necessary. The proposed solution is a multidimensional gradual classification scheme which is presented in the following.

In contrast to observing "agency *in medias res*" (Rammert 2011, 15), the potential of smart, autonomous technology is the focus of this model. The engineering perspective makes it possible to design the potential and realize the actuality of computer-mediated artefacts. While technographs such as Latour strive to observe and analyse the interactions without prejudices by "opening up black boxes", this paper advocates making use of computational science and engineering know-how in order to enhance the understanding of socio-technical environments. Thus Latour's flat and symmetric concept of agency, which he applies to both humans and non-humans, is not used. Rammert's fiction of technical agency (Rammert 2011) is substituted by Floridi's "method of levels of abstraction" (Floridi 2008, 2011). Thus the underdetermined so-called "pragmatic fiction of technical agency" need not be contrasted with the "reality of distributed agency" (Rammert 2011) in socio-technical environments. Both the potential of individual and distributed agency and its actualization may be described by domain-specific levels of abstraction. A multidimensional perspective on the individual and joint capabilities of human and non-human actors replaces the one-dimensional layered model of Rammert and Schulz-Schaeffer (2002).



As Rammert states, “agency really is built into technology” but – as demonstrated above – not “as it is embodied in people” (Rammert 2011, 6) but by intelligent design performed by engineers and computer scientists. In order to demonstrate the potential for agency, not only the activity levels of any entities but also their potential for adaptivity, interaction, personification of others, individual action and joint action has to be taken into account.

Being at least (re)active is the minimal requirement for being an agent. Higher activity levels allow the environment to be influenced. Being able to adapt is a gradual faculty. It starts with primitive adaption to environment changes and ranges up to the adaption of long-term strategies and the corresponding goals based on past experiences and (self-reflective) reasoning of human beings. As shown below, acting may be discerned from just behaving based on activity levels and on being able to adapt in a “smart” way.

The potential for interaction is a precondition of any collaborative performance. The potential of the personification of others enables agents to integrate predicted effects of own and other actions. “Personification of non-humans is best understood as a strategy of dealing with the uncertainty about the identity of the other ... Personifying other non-humans is a social reality today and a political necessity for the future” (Teubner 2006, 497). Personification is similar to Dennett’s intentional stance (1995) since it is a pragmatic attribution. It starts with the attribution of simple dispositions up to perceiving the other as a human-like actor. This capability may affect any tactically or strategically motivated individual action. Moreover, it is a prerequisite of any form of defining joint goals and joint (intentional) commitment in any ensemble of agents. The capabilities for individual action and joint action may be defined based on activity levels, the potential for adaptivity, interaction and personification of others possessed by the involved actor(s).

Any type of an agential entity may be classified according to its characteristics in these dimensions. For any entity type the maximum potential (in these dimensions) is defined by a distinct value tuple. It may be depicted by a point in the multidimensional space spanned by the dimensions introduced above.

Any instantiation of an agent may be characterized by a distinct value tuple at a moment in time, i.e. by its actual time-stamped value. In agent-based systems, the changes over time correspond to changes of state of each agent.

Note that in the following the granularity on the different axes is only used as an example and can be adjusted according to the systems to be analysed and/or compared.

### ***3.5.2 The Multidimensional Framework for Individual and Distributed Agency***

The conceptual framework for agency and action offers a multidimensional gradual classification scheme for the observation and interpretation of scenarios where humans and non-humans interact. The “activity level” axis aims at the dimension

“technology in action” (Rammert 2008) providing a scale for the grade of active behaviour a technical object may display. The degree of adaptivity describes the plasticity of the phenotype. Individual agency may be described based on the potential for activity and adaptivity. Interaction is needed for coordination and control via communication. The personification of others may serve as a basis for joint action. Joint agency may be defined based on these dimensions.

The activity level allows individual behaviour to be characterized depending on the degree of self-inducible activity potential. It starts with passive entities such as road bumpers, hammers, and nails. Entities that display a certain predefined behaviour once they are started may be called semi-active (Rammert 2011, 7) or active without alternatives. Examples include hydraulic pumps or software artefacts such as algorithms searching in batch mode, compilers or basic help assistants. Reactive objects demonstrate the next level. These technical elements display identical output to identical input, e.g. realized by simple feedback loops or other situated reactions such as heating systems, swarm intelligence systems or ant colony optimization algorithms. Active entities permit individual selection between alternatives resulting in changes in the behaviour. From an internal perspective, this corresponds to Rammert’s level of contingency, ranging from choosing between preselected options to self-generated actions (Rammert 2011, 9). The minimal requirement for active entities is: perceive-plan-act. Examples are to be found in robotics, in sophisticated software agents like automatic bid agents in high-frequency trading systems, and in certain multi-agent systems realizing e-negotiation as well as in cyber-physical systems. Proactive entities try to anticipate the behaviour of other entities and act accordingly. The minimal requirement for their internal organization is: perceive-predict-(re)plan-act. Such technical modules are part of many cyber-physical systems where processes are controlled, e.g. in traffic control systems. Multi-agent systems in the above-mentioned emergency control systems may also display proactive behaviour in a dynamically changing environment. The next level corresponds to the ability to set one’s own goals and pursue them. It requires self-regulation based on self-monitoring and self-control. Intrinsic motivation may support such a process management, at least in humans. For the foreseeable future, self-conscious intentional behaviour will be reserved for humans.

These capabilities depend on an entity-internal system for information processing linking input to output. In the case of humans, it equals a cognitive system connecting perception and action. For material artefacts or software agents, an artificial “cognitive” system couples (sensor) input with (actuator) output.

Based on such a system for (agent-internal) information processing, the level of adaptivity may be defined. It characterizes the plasticity of the phenotype, i.e. the ability to change one’s observable characteristics including any traits that may be made visible by a technical procedure, in correspondence to changes in the environment. Models of adaptivity and their corresponding realizations range from totally rigid to simple conditioning up to impressive cognitive agency, i.e. the ability to learn from past experiences and to plan and act accordingly. A wide range of models coexist enabling study and experimentation with artificial “cognition in action”. This dimension is important to all who define agency as situation-appropriate

behaviour and who deem the plasticity of the phenotype as an essential assumption of the conception of man.

The potential for interaction, i.e. coordination by means of communications, is the basis of most if not all social computing systems and approaches to distributed problem solving. It may range from non-existent to informing others via bulletin boards such as those used in social networks and other forms of asynchronous communication to hard-wired cooperation mechanisms. Structured communication based on predefined scripts constitutes the next level. Examples are found in electronic auctioning. Unstructured, ad hoc communication of arbitrary elements may be found in demand-oriented coordination.

Inter-agency between technical agents and humans ranges from fixed instrumental relations, as found in cyborgs, to temporary instrumental relations like those between a human and an exoskeleton, a virtual servant or a robotpet. Principal-agent relations are realized when tasks are delegated to others. Simple duties may be delegated to primitive software assistants such as mailbots, sellbots or newsbots. More complex ones may be performed by robots or software agents that are bound by directives. Flexible partnerships may be seen in hybrid multi-agent systems, where humans interact with technical agents. Technical agents are currently not stakeholders to whom legal or moral responsibility is delegated. Nevertheless, these options and their implications for our legal systems are already discussed in the literature (Pagallo 2013; Chopra and White 2011).

Primitive mechanisms for coordination as in swarm intelligence systems do not need the personification of interaction partners. Ad hoc cooperation is a different case. The personification of others lies in the foundation for interactive planning, sharing strategies and adapting actions. This capability is non-existent in most material and software agents.

Assuming that another agent possesses a certain disposition to behave or act may be considered as the most fundamental level of personification. It may be found in theoretic game approaches or in so-called minimal models of the mind (Butterfill and Apperly 2013). Such models are used in robotics, e.g. in Hiatt et al. (2011). Many technical tools display purely passive or only reactive behaviour without the ability to adapt to changes in the environment. However, many technical agents such as automatic bid agents of electronic auctioning systems or cars on autopilot are able to learn from past experiences. The dynamics of social interaction and action-based learning and concept forming may lie at the foundation for “bootstrapping the cognitive system” of robots. Cangelosi and colleagues present a “roadmap for developmental robotics” based on such an approach (Cangelosi et al. 2010). Dominey and Warneken aim to explore “the basis of shared intentions in human and robot cognition”. They demonstrate how “computational neuroscience”, robotics and developmental psychology may stimulate each other (2011). These research projects may serve as circumstantial evidence for an evolutionary path in robotics.

The work done by Tomasello and colleagues motivates engineers to strive towards lessening the gap between the cognitive and agential capabilities of current technical agents and humans. According to the latest scientific findings, “chimpanzees understand others in terms of a perception-goal psychology, as

opposed to a full-fledged, human-like belief-desire psychology” (Call and Tomasello 2008, 187). This provides the basis for topic-focused group decision making based on egoistical behaviour: “they [the chimpanzees] help others instrumentally, but they are not so inclined to share resources altruistically and they do not inform others of things helpfully” (Warneken and Tomasello 2009, 397). Thus great apes may display so-called joint intentionality (Call and Tomasello 2008). In contrast, young children seem to have a “biological disposition” for helping and sharing. It may even be shown that “collaboration encourages equal sharing in children but not in chimpanzees” (Hamann et al. 2011). Understanding the other as an intentional agent allows even infants to participate in so-called shared actions (Tomasello 2008). Understanding others as mental actors lays the basis for interacting intentionally and acting collectively (Tomasello 2008). Engineers do not expect technical agents to evolve as humans did but they may profit from the insights gained in evolutionary anthropology.

Currently there is quite a gap between non-human actors and human ones in terms of their ability to take the “shared point of view” (Tuomela 2007) and to interact intentionally. This strongly limits the scope of social computing systems when they are used to predict human behaviour or if they are aimed at engineering and simulating future environments.

### ***3.5.3 Individual Agency and Inter-agency***

The potential for both individual action and for joint action may be defined based on the above-mentioned capabilities for activity, adaptivity, interaction and personification of others. The individual agency ranges from the individual potential for behaving to the individual potential for acting: disjunct levels may be defined based on the dimensions “activity level” and ability to adapt. The activity level of technical artefacts defines whether an artefact may only be used as a tool or may actively interact with its environment. Adaptivity is crucial for the individual regulation of the behaviour and subtle execution control.

In order to stress the communalities between human and non-human agents, an agent is counted as being capable of acting (instead of just behaving) if the following conditions concerning its ontogenesis hold: “the individual actor [evolves] as a complex, adaptive system (CAS), which is capable of rule-based information processing and based on that able to solve problems by way of adaptive behaviour in a dynamic process of constitution and emergence” (Kappelhoff 2011, 320). Thus the ability to learn and adapt is deemed crucial for acting.

Disjunct levels of inter-agency and distributed agency may be characterized based on activity levels and the corresponding range of the capability for interaction. Current technical agents mostly interact based on predefined scripts. Ad hoc communications often remain a technical challenge. However, the robotics learns from evolutionary anthropologists so that service robots can be taught to move in households and participate in basic cooperative actions (CoTeSys 2014) or be taught

verbal interaction based on exemplary communication events (Fischer et al. 2011). These projects may provide the first tentative examples of “shared cooperative activity” (Bratman 1992), a kind of shared activity which is based on mutual responsiveness, commitment to joint activity and commitment to mutual support. Rational, self-organising teams consisting of collaborating humans, robots and software agents can be found in current research projects as self-organizing production systems or managed health care systems. Such teams display a “modest sociality” (Bratman 2014) emerging from structures of interconnected planning agency.

Constellations of inter-agency and distributed agency in social computing systems or hybrid constellations, where humans, machines and programs interact, may be described, examined and analysed using the above-introduced classification scheme for agency and action. These constellations start with purely virtual systems like swarm intelligence systems and fixed instrumental relationships between humans and assistive software agents where certain tasks are delegated to artificial agents. They continue with flexible partnerships between humans and software agents. They range up to loosely coupled complex adaptive systems. The latter may model such diverse problem spaces as predator–prey relationships of natural ecologies, legal engineering scenarios or disaster recovery systems. Their common ground and their differences may be discovered when the above-outlined multidimensional, gradual conceptual framework for agency and action is applied.

A subset of these social computing systems, namely those that may form part of the infrastructure of our world, provides a new form of “embedded governance”. Their potential and limits may also be analysed using the multidimensional agency concept.

### 3.6 Conclusions and Future Work

The proposed conceptual framework for agency and action offers a multidimensional gradual classification scheme for the observation and interpretation of scenarios where humans and non-humans interact. It may be applied to the analysis of the potential of social computing systems and their virtual and real actualizations. The above-introduced approach may also be employed to describe situations where decisions to act are delegated to technical agents. It can be used both by the software engineer and the philosopher when role-based interaction in socio-technical systems is to be defined and analysed during execution.

Proto-ethical agency in social computing systems may be explored by adapting (Moor 2006) to the framework. Profiting from work done by Darwall (2006), the framework could be expanded in order to potentially attribute commitments to diverse socio-technical actors. Shared agency, a “planning theory of acting together” as defined by Bratman (2014), could be investigated in socio-technical contexts where technical elements are not mere tools but interaction partners. Last but not least, social relations to technical agents could be evaluated similarly to (Misselhorn et al. 2013) by making use of the framework when characterizing potentiality and actuality of the technical agents.

## References

- Berthier, Denis. 2004. *Médiations sur le réel et le virtuel*. Paris: Editions L'Harmattan.
- Berthier, Denis. 2007. Qu'est-ce que le virtuel. *La Jaune et la Rouge*, June 2007, <http://denis.berthier.pagesperso-orange.fr/Articles/JR-Qu-est-ce-que-le-virtuel.pdf>. Accessed 27 Apr 2014.
- Bratman, Michael. 1992. Shared cooperative activity. *The Philosophical Review* 101(2): 327–341.
- Bratman, Michael. 2014. *Shared agency: A planning theory of acting together*. Cambridge, MA: Harvard University Press.
- Butterfill, Stephen, and Ian Apperly. 2013. How to construct a minimal theory of mind. *Mind & Language* 28(5): 606–637.
- Call, Joseph, and Michael Tomasello. 2008. Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences* 12(5): 187–192.
- Callon, Michel, and Bruno Latour. 1992. Don't throw the baby out with the bath school! A reply to Collins and Yearley. In *Science as practice and culture*, ed. Andrew Pickering, 343–368. Chicago: Chicago University Press.
- Cangelosi, Angelo, Giorgio Metta, Gerhard Sagerer, Gerhard Nolfi, Chrystopher Nehaniv, Kerstin Fischer, Jun Tani, Tony Belpaeme, Giulio Sandini, Francesco Nori, Luciano Fadiga, Britta Wrede, Katharina Rohlfing, Elio Tuci, Kerstin Dautenhahn, Joe Saunders, and Arne Zesche. 2010. Integration of action and language knowledge: A roadmap for developmental robotics. *IEEE Transactions on Autonomous Mental Development* 2(3): 167–195.
- Chliaoutakis, Angelos, and Georgios Chalkiadakis. 2014. Utilizing agent-based modeling to gain new insights into the ancient Minoan civilization. In *Proceedings of the 2014 international conference on autonomous agents and multi-agent systems*, 1371–1372. Richland: SC.
- Chopra, Sami, and Laurence White. 2011. *A legal theory for autonomous artificial agents*. Ann Arbor: The University of Michigan Press.
- Clark, Andy. 2003. *Naturally born cyborgs*. Oxford: Oxford University Press.
- CoTeSys. 2014. Cluster of excellence cognition in technical systems. <http://www.cotesys.de/>. Accessed 28 Apr 2014.
- Darwall, Stephen. 2006. *The second person standpoint: Morality, respect and accountability*. Cambridge, MA: Harvard University Press.
- Dennett, Daniel. 1995. *The intentional stance*. Cambridge, MA: MIT Press.
- Dominey, Peter, and Felix Warneken. 2011. The basis of shared intentions in human and robot cognition. *New Ideas in Psychology* 29: 260–274.
- Economist. 2010. Agents of change, 22 July 2010. <http://www.economist.com/node/16636121/print>. Accessed 27 Apr 2014.
- Fischer, Silke, Denis Schulze, Pia Borggrebe, Martina Piefke, Sven Wachsmuth, und Katharina Rohlfing. 2011. Multi-modal anchoring in infants and artificial systems. Poster presented at IEEE-ICDL-EPIROB conference 2011, Frankfurt, 24–27 Aug 2011.
- Floridi, Luciano. 2008. The method of levels of abstraction. *Minds and Machines* 18(3): 303–329.
- Floridi, Luciano. 2011. *The philosophy of information*. Oxford: Oxford University Press.
- Giddens, Anthony. 1984. *The constitution of society, outline of the theory of structuration*. Cambridge, UK: Polity Press.
- Hamann, Katharina, Felix Warneken, Julia Greenberg, and Michael Tomasello. 2011. Collaboration encourages equal sharing in children but not in chimpanzees. *Nature* 476: 328–331. doi:10.1038/nature10278.
- Hiatt, Laura, Anthony Harrison, and Gregory Trafton. 2011. Accommodating human variability in human–robot teams through theory of mind. *Proceedings of the International Joint Conference on Artificial Intelligence 2011* 3: 2066–2071.
- Hossain, M. Anwad, and Dewan Ahmed. 2012. Virtual caregiver: An ambient-aware elderly monitoring system. *IEEE Transactions on Information Technology in Biomedicine: A publication of*

- the *IEEE Engineering in Medicine and Biology Society* 16(6): 1024–1031. doi:10.1109/TITB.2012.2203313.
- Hubig, Christoph. 2006. *Die Kunst des Möglichen I – Technikphilosophie als Reflexion der Medialität*. Bielefeld: Transcript Verlag.
- Hubig, Christoph. 2010. Technik als Medium und “Technik” als Reflexionsbegriff. Manuscript. [http://www.philosophie.tu-darmstadt.de/institut/mitarbeiterinnen\\_1/professoren/a\\_hubig/downloadbereich/downloadsprofhubig.de.jsp](http://www.philosophie.tu-darmstadt.de/institut/mitarbeiterinnen_1/professoren/a_hubig/downloadbereich/downloadsprofhubig.de.jsp). Accessed 30 Apr 2014.
- Jennings, Nicolas. 2010. *ALADDIN end of project report*. <http://www.aladdinproject.org/wp-content/uploads/2011/02/finalreport.pdf>. Accessed 27 Apr 2014.
- Kappellhoff, Peter. 2011. Emergenz und Konstitution in Mehrebenenselektionsmodellen. In *Emergenz – Zur Analyse und Erklärung komplexer Strukturen*, ed. Jens Greve and Annette Schnabel, 319–345. Berlin: Suhrkamp Taschenbuch Wissenschaft.
- Latour, Bruno. 1988. Mixing humans and nonhumans together. The sociology of a door-closer. *Social Problems* 35(4): 298–310.
- Latour, Bruno. 1990. Drawing things together. In *Representation in scientific practice*, ed. Michael Lynch and Steve Woolgar, 19–68. Cambridge, MA: MIT Press.
- Latour, Bruno. 1996. On interobjectivity. *Mind, Culture, and Activity* 3: 228–245.
- Latour, Bruno. 2005. *Reassembling the social – An introduction to actor-network-theory*. Oxford: Oxford University Press.
- Laurel, Brenda. 1991. *Computers as theatre*. Reading: Addison-Wesley.
- Leibnizcenter for Law. 2011. Multi-agent PhD position available. [www.leibnitz.org/wp-content/uploads/2011/02/werwerbung.pdf](http://www.leibnitz.org/wp-content/uploads/2011/02/werwerbung.pdf). Accessed 24 Jan 2012.
- Maes, Pattie. 1994. Adaptive autonomous agents. *Artificial Life Journal* 1(1): 135–162.
- Mainzer, Klaus. 2007. *Thinking in complexity. The complex dynamics of matter, mind, and mankind*, 5th ed. Heidelberg: Springer.
- Mainzer, Klaus. 2010. *Leben als Maschine?: Von der Systembiologie zur Robotik und Künstlichen Intelligenz*. Paderborn: Mentis Verlag.
- Misselhorn, Catrin, Ulrike Pompe, and Mogg Stapleton. 2013. Ethical considerations regarding the use of social robots in the fourth age. *GeroPsych* 26(2): 121–133.
- Moor, James. 2006. The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems* 21(4): 18–21. doi:10.1109/MIS.2006.80.
- Muldoon, Ryan, Chiara Liscianra, Christina Bicchieri, Stephan Hartmann, and Jan Sprenger. 2014. On the emergence of descriptive norms. *Politics, Philosophy and Economics* 13(1): 3–22.
- Nealon, John, and Antonio Moreno. 2003. Agent-based health care systems. In *Applications of software agent technology in the health care domain*, ed. John Nealon and Antonio Moreno, 3–18. Basel: Birkhäuser Verlag.
- Pagallo, Ugo. 2013. *The law of robots: Crimes, contracts, and torts*. New York: Springer.
- Rammert, Werner. 2007. *Die Techniken der Gesellschaft: In Aktion, in Interaktivität und in hybriden Konstellationen*. Berlin: The Technical University Technology Studies Working Papers.
- Rammert, Werner. 2008. *Where the action is: Distributed agency between humans, machines, and programs*. Berlin: The Technical University Technology Studies Working Papers.
- Rammert, Werner. 2011. *Distributed agency and advanced technology or: How to analyse constellations of collective inter-agency*. Berlin: The Technical University Technology Studies Working Papers.
- Rammert, Werner, and Ingo Schulz-Schaeffer. 2002. Technik und Handeln: Wenn soziales Handeln sich auf menschliches Verhalten und technische Abläufe verteilt. In *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik*, ed. Werner Rammert and Ingo Schulz-Schaeffer, 11–64. Frankfurt: Campus.
- Ruß, Aaron, Dieter Müller, and Wolfgang Hesse. 2010. Metaphern für die Informatik und aus der Informatik. In *Menschenbilder und Metaphern im Informationszeitalter*, ed. Michael Böcker, Mathias Gutmann, and Wolfgang Hesse, 103–128. Berlin: LIT Verlag.

- Schulz-Schaeffer, Ingo. 2000. Akteur-Netzwerk-Theorie. Zur Koevolution von Gesellschaft, Natur und Technik. In *Soziale Netzwerke. Konzepte und Methoden der sozialwissenschaftlichen Netzwerkforschung*, ed. Johannes Weyer, 187–209. München: R. Oldenbourg Verlag.
- Teubner, Gunther. 2006. Rights of non-humans? Electronic agents and animals as new actors in politics and law. *Journal of Law and Society* 33: 497–521.
- Thürmel, Sabine. 2012. A multi-dimensional agency concept for social computing systems. In *Proceedings of the AISB/IACAP world congress social computing, social cognition, social networks and multiagent systems*, ed. Gordana Dodig-Crnkovic, Antonino Rotolo, Giovanni Sartor, Judith Simon, and Clara Smith, 87–91.
- Thürmel, Sabine. 2013. Die partizipative Wende: Ein multidimensionales, graduelles Konzept der Handlungsfähigkeit menschlicher und nichtmenschlicher Akteure. Dissertation, München: Technische Universität München.
- Tomasello, Michael. 2008. *Origins of human communication*. Cambridge, MA: MIT Press.
- Tuomela, Raimo. 2007. *The philosophy of sociality: The shared point of view*. Oxford: Oxford University Press.
- Turkle, Sherry. 2006. A nascent robotics culture: New complicities for companionship, *American Association for Artificial Intelligence (AAAI) technical report series*, July 2006, 21st national conference on artificial intelligence human implications of human-robot interaction (HRI), Boston July 2006.
- Warneken, Felix, and Michael Tomasello. 2009. Varieties of altruism in children and chimpanzees. *Trends in Cognitive Science* 13(9): 397–402.
- Wedde, Italy, Horst, Sebastian Lehnhoff, Christian Rehtanz, and Olav Krause. 2008. Bottom-up self-organization of un-predictable demand and supply under decentralized power management. In *Proceedings of the 2nd IEEE international conference on self-adaptation and self-organization (SASO'08)*, 10–20. Venice: Italy.
- Woolridge, M. 2009. *An introduction to multiagent systems*, 2nd ed. Italy: Venice.