From Human Interaction to Human-Robot Interaction: A Possible Model

Francesca A. Lisi

Dipartimento di Informatica, Università degli Studi di Bari “Aldo Moro”, Italy
FrancescaAlessandra.Lisi@uniba.it

Abstract. This paper presents the Elementary Pragmatic Model, originally developed for studying human interaction and widely used in psycotherapy, as a starting point for novel AI applications where humans interact with AI agents.

1 Introduction

Artificial Intelligence (AI) and Psychology are deeply interconnected and have been influencing each other's development since the very beginning. In particular, AI researchers need a deeper understanding of human psychology in order to develop AI agents better at interacting with humans.

This paper proposes the *Elementary Pragmatic Model* (EPM) [10], originally developed for studying human interaction and widely used in psycotherapy, as a starting point for novel AI applications where human-computer interaction is particularly challenging such as in *Human-Robot Interaction* (HRI) [7]. In this preliminary investigation issues of *Machine Ethics* [1] are also taken into account.

The paper is structured as follows. Section 2 briefly introduces the EPM whereas Section 3 sketches possible applications of EPM in AI.

2 The Elementary Pragmatic Model

The EPM is based on the analysis of the interaction between two subjects by means of an elementary binary model. The communication is modeled according to information theory principles and gives rise to triplets inspired to the ‘Shannon's triads’. As illustrated in Figure 1 every triplet comprises: (i) the first interacting subject's world (A); (ii) the second interacting subject's world (B); (iii) the result of the interaction, which means how A has been transformed by the interaction with B thus giving rise to a new subject (A').

Triads are then the starting point for the development of the EPM which is structured into four levels of growing complexity as shown in Figure 2: an elementary level (*i.e.*, the level of the *triads*); a second level (*i.e.*, the level of the four *co-ordinates* of interaction); a third level (*i.e.*, the level of the sixteen functions or rather of the sixteen *relational styles*); a fourth level (*i.e.*, the level of the *interaction*).

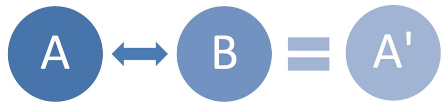

Fig. 1: Triad in the EPM.

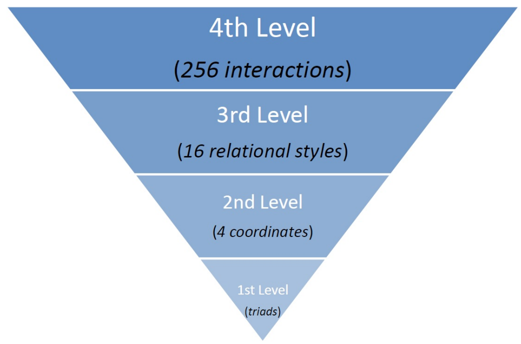

Fig. 2: Levels of the EPM.

Second Level The triplets make it possible for the graphic definition of four spaces of dyadic interaction, by means of a Venn's diagram (see Figure 3) in which the circles labeled with A and B represent the worlds of the two interacting subjects; the intersection between the two circles represent the elements that are common to both the subjects' worlds; everything off the circles is what is outside the interacting worlds. Each of these components can be full or empty. Thus four co-ordinates of the interactions are defined: *acceptance*, *maintenance*, *sharing*, and *antifunction*.

Third Level The combination of the four co-ordinates of A with those of B gives rise to sixteen functions that model different relational styles in human interaction, such as: Who obliterates him/her-self in the relation (f_0); Who maintains his/her own world (f_3); Who accepts the other's world (f_5); Who accepts what only exists, or does not exist, in his/her own and in the other's world (f_9); Who accepts everything (f_{15}). These functions can be graphically represented by Venn's diagrams.

Fourth Level The interaction of the sixteen relational styles among them leads to 256 possibilities, as shown in Table 1. For illustrative purpose let us consider the case of psychoterapy where A is the patient, and B the therapist. If A adopts a certain relational style and B adopts another one, by crossing the two functions in Table 1 we shall read how the patient was transformed by the interaction with the therapist. For instance, let us suppose that the patient is pursuing a goal creatively (*i.e.*, A is in the state determined by f_9) and the therapist shows sympathy with the patient on his/her accomplishments (*i.e.*, B adopts the style f_5). The interaction between the two triggers a mental state transition in the patient from f_9 to f_3.

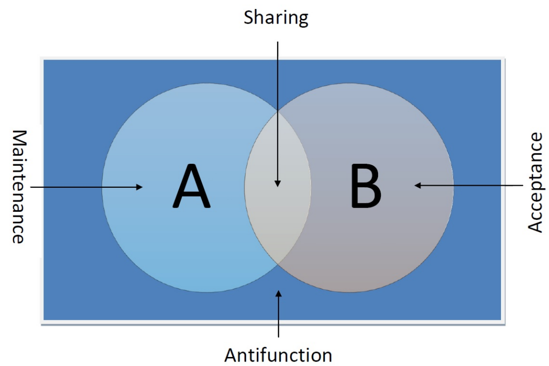

Fig. 3: Co-ordinates of the dyadic interaction in the EPM.

Transitions can be enforced by means of so-called SISCI sentences, *i.e.* sentences with a strong psychological impact [5]. These sentences act like mantras. The more often one pronounces them the stronger is their effect as mental modifiers. For instance, a sentence like “If you pursue a goal creatively, and the outcome is positive for you, this will make you stronger” can be used by the therapist to induce the transition from f_9 to f_3 in the mind of the patient in the example above. Note that f_0, f_3 and f_{15} are considered as *final states* in the interaction according to EPM. This means that A either ends up obliterating him/her-self in the relation with B or keeps maintaining his/her own world or becomes unable to choose (chaos).

Table 1: Dyadic interactions in the EPM.

$B \backslash A$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	0	2	3	0	0	6	7	7	6	15	15	3	2	15	15
1	0	1	2	3	1	1	7	7	6	7	14	15	3	3	15	15
2	0	0	0	3	2	2	4	7	5	4	13	15	3	2	13	15
3	0	1	0	3	3	3	5	7	4	5	12	15	3	3	13	15
4	0	0	2	3	0	4	2	7	3	2	11	11	3	6	11	15
5	0	1	2	3	1	5	3	7	2	3	10	11	3	7	11	15
6	0	0	0	3	2	6	0	7	1	0	9	11	3	6	9	15
7	0	1	3	3	2	7	1	7	0	1	8	11	3	7	9	15
8	0	0	2	3	8	8	14	15	7	14	7	15	3	10	7	15
9	0	1	2	3	9	9	15	15	6	15	6	15	3	11	7	15
10	0	0	0	3	10	10	12	15	5	12	5	15	3	10	5	15
11	0	1	0	3	11	11	13	15	4	13	4	15	3	11	5	15
12	0	0	2	3	8	12	10	15	3	10	3	11	3	14	3	15
13	0	1	2	3	9	13	11	15	2	11	2	11	3	15	3	15
14	0	0	0	3	10	14	8	15	1	8	1	11	3	14	1	15
15	0	1	0	3	11	15	9	15	0	9	0	11	3	15	1	15

3 Towards EPM-based AI applications

A natural AI application of EPM is the development of virtual AI agents for the treatment of mental illness. In this case the agent plays the role of the therapist (subject B). A little step towards this direction is a program for automatically choosing SISCI sentences developed by the same authors of the EPM [6]. Similar in the structure is the class of virtual AI agents designed as companions or personal assistants for elderly or adults with cognitive impairment. However, substantial work should be done in order to develop real AI applications from the seminal ideas here presented.

A major problem is the reliable injection of ethical principles, *e.g.* the *Principles of Biomedical Ethics* of Beauchamp and Childress [4], into the AI agent (whether it is virtual or physical). Indeed, in the healthcare domain, the AI agent should be able to deal with variants of the following ethical dilemma (as highlighted in, *e.g.*, [3,2]): *A healthcare professional has recommended a particular treatment for her competent adult patient, but the patient has rejected it. Should the healthcare professional try to change the patient's mind or accept the patient's decision as final?*

Only two actions are possible when facing this dilemma, namely: (1) accepting a patient's decision to reject a treatment; and (2) trying to convince him/her to change his/her mind. Furthermore, the cases affected by such ethical issue are constrained to three of the four duties reported in [4], viz., respecting the autonomy of the patient, not causing harm to the patient (*non-maleficence*), and promoting patient welfare (*beneficence*). A decision procedure is then necessary to choose the best action to take, which is compliant with the given duties.

One may also think of applications where A is the AI agent and B is the human. In particular, in HRI it is important to interface mutual 'mental' models in order to avoid working at cross-purposes [9]. However, eliciting mental models from humans of what robots can or should do, and combining that knowledge with the computer's model for purposes of planning and conflict avoidance, remains an HRI challenge. We suggest that the EPM could be used as a means for programming and/or guiding the robot. Therefore humans could change the 'mental' model of the robot by voice commands (chosen from a predefined set of sentences, inspired to SISCI sentences, which then builds up a novel *robot programming language*). In particular, the sequence of voice commands given by a human B to a robot A during their interaction should aim at bringing A to the most appropriate final state according to the context. For instance, a command like the SISCI sentence reported in the therapist-patient example above acts like a reinforcement mechanism for the well-behaving robot (*i.e.*, the robot consolidates its 'mental model' upon approval from the human).

A deep investigation is needed to figure out which functions of the EPM are actually relevant and/or desirable in HRI so that machine ethics issues are properly addressed. A major approach followed by current studies in the field is the *theory of prima facie duties* [8] of which the abovementioned principles of biomedical ethics are an instantiation for the healthcare domain. An interesting direction for the research in AI and Robotics can be then the combination of this theory with the EPM in order to define a cognitively-inspired programming language for ethical machines.

Acknowledgements The author would like to warmly thank Piero De Giacomo for having introduced her to the Elementary Pragmatic Model.

References

1. Allen, C., Wallach, W., Smit, I.: Why machine ethics? IEEE Intelligent Systems **21**(4), 12–17 (2006). https://doi.org/10.1109/MIS.2006.83, `https://doi.org/10.1109/MIS.2006.83`
2. Anderson, M., Anderson, S.L.: ETHEL: toward a principled ethical eldercare system. In: AI in Eldercare: New Solutions to Old Problems, Papers from the 2008 AAAI Fall Symposium, Arlington, Virginia, USA, November 7-9, 2008. AAAI Technical Report, vol. FS-08-02, pp. 4–11. AAAI (2008), `http://www.aaai.org/Library/Symposia/Fall/2008/fs08-02-002.php`
3. Anderson, M., Anderson, S.L., Armen, C.: Medethex: A prototype medical ethics advisor. In: Proceedings, The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, July 16-20, 2006, Boston, Massachusetts, USA. pp. 1759–1765. AAAI Press (2006), `http://www.aaai.org/Library/AAAI/2006/aaai06-292.php`
4. Beauchamp, T., Childress, J.: Principles of Biomedical Ethics. Oxford University Press (1979)
5. De Giacomo, P., L'Abate, L., Margari, F., De Giacomo, A., Santamato, W., Masellis, R.: Sentences with strong psychological impact in psychotherapy: Research in progress. Journal of Contemporary Psychotherapy **38**(2), 65–72 (2008). https://doi.org/10.1007/s10879-007-9069-6, `https://doi.org/10.1007/s10879-007-9069-6`
6. De Giacomo, P., L'Abate, L., Margari, F., Santamato, W., De Giacomo, A., Masellis, R., D'Agostino, V.: A program to select sentences with a strong psychological impact. Journal of Contemporary Psychotherapy **39**(3), 165–169 (2009). https://doi.org/10.1007/s10879-009-9112-x, `https://doi.org/10.1007/s10879-009-9112-x`
7. Goodrich, M.A., Schultz, A.C.: Human-robot interaction: A survey. Foundations and Trends in Human-Computer Interaction **1**(3), 203–275 (2007). https://doi.org/10.1561/1100000005, `https://doi.org/10.1561/1100000005`
8. Ross, W.: The Right and the Good. Oxford University Press (1930)
9. Sheridan, T.B.: Human-robot interaction: Status and challenges. Human Factors **58**(4), 525–532 (2016). https://doi.org/10.1177/0018720816644364, `https://doi.org/10.1177/0018720816644364`
10. Silvestri, A., De Giacomo, P., Pierri, G., Lefons, E., Pazienza, M.T., Tangorra, F.: A basic model of interacting subjects. Cybernetics and Systems **11**(1-2), 115–129 (1980). https://doi.org/10.1080/01969728008960230, `https://doi.org/10.1080/01969728008960230`