

A Metrics Framework for Quantifying Autonomy in Complex Systems

Christopher-Eyk Hrabia^(✉), Nils Masuch, and Sahin Albayrak

Faculty of Electrical Engineering and Computer Science, DAI-Labor,
Technische Universität Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany
{christopher-eyk.hrabia,nils.masuch,sahin.albayrak}@dai-labor.de

Abstract. Autonomous systems, often realized as multi-agent systems, are envisioned to deal with uncertain and dynamic environments. They are applied in dangerous situations, e.g. as rescue robots or to relieve humans from complex and tedious tasks like driving a car or infrastructure maintenance. But in order to further improve the technology a generic measurement and benchmarking of autonomy is required. Within this paper we present an improved understanding of autonomous systems. Based on this foundation we introduce our concept of a multi-dimensional autonomy metric framework that especially takes into account multi-system environments. Finally, our approach is illustrated by means of an example.

Keywords: Artificial intelligence · Autonomous systems · Autonomous robots · Metrics · Intelligent systems · Adaptive systems · Multi-agent systems

1 Introduction

Modern computer and robotic systems assist humans in almost every situation in workaday life. It is always the goal to simplify human life by delegating either unpleasant or dangerous tasks to artificial systems. These tasks are becoming more and more complex, facing uncertainty in their execution environment and require robust and flexible solutions. Examples range from robotic vacuum cleaners [29] to artificial personal assistants for task and time management [18] to military drones in combat.

Designing systems for uncertain environments is a challenging task. Especially because engineers are unable to foresee all conditions, interactions and influences a system will have to deal with in advance during specification and development. Sometimes there is even a general lack of information.

The term *autonomous systems* is widely used for systems that are able to deal with such situations. Many researchers are trying to improve concepts and technologies behind. As objectives of autonomous systems perfectly match to the ones of multi-agent systems the latter's paradigm is often used for realization.

In order to properly design and develop such systems we need to identify the important aspects and means to create measurable goals we can improve

in future. For an iterative advancement it is required to benchmark and measure these systems during the development process in their application context. Furthermore, selecting an appropriate system for a required autonomy level in a particular scenario requires a quantification of autonomy, too. In our opinion there is still a lack of a widely accepted concept of autonomous systems or robots that contains a detailed definition according to different aspects and a clear distinction to other system concepts. Moreover, the consequences of bringing several different systems together are unclear.

The remainder of the paper is structured as follows. In Sect. 2, we give an overview about the variety of definitions and understandings of the term *autonomous system* and take a deeper look on metrics that are being proposed to measure such systems. Subsequently, in Sect. 3, we classify autonomous systems in relation to other system types and present our own definition of an autonomous systems. A context-specific multi-dimensional metric to control and benchmark autonomous systems is proposed in Sect. 4. In order to apply our multi-dimensional metric framework Sect. 5 highlights necessary considerations, whereas Sect. 6 illustrates the application with an artificial example. Finally, in Sect. 7, we summarize our work and outline open issues and future steps we want to work on.

2 Related Work

In order to come up with a definition of autonomous systems we will give an overview about state-of-the-art definitions and further focus on the particular metrics on which they rely.

2.1 Autonomous Systems Definitions

A review over existing literature about autonomous systems and agents reveals that there still does not exist one major agreement on definition or understanding of a concept of autonomy in computer and software systems. The developed definitions are usually depending on a specific problem or application domain and therefore are only looking on the concept from a limited perspective.

A simple and common understanding of autonomous systems derives directly from the literal translation of the word autonomy, which has ancient Greek origin and means self-government. This interpretation is used by Castelfranchi [8], often extended with a general independence from other entities [12]. In [10] the authors mention unpredictability and goal-directedness as the main aspects of autonomous systems. Moreover, they define self-directedness as the general substance of autonomy. Luck et al. [16] are also considering self-directedness or self-government, but propose the ability of goal generation based on an inner motivation or drive as the key component of an autonomous system. Further, motivation is seen as a higher-level non-derivative component that characterizes the nature of agents. The definition of autonomous systems as entities generating own goals suits also the concept of self-government.

Close to the motivation of autonomous systems that are supposed to deal with unforeseen environmental conditions is the concept of adaptation. In [1] it is stated that autonomy and adaptability are interconnected and decision making requires adaptation to the environment. These ideas are in accordance with the visions of the promoters of autonomic computing [15] and organic computing [25] as well. They understand self-adaptability as major property for the realization of software that can manage itself at runtime. Furthermore, self-adaptability is considered as the foundation of other envisioned self-* properties.

A different approach is presented by Barber et al. [3]. They have identified three distinct types of intervention in an autonomous agent: modification of environment, influence over beliefs and intervention in the decision-making process. In their opinion only the two last types have to be considered for autonomy-altering. Hence, they focus on independence on decision-making to pursue an agent's goal as well as independence of control over its own belief state [3,4].

One major concern of autonomy concepts is the requirement for an adjustment of the granted independence in general or a specification of a context defined by other systems or humans. As a consequence, some work is focused on the so called "adjustable autonomy" [11]. This adjustment is always seen as an outer influence on autonomous systems, though reducing the independence of systems from other entities. A common idea is the usage of policies and rules for different autonomy related properties. Scerri et al. [23] and Tambe et al. [28] focus on the aforementioned decision making capabilities. Further extensions aggregate all restrictions in role models [31]. Bradshaw et al. [5] present a formal framework for the description of action policies. They differentiate in possible, available and obligated actions that are arranged on a prescriptive and descriptive layer, representing self-directedness and self-sufficiency. Another perspective on policies is given by Myers et al. [19] as a specification for the autonomy context. This also includes possibilities of human consultation from the system's perspective. Additionally, the importance of such kind of behaviour is considered in [17], too.

Some research focusing on independence has created concepts of different areas of autonomy. An explicit external perspective of autonomous systems in a complete relational structure with classifications for user, social-dependence, norm, environment and self-autonomy is presented by Carabelea et al. [6]. Moreover, the authors mention the existence of an inner layer of autonomy focusing on the decision making process. Another work also focuses on the relational structure of independence in a hierarchical holonic agent organisation and hence classifies it into different kinds of autonomy [24]. The existing classes are skill and resource, goal, representational, deontic, planning, income, exit and processing autonomy. Verhaegen et al. [30] distinguish between natural and artificial agents, but focus as well on the aspect of independence in different areas like norms, external stimuli and motivations.

The presence of different layers of autonomy in an autonomous system is discussed by Castelfranchi et al. [9] and Maheswaran et al. [17] and was as well mentioned by Carabelea et al. [6]. In the context of adjustable autonomy there

exists control from outside over a system itself and control from the perspective of the system on how or if it decides to transfer decision-making control to other entities [6]. Then again, Castelfranchi et al. [9] understand autonomy as a matter of power, with internal and external aspects. External aspects describe conditions for actions or resources, internal aspects define system abilities, skills and resources.

Further, most of the shown literature has a strong focus on single systems for their consideration of autonomy. Exceptions are made by Carabelea et al. [6], who show the concept of delegation of decision-making capabilities, as well as in the work of Scerri et al. [23] and Tambe et al. [28]. Moreover, the importance of dependence from other entities for the creation of a belief model is demonstrated [4], too.

At least, it seems there is an existing agreement, that autonomy needs to be evaluated and compared in a specific context or with respect to some goal, because a system could behave autonomously in one situation and non-autonomous in a different context [3, 5, 6, 27].

In summary, the presented literature shows that both self-government as well as self-directedness based on an inner motivation are important aspects of an autonomy concept. Further, a general independence of the system, especially for decision making, belief management, actions and resources accessibility has to be considered. In addition, the capability of adapting to the environment is strongly interconnected with autonomy. Moreover, we see that an autonomous system seems to have different layers, usually seen as inner and outer parts. Also, an autonomous system is always related to other systems; therefore the influence from other systems, as well as the delegation of the system itself needs to be examined. However, we see that all these aspects have not yet been combined into a single concept.

2.2 Autonomous Systems Metrics

Existing publications on autonomous systems come up with a variety of properties that can affect the degree of autonomy. In the following we subsume these aspects under different metrical categories of autonomy.

Interaction: most of the publications describe the degree of necessary interaction with or observation by external entities in order to fulfil the implicit task of a system as an important category for autonomy. The authors in [7] discuss the term of autonomy for spacecrafts and deduce that it depends on the tracking intensity and the amount of communication between vehicle and ground. [32] also states that autonomy mostly represents the ability to assign the system's goals without any or with only minimal external intervention. The authors in [3] are more precise and state that they do not see the principal interaction with the environment as autonomy-altering. In their opinion it is more important to look whether there are instances that can change the environment in a way that the system changes its behaviour, which would be an indirect influence and autonomy-limiting factor.

Permissions, Norms, Obligations: these aspects are a refinement of the interaction category. In [5] the authors state that autonomy largely depends on the attribute of self-directedness, which means that the system can decide without external influences. In general, the freedom of an autonomous system can either be limited by explicit restrictions (norms or obligations) or extended by permissions which extend a principal behavioural restriction.

Quality: the degree of quality represents a level-based scheme for defining autonomy. In [4] the authors do so by setting different levels on the system's belief autonomy, which can either be manipulated by external entities or by the quality of the system's perception.

Uncertainty: an autonomous system implicitly sets the expectation of being able to at least fulfil its intended goals. Since in most cases the environment is dynamic and sometimes also unpredictable, uncertainty is a category that autonomous systems have to deal with. As a consequence, it can be stated that the higher the level of uncertainty of the environment is the more autonomous the system (always under the assumption of fulfilling its goals) [32].

Technical Aspects: some of the publications come up with pragmatic measurements like the time of ignorance [7, 20], which represents the time a system can be ignored by an external observer while still acting productively. Technical software measures are used by Alonso et al. [2]. The authors measure their determined key attributes of autonomy, which are self-control, functional independence and evolution capability, based on static and dynamic code analysis with properties like complexity of pointers and references, number of variables describing the internal state and state update frequency. For example, the functional independence of a system is being measured by using an executive message ratio $EMR = 1 - \frac{ME}{MR}$. MR defines the total number of messages the software agent receives, whereas ME represents all messages the software agent is obliged to respond or react to (e.g. because they were sent by the user the agent represents). According to this definition the system is more autonomous the less instructions the agent becomes via messages.

In conclusion, a lot of aspects have been proposed so far which play a significant role in defining the system's autonomy. However, most of the publications try to focus on very few attributes to define a metric in a particular domain, whereas a comprehensive view that is considering all different aspects together cannot be found so far.

3 Definition and Classification

As the motivation of this paper implies, autonomous systems can be quite different according to the degree of autonomy. None of the systems that exist so far are fully autonomous, so many systems fulfil only parts of existing autonomy definitions and can still be interpreted as autonomous systems. In turn, this makes it quite difficult to define exact bounds that characterize an autonomous

system. For this reason, we will define different characteristics that are relevant for deciding whether we are dealing with a system that is autonomous or not.

There is not only the criteria of autonomy when talking of modern system structures. Many other aspects exist and often they overlap each other. In order to clarify our understanding of autonomous systems we will first try to set them in relation to other concepts. Afterwards, we will come up with a comprehensive discussion about autonomous systems themselves.

3.1 Classification

When talking about practical realization of Artificial Intelligence (AI) the term intelligent system is often being mentioned. Rudas et al. [21] define an intelligent system as a system that “*emulates some aspects of intelligence exhibited by nature. These include learning, adaptability, robustness across problem domains, improving efficiency (over time and/or space), information compression, extrapolated reasoning*”. So we can deduce that an intelligent system is an application of AI that is specialized to some sort of challenge and does not have to offer general intelligence. A subclass of intelligent systems are adaptive systems [22], which are able to react to changes of the environment that the developer cannot foresee completely at design time or are even able to find solutions to modified goals.

Autonomous systems in fact can also be intelligent and adaptive systems but do not necessarily have to be so. Automation systems are often classified as a subcategory of an autonomous system. An automation system realizes processes from start to end without human intervention with a clear focus on independence. However, usually the environment may not make changes, which the system developer could not foresee. An example for this would be a thermostat that fully automatically adapts the cooling or heating in order to reach a desired temperature. The state space in the system is clearly defined and it is only able to react on this range. Furthermore, the same input will lead to the same output. So we state that an automation system is autonomous but is usually not an intelligent system. Another important subcategory of autonomous systems are autonomic systems [13,15], whose focus is self-management, with the goal of configuring, healing, optimizing, and protecting itself in order to recover from failures or optimize for changed conditions.

In our opinion the real benefit and also challenge lies within autonomous systems that are simultaneously adaptive systems. In turn, adaptivity infers the necessity to learn from new experiences. These can be external motivations either to change the goals of the system or to find new or better processes to fulfil existing goals. Even more, an autonomous system, which is not considering adaptivity as a key feature of itself, is not able to solve real world problems in dynamic environments. For this reason we will focus on autonomous systems that are as well adaptive systems in the following sections.

3.2 Autonomous System

In the former sections we have argued that adaptability within dynamic environments is the main justification for autonomous systems. Further, we were able to distinguish between adaptive and autonomous systems, with the result that actually all purposeful autonomous systems are adaptive systems, too. Beyond that, an autonomous system from the engineer’s point of view does only make sense, if there still exists the possibility of having some kind of influence on the system. In addition, we think that complete autonomy cannot exist at all, because at least the innermost motivation of a system needs to be defined or controlled by something else. On the other hand an autonomous system in the real world is always interacting with other systems and entities, so it is necessary to consider relational aspects in a sufficient definition of autonomous systems. Thus, our definition of autonomous systems includes two layers, an adaptive layer, including all capabilities required to adapt to changing environment and a relational layer, which specifies delegation and dependence on other systems. This corresponds with the existing concepts of self-sufficiency as the adaptive layer and self-directedness as the relational layer [16]. As a consequence we propose a definition of autonomous systems as adaptive systems extended with relational aspects. Within the relational layer there exist two different aspects, the influence from outside and the delegation of duties to other systems (see Fig. 1).

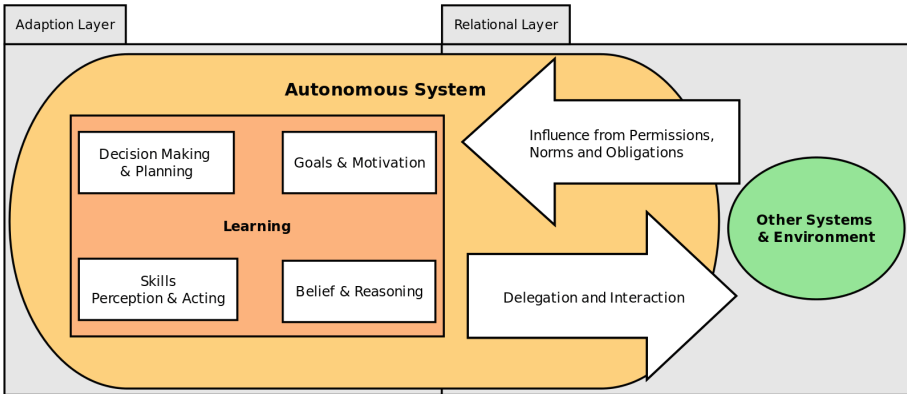


Fig. 1. Layers and capabilities of an autonomous system

The influence from outside corresponds to the popular concept of adjustable autonomy [5, 17, 19, 23, 28, 31], describing influences on the system independence from another system by restricting available resources and capabilities with the usage of policies, norms or obligations. In this sense a system’s autonomy is represented as a percentage of independence in the use of its available capabilities. The other relation direction is more complex, because it is not only a matter if a system delegates duties or not. It has to be considered if the system has just chosen to or if it was necessary to delegate, because it is not having the

required capabilities on its own. Another question is, whether the delegation to another system is reliable or not. For example, a system would lose autonomy if it delegates to another system, which it does not fully control, saying it has a reduced reliability. If it delegates to another system which is just a simple automata having 100% reliability in execution of the delegated task (not included in the reliability are external factors, which the other system cannot control), it would obtain its autonomy. The influence of other entities restricts the autonomy related capabilities of the system.

The inner layer of an autonomous system contains all capabilities that are required to adapt to the environment. Based on the literature, we came up with the following required capabilities: decision making, goal generation or motivation, belief generation and skills for perception and acting. In our opinion not all these capabilities are required to achieve sufficient adaptation. The very important aspect of decision making has to be extended by a more long-term consideration of planning. Belief generation depends on reasoning capabilities, thus it has to be considered as well. Finally, the capability of learning, which is mentioned in some of the discussions about autonomous systems, needs to be added, because in our opinion it is crucial for the whole system adaptability. In the following each of the capabilities is described in further detail.

The available *perception and acting skills* or actions to interact with the environment are important in terms of diversity. For example, a system with a less diverse set of moving capabilities (e.g., only walking besides crawling, driving or swimming) is probably less able to adapt to different ground surfaces, even if it does have a huge number of different leg moving styles. Thus, it is not just the quantity of available skills, in fact the diversity of skills is important.

Like already clarified from the literature, *decision making* is a core aspect of autonomy, but it is not only the resulting independence which makes it necessary. Rather it is the purpose of making the decision to adapt to the changing environment. The capability of planning is closely related, it allows the system to reach its goals factoring in environmental conditions. A system with more advanced planning and decision making features is able to consider changing conditions and as a consequence is able to adapt more quickly and appropriate.

Each system is following some *goals*. These goals are based on a motivation and are somehow entered initially from outside. The difference in the sense of autonomy and adaptability results from the level of abstraction in the goal description. The ability of creating new own goals, or to adapt existing goals is another consideration for improved adaptability.

The *belief state* of a system and its *reasoning capabilities* belong to each other. The system needs to reason from raw input information to create its beliefs. In its simplest form the system consists of reflexive mechanisms in which a rule set is triggered based on the perception. The resulting actions which modify the environment lead to a new perception and therefore a new belief state of the system. However, more complex reasoning capabilities allow for better view of the environment, which is the fundament of each adaptation mechanism.

Learning was already highlighted as a crucial capability for an autonomous and therefore adaptive system. The reason for this is the superior character of this capability, enabling the system to improve all other capabilities over time. The other presented capabilities perception and acting skills, decision making, interpretation and creation of goals and reasoning and belief creation make adaptation in a changing environment possible, even without learning, but learning will generate enhanced adaptability.

Combining everything our definition of an autonomous system is as follows and as well visualized in Fig. 1:

An autonomous system follows an innermost motivation in an uncertain and dynamic environment by adapting its capabilities in order to fulfil the inferred goals. It has capabilities of perception and acting, decision making and planning, interpretation and creation of goals, reasoning and belief creation for being able to adapt. Learning is superior to the other capabilities, but non mandatory. It can have relational dependence on other systems. If it is delegating its duties to them, its autonomy is depending on their reliability. Further, the relational dependence can lead to restrictions in the use of its capabilities and therefore on its autonomy.

4 Metric

In the last section an improved understanding of autonomous systems was created, which is useful to distinguish from other system concepts. Further, it clarifies the justification and the core motivation and it points out the core characteristics. These core characteristics give a direction for further improvements in making autonomous systems more autonomous in future. On the other hand, it was already shown, that it is not the actual goal to strive for an absolute 100% autonomous system, rather it is required to create autonomous systems with a clearly defined scope of autonomy and options for external control. For this reason we have developed a multi-dimensional metric for autonomous systems that considers the core capabilities of our understanding of autonomous systems. In order to direct development efforts this metric allows for an relative estimation of the system's development progress by comparing the different states or differences to other existing systems. Further, the metric indicates which capabilities have to be considered and ultimately controlled for external adjustment of the system's autonomy. For example, it points out which characteristics of an autonomous system can be controlled in which range by a human operator in order to enable system operation in the required borders. Similar to the consensus in literature this metric and resulting ratings do only make sense in a specified context.

Our metric introduces several scales based directly on the elaborated core capabilities, namely decision making and planning, goal generation and motivation, belief and reasoning, available skills and learning. Further, it takes into account the two layer concept with external influence and delegation from inside.

As shown in Fig. 1 other systems can have influence on the autonomy of a system. If a system contains one or more subsystems it is dependent on the

current scope. Other systems can restrict capabilities to some percentage based on permissions, obligations and norms. Likewise, a system can lose autonomy depending on the reliability of other systems if the system decides to delegate some of its capabilities. For example, another system could restrict the decision making capability in a way that it needs to consult another system on every decision related to its task execution order, resulting in a decreased autonomy of the system. Additionally, the system could delegate its reasoning to another system with perfect communication and 100% reliability, resulting in no degradation of autonomy (provided that the delegating instance always remains the control to withdraw the delegation). Further, similar to Johnson et al. [14], a measure of autonomy can never describe a system's performance, it only describes the capability of independent adaption to volatile environments, while striving for its goals. Moreover, it is important to point out that it is necessary to determine the weighting between the scales of the multi-dimensional metric in the observed context of the evaluated systems based on the presented concept. This would usually be done by a domain expert.

Besides the general influence by other systems affecting specific capabilities, metrics for particular capabilities and combinations of them are presented below.

4.1 Perception and Acting Skills

The system's skill set including different perception and acting capabilities influences the autonomy by defining the outer bounds of adaptability. If a system does not have the capabilities to retain operation in a given environment or if these capabilities are restricted, it has no chance to adapt. Comparing different system adaption opportunities, the most important aspect is diversity. The quality of the capability is not important, as long as it provides a sufficient quality to achieve the system's goals in general, because we do not want to measure the performance of the system. An ideal system needs a broad range of very different capabilities. For example, a robot with ultrasonic range finders and laser range scanners is able to adapt to environments with transparent surfaces or high frequency sound noise. Having only one of these capabilities would lead to problems in one of these environmental conditions.

As a consequence of this, a dimension for diversity is required. A sufficient option is the Shannon-Index, a well known approach from information theory [26]. It is a quantitative measurement of the number of different available types in relation to the evenness of distribution. A higher number of types together with an even distribution among the skills has the highest diversity. A type refers to a group of sensors or actors providing the same kind of information or realising similar actions. The calculation is shown in Eq. 1, where p_i represents the proportion of capability belonging to the i th type of N possible capabilities in the particular context, with n_i belonging to a particular type of capabilities. H' is the diversity index, a higher number corresponds to a higher diversity.

$$H' = - \sum_{i=1}^n p_i \cdot \ln(p_i) \quad \text{where} \quad p_i = \frac{n_i}{N} \quad (1)$$

The classification of types has to be adjusted in consideration of the context. The Shannon-Index can be applied for both the perception and the acting skills leading to the functions $H'(P_{Norm})$ and $H'(A_{Norm})$, which are normalized values based on Eq. 9. The total autonomy degree for these capabilities can then be defined as:

$$PA_{SCORE} = w_1 \cdot H'_{P_{Norm}} + w_2 \cdot H'_{A_{Norm}} \quad (2)$$

Both values are weighted according to the context by using w_1 and w_2 with $w_1 + w_2 = 1.0$.

4.2 Belief and Reasoning

The belief describes the state of the environment from the system's point of view which might be imperfect. For that reason it is critical to evaluate the environmental conditions to adapt sufficiently. In order to compare the quality of the belief and reasoning capabilities, different attributes have to be measured: the amount of information the system is able to reason from and the update rate of its belief generation.

The applied measurement methods and units have to be specified for the evaluation context, e.g. number of reasoning input sources multiplied with bandwidth, maximum storage complexity of the belief state and update rate per second.

We propose the following metric as a generic measurement to evaluate the system's belief autonomy which in turn affects the overall autonomy:

$$BR_{SCORE} = w_1 \cdot BIA_{Norm} + w_2 \cdot BUR_{Norm} \quad (3)$$

The belief and reasoning autonomy BR_{SCORE} depends on the amount of belief information BIA (amount of processed perception data) and the belief update rate BUR (frequency of refreshing the belief state) on these information. As it is necessary to normalize these values in order to combine both parameters we propose to utilize BIA_{Norm} and BUR_{Norm} , which are computed using Eq. 9. Both parameters are aggregated using weights w_1 and w_2 with $w_1 + w_2 = 1.0$.

4.3 Learning

In this context the focus of learning is the ability of long-term improvement of single capabilities with the goal of improving the system's adaptability. This is very difficult to measure, because the advantage of learning could only be determined in a direct comparison. Because learning is strongly related to reasoning the proposed measurements can be applied in a similar manner. Especially the amount of information to reason from is relevant for learning in the context of

adaptability. By storing historical data the system is able to learn e.g. typical behaviour patterns of the environment. This consideration needs to be distinguished from machine learning performance metrics like precision, recall and accuracy.

$$L_{SCORE} = w_1 \cdot LIA_{Norm} + w_2 \cdot LUR_{Norm} \quad (4)$$

The L_{SCORE} for measuring the learning capability is similar to the BR_{SCORE} in Sect. 4.2. Thus, the single measurement values need to be normalized for the combination as well, see Sect. 4.5. The weights w_1 and w_2 are limited to $w_1 + w_2 = 1.0$, too. In distinction the L_{SCORE} considers the stored historic information used for learning LIA (Learning Information Amount) and the update rate of the possible temporal repeating of the learning method LUR (Learning Update Rate). A temporal repeated learning update on the updated LIA is crucial for an autonomous system for keeping track with the volatile environment.

4.4 Motivation, Goals, Planning and Decision Making

Goals, planning and decision making are fundamental attributes of intelligent systems. Without them a system would be completely static and could therefore not adapt to changing situations.

A system has more possibilities to adapt and is less influenced if it can cope with high level goal descriptions, because in this case it has more degrees of freedom on how to achieve its goals and adapt to changing environments. This is also valid for motivation, which we understand as a very general description of a goal. The ability of decomposing goals into sub-goals and atomic tasks is part of planning. Decision making is the selection of computed plans or alternatives during planning. This in turn leads to the statement that the quality of goal generation, planning and decision making affects the degree of autonomy.

Hence, from the inner perspective of the system these attributes depend heavily on each other. It is not possible to determine if a particular decision fosters adaptability on its own, as long as the system is being able to decide in general. This means, that decision making is only an on/off attribute for the adaption layer and can hardly be quantified more precisely. Thus, decision making is more important for the relational layer of our metric concept, which defines the relation to other systems in terms of independence and reliability. Due to that the adaption layer has to focus on the planning ability of creating a wide range of behaviour possibilities by using task decomposition. Therefore, the aim is to measure the ability of decomposing tasks or goals in as many atomic actions as possible or measuring the level of abstraction in goal descriptions the system is able to understand.

A sufficient measure could determine the mean number of atomic system actions resulting from a given goal. For example, a less adaptable system like a 100% remote controlled robot has already atomic actions, like “move forward” and “turn right”, in its goal description. On the contrary a robot that receives the destination position as a goal can decompose the goal in different ways, enabling the choice of e.g. way points, velocity and locomotion style.

Formally, a goal G can be decomposed into a set of predicates $Pred$ and Tasks T as shown in Eq. 5. A predicate, which is in fact a subgoal, again consists of a set of elements that are either predicates or tasks (Eq. 6).

$$G = \{Pred \cup T\} \quad (5)$$

$$Pred = \{x_1 \dots x_n \mid x_i \in Pred \cup T\} \quad (6)$$

Because of its recursive structure, each goal G can finally be decomposed into a set of atomic tasks. Therefore, $t \in T$ is defined as a task which is atomic and consequently not decomposable. T_G is defined as the amount of all atomic tasks that – in some combination – help fulfilling G .

$$T_G = |\{t \mid t \in T, t \xrightarrow{\text{partiallyFulfills}} G\}| \quad (7)$$

Equation 8 shows our suggestion of defining the goal and planning autonomy GP_{SCORE} .

The term describes the average fraction between the tasks the system was able to decompose out of the goal (DT_G) and all possible tasks related to a goal for each potential step i . If DT_G contains the same tasks as T_G , then the system has a complete view on the possible task sets for reaching G . In consequence this part reflects the decomposing and planning ability of the system.

$$GP_{SCORE} = \frac{\overline{DT_{G_i}}}{T_{G_i}} \quad (8)$$

In some cases the goal of a system might change over time. The GP_{SCORE} will then be dynamic, which would have to be considered in the evaluation of the system. For instance this can be achieved by aggregating measurements at each state where the system's goal changed at runtime.

4.5 Scaling and Aggregation of the Capabilities

The capability scores together with relational characteristics result in a set of measurements. C_i represents one of the capability scores. In order to interpret and evaluate the results the capability scores have to be scaled and aggregated. Therefore it is required that for particular measurements a greater value corresponds to extended autonomy support and all values are greater than 0. If necessary the measurements have to be quantified and rescaled.

Further, all measurements need to be normalized for being able to compare them with each other, as well as with other systems. We applied unity normalization, shown in Eq. 9. The same normalization approach is used for the normalization of the multiple measurements as part of the capability scores of perception and acting, belief and reasoning and learning. All normalizations are applied before the weighting.

$$C_{\text{Norm}_i} = \frac{C_i - \min(C_i)}{\max(C_i) - \min(C_i)} \quad (9)$$

with C_i as a single capability measurement, $C_i > 0$ and a higher value corresponding to extended autonomy support.

One important aspect is the definition of the maximum and minimum values for these measurements. If several systems or system states are compared in a given context, maximum and minimum are at least defined by the measurement result range of the evaluated systems. If only a single system is evaluated the bounds have to be defined based on domain and system knowledge, planned development roadmaps or envisioned future upgrades. These definitions can also be used to extend the range given by different system measurements in case of a system comparison in order to extend the considered context.

In a next step each capability score C_i is combined with its delegation reliability R_i and restriction influence I_i , as well as a general weighting factor w_i , see Eq. 10. R_i and I_i express the dependence on a percentage basis. If there is no information about restricting influences, we define $R_i = 1.0$ and $I_i = 0.0$ corresponding to neutral values. If the situation is unknown or too complex to estimate an approximation should be used. The weighting factor w_i can be adjusted according to the context and problem domain, but needs to be the same for a comparison amongst several systems. The default configuration weights all capabilities equally within a score, as well as between the scores.

$$C_{\text{Score}_i} = C_{\text{Norm}_i} \cdot w_i \cdot R_i \cdot (1 - I_i) \quad (10)$$

with $0 < C_i, w_i, R_i, I_i < 1$ and $\sum_{i=1}^n w_i = 1.0$

At the end all capability scores C_{Score_i} can be aggregated into a single autonomy score A_{Score} , shown in Eq. 11. This autonomy score with a range from 0..1 allows for an absolute context specific comparison of the considered systems or system configurations. Nevertheless, a detailed comparison considering all capability scores should be preferred most of the time.

$$A_{\text{Score}} = \sum_{i=1}^n C_{\text{Score}_i} \quad (11)$$

5 Metric Application

In the following we discuss required decisions to be made and points to consider when utilizing our metric framework approach.

Context Selection. First, the context of the evaluation has to be defined. This includes the application of the systems, the corresponding environment and the bounds of the studied systems, if they are taken out of a larger context.

Even though the metric framework would support context independent comparisons amongst several systems because of applied normalizations, we do not recommend it. The reason is that some measurements of single capabilities are

strongly dependent on the context, like the set of possible tasks belonging to the mission goal or the set of available sensors and actors.

Weight and Capability Selection. Next, the domain expert has to define weights for the relation between capabilities as well as amongst the measurement of single capabilities. This can be used to express priority or importance of a capability in a given context. Induced by the normalization an equal weighting provides a reasonable starting point. Here it would also be possible to omit or extend capabilities and their corresponding measures in correspondence to special requirements, as long as the general concepts of the metric, like normalization, weighting and focusing on adaption, are not violated. It should always be revised, that this metric framework does not try to evaluate the performance of systems, but rather possibilities and capabilities valuable for adaption to uncertain environments.

Adaption Layer: In case of the adaption layer the following important aspects of each capability have to be considered.

Perception and Acting: Available sensors and actors need to be grouped to types. An actor or sensor belongs to one type if it is addressing similar means.

Planning and Goals: T_G and DT_G are abstracted values for a given context, if an abstraction is not possible they can be determined by measuring and averaging during execution in reality or simulation.

Belief and Reasoning: Defining suitable units for information amount and update rate. For instance using the size of the state space or the amount of processed bytes for the *BIA* and frequency in relation to global time or relative to the refresh frequency of the sensory input for the *BUR*.

Learning: The learning measures are based on the belief and reasoning, for this reason the same considerations are required. The difference is just the consideration of historic data and the capability of repeated learning on it.

Decision Making: Since decision making can not be evaluated in connection with adaptability, it is only necessary to determine if it is available or not.

All measurements have in common that a greater value corresponds to an extended autonomy support. If necessary the measurements have to be quantified and rescaled for all systems in the same manner. If it is only the intention to evaluate the progress of a single system it is necessary to envision possible future extensions defining the range of single measurements.

Relational Layer. For the relational layer it is necessary to examine if the system's capabilities are influenced by others or if it is known to delegate its duties to other systems. Furthermore, the amount of this relations has to be estimated and discretised to a percentage value. These estimations can be based on averaged frequencies of influence or length of time periods for I_i and error

rates or confidence measures for R_i . If there are no restrictions known or can not be determined from historical data external influence is 0% ($I_i = 0.0$) and reliability is 100% ($R_i = 1.0$).

6 Example

In this section we illustrate the application of our multi-dimensional metric with an artificial example. It was the intention to keep it simple and comprehensible.

The example scenario is represented by different robots that are crawling through a garbage dump in an autonomous multi-robot recycling system. They pursue the goal of recycling as much material as possible. Therefore the robots need to detect valuable and recyclable materials and need to plan collection and transport to the disassembly unit.

In this simple example we have two different types of robotic systems. System A is a wheeled robot with one arm and one gripper. It has a standard 2D-vision, a thermal camera vision and a medium sized multi-purpose computing architecture. Further, it delegates object recognition to a remote web-service with 80% accuracy. Moreover, it supports to learn object recognition from the 100 last recognized items. Learning is performed every 60s. System B is a wheeled robot with two arms and two grippers. It has a standard 2D-vision and small size multi-purpose competing architecture. Furthermore, it is obligated to validate its self-made recognitions by a human 20% of the time.

Table 1 shows the domain and scenario specific application of our metric with quantified, adjusted scales and units. For clarification, we have assumed that a less powerful computing architecture results in degraded belief, reasoning and planning capabilities. Hence, a small computing architecture can process smaller information amounts with a lower update rate, as well as less average decomposed subtasks for a given goal. This also illustrates the application of a capability score if not enough information for a full evaluation is available. Here, medium and small are directly taken from the computing architecture description and quantified in the range of 0–4 (extra small, small, medium, large). This exposes further the definition of a custom measurement range. Perception and acting are calculated with the Shannon-Index, see Eq. 1. Because both systems are capable of decision making ($1 = \textit{capable}$, $0 = \textit{not capable}$) the capability is set for both to $C_{\text{Decision Making}} = 1$. This ensures also a valid calculation for the relational component of the capability score C_{Score_i} . For the relational layer we have only the two mentioned statements, highlighted in red, all other cases are either 100% = 1.0 reliable or have 0% = 0.0 influence by restriction.

The normalized measurements, the applied default weights for the main capabilities and the measurements as part of their scores and the calculated results for the capability and autonomy score are presented in Table 2. The measurements “Learning Update Rate” and “Belief Update Rate” were rescaled with $\text{Update Rate} = 1/\text{Update Rate}$. It has to be highlighted that C_{norm_i} of perception and acting are 0 for one of each configurations because of unity normalisation and the used minimal value range.

Table 1. Autonomy metric values of the presented example for both systems

Capabilities	Adaption-Layer	Relational-Layer					
		C_i		Reliability of Delegation		Influence by Restriction	
		System A	System B	System A	System B	System A	System B
Perception and Acting	PA_{Score}						
Perception Diversity		0.35	0.69	1.0	1.0	0.0	0.0
Acting Diversity		0.69	0.35	1.0	1.0	0.0	0.0
Belief and Reasoning	BR_{Score}						
Belief Information Amount		2.0 (Medium)	1.0 (Small)	1.0	1.0	0.0	0.0
Belief Update Rate		2.0 (Medium)	1.0 (Small)	1.0	1.0	0.0	0.0
Decision Making		1 (available)	1 (available)	0.8	1.0	0.0	0.2
Goals and Planning	GP_{Score}	2.0 (Medium)	1.0 (Small)	1.0	1.0	0.0	0.0
Learning							
Learning Information Amount	L_{Score}	100 items	0 items	1.0	1.0	0.0	0.0
Learning Update Rate			1/60 s	∞	1.0	1.0	0.0

Based on such an evaluation result a system designer can compare two systems in a specific scenario or evaluate which capabilities can be improved in order to increase its autonomy. In the presented example with overall equal weights system A achieves a higher autonomy score as system B, because it has higher or equal capability scores for all capabilities except “perception”. This result is even more obvious in the visual representation of the multi-dimensional metric in Fig. 2, where the larger covered area corresponds with a higher autonomy score. In consequence it is possible to select a system with appropriate autonomy based on the score for a scenario or to determine if a planned or implemented extension gives the intended enhancement on capability level or in aggregation.

The presented example has illustrated the process of applying our generic multi-dimensional autonomy metric framework. This, as well as the proposed capability measurements, can be used as a guideline to evaluate and compare autonomous systems in a particular context. Indeed it can be sufficient to adjust

Table 2. Normalized autonomy metric values with default weights, calculated results and aggregation

Capabilities	Weight	Adaption-Layer		Relational-Layer				C_{Score}		
		C_{norm_i}		Reliability of Delegation		Influence by Restriction				
		A	B	A	B	A	B	A	B	
Perception and Acting	PA_{Score}	0.2	0.50	0.50	1.00	1.00	0.00	0.00	0.10	0.10
		0.5	0.00	1.00	1.00	1.00	0.00	0.00		
		0.5	1.00	0.00	1.00	1.00	0.00	0.00		
Belief and Reasoning	BR_{Score}	0.2	0.50	0.25	1.00	1.00	0.00	0.00	0.10	0.05
Belief Information Amount		0.5	0.50	0.25	1.00	1.00	0.00	0.00		
Belief Update Rate		0.5	0.50	0.25	1.00	1.00	0.00	0.00		
Decision Making		0.2	1.00	1.00	0.80	1.00	0.00	0.20	0.16	0.16
Goals and Planning	GP_{Score}	0.2	0.50	0.25	1.00	1.00	0.00	0.00	0.10	0.05
Learning										
Learning Information Amount	L_{Score}	0.2	1.00	0.00	1.00	1.00	0.00	0.00	0.20	0.00
Learning Update Rate			0.5	1.00	0.00	1.00	1.00	0.00	0.00	
		0.5	1.00	0.00	1.00	1.00	0.00	0.00		
Autonomy Score	A_{Score}								0.66	0.36

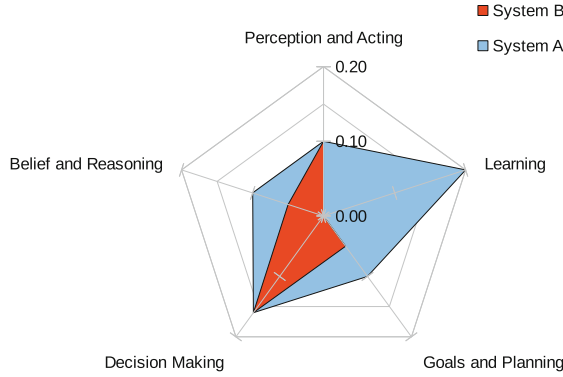


Fig. 2. Multi-dimensional autonomy system comparison with a spider chart

certain capability measurements to the context and the available information, while still respecting all main capabilities and the developed relational concept. Moreover, it has to be incorporated that autonomy values for just one system are not meaningful. They have always to be considered in comparison to other systems or with itself during the development process.

7 Conclusion

In this work we presented an extended understanding of autonomous systems. Moreover, we have differentiated our concept from other system concepts like adaptive systems and automation systems. The core of our autonomy concept is that an autonomous system is always striving for its innermost goal or motivation while it is adapting to the uncertain and dynamic environment. Further, important capabilities were exposed and specified, namely perception and acting, decision making and planning, interpretation and creation of goals, reasoning and believe creation and learning in general. Furthermore, we have introduced a layer concept that contains an adaptation and relational layer. The adaption layer consists of mentioned capabilities and the relational layer models the interaction with other systems. The explicit integration of reliability and independence for modelling the multi-system interaction is an important contribution and clarifies the role of autonomous systems in the context of multi-agent systems.

Based on the definition a generic multi-dimensional metric framework for classification and benchmarking of autonomous systems in a specified evaluation context or domain was developed. This metric allows for a quantified inter-system comparison, controlling and goal specification during the development process.

In the future we are looking for options of defining valid presets of scale combination weightings that can be used as a base for domain experts. Further, we are planning a comprehensive evaluation of the whole concept in several robotic research and development projects within different application domains.

References

1. Alonso, E., Mondragón, E.: Agency, learning and animal-based reinforcement learning. In: Nickles, M., Rovatsos, M., Weiss, G. (eds.) *AUTONOMY 2003*. LNCS (LNAI), vol. 2969, pp. 1–6. Springer, Heidelberg (2004)
2. Alonso, F., Fuertes, J.L., Martínez, L., Soza, H.: Towards a set of measures for evaluating software agent autonomy. In: *Eighth Mexican International Conference on Artificial Intelligence, MICAI 2009*, pp. 73–78. IEEE (2009)
3. Barber, K.S., Martin, C.E.: Agent autonomy: specification, measurement, and dynamic adjustment. In: *Proceedings of the Autonomy Control Software Workshop at Autonomous Agents*, vol. 1999, pp. 8–15. Citeseer (1999)
4. Barber, K.S., Park, J.: Agent belief autonomy in open multi-agent systems. In: Nickles, M., Rovatsos, M., Weiss, G. (eds.) *AUTONOMY 2003*. LNCS (LNAI), vol. 2969, pp. 7–16. Springer, Heidelberg (2004)
5. Bradshaw, J.M., Feltovich, P.J., Jung, H., Kulkarni, S., Taysom, W., Uszok, A.: Dimensions of adjustable autonomy and mixed-initiative interaction. In: Nickles, M., Rovatsos, M., Weiss, G. (eds.) *AUTONOMY 2003*. LNCS (LNAI), vol. 2969, pp. 17–39. Springer, Heidelberg (2004)
6. Carabelea, C., Boissier, O., Florea, A.: Autonomy in multi-agent systems: a classification attempt. In: Nickles, M., Rovatsos, M., Weiss, G. (eds.) *AUTONOMY 2003*. LNCS (LNAI), vol. 2969, pp. 103–113. Springer, Heidelberg (2004)
7. Carraway, G.F., Squibb, J.B.: Autonomy metrics. In: *Proceedings of 4th International Symposium on Space Mission Operations and Ground Data Systems*, vol. 2, Munich (1996)
8. Castelfranchi, C.: Guarantees for autonomy in cognitive agent architecture. In: Wooldridge, M.J., Jennings, N.R. (eds.) *Intelligent Agents*. Lecture Notes in Computer Science, vol. 890, pp. 56–70. Springer, Heidelberg (1995)
9. Castelfranchi, C., Falcone, R.: From automaticity to autonomy: the frontier of artificial agents. In: Hexmoor, H., Castelfranchi, C., Falcone, R. (eds.) *Agent Autonomy. Multiagent Systems, Artificial Societies, and Simulated Organizations*, vol. 7, pp. 103–136. Springer, US (2003)
10. Delaroche, L., Fogel, D.F.L., Freeman, W., Grossberg, S., Lee, S., Lima, P., Pouchard, L., Schultz, A.: Measuring performance of systems with autonomy: metrics for intelligence of constructed systems. In: *Measuring the Performance and Intelligence of Systems: Proceedings of the 2000 PerMIS Workshop*, 14–16 August 2000, vol. 970, p. 4. The Institute (2001)
11. Dorais, G., Bonasso, R.P., Kortenkamp, D., Pell, B., Schreckenghost, D.: Adjustable autonomy for human-centered autonomous systems. In: *Working notes of the Sixteenth International Joint Conference on Artificial Intelligence Workshop Adjustable Autonomy Systems*, pp. 16–35 (1999)
12. Gouaich, A.: Requirements for achieving software agents autonomy and defining their responsibility. In: *Proceedings of Autonomy Workshop at AAMAS 2003*, vol. 236 (2003)
13. Hinchey, M.G., Sterritt, R.: Self-managing software. *Computer* **39**(2), 107–109 (2006)
14. Johnson, M., Bradshaw, J., Feltovich, P., Jonker, C., van Riemsdijk, B., Sierhuis, M.: Autonomy and interdependence in human-agent-robot teams. *IEEE Intell. Syst.* **27**(2), 43–51 (2012)
15. Kephart, J.O., Chess, D.M.: The vision of autonomic computing. *Computer* **36**(1), 41–50 (2003)

16. Luck, M., D'Inverno, M., Munroe, S.: Autonomy: variable and generative. In: Hexmoor, H., Castelfranchi, C., Falcone, R. (eds.) *AUTONOMY 2003. Multiagent Systems, Artificial Societies, and Simulated Organizations*, vol. 7, pp. 11–28. Springer, Heidelberg (2003)
17. Maheswaran, R.T., Tambe, M., Varakantham, P., Myers, K.: Adjustable autonomy challenges in personal assistant agents: a position paper. In: Nickles, M., Rovatsos, M., Weiss, G. (eds.) *AUTONOMY 2003. LNCS (LNAI)*, vol. 2969, pp. 187–194. Springer, Heidelberg (2004)
18. Myers, K., Berry, P., Blythe, J., Conley, K., Gervasio, M., McGuinness, D., Morley, D., Pfeffer, A., Pollack, M., Tambe, M.: An intelligent personal assistant for task and time management. *AI Mag.* **28**(2), 47–61 (2007)
19. Myers, K.L., Morley, D.N.: Policy-based agent directability. In: Hexmoor, H., Castelfranchi, C., Falcone, R. (eds.) *Agent Autonomy. Multiagent Systems, Artificial Societies, and Simulated Organizations*, vol. 7, pp. 185–209. Springer, Heidelberg (2003)
20. Olsen, D.R., Goodrich, M.A.: Metrics for evaluating human-robot interactions. In: *Proceedings of PERMIS*, vol. 2003, p. 4 (2003)
21. Rudas, I.J., Fodor, J.: Intelligent systems. *Int. J. Comput. Commun. Control* **3**(3), 132–138 (2008)
22. Salehie, M., Tahvildari, L.: Self-adaptive software: landscape and research challenges. *ACM Trans. Auton. Adapt. Syst.* **4**(2), 1–42 (2009)
23. Scerri, P., Pynadath, D., Tambe, M.: Adjustable autonomy in real-world multi-agent environments. In: *Proceedings of the Fifth International Conference on Autonomous Agents, AGENTS 2001*, pp. 300–307. ACM, New York (2001)
24. Schillo, M., Fischer, K.: A taxonomy of autonomy in multiagent organisation. In: Nickles, M., Rovatsos, M., Weiss, G. (eds.) *AUTONOMY 2003. LNCS (LNAI)*, vol. 2969, pp. 68–82. Springer, Heidelberg (2004)
25. Schmeck, H.: Organic computing - a new vision for distributed embedded systems. In: *8th IEEE International Symposium on Object-Oriented Real-Time Distributed Computing (ISORC)*, pp. 201–203, May 2005
26. Shannon, C.: A mathematical theory of communication. *Bell Syst. Tech. J.*, **27**, 379–423, 623–656, July, October 1948
27. Sierhuis, M., Bradshaw, J.M., Acquisti, A., Van Hoof, R., Jeffers, R., Uszok, A.: Human-agent teamwork and adjustable autonomy in practice. In: *Proceedings of the Seventh International Symposium on Artificial Intelligence, Robotics and Automation in Space (I-SAIRAS)* (2003)
28. Tambe, M., Scerri, P., Pynadath, D.V.: Adjustable autonomy for the real world. *J. Artif. Intell. Res.* **17**(1), 171–228 (2002)
29. Ulrich, I.R., Mondada, F., Nicoud, J.D.: Autonomous vacuum cleaner. *Robot. Auton. Syst.* **19**, 4–233 (1997)
30. Verhagen, H.: Autonomy and reasoning for natural and artificial agents. In: Nickles, M., Rovatsos, M., Weiss, G. (eds.) *AUTONOMY 2003. LNCS (LNAI)*, vol. 2969, pp. 83–94. Springer, Heidelberg (2004)
31. Weiß, G., Rovatsos, M., Nickles, M.: Capturing agent autonomy in roles and xml. In: *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS 2003*, pp. 105–112. ACM, New York (2003)
32. Yavnai, A.: Entropy-based criteria for intelligent autonomous systems. In: *Proceedings of the 1991 IEEE International Symposium on Intelligent Control*, pp. 55–60, August 1991