Agents, Intelligence and Tools

Andrea Omicini, Michele Piunti, Alessandro Ricci, and Mirko Viroli

Abstract This chapter investigates the relationship among agent intelligence, environment and the use of tools. To this end, we first survey, organise and relate many relevant approaches in the literature, coming from both within and without the fields of artificial intelligence and computer science. Then we introduce the A&A meta-model for multiagent systems (MAS), where *artifacts*, working as tools for agents, are used as basic building blocks for MAS modelling and engineering, and discuss the related metaphor of the *Agens Faber*, which promotes a new, principled way to conceive and build intelligent systems.

1 Introduction

The role of *tools* beyond language is variously exploited in human activities and societies. Organised workspaces based on artifacts and tools of diverse nature are ubiquitous in human *environments*. A tool can be conceived and explicitly built to achieve a specific goal (embedding a specific goal), stored for repeated and iterated use and exploited for building new tools. In general, a tool requires expertise of its users, their awareness of the domain problem, as well as their expertise in problem solving [14]. In all, the ability to use and make tools is as essential as symbolic language skills in defining intelligence of human beings and is typically used by ethologists to understand and measure animal intelligence.

Given the straightforward anthropomorphic interpretation of agents as human representatives in computational systems, even in the trivial acceptance of *personal assistants* [13], the need for a definition of a notion of tool in the fields of MAS (multiagent systems) and AI is quite obvious. In particular, it seems essential for the very notion of intelligence in MAS and in general for the notion of intelligent

Andrea Omicini, Michele Piunti, Alessandro Ricci, Mirko Viroli
ALMA MATER STUDIORUM, Università di Bologna, Cesena, Italy, e-mail: {andrea.omicini, michele.piunti, a.ricci, mirko.viroli}@unibo.it

system to provide agents with the conceptual instruments to perceive and affect the environment where they live and interact, going beyond the well-explored issues of an agent's internal architecture and ability to speak symbolic languages.

In this chapter we first survey some of the most relevant conceptual frameworks in human sciences where the notions of environment and tools are suitably developed and related to the issue of human intelligence (Section 2). Then, we shift our focus on artificial systems (Section 3), discussing some of the most relevant literature on the relationship between intelligence and environment. Drawing from the results of the previous sections, in Section 4 we present the A&A meta-model for MAS, which introduces a further dimension besides agent rationality in the context of intelligent agent systems, i.e. the dimension of the artifacts and tools conceived and designed to support agent rationality and activity. The notion of *Agens Faber* is discussed, and we elaborate on its impact on the notion of agent and system intelligence, as well as on the construction of intelligent systems. Finally, Section 5 concludes the chapter.

2 The Role of Environment, Artifacts and Tools in Human Cognitive Systems

The history of human evolution is characterised by development of forms of complex social life accompanied by growing cognitive abilities in using and making complex tools communicating in complex ways [7]. However, as remarked by Norman [16], the power and importance of culture and artifacts to enhance human abilities are ignored within much of contemporary cognitive science despite the heavy prominence given to their importance in the early days of psychological and anthropological investigation. The field has a sound historical basis, starting at least with Wundt [29], nurtured and developed by the Soviet social-historical school of the 1920s [26, 12, 11, 27], and still under study by social scientists, often unified by titles such as *activity theory*, *action theory*, *situated action*, with most of the research centered in Scandinavia, Germany, and the former Soviet Union.

In the early part of the 1900s, American psychology moved from its early interest in mental functioning to the behavioral era, in which studies of representational issues, consciousness, mind, and culture were almost neglected. As a result, the historical continuity with the earlier approaches as well as with European psychology had been lost. With the end of the behavioral era, American cognitive psychology had to recreate itself, borrowing heavily from British influences. The emphasis was on the study of the psychological mechanisms responsible for memory, attention, perception, language, and thought within the single, unaided individual, studied almost entirely within the research laboratory. There was little or even no emphasis on group activities, on the overall situation in which people accomplished their normal daily activities, or on naturalistic observations, and then little thought was given to the role of the environment (whether natural or artificial) in the study of human cognition.

The field only recently returned to pay serious attention to the role of the situation, other people, natural and artificial environments, and culture. In part, this change has come about through the dedicated effort of the current researchers, in part because the current interest in disciplines such as Computer Supported Cooperative Work, Human Computer Interaction and Distributed Artificial Intelligence has forced consideration of the role of real tasks and environments, and therefore of groups of cooperating individuals, of artifacts, and of culture.

In the remainder of the section we discuss the main points that characterise the notion and role of environment, artifacts and tools within human cognitive systems, by recalling some of the main concepts developed in the context of the studies and disciplines that mostly focussed on such aspects.

2.1 Context and Tools in Human Activities: Activity Theory and Distributed Cognition

The basic underlying principle of Activity Theory (AT) is the *principle of unity and inseparability of consciousness (human mind) and activity*: human mind comes to exist, develops, and can only be understood within the context of a meaningful, goal-oriented, and socially determined interaction between human beings and their material environment. Then, a fundamental aspect for AT has been from its beginning the *interaction* between the individuals and the *environment* where they live, in other terms, their *context*. After an initial focus on the activity of the individuals, AT research has lately evolved toward the study of human collective work and social activities, then elaborating on issues such as the coordination and organisation of activities within human society.

A central point in the AT conceptual framework is the fundamental role of *artifacts* and *tools* in human activities: according to AT every non-trivial activity is mediated by some kind of artifact. More precisely, every activity is characterised by a *subject*, an *object* and by one or more *mediating artifacts*:

- a *subject* is an agent or group engaged in an activity;
- an *object* (in the sense of *objective*) is held by the subject and motivates the activity, giving it a specific direction (the objective of the activity); the object of activity could range from mental objectives (e.g. making a plan) to physical ones (e.g. writing a paper);
- the *mediation artifacts*, which are the tools that enable and mediate subject actions toward the object of the activity. The mediating artifacts could be either physical or abstract / cognitive, such as symbols, rules, operating procedures, heuristics, scripts, individual / collective experiences, and languages.

The definition is clearly oriented to bringing into the foreground not only individuals (subjects) and their cognitive aspects, but also the context where they play, and the continuous dynamic process that links subjects to the context.

Mediation, along with *mediated interaction*, is a central point of the definition. This reflects one of the main conceptual cornerstores coming from Soviet Psychology (SP): there, a fundamental feature of human development is the shift from a *direct* mode of acting on the world to one *mediated* by some external tool; whereas Marx focused on physical mediating tools, Vygotsky extended the concept toward psychological tools. In both cases a human actor, according to the vision, is not reacting directly to the world, as an animal does, but always by means of a mediating artifact of some sort.

According to AT, mediating tools have both an *enabling* and a *constraining* function: on the one hand, they expand out possibilities to manipulate and transform different objects, but on the other hand the object is perceived and manipulated not 'as such' but within the limitations set by the tool. Mediating artifacts shape the way human beings interact with reality. According to the principle of internalisation / externalisation, shaping external activities ultimately results in shaping internal ones. Then, artifacts embody a set of social practices, and their design reflects a history of particular use They usually reflect the experiences of other people who have tried to solve similar problems at an earlier time and invented / modified the tool to make it more efficient. Experience is accumulated in the structural properties of the tools (shape, material,..), and in the knowledge of how the tools should be used as well The term *appropriation* is used to indicate the process of learning these properties and knowledge [26]. Finally, mediating tools are created and transformed during the development of the activity itself, then they carry on a given culture, the historical remnants of that development. So, the use of tools is a means for the *accumulation* and *transmission* of *social knowledge*: tools influence not only the external behaviour, but also the mental functioning of individuals using them.

The notion of artifact and tool are also at the core of Distributed Cognition (DCog), a theory of psychology recently developed by Edwin Hutchins [9], focussing on the social aspects of cognition. The core idea of DCog is that human knowledge and cognition are not confined to the individual: instead, they are distributed by placing memories, facts, or knowledge on the objects, individuals, and tools in our environment. Accordingly, social aspects of cognition are understood and designed by putting emphasis on the environment where individuals are situated, which provides opportunities to reorganise the distributed cognitive system to make use of a different set of internal and external cognitive processes. A system is conceived as a set of representations, which could be either in the mental space of the participants, or external representations available in the environment. Here, a main aspect concerns the interchange of information between these representations. For this purpose, DCog proposes a framework where the co-ordination between individuals and artifacts is based on explicitly modelling the representations where information is held in and transformed across, as well as the processes by which representations are co-ordinated with each other.

2.2 Intelligence as Active Externalism

In [24], Sterenly illustrates how many organisms are epistemic and ecological engineers: their evolutionary capabilities to alter systems are part of their intelligence, and often have fitness effects that get strengthened across generations. In particular, cognitive agents that have the particular epistemic ability to intentionally act for changing the informational character of their environment are used to facilitate their activities. Humans continually modify and arrange their environment not only in order to achieve some personal goal, but also for additional practical aims. By modifying environments, they ease their tasks aiding memory and computational burdens.

Clark's *Active Externalism* [5] described the active role of environment in driving cognitive processes, blurring the boundaries between internal and external cognition. In particular, he enlightened the special relation between mind and artifacts. Both internal and external resources are viewed as an *extended mind* engaged in a larger coupled interaction where the agent (the subject) interacts with an external device (the tool). Sometimes such boundaries are indiscernible, and the aggregate agent-plus-environment may be seen as generating a unique extended cognitive system. An example was brought by Simon [23], according to whom human internal reasoning relies on internal resources such as brain memory 'as if' they were external devices, hence search information in personal memory is not different in essence from search in the external environment.

Besides, Sterenly [24] stressed that the effective use of artifact and epistemic tools relies on informational resources internally present to the agent. He noticed that the use of epistemic tools in common and contested space, where artifacts are jointly and repeatedly modified, created and used by a society of agents, *remains* a cognitively demanding process. Agents require a rich information base to fully exploit artifacts: they acquire this information piecemeal, often with the need to integrate different properties and uses across different information domains. Moreover, the use of artifacts in society implies the need to resolve problems of social and normative coordination (negotiation, division of tasks and resources).

2.3 Language vs. Tool, or The Language as a Tool

Gordon H. Hewes [8] observed how only in recent years has the relation between language and tools in human evolution finally been perceived as a single coherent problem. Language is an example of how subjective, inner cognitive abilities have been externalised in the outside world. Language evolved to enable extensions of human cognitive resources in the context of social, actively coupled systems: it clearly allows humans to properly spread some of their cognitive burden to others. In doing so, language serves as a special mediating tool whose role is to extend the bounds of subjective cognition. One may envisage the strict relation between use of language as a particular instance of the more general use of tools, both expanding the boundary between organisms (or agents) and the environment where they live.

In this context, along with language, tools can be considered as a distinctive expression of intelligence, powerful amplifiers of the (both individual and social) animal ability to affect the environment either to adapt to changes, i.e. preventing risks, or to meet purposes and achieve goals. Ethologists commonly assess the boundaries meant for intelligence by observing the ability of animals in facing problems that require the use of tools to be solved (see for instance [2, 21]). Even more interestingly, a *tool-equivalent* of the Turing test has been originally proposed by philosopher Ronald Endicott, and then refined by R. St. Amant and A.B. Wood in [2]. The "Tooling Test for Intelligence" is aimed at evaluating intelligence in terms of the ability to use, build and share tools - see also [2] for a detailed description. In their childhood, humans learn to manipulate objects even before evolving language abilities. Their linguistic and technical skills merge together in the processes of cognitive development. Speech, manual gesture, and other forms of communication evolve side by side to tool-using skills and imitative abilities [7]. Accordingly, artifacts and tools are objects that invite humans to develop new (non-lexical-based) interactions in their environment, hence improving skills of use, affordance, recognition, and manipulation.

3 The Role of Environment, Artifacts and Tools in Artificial Cognitive Systems

In the sciences of the artificial, the notion of environment has essentially followed the same conceptual trajectory as in human sciences. Neglected in the very beginning by the developments of the "Symbolic AI" traditional approach, the notion of environment was subsequently strongly developed by Brooks' situated agents - however, essentially refusing cognitive aspects of intelligent behaviour. Agre & Horswill [1] paved the way for a full-fledged notion of environment to be integral to the agent's reasoning loop, while Kirsh [10] pointed out the fundamental principles for the intelligent use of the environment.

After summarising such influential approaches to the notions of environment and tool in the remainder of this section, the next section presents the *Agens Symbolicus plus Agens Faber* vision, which extends the analytic approaches by Agre & Horswill and Kirsh towards the synthetic issues of the construction and engineering of intelligent systems, by introducing and exploiting the A&A meta-model for MAS.

3.1 Revisiting The Concept of Environment in AI and Robotics

The definition of the notion of environment in AI and robotics is somehow fuzzy, and essentially derives from the subjective view of agents that see the environment as whatever is "outside their skin". In general, the environment in artificial systems is defined, in a sense, dually with respect to the active entities in the system, and typically includes everything relevant (places, objects, circumstances, ...) that surrounds some given agent. The environment is where agents live, it is either the target or the means of their actions, and as such it determines their effects. In natural / social systems, as well as in artificial systems, a model of the environment is required, which could account for its structures and dynamics, and could work as a stable basis for an agent's practical reasoning.

The distinction between agent and environment has been somehow blurred in early AI: the world where agents live and interact is typically oversimplified, and can typically be reduced to the subjective perception and individual action of the agent. The notion of "task environment" by Newell & Simon [15] is on the one hand perfectly functional as a framework for analysis of action and environment, on the other hand it is also nothing more than a formal representation of the space of choices and outcomes that hides the complexity of the physical world, and also essentially reduces world to a subjective agent construct. As a result, for instance, two agents co-existing within the same environment but featuring different abilities to reason and act will basically result in inhabiting different task environments, even though their physical surroundings and goals might be identical. While this bias may bring no problems in cases such as theorem-proving and chess, it turns out to be crucial whenever agent's actions have uncertain outcomes.

Newell and Simon's phenomenological approach to task analysis was overcome by Agre and Horswill [1], who explicitly introduce a notion of environment out of the agent mind to support agent reasoning. In particular, the concept of *lifeworlds* introduces a structured vision of the environment, organised so as to promote the activities that take place within it. In particular, similar to [10], lifeworlds organise and arrange tools in the environment so as to make agent's activities simpler, and reduce the cognitive burdens of agents. Tools mediate activities and spread suitable solutions to given problems, and their use promotes repeated and customary activities. Environments may contain artifacts that have been *specifically evolved*, and thus designed, to support agent activities. Having in mind the precise dynamics of agent activities it becomes possible to design suitable machinery (such as tools, artifacts) that are consistent with a given pattern of interaction. Thus, principled characterisation of interactions between agents and their environment is then explicitly used to guide explanation and design.

3.2 The Intelligent Use of the Environment

To explain how the structure of environments may provide an empowerment in cognitive abilities, Kirsh adopts the metaphor of the environmental 'oracle' [10]. Accordingly, the environment can suitably be used as an external memory, for example for reminding the system which tasks still have to be performed, for spatially clustering the task in sub-problems, or for organising the tools and the artifacts according to some ordering principle. Thus, tool discovery, selection and use can be considered as a fundamental part of agent intelligence, as well as a pivotal aspect for defining agent activities. In [10], David Kirsh illustrated a number of reasons to shift attention to the particular role of environment for complex and distributed problem solving. By analysing agents and environments in their *relation of use*[1], Kirsh noticed many important benefits for agents.

Using environment simplifies choice — Intelligent use of space speeds up the creation of a problem space representation. Besides, this helps to contract the complexity and the cognitive costs by reducing the *fan-out* of the feasible actions and the number of times at which to take a decision. Exploiting tools ameliorates the average branching factor, hence the complexity of the decisions that was made less complex by information that could be read off from environment. It eliminates decision points, creating more rational decision alternatives[2]. Moreover it allows better heuristics of choice, driven by the affordances of tools and their perceived utilities.

Using environment simplifies perception — Space arrangements have a direct influence on costs of perception, i.e. making it possible to notice properties that have not been noticed yet, facilitating the discovery of relevant information. Simple techniques *regionalizing* environments and clustering objects according to their similarities can be used to highlight their differences thus creating categories of use and restricting the kind of actions an agent may take. Exploiting landmarks or pro-actively marking objects can be a practice for reminding activities and information not to be lost.

Using environment saves computation — Particular setting of environments can be suitably exploited to save internal computation, for instance creating a visual cue in order to make the relevant information more explicit.[3]

4 Agents & Artifacts as Abstractions for Engineering Intelligent Systems

In the context of autonomous agents and multi-agent systems (MAS), the notion of environment recently gained attention for the design and engineering of intelligent systems [28]. Besides agents, as autonomous pro-active entities, either software or physical, designed to accomplish some kind of goal or perform some activity, the

[1] A widely accepted definition of tool use is due to Beck [3]*"tool use is the external employment of an unattached environmental object to alter more efficiently the form, position or condition of another object, another organism, or the user itself when the user holds or carries the tool during or just prior to use and is responsible for the proper and effective orientation of the tool"*.

[2] Theoretically, by modifying the description of state to allow a single action selection at each decision point reduces the complexity of search from b^n to n, where n is the average depth to the goal node and b the average branching factor

[3] Kirsh proposed here a Tetris experiment, showing how the possibility of rotating zoids in the game takes less time than rotating them mentally.

environment can be considered a *first-order entity* defining the context where agents are situated, designed so as to embed some functionalities and features to be exploited by agents for supporting their individual and collective work.

Among others, the Agents and Artifacts (A&A) approach [22] introduces the notion of *artifact* as a first-class abstraction for modelling and designing MAS working environments, i.e. that part of the MAS designed by MAS engineers and then *used* by agents at runtime for their goals or tasks. The conceptual and engineering framework introduced by A&A draws its inspiration directly from the concepts introduced by Activity Theory and more generally from the theories and methodologies introduced in previous sections.

The A&A perspective introduces in the context of intelligent agent systems a further dimension besides agent rationality, i.e. the dimension of the artifacts and tools to be conceived and designed to support agent rationality and activity. Apparently, this introduces a dualism between *Agens Faber* and *Agents Symbolicus*, analogous to the duality between Homo Faber or Homo Symbolicus [4] - who comes first? [19]. Such a dualism obviously has to be solved without a winner: then, why should we choose between an Agens Faber and an Agens Symbolicus while we aim at intelligent agents? Accordingly, adopting an evolutionary perspective over agents, and carrying along the analogy with the development and evolution of human intelligence, we claim here that a theory of agent intelligence should not be limited to the modelling of inner, rational processes (as in BDI theory), and should instead include not only the basics of practical reasoning, but also a suitable theory of the artifacts and the means for their rational use, selection, construction and manipulation. This is in fact the idea behind the *Agens Faber* notion: agent intelligence should not be considered as separated by the agent's ability to perceive and affect the environment - and so, that agent intelligence is strictly related to the artifacts that enable, mediate and govern any agent (intelligent) activity.

Along this line, in the remainder of this section we first collect some considerations of ours about the conceptual relation between agents and artifacts, then we briefly describe a first model and taxonomy of artifacts introduced by the A&A approach.

4.1 On the Relation between Agents and Artifacts

4.1.1 Goals of Agents and Use of artifacts

By considering the conceptual framework described in [6], agents can be generally conceived as *goal-governed* or *goal-oriented* systems. Goal-governed systems refer to the strong notion of agency, i.e. agents with some form of cognitive capabilities, which make it possible to explicitly represent their goals, driving the selection of agent actions. Goal-oriented systems refer to the weak notion of agency, i.e. agents whose behaviour is directly designed and programmed to achieve some goal, which is not explicitly represented. In both goal-governed and goal-oriented systems, goals are *internal*. *External goals* instead refer to goals which typically belong to the social context or environment where the agents are situated. External goals are sorts of regulatory states which condition agent behaviour: a goal-governed system follows external goals by adjusting internal ones.

This basic picture is then completed by systems which are not goal-oriented. This is the case of passive objects, which are characterised by the concept of *use*: they have not internal goals, but can be *used* by agents to achieve their goals. *Artifacts* are objects explicitly designed to provide a certain *function*[4], which guides their use. The concept of *destination* is related but not identical to the concept of function: it is an external goal which can be attached to an object or an artifact by users, in the act of using it. Then an artifact can be used according to a destination which is different from its function.

An interesting distinction has been proposed, concerning agents / artifacts relationships, between *use* and *use value* [6]: there, use value corresponds to the evaluation of artifact characteristics and function, in order to *select* it for a (future) use. The distinction corresponds to two different kinds of external goals attached to an artifact: *(i)* the use-value goal, according to which the artifact should allow user agents to achieve their objective, such an external goal drives the agent selection of the artifact; *(ii)* the use goal, which directly corresponds to the agent internal goal, which guides the actual usage of the artifact. From the agent point of view, when an artifact is selected and used it has then a use-value goal that somehow matches its internal goal.

By extending the above considerations, the classical tool-using / tool-making distinction from anthropology can be articulated along three main distinct aspects, which characterise the relationship between agents and artifacts:

- use
- selection
- construction and manipulation

While the first two aspects are clearly related to use and use value, respectively, the third is the rational consequence of a failure in the artifact selection process, or in the use of a selected artifact. Then, a new, different artifact should be constructed, or obtained by manipulation of an existing one.

4.1.2 Agents Reasoning about Artifacts

One of the key issues in the Agens Faber approach is how artifacts can be effectively exploited to improve agent ability to achieve individual as well as social goals. The main questions to be answered are then: How should agents reason to use artifacts in the best way, making their life simpler and their action more effective? How can agents reason to select artifacts to use? How can agents reason to construct or adapt artifact behaviour in order to fit their goals?

[4] The term "function" here refers to the functionality embodied by an artifact, and should not be confused with the same term as used e.g. in mathematics or in programming languages

On the one hand, the simplest case concerns agents directly programmed to use specific artifacts, with usage protocols directly defined by the programmer either as part of the procedural knowledge / plans of the agent for goal-governed systems, or as part of agent behaviour in goal-oriented systems. In spite of its simplicity, this case can bring several advantages for MAS engineers, exploiting separation of concerns for programming simpler agents, by charging some burden upon specifically-designed artifacts. On the other hand, the intuition is that in the case of fully-open systems, the capability of the artifact to describe itself, its function, interface, structure and behaviour could be the key for building open MASs where intelligent agents dynamically look for and select artifacts to use, and then exploit them for their own goals.

At a first glance, it seems possible to frame the agent's ability to use artifacts in a hierarchy, according to five different cognitive levels at which the agent can use an artifact:

unaware use — at this level, both agents and agent designers exploit artifacts without being aware of it: the artifact is used implicitly, since it is not denoted explicitly. In other words, the representation of agent actions never refer explicitly to the execution of operation on some kind of artifacts.

embedded / programmed use — at this level, agents use some artifacts according to what has been explicitly programmed by the designer: so, the artifact selection is explicitly made by the designer, and the knowledge about its use is implicitly encoded by the designer in the agent. In the case of cognitive agents, for instance, agent designers can specify usage protocols directly as part of the agent plan. From the agent point of view, there is no need to understand explicitly artifact operating instructions or function: the only requirement is that the agent model adopted could be expressive enough to model in some way the execution of external actions and the perception of external events.

cognitive use — at this level, the agent designer directly embeds in the agent knowledge about what artifacts to use, but how to exploit the artifacts is dynamically discovered by the agent, reading the operating instructions. artifact selection is still a designer affair, while how to use it is delegated to the agent's rational capabilities. So, generally speaking the agent must be able to discover the artifact function, and the way to use it and to make it fit the agent goals. An obvious way to enable agent discovery is to make artifacts explicitly represent their function, interface, structure and behaviour.

cognitive selection and use — at this level, agents autonomously select artifacts to use, understand how to make them work, and then use them: as a result, both artifact selection and use are in the hands of the agents. It is worth noting that such a selection process could also concern sets of cooperative agents, for instance interested in using a coordination artifact for their social activities.

construction and manipulation — at this level, agents are lifted up to the role of designers of artifacts. Here, agents are supposed to understand how artifacts work, and how to adapt their behaviour (or to build new ones from scratch) in order to devise out a better course of actions toward the agent goals. Because of its complexity, this level more often concerns humans: however, not-so-complex agents can be adopted to change artifact behaviour according to some schema explicitly pre-defined by the agent designers.

4.2 A Model of Artifacts for MAS

In order to allow for its rational exploitation by intelligent agents, an artifact for MAS possibly exposes *(i)* a *usage interface*, *(ii) operating instructions*, and *(iii)* a *service description*. On the one hand, this view of artifacts provides us with a powerful key for the interpretation of the properties and features of existing non-agent MAS abstractions, which can be then catalogued and compared based on some common criteria. On the other hand, it is also meant to foster the conceptual grounding for a principled methodology for the engineering of MAS environment, where artifacts play the role of the core abstractions.

Usage Interface — One of the core differences between artifacts and agents, as computational entities populating a MAS, lays in the concept of *operation*, which is the means by which an artifact provides for a service or function. An agent executes an action over an artifact by invoking an artifact operation. Execution possibly terminates with an *operation completion*, typically representing the outcome of the invocation, which the agent comes to be aware of in terms of perception. The set of operations provided by an artifact defines what is called its *usage interface*, which (intentionally) resembles interfaces of services, components or objects, in the object-oriented sense of the term.
In MASs, this interaction schema is peculiar to artifacts, and makes them intrinsically different from agents. While an agent has no interface, acts and senses the environment, encapsulates its control, and brings about its goals proactively and autonomously, an artifact has instead a usage interface, is used by agents (and never the opposite), is driven by their control, and automatises a specific service in a predictable way without the blessing of autonomy. Hence, owning an interface strongly clearly differentiates agents and artifacts, and is therefore to be used by the MAS engineer as a basic discriminative property between them.

Operating Instructions — Coupled with a usage interface, an artifact could provide agents with *operating instructions*. Operating instructions are a description of the procedure an agent has to follow to meaningfully interact with an artifact over time. Most remarkably, one such description is history dependent, so that actions and perceptions occurring at a given time may influence the remainder of the interaction with the artifact. Therefore, operating instructions are basically seen as an exploitation protocol of actions / perceptions. This protocol is possibly furthermore annotated with information on the intended preconditions and effects on the agent mental state, which a rational agent should read and exploit to give a meaning to operating instructions. Artifacts being conceptually similar to devices used by humans, operation instructions play a role similar to a manual, which a human reads to know how to use the device on a step-by-step basis, and depending on the expected outcomes he/she needs to achieve. For instance, a digital camera provides buttons and panels (representing its usage interface), and therefore comes with a manual describing how to use them, e.g. which sequence of buttons are to be pushed to suitably configure the camera resolution.

Function Description — Finally, an artifact could be characterised by a *function description* (or service description). This is a description of the functionality provided by the artifact, which agents can use essentially for artifact selection. In fact, differently from operating instructions, which describe *how* to exploit an artifact, function description describes *what* to obtain from an artifact. Clearly, function description is an abstraction over the actual implementation of the artifact: it hides inessential details over the implementation of the service while highlighting key functional (input/output) aspects of it, to be used by agents for artifact selection. For instance, when modelling a sensor wrapper as an artifact, we may easily think of the operations for sensor activation and inspection as described via usage interface and operations instructions, while the information about the sensory function itself being conveyed through function description of the sensor wrapper.

Besides this model, some basic properties and features can be identified for artifacts, which possibly enhance agent ability to use them for their own purposes:

Inspectability — The state of an artifact, its content (whatever this means in a specific artifact), its usage interface, operating instructions and function description might be all or partially available to agents through *inspectability*. Whereas in closed MASs this information could be hard-coded in the agent - the artifact engineer develops the agents as well - in open MASs third-party agents should be able to dynamically join a society and get aware at run-time of the necessary information about the available artifacts. Also, artifacts are often in charge of critical MAS behaviour [17]: being able to inspect a part or the whole of an artifact features and state is likely to be a fundamental capability in order to understand and govern the dynamics and behaviour of a MAS.

Controllability — Controllability is an obvious extension of the inspectability property. The operational behaviour of an artifact should then not be merely inspectable, but also controllable so as to allow engineers (or even intelligent agents) to monitor its proper functioning: it should be possible to stop and restart an artifact working cycle, to trace its inner activity, and to observe and control a step-by-step execution. In principle, this would largely improve the ability of monitoring, analysing and debugging at execution time the operational behaviour of an artifact, and of the associated MAS social activities as well.

Malleability — Also related to inspectability, malleability (also called *forgeability*) is a key-feature in dynamic MAS scenarios, when the behaviour of artifacts could be required to be modified dynamically in order to adapt to the changing needs or mutable external conditions of a MAS. Malleability, as the ability to change the artifact behaviour at execution-time, is seemingly a crucial aspect in on-line engineering for MASs, and also a perspective key issue for self-organising MASs.

Predictability — Differently from agents, which as autonomous entities have the freedom of behaving erratically, e.g. neglecting messages, usage interface, operating instructions and function description can be used as a contract with an artifact by an agent. In particular, function description can provide precise details of the outcomes of exploiting the artifact, while operating instructions make the behaviour of an artifact predictable for an agent.

Formalisability — The predictability feature can be easily related with *formalisability*. Due to the precise characterisation that can be given to an artifact's behaviour, until reaching a full operational semantics model, for instance, as developed for coordination artifacts in [20], it might be feasible to automatically verify the properties and behaviour of the services provided by artifacts, for this is intrinsically easier than services provided by autonomous agents.

Linkability — artifacts can be used to encapsulate and model reusable services in a MAS. To scale up with complexity of an environment, it might be interesting to compose artifacts, e.g. to build a service incrementally on top of another, by making a new artifact realising its service by interacting with an existing artifact. To this end, artifacts should be able to invoke the operation of another artifact: the reply to that invocation will be transmitted by the receiver through the invocation of another operation in the sender.

Distribution — Differently from an agent, which is typically seen as a point-like abstraction conceptually located to a single node of the newtwork, artifacts can also be distributed. In particular, a single artifact can in principle be used to model a distributed service, accessible from more nodes of the net. Using linkability, a distributed artifact can then be conceived and implemented as a composition of linked, possibly non-distributed artifacts, or vice versa, a number of linked artifacts, scattered through a number of different physical locations could be altogether seen as a single distributed artifact. Altogether, distribution and linkability promote the layering of artifact engineering, as sketched in Section 4.3.

As a final remark, it should be noted that all the artifact features presented above play a different role when seen from the different viewpoints of agents and of MAS engineers. For instance, operating instructions are mostly to be seen as a design tool for engineers, as well as a run-time support for rational agents. Instead, features like inspectability and malleability gain particular interest when the two viewpoints can be made one: when an intelligent agent is allowed to play and is capable of playing the role of the MAS engineer, it can in principle understand the state and dynamics of the MAS by observing the artifacts, then possibly working as an Agens Faber: that is, by re-working its tools (the artifacts) in order to suitably change the overall MAS behaviour.

4.3 A Basic Taxonomy of Artifacts for MAS

Many sorts of different artifacts can populate a MAS, providing agents with a number of different services, embodying a variety of diverse models, technologies and tools, and addressing a wide range of application issues. Correspondingly, a huge variety of approaches and solutions are in principle available for MAS engineers when working to shape the agent environment according to their application needs. So, the mere model of artifacts for MAS is no longer enough: a *taxonomy* of artifacts comes to be useful, which could help MAS engineers first defining the basic classes of artifacts, their differences and peculiarities, then classifying known artifacts, to understand and possibly compare them.

Among the many possible criteria for a taxonomy, we find it useful to focus on the mediation role of the artifact, and then discriminate artifacts based on the sort of (non-artifact) MAS entities they are meant to tie together. According to the pictorial representation in Fig. 1, our first proposal here divides artifacts into *individual artifacts*, *social artifacts*, and *resource artifacts*.

Individual artifacts are artifacts exploited by one agent only, in other terms, an individual artifact mediates between an individual agent and the environment. Individual artifacts can serve several purposes, including externally enhancing agent capabilities, such as e.g. adding a private external memory, enacting a filtering policy of the agent actions toward other artifacts (as in the case of agent coordination contexts [18]), providing individual agents with useful information on the organisation, and so on. In general, individual artifacts are not directly affected by the activity of other agents, but can, through linkability, interact with other artifacts in the MAS.

Social artifacts are instead artifacts exploited by more than one agent. In other terms, a social artifact mediates between two or more agents in a MAS. In general, social artifacts typically provide a service which is in the first place meant to achieve a social goal of the MAS, rather than an individual agent goal. For instance, social artifacts might provide a coordination service [25], governing the activities of two or more agents, as for example in multi-party protocols, but can also realise global knowledge repositories, shared ontologies, or organisation abstractions containing information on roles and permissions.

Finally, resource artifacts are artifacts that conceptually wrap external resources, in other terms, a resource artifact mediates between a MAS and an external resource. External resources can be either legacy components and tools, applications written with non-agent technologies because of engineering convenience, such as Web Services, or physical resources which the agents of a MAS might need to act upon and sense. In principle, resource artifacts can be conceived as a means to raise external MAS resources up to the agent cognitive level. In fact, they provide external resources with an usage interface, some operating instructions, and a service description, and realise their task by dynamically mapping high-level agent interactions upon lower-level interactions with the resources, using e.g. specific transports such as object-oriented local or remote method calls, HTTP requests, and the like.

Altogether, individual, social and resource artifacts can be used as the basis for building the glue keeping agents together in a MAS, and for structuring the environment where agents live and interact. In fact, our taxonomy, as apparent from Fig. 1, defines a structured, *layered* view over the MAS environment, and implicitly suggests a model for organising agent interaction within a MAS. As such, the artifact

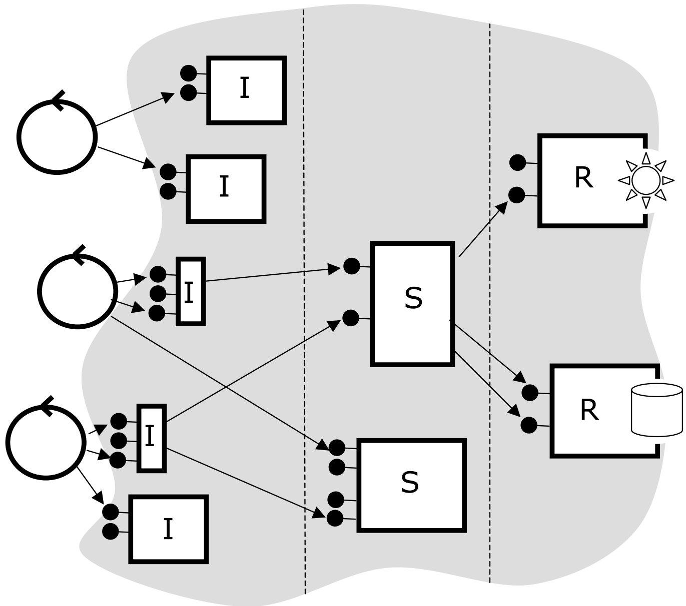

Fig. 1 Individual, social, and resource artifacts: a layered view of artifacts for MAS, from [19].

taxonomy could lead to a well-principled foundation for a general agent-oriented methodology for the engineering of the agent environment as a first-class entity.

5 Conclusion

By drawing an analogy between intelligence in MASs and the development and evolution of human intelligence, in this chapter we elaborated on MAS environment, agent tools, and their relationship with agent intelligence. Our metaphor of the Agens Faber comes to say that a theory of agent intelligence should not be limited to modelling the inner rational process of an agent, but should instead include not only the basics of practical reasoning, but also a theory of the agent *artifacts*, as defined in the A&A meta-model for MAS, providing agents with the means for the rational use, selection, construction, and manipulation of artifacts.

References

1. P.E. Agre and I. Horswill. Lifeworld analysis. *Journal of AI Research*, 6(1):111–145, 1997.
2. Robert St. Amant and Alexander B. Wood. Tool use for autonomous agents. In *Proceedings of the National Conference on Artificial Intelligence (AAAI-05)*, 2005.

3. Benjamin B. Beck. Animal tool behavior. the use and manufacture of tools by animals. *The Quarterly Review of Biology*, 56 (2):231, 1981.
4. Kristina Berggren. Homo Faber or Homo Symbolicus? The fascination with copper in the sixth millennium. *Transoxiana: Journal Libre de Estudios Orientales*, 8, June 2004.
5. Andy Clark and Dave Chalmers. The extended mind. *Analysis*, 58: 1:7–19, 1998.
6. Rosaria Conte and Cristiano Castelfranchi, editors. *Cognitive and Social Action*. University College London, 1995.
7. Kathleen R. Gibson and Tim Ingold. *Tools, Language and Cognition in Human Evolution*. Cambridge University Press, 1995.
8. G. W. Hewes. *Tools, Language and Cognition in Human Evolution*, chapter A history of speculation on the relation between tools and languages, pages 20–31. Cambridge University Press, 1993.
9. Edward Hutchins. *Cognition in the Wild*. MIT Press, 1995.
10. David Kirsh. The intelligent use of space. *Artificial Intelligence*, 73(1-2):31–68, 1995.
11. A. N. Leont'ev. *The making of mind: A personal account of Soviet psychology*. Harvard University Press, 1979.
12. A. N. Leont'ev. *Problems of the development of mind*. Progress Publishers, 1981.
13. Pattie Maes. Agents that reduce work and information overload. *Commun. ACM*, 37(7):30–40, 1994.
14. G. Martelet. *Evolution et Creation, tome 1*. Paris, 1998.
15. Allen Newell and Herbert Alexander Simon. *Human Problem Solving*. Prentice-Hall, Englewood Cliffs, NJ, USA, 1972.
16. Donald A. Norman. Cognitive artifacts. In John Carroll, editor, *Designing interaction: psychology at the human-computer interface*, Cambridge Series On Human-Computer Interaction archive, pages 17–38. Cambridge University Press, 1991.
17. Andrea Omicini, Sascha Ossowski, and Alessandro Ricci. Coordination infrastructures in the engineering of multiagent systems. In Federico Bergenti, Marie-Pierre Gleizes, and Franco Zambonelli, editors, *Methodologies and Software Engineering for Agent Systems: The Agent-Oriented Software Engineering Handbook*, chapter 14, pages 273–296. Kluwer Academic Publishers, June 2004.
18. Andrea Omicini, Alessandro Ricci, and Mirko Viroli. Formal specification and enactment of security policies through Agent Coordination Contexts. *Electronic Notes in Theoretical Computer Science*, 85(3), August 2003. 1st International Workshop "Security Issues in Coordination Models, Languages and Systems" (SecCo 2003), Eindhoven, The Netherlands, 28–29 June 2003. Proceedings.
19. Andrea Omicini, Alessandro Ricci, and Mirko Viroli. *Agens Faber*: Toward a theory of artefacts for MAS. *Electronic Notes in Theoretical Computer Sciences*, 150(3):21–36, 29 May 2006. 1st International Workshop "Coordination and Organization" (CoOrg 2005), COORDINATION 2005, Namur, Belgium, 22 April 2005. Proceedings.
20. Andrea Omicini, Alessandro Ricci, Mirko Viroli, Cristiano Castelfranchi, and Luca Tummolini. Coordination artifacts: Environment-based coordination for intelligent agents. In Nicholas R. Jennings, Carles Sierra, Liz Sonenberg, and Milind Tambe, editors, *3rd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2004)*, volume 1, pages 286–293, New York, NY, USA, 19–23 July 2004. ACM.
21. J. Povinelli, D. *Folk Physics for Apes: The Chimpanzees Theory of How the World Works*. Oxford University Press, 2000.
22. Alessandro Ricci, Mirko Viroli, and Andrea Omicini. "Give agents their artifacts": The A&A approach for engineering working environments in MAS. In Edmund Durfee, Makoto Yokoo, Michael Huhns, and Onn Shehory, editors, *6th International Joint Conference "Autonomous Agents & Multi-Agent Systems" (AAMAS 2007)*, pages 601–603, Honolulu, Hawai'i, USA, 14–18 May 2007. IFAAMAS.
23. Herbert Alexander Simon. *The Sciences of the Artificial*. MIT Press, Cambridge, Mass., 1981.
24. Kim Sterenly. *The Externalist Challenge: New Studies on Cognition and Intentionality*, chapter Externalism, Epistemic Artefacts and the Extended Mind. Berlin and New York: De Gruyter, 2003.

25. Mirko Viroli and Andrea Omicini. Coordination as a service. *Fundamenta Informaticae*, 73(4):507–534, 2006. Special Issue: Best papers of FOCLASA 2002.
26. L. S. Vygotsky. *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, November 1980.
27. J. V. Wertsch. *Vygotsky and the social formation of mind*. Harvard University Press, 1979.
28. Danny Weyns and H. Van Dyke Parunak, editors. *Journal of Autonomous Agents and Multi-Agent Systems. Special Issue: Environment for Multi-Agent Systems*, volume 14(1). Springer Netherlands, 2007.
29. Wilhelm Wundt. *Elements of folk psychology: Outlines of a psychological history of the development of mankind*. Allen and Unwin, 1916.