Coordination and Trust in MAS Towards Intelligent Socio-technical Systems

Andrea Omicini
Alma Mater Studiorum—Università di Bologna
andrea.omicini@unibo.it

Alessandro Ricci
Alma Mater Studiorum—Università di Bologna
a.ricci@unibo.it

Abstract

The discipline of intelligent systems engineering is slowly moving towards its presumable final destination – that is, agent-oriented software engineering –, whereas multi-agent systems emerge as the apparent source for abstractions, technologies, and methods for intelligent socio-technical systems. There, fundamental notions such as *trust* and *coordination* play a key role in the modelling and engineering of multi-agent systems: yet, even though they both insist on the same conceptual space – the space of agent and MAS *interaction* – the intrinsic nature of their relationship has not been explored deeply enough till now.

We try and provide a brief look at the two concepts, as they developed in the literature in the last decades, then focus on some possible dimensions around which their understanding and definition could possibly be framed in a comprehensive and coherent way; finally, we discuss a case study of a blockchain-based infrastructure providing for *trustworthy coordination*.

This paper was supported by the "ENGINES – ENGineering INtElligent Systems around intelligent agent technologies" project funded by the Italian MUR program "PRIN 2022" under grant number 20229ZXBZM.

1 Introduction

When looking back at the many works on both *trust* and *coordination* in the field of multi-agent systems (MAS), where the two notions mostly dwell and strive, clarity might not be the first thing to come to mind. On the one hand, largely diverse understanding of both terms have been widely used in the technical literature over the decades, in spite of the many efforts – e.g., [1, 2] for trust, [3, 4] for coordination, among the many others – to provide a common general definition, upon which computer scientists and engineers could rely for the modelling and design of their (intelligent) systems. On the other hand, both notions are fundamental, even essential, for the engineering of intelligent systems, in particular when *socio-technical system* (STS) henceforth) are concerned [5].

Three things, however, clearly emerge from a preliminary cursory examination of the literature. First, both trust and coordination are first of all non-technical terms, with a widespread acceptation, commonly used with relevant meaning in the context of human social systems. Then, they both have operational acceptations in the MAS field, which are essential not just for agent-based systems, but more generally for intelligent STS. Finally, whatever acceptations of the terms one might adopt, they are largely related notions, since they both deal with the dimension of *interaction* within MAS.

Accordingly, in the following we explore some of the intersections between the notions of trust and coordination in MAS and STS, pointing out some fundamental dimensions that could help illustrate their mutual strict relationship.

2 Trust and Coordination

2.1 The Many Ways of Coordination

Many are the viewpoints one could use to look at the issue of coordination, even when focusing on artificial systems only. One of the most general definitions originates from organisational science, where coordination is defined as the *management of dependencies between* *organisational activities* [6]: for instance, one workflow in an organisation may involve a secretary writing a letter, an official signing it, and another employee sending it to its final destination—each activity depending on each other for the success of the whole coordinated process. This notion straightforward generalises to coordination in MAS, where *agents* are the subjects whose activities need to be coordinated – the *coordinables* in [7] –, encapsulating the entities for which dependencies arise—like goals, actions, and plans. According to [8], the process of coordination involves two major tasks: *(i)* a detection of dependencies, *(ii)*, a decision about which coordination action to be taken. A coordination *mechanism* shapes the way that agents perform these tasks [9].

Shifting from a *modelling* to a *design* perspective, coordination is probably best conceived as the endeavour of *managing the space of interaction* [10] within a MAS. The design stance on coordination basically amounts at translating dependency detection and decision tasks into suitable interaction patterns. Diverse approaches such as planning, negotiation, norms, reputation management, and the like, aims at suitably *shaping* the space of interaction within a MAS by either *directly* making assumptions on agent behaviours or by *indirectly* modifying the environment around agents themselves.

Looking at the MAS interaction space, two distinct viewpoints can be taken: the individual agent one, and the external observer one [11]. The first perspective is what is called *subjective viewpoint* over interaction, where an agent looks at the MAS overall dynamics (other agents and environment), aiming at understanding and predicting the relevant behaviours there, so as to be able to coordinate its own behaviour with the rest of the system towards the achievement of its own goals. The second perspective is what is called *objective viewpoint*, where the space of interaction is basically seen as composed by the observable behaviour of all MAS agents and environment, as well as by all their interaction histories [12]. The objective viewpoint is typically identified as the *engineering perspective*, adopted by MAS designers to map out both individual and social interaction protocols and patterns so as to drive the overall system behaviour towards the achievement of the overall system goal. As a simple example of subjective vs. objective coordination, it is enough to consider the simple setting of the well-known *blocks world* [13] with many agent involved, as described in [14]. There, the subjective viewpoint of each agent pushes each of them at devising out the individual plan that would most likely allow the agent to reach its own goal, whatever other agents would do; whereas, the objective viewpoint favours the multi-plan that would most likely drive the MAS towards the achievement of its global goal—for instance, by maximising the overall outcome for all planning agents in the MAS.

In particular, the ability to act on the MAS interaction space without affecting agents is a premise to any form of objective coordination. Thus, objective coordination first of all deals with MAS *environment*, and depends on the availability of suitable models of the agent environment, as well as on their proper embodiment within MAS *infrastructures* [15]. There, objective coordination takes the form of *coordination abstractions*, provided as run-time *coordination services* by the agent infrastructure [16].

2.2 The Many Facets of Trust

The common acceptation of the term *trust* is basically too wide to allow for a definition that could be at the same time comprehensive at the conceptual level and operable at the technical level. In [17], Castelfranchi e Falcone design a general framework accounting for the layered notion of trust, encompassing most (if not all) the many different acceptations of the term proposed and adopted in the literature. Among the many other results, the authors point out how "it is crucial to distinguish at least between two kinds and meanings of trust: (a) trust as psychological attitude of X towards Y relative to some possible desirable behavior or feature, (b) trust as the decision and the act of relying on, counting on, depending on Y." There, it is quite clear that a fundamental distinction has to be made between the inner, individual notion of agent trust, and its emergence in terms of the observable agent behaviour.

Even though, starting from [1], many definitions of trust mostly refer to the psychological viewpoint – which is also the nearest to the "folk acceptation" of the term –, a number of acceptations of "trust" either explicitly or implicitly refer to a "social" or "infrastructural" interpretation of the term. For instance, trust is quite a typically relevant notion in the security area. There, notions such as *trusted computing* [18] or *trusted third parties* [19] mostly refer to a general notion of *trusted infrastructure* or *trusted platform*, basically deferring trust to some external technological element (either hardware or software) rather than to software components (like agents).

More generally, while referring the reader to Chapter 12 of [17] for a comprehensive discussion of the many articulations of the relationship between trust and technology, the shared space where trust meets the technological infrastructure is especially relevant in the context of MAS, in particular in the field of electronic institutions [20], or *e-institutions*. There, the observable behaviours of autonomous agents are constrained by the rules enforced by the e-institution around which the global MAS behaviour is designed, and which is the implicit *locus* of trust for individual agents and MAS designers. Other approaches exist that deal with open MAS via non-centralised techniques, such as [21].

3 Trust vs. Coordination

Whereas almost any class or course on MAS or STS typically quotes trust and coordination as two fundamental and strictly related concepts, typically placing them in the areas of collaborative / cooperative approaches or *agreement technologies* [22], the number of scientific and technical works focussing specifically on both of them, and on their relationship, is relatively low. Works focussing on trust typically mention coordination as a sort of post-product of some form of trust between software or human components in a STS, whereas coordination approaches dealing with trust mostly deal with trust as a critical prerequisite of any agreement process [23]. The fact is that both notions are so wide and comprehensive that any attempt to systematically define their mutual relationship is doomed to failure when attempting to capture any sort of meaningful generality. On the other hand, a more fruitful approach could be undertook by drawing some significant parallels between the two notions, or, by devising out some shared criteria for their understanding—which is what we will do in the remainder of this section.

3.1 Dependence

Interdependency and coordination have been lasting topics in organisational research—and that has brought about the most widespread definition of coordination in MAS and STS as the the *management of dependencies* [6]. So, as we mentioned above, devising out the dependencies between agents – in terms of tasks, goals, actions, ... – is the conceptual and technical pre-condition for the coordination of both human and software agents.

In the dichotomy between modelling and design stances, the individual agent perception of dependence w.r.t. other agents is the premise to subjective coordination, whereas the overall network of dependencies among agents, and between individual agents and the whole system, accounts for the dimension of objective coordination .

The notion of dependence is clearly essential for trust, too. The notion of *dependence* and *trust networks* are discussed in [24] as the conceptual and technical representation of social dependence, whereas the role of dependence networks in the interaction between agents, with particular interest in the relationship between dependence and trust, are investigated in [25] There, different forms of *objective* and *subjective* dependence are defined and formalised, then used as the grounding for agent reasoning – in particular, in terms of negotiation power –, and for establishing operable trust relationships between agents. Even though coordination is not even mentioned in neither [24] nor [25], the same notions of dependence and trust networks could be easily seen as technical foundations for the whole coordination discourse in both MAS and STS.

Overall, the concept of dependence is at the very core of both the notions of trust and coordination. So, when dealing with the engineering of intelligent STS built out of autonomous components, modelling dependencies – e.g., in terms of dependence networks [25] – is definitely a necessary prerequisite for the management of both trust and coordination.[1]

3.2 Subjective vs. Objective Trust

As we mentioned above, the distinction between the subjective and the objective viewpoint is essential for understanding the many ways in which coordination articulates.[11]. On the one side, the subjective viewpoint of an autonomous agent takes on the system interaction space from the individual agent perspective so as to coordinate its own actions towards the achievement of its own goals—so it deals with the individual aspects of agent autonomy and intelligence. On the other side, the objective viewpoint oversees every interaction within a system, including both agent-to-agent and agent-to-environment interactions, and relates them to the global system behaviour—so it deals with the social aspects of system autonomy and intelligence through the system infrastructure (e.g., through the *coordination middleware* [15].

When adopting the design stance over intelligent STS, this distinction basically amounts at separating concerns: in short, when designing intelligent autonomous agents, engineers should adopt the individual subjective viewpoint over coordination; when designing intelligent STS overall, engineers should instead shift towards the global objective viewpoint, so as to shape the social interaction and the system infrastructure according to the system goal.

Whereas, to the best of our knowledge, a similar distinction has not explicitly emerged in the field of trust yet, it is however quite easy to set a line between the many works in the field mostly dealing with the modelling of trust within individual agents, and those focussing on the social or infrastructural aspects of trust. More specifically, most of

[1] In the literature, both *dependence* and *dependency* are widely adopted with no easy-to-discern rationale behind their use. In this paper, we choose to use the (uncountable) term "dependence" to label the conceptual relationship between distinct elements in a system, and "dependencies" to denote the diverse analytical forms that dependence can take within a system at either design or execution time.

the approaches dealing with the conceptualisation and formalisation of the notion of trust typically focus on subjective aspects—"X trusts Y" sorts of models. On the other hand, many approaches dealing with the technical and operable aspects of trust mostly focus on the technological infrastructure [18, 19], where trust-related issues can be faced at the system level.

It is then quite tempting to draw inspiration from coordination research, and observe that a suitably-shaped notion of *subjective trust* would easily help in modelling and engineering individual aspects of trust for autonomous agents, whereas a well-defined notion of *objective trust* would definitely provide a neat conceptual framework for modelling and engineering systemic aspects of trust within intelligent STS. Finally, this also suggests that middleware for MAS and STS should be designed so as to account for the objective viewpoint of both coordination and trust, handling social and environment issues in a coherent and comprehensive way.

4 Trustworthy Coordination: A Case Study

As a case study for our discussion, in the following we shortly illustrate the case of *trustworthy coordination*—that is, a coordination model implemented over a distributed infrastructure for trusted computing [18].

In the last years, *distributed ledger technologies* (DLT [26]) have emerged as the most reliable technology standard for trusted infrastructure for open distributed systems. In particular, *blockchain* technologies (BCT [26]) denote a heterogeneous number of technologies providing a novel way of dealing, essentially, with distributed asset tracking or identity management—whose most prominent use case is represented by cryptocurrencies. BCT provide secure decentralisation of both data and control in distributed systems, by effectively combining several results from cryptography, distributed consensus, game theory, and state machine replication. In particular, blockchain provides *untrusted parties* (such as autonomous agents in an open MAS) with a layer of trust enabling secure and trusted records along with secure and trusted transactions. Since trust, secured communications, and data consistency are critical issues for MAS engineering, BCT have gained traction as a straightforward way for injecting MAS with those features. Overall, as a middleware for MAS and STS, BCT currently represent one of the most prominent examples of infrastructure for objective trust.

BCT are also appealing from the more specific perspective of coordination [27] since they provide MAS with highly-desirable properties such as total ordering of events, data consistency, accountability of actions, identity management, and fault tolerance. This is why in [28] *smart contracts* [29] are exploited to implement *tuple-based* coordination [30] over Ethereum BCT [31]. More generally, in [27] the authors take a general look at the mainstream BCT implementations, and show how the general pattern of *coordination as a service* [16] can be implemented over BCT. The proposed computational model for smart contracts is there implemented on a custom blockchain called Tenderfone [32] that embodies both the conceptual and technical requirements for both objective coordination and objective trust.

5 Conclusion

In this work we take a fast look at two of the key issues in the modelling and engineering of intelligent STS—that is, *trust* and *coordination*. We try and overcome the fuzziness of the current definitions of the two notions, by focussing on their essential role in the management of *interaction* in complex intelligent systems, and on the common issues of *dependence* and *objective vs. subjective* viewpoints. Finally, we illustrate a case study where a coordination service is built over a trusted BCT so as to implement a *trustworthy coordination* infrastructure.

In the end, we expect that when AI research will rightly shift its focus back towards *interaction* as a core issue for intelligent system, both trust and coordination will be easily acknowledged as two of the key aspects in the area of intelligent systems engineering. Then, some general frameworks accounting for them both in a coherent and comprehensive way will be needed—expressing dependencies between components in a systematic way, and consistently capturing both subjective and objective viewpoints over trust and coordination.

Along that line, for instance, the A&A meta-model [33] could work as the source for an coherent notion of *trust artefact* as the core abstraction for objective trust provided by a suitably-shaped infrastructure—in the same way as *coordination artefacts* [34] work for objective coordination, and on the same line of negotiation and argumentation artefacts for MAS [35]. We are confident that future research will shed some light on that subject.

References

[1] Diego Gambetta, editor. *Trust: Making and Breaking Cooperative Relations.* Basil Blackwell, 1988.

[2] Cristiano Castelfranchi and Rino Falcone. Principles of trust for MAS. cognitive anatomy, social importance, and quantification. In *Proceedings of the 3rd International Conference on Multi Agent Systems (ICMAS 1998)*, pages 72–79. IEEE Computer Society, 1998.

[3] George A. Papadopoulos and Farhad Arbab. Coordination models and languages. In Marvin V. Zelkowitz, editor, *The Engineering of Large Systems*, volume 46 of *Advances in Computers*, pages 329–400. Academic Press, 1998.

[4] Andrea Omicini, Franco Zambonelli, Matthias Klusch, and Robert Tolksdorf, editors. *Coordination of Internet Agents.* Springer Berlin Heidelberg, March 2001.

[5] Brian Whitworth. Socio-technical systems. In Claude Ghaou, editor, *Encyclopedia of Human Computer Interaction*, pages 533–541. IGI Global, 2006.

[6] Thomas W. Malone and Kevin Crowston. The interdisciplinary study of coordination. *ACM Computing Surveys*, 26(1):87–119, 1994.

[7] Paolo Ciancarini. Coordination models and languages as software integrators. *ACM Computing Surveys*, 28(2):300–302, June 1996.

[8] Frank von Martial. *Co-ordinating Plans of Autonomous Agents*, volume 661 of *Lecture Notes on Artificial Intelligence.* Springer-Verlag, Berlin, 1992.

[9] Sascha Ossowski. *Co-ordination in Artificial Agent Societies. Social Structures and Its Implications for Autonomous Problem-Solving Agents*, volume 1535 of *Lecture Notes in Artificial Intelligence*. Springer-Verlag, 1999.

[10] Nadia Busi, Paolo Ciancarini, Roberto Gorrieri, and Gianluigi Zavattaro. Coordination models: A guided tour. In Omicini et al. [4], chapter 1, pages 6–24.

[11] Michael Schumacher. *Objective Coordination in Multi-Agent System Engineering. Design and Implementation*, volume 2039 of *LNCS*. Springer, April 2001.

[12] Peter Wegner. Why interaction is more powerful than algorithms. *Communications of the ACM*, 40(5):80–91, May 1997.

[13] Earl D. Sacerdoti. *A Structure for Plans and Behavior*, volume 3 of *Artificial Intelligence Series*. Elsevier, 1977.

[14] Andrea Omicini and Sascha Ossowski. Objective versus subjective coordination in the engineering of agent systems. In Matthias Klusch, Sonia Bergamaschi, Peter Edwards, and Paolo Petta, editors, *Intelligent Information Agents: An AgentLink Perspective*, volume 2586 of *Lecture Notes in Computer Science*, pages 179–202. Springer Berlin Heidelberg, 2003.

[15] Andrea Omicini, Alessandro Ricci, Mirko Viroli, Marco Cioffi, and Giovanni Rimassa. Multi-agent infrastructures for objective and subjective coordination. *Applied Artificial Intelligence*, 18(9–10):815–831, October–December 2004. Special Issue: Best papers from EUMAS 2003: The 1st European Workshop on Multi-agent Systems.

[16] Mirko Viroli and Andrea Omicini. Coordination as a service. *Fundamenta Informaticae*, 73(4):507–534, 2006. Special Issue: Best papers of FOCLASA 2002.

[17] Cristiano Castelfranchi and Rino Falcone. *Trust Theory: A Socio-Cognitive and Computational Model.* Wiley Publishing, 1st edition, 2010.

[18] Chris Mitchell, editor. *Trusted Computing.* The Institution of Engineering and Technology (IET), November 2005.

[19] Peter J. Skevington and Timothy P. Hart. Trusted third parties in electronic commerce. *BT Technology Journal*, 15(2):39–44, April 1997.

[20] Pablo Noriega and Carles Sierra. Electronic institutions: Future trends and challenges. In Matthias Klusch, Sascha Ossowski, and Onn Shehory, editors, *Cooperative Information Agents VI*, pages 14–17, Berlin,

Heidelberg, 2002. Springer Berlin Heidelberg.

[21] Yosi Mass and Onn Shehory. Distributed trust in open multi-agent systems. In Rino Falcone, Munindar Singh, and Yao-Hua Tan, editors, *Trust in Cyber-societies. Integrating the Human and Artificial Perspectives*, volume 2246 of *Lecture Notes in Artificial Intelligence*, pages 159–174. Springer Berlin Heidelberg, 2001.

[22] Sascha Ossowski, editor. *Agreement Technologies*, volume 8 of *Law, Governance and Technology Series*. Springer Netherlands, 2012.

[23] Sascha Ossowski. Coordination in multi-agent systems: Towards a technology of agreement. In Ralph Bergmann, Gabriela Lindemann, Stefan Kirn, and Michal Pěchouček, editors, *Multiagent System Technologies*, volume 5244 of *Lecture Notes in Artificial Intelligence*, pages 2–12, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[24] Cristiano Castelfranchi and Rino Falcone. Grounding human machine interdependence through dependence and trust networks: Basic elements for extended sociality. *Frontiers in Physics*, 10:946095, 2022.

[25] Rino Falcone and Alessandro Sapienza. Dependence networks and trust in agents societies: Insights and practical implications. In Rino Falcone, Cristiano Castelfranchi, Alessandro Sapienza, and Filippo Cantucci, editors, *24th Workshop "From Objects to Agents" (WOA 2023)*, volume 3579, pages 106–122. CEUR-WS.org, November 2023.

[26] ISO/TC 307. Blockchain and distributed ledger technologies — vocabulary. ISO 22739:2020, International Organization for Standardization, 2020.

[27] Giovanni Ciatto, Stefano Mariani, Alfredo Maffi, and Andrea Omicini. Blockchain-based coordination: Assessing the expressive power of smart contracts. *Information*, 11(1):1–20, January 2020. Special Issue "Blockchain Technologies for Multi-Agent Systems".

[28] Giovanni Ciatto, Stefano Mariani, and Andrea Omicini. Blockchain for trustworthy coordination: A first study with Linda and Ethereum. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 696–703, December 2018.

[29] Loi Luu, Duc-Hiep Chu, Hrishi Olickel, Prateek Saxena, and Aquinas Hobor. Making smart contracts smarter. In *2016 ACM SIGSAC Conference on Computer and Communications Security (CCS'16)*, pages 254–269. ACM Press, 2016.

[30] Davide Rossi, Giacomo Cabri, and Enrico Denti. Tuple-based technologies for coordination. In Omicini et al. [4], chapter 4, pages 83–109.

[31] Gavin Wood. Ethereum: a secure decentralised generalised transaction ledger, 2014.

[32] Giovanni Ciatto, Alfredo Maffi, Stefano Mariani, and Andrea Omicini. Towards agent-oriented blockchains: Autonomous smart contracts. In Yves Demazeau, Eric Matson, Juan Manuel Corchado, and Fernando De la Prieta, editors, *Advances in Practical Applications of Survivable Agents and Multi-Agent Systems: The PAAMS Collection*, volume 11523 of *Lecture Notes in Computer Science*, pages 29–41. Springer International Publishing, June 2019.

[33] Andrea Omicini, Alessandro Ricci, and Mirko Viroli. Artifacts in the A&A meta-model for multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 17(3):432–456, December 2008. Special Issue on Foundations, Advanced Topics and Industrial Perspectives of Multi-Agent Systems.

[34] Andrea Omicini, Alessandro Ricci, and Mirko Viroli. Coordination artifacts as first-class abstractions for MAS engineering: State of the research. In Alessandro F. Garcia, Ricardo Choren, Carlos Lucena, Paolo Giorgini, Tom Holvoet, and Alexander Romanovsky, editors, *Software Engineering for Multi-Agent Systems IV: Research Issues and Practical Applications*, volume 3914 of *Lecture Notes in Computer Science*, pages 71–90. Springer Berlin Heidelberg, April 2006. Invited Paper.

[35] Enrico Oliva, Peter McBurney, Andrea Omicini, and Mirko Viroli. Argumentation and artifacts for negotiation support. *International Journal of Artificial Intelligence*, 4(S10):90–117, Spring 2010. Special Issue on Negotiation and Argumentation in Artificial Intelligence.