Article Identifier
DOI: 10.22032/dbt.58917
URN: urn:nbn:de:gbv:ilm1-2023isc-133:0

60th ILMENAU SCIENTIFIC COLLOQUIUM
Technische Universität Ilmenau, 4 – 8 September 2023
URN: urn:nbn:de:gbv:ilm1-2023isc:1

USING THE DIAGRAMMATIC IMAGE SCHEMA LANGUAGE FOR JOINT HUMAN-MACHINE COGNITION

Till Mossakowski[1], Maria Hedblom[2], Fabian Neuhaus[1],
Stephanie Arevalo Arboleda[3], Alexander Raake[3]

[1]Otto-von-Guericke-Universität Magdeburg, Germany
[2]Jönköping AI Lab (JAIL), Jönköping University, Sweden
[3]Technische Universität Ilmenau

Joint human-machine cognitive systems involve cooperation of humans and machines in perception, communication, decision-making and problem-solving tasks [1]. Recently, joint cognition has received increasing attention in the context of productive teaming, which enhances cooperation between humans and machines in the context of production [2,3]. In productive teaming, humans and machines cooperate in a way that enables them to complement each other's strengths and weaknesses, with the goal of achieving better outcomes than either could achieve alone. For this purpose, a crucial component is the development of common goals and shared mental models. Hence, the development of a joint cognition will be helpful. While humans are naturally equipped with cognition, machines usually lack a deeper cognitive understanding of both situations and human's intentions. On the way towards a true joint cognition, thus it will be necessary to equip machines with some basic tools for grasping and representing (parts of) human cognition. In this work, handover tasks are an example for discussing approaches towards a joint human-machine cognition.

Handover tasks involve three components: a giver, a taker, and an object that will be transferred wherein each component has a current status and a specific intention. Here, it is crucial to have a mutual understanding (awareness) of the current status, role and goal of each element involved in a handover activity. The synergy of understanding these is part of a shared human-robot cognition. In human-robot handover tasks, sharing awareness of each other's roles and intentions is the key to a successful handover [4]. To this aim, humans and robots should be able to interpret situational updates, understand each other's objectives and be aware of their roles [5, 6]. Here, a shared mental model between humans and robots becomes crucial. Barnes et al. [7] presented a model of the shared decision space between humans and robots. Their model describes humans' decisions as influenced by emotions, situation awareness, meta and domain knowledge, ethics, and intent; whereas robot's decisions are based on etiquettes, sensors' data, world model, domain knowledge, rules, and intent inference.

We propose to use image schemas [8] as a tool towards this (partial) joint cognition. Image schemas are mental patterns that infants learn from perceptual experiences in an early phase. They represent conceptual forms of embodied, spatiotemporal relationships. Common image schemas are CONTAINMENT, CONTACT, and SUPPORT. Despite their name, image schemas are not images, but rather semantic constructions. Image schemas can be used for grounding language into embodied experience.

A first step in this research plan is the identification of image schemas that are relevant for joint cognition in productive teaming. While this certainly will depend on the application at hand, we suppose that there are image schemas that will reoccur in many different production contexts. In order to be machine-processable, these image schemas then need to be constructed and represented in a formal way. The Diagrammatic Image Schema Language (DISL) [9] is proposed as a solution here. It allows the systematic and structured construction of image schemas using human-readable comic strip-like representations. Moreover, it comes with a formal machine-processable interchange format.

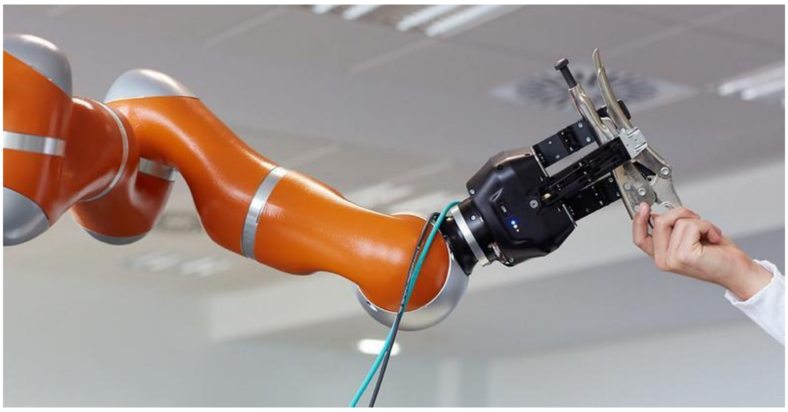

Figure 1: Picture showing handover of object between robot arm and human user. Tecnalia/Flickr, CC BY-NC-ND

As an example, consider a robot that hands over a tool to a human (cf. Fig. 1). While this is a static picture, one can imagine situations before and after that, illustrating the dynamics of the situation. We now abstract this process of handing over with a 3-scene comic strip (Fig. 2). Due to proximity cues, the object is visually grouped by the human reader of the scheme to be held by the hand closer to it. Here, we can identify two roles, giver and taker. The giver's goal is to release the object while the taker's goals are reaching and grasping. In terms of the current status, we can identify two statues: hand open and hand closed.

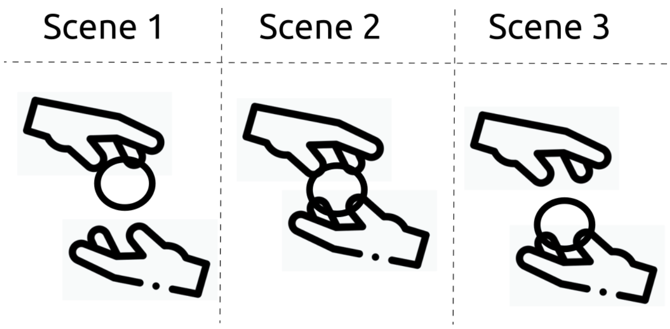

Figure 2: Human-readable comic strip for the example of handing over an object.

Fig. 3 shows how the same process is represented in DISL, reaching a higher level of abstraction. In this strip, all objects are abstracted to boxes, focussing entirely on the essentials of the involved image schemas. An arrow above an object indicates that it is in motion. Objects touching each other are in contact. Moreover, objects can actively exert force, indicated by two triangles directed towards the object borders, or passively suffer force, indicated by two triangles directed towards the center of the object. Note that both notions of force do not necessarily involve movement. In particular, scene 1 might look like the first object pushing the second one. This is possible, but not necessary here. In our context, the exerted force just means that the first (active) object holds the second (passive) object. Directionality cannot be expressed through force, but only through movement, see scenes 2 and 4.

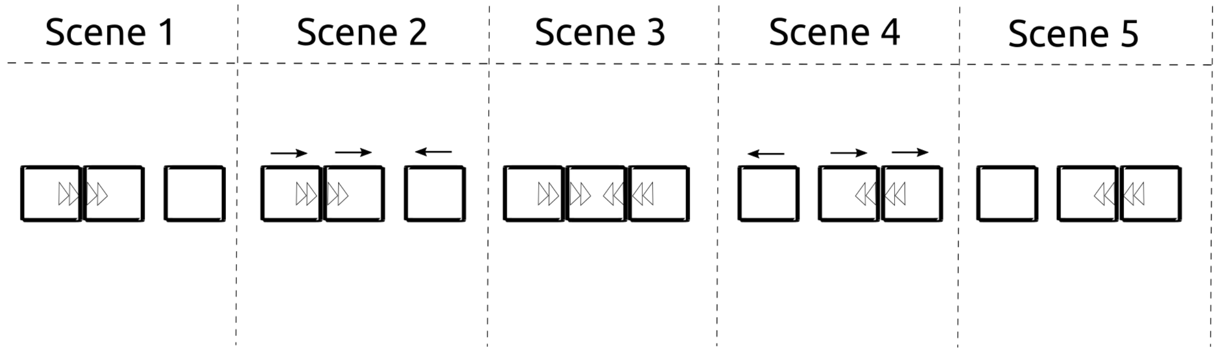

Figure 3: Illustration of DISL with higher level of abstraction, indicating force and direction (triangles) and movement and direction (arrows). See text for details.

While both strips are human-readable representations, the strip in Fig. 3 can be better transformed to a machine-readable representation. The translation of Fig. 3 into the DISL interchange format is as follows:

Strip: Robot hands over tool to human
Scene1: Object(robot-hand), Object(human-hand), Object(tool), Contact(robot-hand,tool), Active_Force(robot-hand), Passive_Force(tool)
Scene2: Object(robot-hand), Object(human-hand), Object(tool), Contact(robot-hand,tool), Active_Force(robot-hand), Passive_Force(tool), Motion(robot-hand), Motion(tool), Motion(human-hand), Move_Towards(robot-hand,human-hand), Move_Towards(tool,human-hand)
Scene3: Object(robot-hand), Object(human-hand), Object(tool), Contact(robot-hand,tool), Contact(human-hand,tool), Active_Force(robot-hand), Passive_Force(tool), Motion(robot-hand)
Scene4: Object(robot-hand), Object(human-hand), Object(tool), Contact(human-hand,tool), Active_Force(human-hand), Passive_Force(tool), Motion(robot-hand), Motion(tool), Motion(human-hand), Move_Away_From(robot-hand,human-hand), Move_Away_From(tool,robot-hand)
Scene5: Object(robot-hand), Object(human-hand), Object(tool), Contact(human-hand,tool), Active_Force(human-hand), Passive_Force(tool)

Figure 4: Translation of the strip in Fig. 3 to the DISL interchange format.

This symbolic description abstracts away from many representational details and captures the essence of the dynamics of the "handing over" image schema that can be found in many different situations. It hence can provide a useful grounding of the notion of "handing over". More specifically, in the first place, the natural language term "handing over" can be understood by humans only. Humans can refine this notion to the strips shown in Figs. 2 and 3 above. The strip in Fi.g 3 can then be formalized within the DISL interchange format (Fig. 4), which expresses the term "handing over" in terms of more specific image-schematic predicates like OBJECT and CONTACT. Now predicates like OBJECT and CONTACT used in DISL are amenable to grounding in sensor data, using object and situation recognition methods, as well as qualitative abstraction methods. We expect that a combination of symbolic methods and neural networks will be useful here, leading to a neuro-symbolic approach [10]. DISL can hence provide an intermediate layer between abstract image schemas and sensorimotor data.

One intended scenario that enables us to test this theory in practice will be a virtual reality simulation of a production process. The situations and dynamics of this simulation shall be mapped into DISL's qualitative terms, as exemplified by the hand-over task above. This

representation layer will provide an intermediate language understood by human and machine and is a step towards a shared representation.
Communication between a human and a (simulated) machine will take place via an existing dialogue system. In the simplest form, there will be command language, e.g. “drive 2m to the right”, “lower the grappler" etc. There could be a controlled language, or full natural language could be used. This language will use concepts from DISL, as well as further concepts (preferably leading to some extension of DISL) needed for expressing statuses and goals. The dialogue system has a mental model, with a (semiotic) mapping of abstract DISL terms to referents (i.e. objects or actions in the VR environment). The deviation between the VR and the mental model can then be learned.
In such a setting, the process and quality of the dialogue can be measured. This allows indications of the cognitive adequacy [11] of the representation in DISL.
A challenge is that users also have some domain knowledge about the machine. This context must also be taken into account. Another challenge is to obtain test persons to enable the collection of dialogues and evaluate the human-system interaction and thus proposed scheme. Such tests can either be carried out with laypersons, covering the basic common sense of productive teaming (which already will be a challenge), or with experts, covering fine-grained domain-specifc knowledge about the involved machine(s). Here, approaches to evaluate the literacy of the human users with regard to the scheme will be needed, that also can accommodate the learning involved when using the system, as illustrated in Fig. 5.

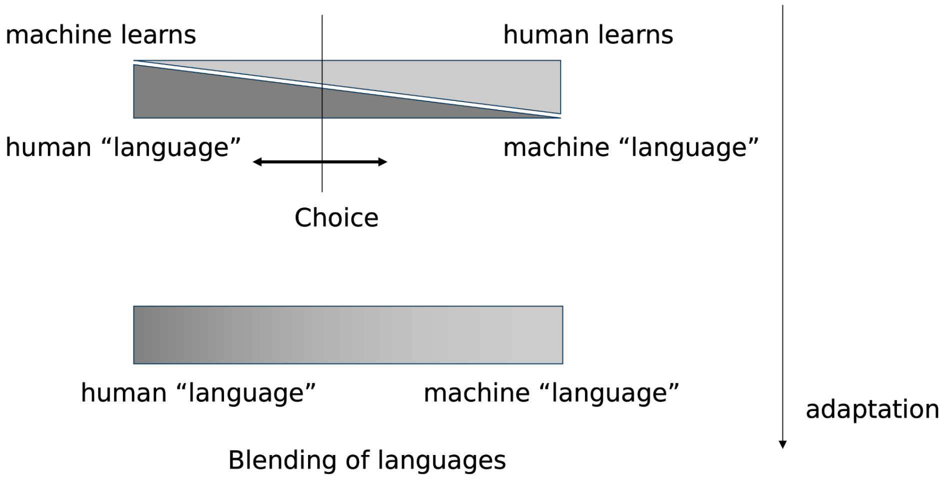

Figure 5: Illustration of assumed learning and adaptation process with regard to language used in human-robot interaction, leading to a “blending” of approaches. The indicated lower, adapted case is to be considered as an abstract view, where both human and robot system are assumed to use language from a common continuum during practical operation.

The long-term goal is to use DISL to achieve shared mental models between human and machine agents, which can be enhanced to support a more effective task allocation and communication, as well as a more accurate decision-making within productive teaming.

REFERENCES

[1] Hollnagel, E., & Woods, D. D. (2005). Joint cognitive systems: Foundations of cognitive systems engineering. CRC press.
[2] Ortmeier, F. & Ragni, M. "Productive Teaming translates to both humans and machines sharing common goals and the mutual dependence required to achieve these." https://de.wikipedia.org/wiki/Productive_Teaming (original citation unavailable)

[3] Johnson, M., & Bradshaw, J.M. (2021). How Interdependence Explains the World of Teamwork. Engineering Artificially Intelligent Systems 2021: 122-146.
[4] Schaefer, K. E., Straub, E. R., Chen, J. Y. C., Putney, J., & Evans, A. W. (2017). Communicating intent to develop shared situation awareness and engender trust in human-agent teams. Cognitive Systems Research, 46, 26–39.
[5] Chen, J. Y. C., & Barnes, M. J. (2014). Human–agent teaming for multirobot control: A review of human factors issues. IEEE Human-machine Systems, 44, 13–29.
[6] Chen, J. Y. C., Lakhmani, S. G., Stowers, K., Selkowitz, A., Wright, J., & Barnes, M. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. Theoretical Issues in Ergonomics Science, 19, 259–282.
[7] Barnes, M. J., Chen, J. Y. C., & Hill, S. D. (2017). Humans and autonomy: Implications for shared decision making in military operations (Tech Rep. ARL-TR-7919). Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.
[8] Mark Johnson. The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason. The University of Chicago Press, Chicago and London, 1987.
[9] Anonymous authors (2023). The Diagrammatic Image Schema Language (DISL). Submitted. https://osf.io/download/pxr6a/?view_only=e24f6ff6fc1c4ce0bce16e328e6d49ad
[10] Pascal Hitzler, Md. Kamruzzaman Sarker: Neuro-Symbolic Artificial Intelligence: The State of the Art. Frontiers in Artificial Intelligence and Applications 342, IOS Press 2021
[11] Markus Knauff: The cognitive adequacy of Allen's interval calculus for qualitative spatial representation and reasoning. Spatial Cogn. Comput. 1(3): 261-290 (1999)

CONTACTS

Prof. Dr. Till Mossakowski
email: till.mossakowski@ovgu.de
ORCID: https://orcid.org/0000-0002-8938-5204

Prof. Dr. Maria Hedblom
email: maria.m.hedblom@gmail.com
ORCID: https://orcid.org/0000-0001-8308-8906

Dr. Fabian Neuhaus
email: fneuhaus@ovgu.de
ORCID: https://orcid.org/0000-0002-1058-3102

Dr. Stephanie Arevalo Arboleda
email: stephanie.arevalo@tu-ilmenau.de
ORCID: https://orcid.org/0000-0001-5577-4407

Prof. Dr. Alexander Raake
email: alexander.raake@tu-ilmenau.de
ORCID: https://orcid.org/0000-0002-9357-1763