Clock-Model-Assisted Agent's Spatial Navigation

Joanna Isabelle Olszewska
School of Computing and Technology, University of Gloucestershire, The Park, Cheltenham, U.K.

Keywords: Qualitative Spatial Relations, Clock Model, Knowledge Engineering, Human-robot Interaction, Robotics, Intelligent Agents, Assistive Aid.

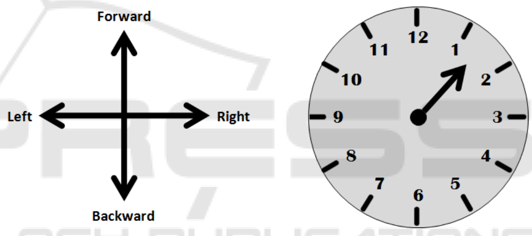
Abstract: Intelligent agent's navigation remotely controlled by means of natural language commands is of great help for robots operating in rescue activities or assistive aid. Whereas full conversation between the human commander and the agent could be limited in such situations, we propose thus to build human/robot dialogues based directly on semantically meaningful instructions like the directional spatial relations, in particular represented by the clock model, to efficiently communicate orders to the agent in the way it successfully gets to a target's position. Experiments within real-world, simulated scenario have demonstrated the usefulness and effectiveness of our developed approach.

1 INTRODUCTION

With the increasing number of robots involved in inspections, explorations, and interventions (Bischoff and Guhl, 2010), the efficient communication between a human commander and his/her intelligent agent is of prime importance, especially to help agent's navigation.

Using natural language for this purpose has been proven to be a promising approach (Summers-Stay et al., 2014). However, human/robot interaction involving natural language is a very difficult process (Tellex et al., 2011), (Baskar and Lindgren, 2014). Indeed, such dialogue should be situated (Pappu et al., 2013) , (Olszewska, 2016) and grounded (Olszewska and McCluskey, 2011), (Olszewska, 2012), (Boularis et al., 2015). The natural language commands should be mapped to low-level instructions (Lauria et al., 2002), (Choset, 2005) in order the agent moves physically in real-time within its environment, which could be ground, air, or underwater, depending of the robot type (Balch and Parker, 2002). Moreover, the intelligent agent needs to have some knowledge to understand the commander's orders (Wooldridge and Jennings, 1995), (Bateman and Farrar, 2004), (MacMahon et al., 2006), (Lim et al., 2011), (Schlenoff et al., 2012), (Bayat et al., 2016).

Hence, in this paper, we propose to study the use of qualitative spatial relations in natural-language dialogues generated by the commander/agent team and constrained by the communication channel availability and occupancy.



(a) (b)

Figure 1: Directional Spatial Relations: (a) LRFB model; (b) clock model.

Qualitative spatial relations are both semantic and symbolic representations of the perceived space in order to describe it and reason about it (Cohn and Renz, 2007). In particular, directional spatial relations formalize the relative positions between different objects of a scene (Renz and Nebel, 2007).

There is a diversity of models codifying directional spatial relations. Common concepts include *'left to'*, *'right to'*, *'front of'*, *'back of'* relations (Bateman and Farrar, 2004), defining the LRFB model (Fig. 1.(a)), and leading to primitive functions such as *'turn to the right'* or *'move forward'* (Marge and Rudnicky, 2010). Other models rely on the cardinal relations, i.e. the *'North'*, *'South'*, *'West'*, and *'East'* concepts (Skiadopoulos et al., 2005). However, this approach requires intrinsically the knowledge of some global reference point, e.g. the *'North'* direction which is not always available in all scenarios. Advanced models (Cohn and Renz, 2007), such as the STAR model, define the directions through

Olszewska J.
Clock-Model-Assisted AgentâĂŹs Spatial Navigation.
DOI: 10.5220/0006253706870692
In *Proceedings of the 9th International Conference on Agents and Artificial Intelligence (ICAART 2017)*, pages 687-692
ISBN: 978-989-758-220-2

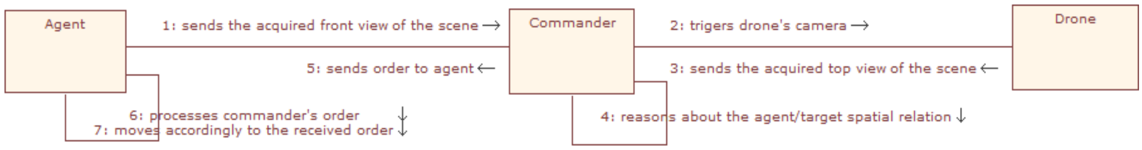


Figure 2: Collaboration diagram of our process.

any number and orientation of intersecting lines, but do not attach any semantically meaningful notion to this kind of representation. A further model called the Clock model (Olszewska, 2015) represents the space as a clock face partitioned in the twelve regions, each corresponding to an hour (Fig. 1.(b)). The directional spatial relations are thus conceptualized in terms of the twelve hours, allowing natural language commands such as *'is at 2 o'clock of'*.

In this work, we chose directional spatial models as basis of human/robot dialogues aiming to guide the agent within its environment to reach a target's position. In this context, we integrate in turns the LRFB model and the clock model into the system generating the dialogues between the two conversational agents, to study the impact of each spatial model on the overall process, where the human operator remotely controls the robot by natural-language commands.

The human agent's orders are based on the reasoning about the objects' spatial relations within images acquired by the on-board camera of the artificial agent as well as any additional agent's view. However, the human's commands sent to the intelligent agent are always formulated in the unique view which is identical for the human/robot team at all time. This tactic automatically generates robot-centric spatial references (Bugmann et al., 2004), directly appropriate for the use by a system controlling the robot using information from its on-board camera.

Hence, specifying that the agent giving the orders is a human which has access to the same view of the scene than the artificial agent implies the conversation is intrinsically situated (Bugmann et al., 2004), and thus, there is no need to further coordinate the robot/human perspectives, avoiding any distortion of their natural expressions about the environment.

On the other hand, the agent's navigation environment could be a schematic, virtual, or natural ground (Marge and Rudnicky, 2010), such as illustrated in Fig. 4(a), Fig. 4(b), and Fig. 4(c), respectively. For this study, we focus on an unfamiliar, flat, non-schematic, real-world type of ground.

The main contribution of this work is the proposed communication system between a human being (called *commander*) and a robot (or *agent*) by means of natural language using the clock-modeled directional spatial relations in order to assist the intelligent agent's navigation to reach a target's position.

The paper is structured as follows. In Section 2, we explain the developed process to guide intelligent agent's navigation using the semantically meaningful clock model to describe the directional spatial relations between this agent and its target's position. This system has been implemented and successfully tested by carrying out simulations of real-world scenarios as reported and discussed in Section 3. Conclusions are drawn up in Section 4.

2 PROPOSED SYSTEM

In this section, we present our system architecture of the humanly assisted, conversational agent's navigation process. It loosely follows the software pipeline structure of a robot consisting of the sensor interface layer, perception layer, planning an control layer, user interface layer, and vehicle interface layer (Russell and Norvig, 2011). Our new approach has seven phases we designed using the Unified Modelling Language (UML) (Booch et al., 2005). In particular, our system's UML collaboration diagram and the UML class diagram are represented in Fig. 2 and Fig. 3, respectively.

As notated in Fig. 2, the artificial agent sends at first the front view of the scene acquired by its on-board camera to the human commander, to establish a situated dialogue between them and to allow reasoning about the scene to be assisted in the navigation.

Secondly, the human commander could use an optional agent. Such additional agent is considered as external to the scene, i.e. it is not present in the views, but it senses the scene (Fig. 3). For example, a drone could obtain additional views of the scene such as the top view without appearing in it (Fig. 5). In case this kind of additional agent is available in the scenario, the commander triggers drone's camera, and then, the drone sends the acquired top view of the scene to the commander.

During the next phase, knowledge about directional spatial relations is invoked (Fig. 3) by applying the LRFB or the clock model and reasoning based on the available views of the scene to find the corresponding directional spatial relation between the object of reference (*agent*) and the relative object (*tar-*

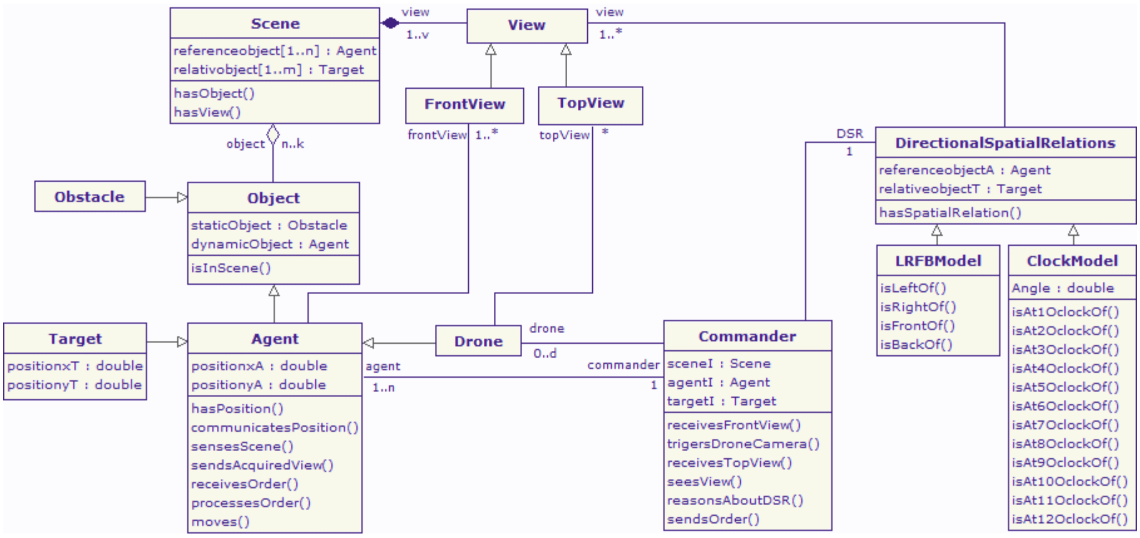


Figure 3: Class diagrams of our system.

get). Once the reasoning about the agent/target spatial relation is performed, this spatial relation is integrated in the natural-language command sent by the commander to the agent (Fig. 2) to assist the agent in reaching its goal.

In case of multiple artificial agents such as the conversational agent and the drone, the different views captured by each of them are merged simultaneously by the commander, before the reasoning phase. However, the commander's order resulting from the reasoning process and sent to the robot is expressed uniquely in the robot's view to generated an intrinsically situated dialogue.

After that, the agent processes the commander's order (Fig. 2) by extracting the provided directional spatial relation from the dialogue line and by mapping it into to low-level instructions ensuring robot motion, i.e. grounding the spatial relation expressed using one directional spatial model by transforming it into the rotation angle around the robot's top-bottom axis and the incremental displacement within the ground plane.

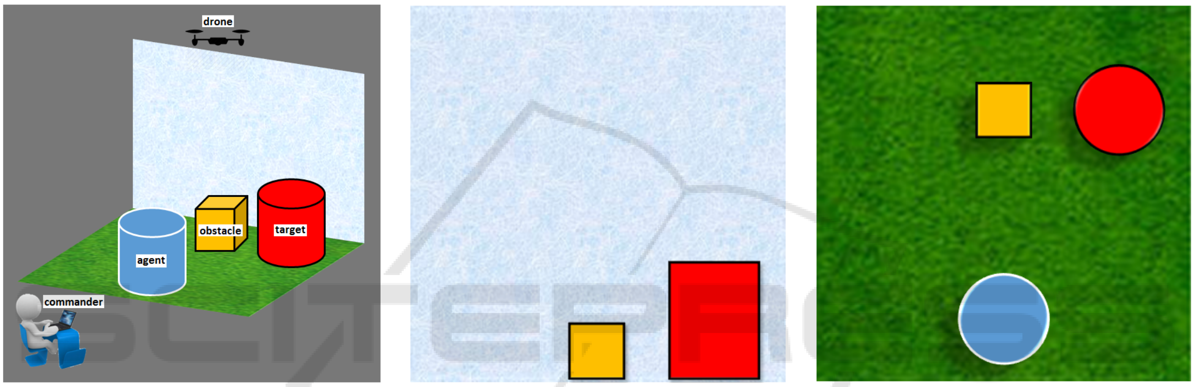
Finally, the robot moves incrementally and accordingly to the received commander order (Fig. 2). In this system, we obviate the need to use the Simultaneous Localization and Mapping (SLAM) process (Saeedi et al., 2015), or map graph building for navigation (Toman and Olszewska, 2014), since the perception/planning/reasoning tasks are performed by the commander based on the environment projections obtained from the acquired view(s) of the sensed scene and the extracted localization of the objects of interest, i.e. the agent, the obstacle, and the target, within it.

Figure 3 models the classes of our navigation system which remotely assists the intelligent agent (*Agent*) in avoiding any *Obstacle* and reaching the position $(positionxT, positionyT)$ of the *Target* by using natural language processing (NLP), in particular integrating *DirectionalSpatialRelations* such as the *LRFBModel* or the *ClockModel*. Such knowledge is used to reason about the *Scene* perceived through *View* such as the *FrontView* and the *TopView* captured by the *Agent* and the *Drone*, respectively. The resulting directional spatial relation between *Object* like the *Agent* and the *Target* is directly sent by the *Commander* to the *Agent* within an automatically situated dialogue. Once this command is processed, the *Agent* moves appropriately, updating its $(positionxA, positionyA)$ to the new one.

It is worth to note the target could be either static or dynamic. In the latter case, it results in an agent/target chase, where both the robot and the target are moving over time on the ground, assuming the agent's speed is equal or greater than the target's one. Hence, in this dynamic problem, the new positions of the target and the robot are obtained thanks to the view(s) acquired after each agent's move. The corresponding new directional spatial relations between the agent and the target is then also recomputed after each agent's move. Thus, the process described in Fig. 2 is repeated iteratively, leading to a dynamic planning updating the spatial dialogue to assist the agent's navigation towards its target.

(a) (b) (c)

Figure 4: Different types of navigation grounds, e.g. (a) artificial ground, (b) virtual ground, (c) real-world ground. Best viewed in colour.



(a) (b) (c)

Figure 5: Simulation of (a) the sensed scene:(b) the front view acquired by the agent; (c) the top view acquired by the drone.

3 EXPERIMENTS AND DISCUSSION

To validate our presented approach, the designed system (Figs. 2-3) has been implemented using C++ programming language, and applied to an intelligent environment (Habib, 2011) benchmark, created by changing the initial position of the agent as well as the target's and obstacle's positions on a virtual, flat ground such as illustrated in Fig. 4(b), within real-world scenario.

Experiments have applied our system in simulations of a real-world scenario consisting all in the human commander's remote assistance of a robot (schematized as the blue cylinder with white contour in Fig. 5) navigation through an intelligent environment to reach the position of a target (represented by the red cylinder with black contour in Fig. 5), avoiding any obstacle such as the yellow cube with black contour in Fig. 5. In this scenario, the target and obstacle are assumed to be static, only the robot is behaving dynamically.

The first experiment does not involves any drone, the reasoning being only based on the scene's front view captured by the robot's on-board camera, whereas the second experiment requires a drone, leading to multiple-view, spatial reasoning. Whereas the commander has access to the additional view of the scene, i.e. the top view sent by the drone, the human agent expresses still all his/her orders to the intelligent agent by referring directly to the common, identical front view; the top view being only used as an additional information to strength the spatial reasoning.

Figure 5 shows a snapshot of the scene and its different views at a given time. It is worth to note that either the commander or the drone are not in the scene itself (Fig. 5(a)), but they are both part of the scenario. Indeed, on one hand, the drone acquires a top view of the scene thanks to its embedded camera. On the other hand, the commander looks remotely at the scene through the agent's camera capturing a front view of the scene (Fig. 5(b)), and sees the top view of the scene generated by the drone's camera (Fig. 5(c)).

In these experiments, the designed system has been tested once with applying the LRFB model and once with integrating the clock model instead. The results are reported in Table 1, while samples of the human/robot dialogues are reported as follows.

An example of these dialogues generated when using the LBFB model in our system is:

HUMAN: The target is right to you. Go!
ROBOT: I have moved.
HUMAN: The target is right to you. Go!
ROBOT: I have moved.
HUMAN: The target is in front of you. Go!
ROBOT: I have moved.
HUMAN: The target is in front of you. Go!
ROBOT: I have moved.
HUMAN: The target is in front of you. Go!
ROBOT: I have reached the target.
HUMAN: Good!

It is worth to note that the commander does not start by sending an order using the *front-of* type relation, to avoid any collision of the agent with the obstacle.

In case of the adoption of the clock model in our system, a sample of the dialogue between the human commander and the intelligent agent is:

HUMAN: The target is at 2 o'clock of you. Go!
ROBOT: I have moved.
HUMAN: The target is at 2 o'clock of you. Go!
ROBOT: I have moved.
HUMAN: Target is at 2 o'clock of you. Go!
ROBOT: I have reached the target.
HUMAN: Good!

Experiment 2 runs the system with using the supplementary agent, namely, the drone. This leads to a multi-robot system (Khamis et al.,), but the conversational agents remain the same than in the experiment 1, i.e. the commander and the robot; the commander communicating with the drone using only machine instructions rather than natural-language dialogues.

Table 1: Navigation accuracy of the intelligent agent reaching the target's position, when the commander formulated orders using the Left/Right/Forward/Backward (LRFB) model and the clock model, respectively.

	LRFB model	clock model
exp1	88.5%	95.4%
exp2	90.6%	97.3%

In Table 1, the clock model is more efficient than the LRFB model in both scenarios. On the other hand, the use of multiple, synchronized views appears to improve the accuracy of the commander's orders and thus, the agent's navigation. In particular, the top view decreases the uncertainty related to the qualitative spatial relations in any front view, resulting in more precise rotation and displacement of the intelligent agent.

The experiments have been carried out on simulated grounds like in (Tellex et al., 2011). We can observe that our system performance is far better than the state-of-art ones (MacMahon et al., 2006), (Tellex et al., 2011), and (Boularis et al., 2015).

From these experiments, we can observe that the clock model refines the knowledge about the directional spatial relations useful for navigation guidance by means of natural language, and leads to a more accurate navigation system. Furthermore, the clock model is 30% faster than the traditional models from the computational point of the view. Thus, the clock model can be widely used in human-agent situated dialogues, since it brings both a gain in precision and speed in the resulting, assisted agent's navigation.

4 CONCLUSIONS

In this work, we proposed a human-robot communication system using directional spatial relations, such as the clock model, which are determined based on the information extracted from the acquired views of the visual scene where the robot evolves in. The corresponding directional spatial relation is then communicated by the commander to the intelligent agent which transforms it by processing the natural language command into motor commands and then moves accordingly to reach the target's position. Our approach could be applied to human-robot interactions (HRI) as well as to remote assistance of autonomous systems' navigation through real-world, unfamiliar environment.

REFERENCES

Balch, T. and Parker, L. (2002). *Robot Teams: From Diversity to Polymorphism*. Taylor & Francis.

Baskar, J. and Lindgren, H. (2014). Cognitive architecture of an agent for human-agent dialogues. In *Proceedings of the International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 89–100.

Bateman, J. and Farrar, S. (2004). Modelling models of robot navigation using formal spatial ontology. In *Proceedings of the International Conference on Spatial Cognition*, pages 366–389.

Bayat, B., Bermejo-Alonso, J., Carbonera, J., Facchinetti, T., Fiorini, S. R., Goncalves, P., Jorge, V., Habib, M., Khamis, A., Melo., K., Nguyen, B., Olszewska, J. I., Paull, L., Prestes, E., Ragavan, S., S.Saeedi, Sanz, R., Seto, M., Spencer, B., Trentini, M., Vosughi, A., and Li, H. (2016). Requirements for building an ontology for autonomous robots. *Industrial Robot: An International Journal*, 43(5):469–480.

Bischoff, R. and Guhl, T. (2010). The strategic research agenda for robotics. *IEEE Robotics and Automation Magazine*, 17(1):15–16.

Booch, G., Rumbaugh, J., and Jacobson, I. (2005). *The Unified Modeling Language User Guide*. Addison-Wesley.

Boularis, A., Duvallet, F., Oh, J., and Stentz, A. (2015). Grounding spatial relations for outdoor robot navigation. In *Proceedings of the IEEE/RSJ International Conference on Robot and Automation (ICRA'15)*, pages 1976–1982.

Bugmann, G., Klein, E., Lauria, S., and Kyriacou, T. (2004). Corpus-based robotics: A route instruction example. In *Proceedings of the International Conference on Intelligent Autonomous Systems*, pages 96–103.

Choset, H. (2005). *Principles of Robot Motion: Theory, Algorithms, and Implementation*. MIT Press.

Cohn, A. G. and Renz, J. (2007). *Handbook of Knowledge Representation*, chapter Qualitative spatial reasoning. Elsevier.

Habib, M. (2011). Collaborative and distributed intelligent environment merging virtual and physical realities. In *Proceedings of the IEEE International Conference on Digital Ecosystems and Technologies*, pages 340–344.

Khamis, A., Hussein, A., and Elmogy, A. Multi-robot task allocation: A review of the state-of-the-art. *Cooperative Robots and Sensor Networks*, pages 59–66.

Lauria, S., Bugmann, G., Kyriacou, T., Bos, J., and Klein, E. (2002). Converting natural language route instructions into robot executable procedures. In *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication*, pages 223–228.

Lim, G. H., Suh, I. H., and Suh, H. (2011). Ontology-based unified robot knowledge for service robots in indoor environments. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 41(3):492–509.

MacMahon, M., Stankiewicz, B., and Kuipers, B. (2006). Walk the talk: Connecting language, knowledge, and action in route intructions. In *Proceedings of the AAAI International Conference on Artificial Intelligence*, pages 1475–1482.

Marge, M. and Rudnicky, A. (2010). Comparing spoken language route instructions for robots across environment representations. In *Proceedings of the ACL Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL'10)*, pages 157–164.

Olszewska, J. I. (2012). Multi-target parametric active contours to support ontological domain representation. In *Proceedings of RFIA Conference*, pages 779–784.

Olszewska, J. I. (2015). 3D Spatial reasoning using the clock model. *Research and Development in Intelligent Systems, Springer*, 32:147–154.

Olszewska, J. I. (2016). Interest-point-based landmark computation for agents' spatial description coordination. In *Proceedings of the International Conference on Agents and Artificial Intelligence (ICAART'16)*, pages 566–569.

Olszewska, J. I. and McCluskey, T. L. (2011). Ontology-coupled active contours for dynamic video scene understanding. In *Proceedings of the IEEE International Conference on Intelligent Engineering Systems*, pages 369–374.

Pappu, A., Sun, M., Sridharan, S., and Rudnicky, A. (2013). Situated multiparty interaction between humans and agents. In *Proceedings of the International Conference on Human-Computer Interaction*, pages 107–116.

Renz, J. and Nebel, B. (2007). *Handbook of Spatial Logics*, chapter Qualitative spatial reasoning using constraint calculi. Springer Verlag.

Russell, S. and Norvig, P. (2011). *Artificial Intelligence: A Modern Approach*. Prentice-Hall.

Saeedi, S., Paull, L., Trentini, M., and Li, H. (2015). Occupancy grid map merging for multiple robot simultaneous localization and mapping. *International Journal of Robotics and Automation*, 30(2):149–157.

Schlenoff, C., Prestes, E., Madhavan, R., Goncalves, P., Li, H., Balakirsky, S., Kramer, T., and Miguelanez, E. (2012). An IEEE standard ontology for robotics and automation. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012)*, pages 1337–1342.

Skiadopoulos, S., Giannoukos, C., Sarkas, N., Vassilliadis, P., Sellis, T., and Koubarakis, M. (2005). Computing and managing cardinal direction relations. *IEEE Transactions on Knowledge and Data Engineering*, 17(12):1610–1623.

Summers-Stay, D., Cassidy, T., and Voss, C. R. (2014). Joint navigation in commander/robot teams: Dialog and task performance when vision is bandwidth-limited. In *Proceedings of the ACL International Conference on Computational Linguistics*, pages 9–16.

Tellex, S., Kollar, T., Dickerson, S., Walter, M. R., Banerjee, A. G., Teller, S., and Roy, N. (2011). Understanding natural language commands for robotic navigation and mobile manipulation. In *Proceedings of the AAAI International Conference on Artificial Intelligence*.

Toman, J. and Olszewska, J. I. (2014). Algorithm for graph building based on Google Maps and Google Earth. In *Proceedings of the IEEE International Symposium on Computational Intelligence and Informatics*, pages 80–85.

Wooldridge, M. and Jennings, N. (1995). Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, 10(2):115–152.