Commitments, Expectations, Affordances and Susceptibilities: Towards Positional Agent Programming

Sileno, G.; Boer, A.; van Engers, T.

DOI
10.1007/978-3-319-25524-8_52

Publication date
2015

Document Version
Final published version

Published in
PRIMA 2015: Principles and Practice of Multi-Agent Systems

Citation for published version (APA):
Sileno, G., Boer, A., & van Engers, T. (2015). Commitments, Expectations, Affordances and Susceptibilities: Towards Positional Agent Programming. In Q. Chen, P. Torroni, S. Villata, J. Hsu, & A. Omicini (Eds.), *PRIMA 2015: Principles and Practice of Multi-Agent Systems: 18th international conference, Bertinoro, Italy, October 26-30, 2015 : proceedings* (pp. 687-696). (Lecture Notes in Computer Science ; Vol. 9387), (Lecture Notes in Artificial Intelligence). Springer. https://doi.org/10.1007/978-3-319-25524-8_52

Commitments, Expectations, Affordances and Susceptibilities: Towards Positional Agent Programming

Giovanni Sileno(✉), Alexander Boer, and Tom van Engers

Leibniz Center for Law, University of Amsterdam, Amsterdam, The Netherlands
{g.sileno,aboer,vanengers}@uva.nl

Abstract. The paper introduces an agent architecture centered around the notions of commitment, expectation, affordance, and susceptibility. These components are to a certain measure at the base of any agent system, however, inspired by research in explanation-based decision making, this contribution attempts to make explicit and start organizing under the same operationalization neglected figures as negative commitment, negative expectation, etc.

Keywords: Cognitive architectures · Positional programming · PACK · Reactive systems · Petri nets

1 Introduction

As the myth tells, even if he knew that all sailors who had done it went lost into the open sea, Ulysses wanted to hear the voices of the Sirens. To achieve this goal, the sail direction set, he put some wax in his companions' ears and asked them to bind him to the mainmast with the strongest rope. He also ordered not to follow any of his requests before destination. Eventually, he succeeded, and we, listening to his story, can understand why. However, are we able to fully represent it with current agent-based languages? The story refers to notions as *conditional persistent commitment* (Ulysses desiring to jump off towards the Sirens, and insisting on trying it even if bound to the mainmast), *positive expectation* (about the fact that the sirens were along that specific path), which find some correspondence in most common BDI representations. However, in modeling those characters, we may easily identify other notions at stake, as e.g. *negative affordance* (associated with the overall plan *preventing* Sirens' effect), *disability* (Ulysses bound to the mast), *negative susceptibility* (the sailors to Ulysses' requests) and *no-susceptibility* (the sailors to Sirens' voices). The purpose of the present work is to identify and consider these "neglected" *positions* as first-class citizens, and to start operationalizing them in practical reasoning terms.

1.1 Background and Motivation

The initial idea behind this contribution grew out from our work in institutional modeling (see e.g. [1]). In a formal institution, each actor is bound to other

Q. Chen et al. (Eds.): PRIMA 2015, LNAI 9387, pp. 687–696, 2015.
DOI: 10.1007/978-3-319-25524-8_52

actors according to the legal relationships derived from the role he is enacting. Hohfeld [2], for instance, identified positions as *duty*, *claim*, *power* and *liability* as the fundamental components to describe the legal configurations holding between two parties.[1] Being our general objective to model complex institutional scenarios, including scenarios of non-compliance, the composition of mere institutional roles is however not satisfactory (e.g. a seem-to-be normal sale may hide a money-laundering scheme). Roles need to be enriched with an explicit intentional component, or, in other words, with an explicit link between *motives* and institutional aspects. This requirement brought us to the exploration of the *agent-role* concept [3].

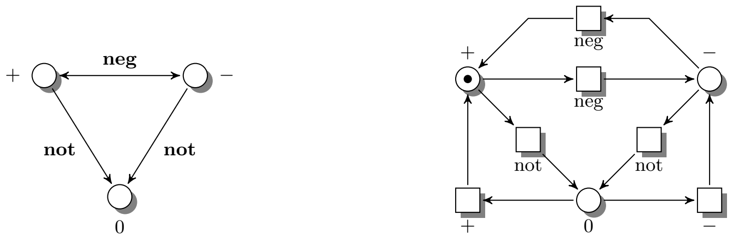
Reflecting on the interactions between the institutional and agent components, we discovered deeper underlying connections between the two domains: *where institutional positions identify extrinsic commitments and abilities, agentic positions identify intrinsic commitments and abilities.*[2] In practice, the *correlativeness* of the institutional positions of two parties can be put in analogy to the correlativeness between agent and environment. This epistemological leap allowed us to explore figures that are usually overlooked in agent modeling, but which are important in the legal domain. For instance, *negative action* – in the two forms of lack of action or actively preventing another outcome – is equally relevant to attribute responsibility [4].

To confirm this representational need, we attempted to review other agent languages/platforms considering the categories we will introduce in the paper. Except Jason [6], built upon AgentSpeak(L) [7], those considered (2APL [8], GOAL [9], ALP [10] and DALI [11]) refer only to default negation and introduce some form of declarative goals. Losing the difference between *null* and *negative* polarity, all negative positions we will introduce conflate. On the other hand, without declarative goals, modeling mutual interactions between goals is much more complex. Even if these and other representational limits may be overcome with adequate extensions, our idea is to approach the issue from an alternative direction: to start from stronger representational requirements, and to construct the reasoning platform on top of those. Rather than rational choice theory, the foundations of our contribution are to be found in cognitive research as e.g. [5], focusing on *explanation-based* decision making.

The paper proceeds as follows. First, we define the foundations of the modeling language, i.e. the notion of position and the two types of negation (§ 2). Second, we analyze in some detail the position types (§ 3). Third, we briefly draft their operationalization in a practical reasoning framework (§ 4). Discussion and future developments end the paper.

[1] These relations bring specific inter-dependencies: e.g. if a party has a duty to perform a certain action, then there is another party that has a claim towards the first. At the same time, if a party is in a certain position (e.g. duty to A), this precludes the same party to be in another position (e.g. no-duty to A).

[2] Extrinsic means that it is the result of social, normative forces, external to the agency: the agent cannot change such position, even if he may still neglect or overlook it. Intrinsic means that the agent has in principle control over it.



(a) *positive*, *negative*, *null* positions and *negation* operators.

(b) positional triple as a Petri net.

Fig. 1. Triangle of contrariety and relative Petri net model.

2 Modeling Language: Positions and Negations

The proposed modeling language is centered around the notion of *position*. In general, a position is a local state of the system that can be related to other positions in dimensional terms. For instance, in a classic logic system, each proposition can be put in relation with its negation. In this framework, however, we consider the dimensional characterization associated to the *triangle of contrariety*.[3] In addition to positive (+) and negative (−) polarities, we consider a *null* (0) polarity. For instance, black is certainly *not* white, but e.g. gray is *not* white as well. Similarly, prohibition is the *opposite* of obligation, but they are both *not* the same as faculty. Different operators can then be considered for those negations: **neg** and **not**, illustrated in Fig. 1a. The first corresponds to classic negation (or *strong negation*), relying on the duality/opposition of two notions (e.g. black/white, obligation/prohibition). The second operator removes the polarity; the null position states a certain qualification cannot be concluded neither positively, nor negatively. In other words, restricting ourselves to the terms proposed by that bipolar frame, *undecidability* holds. To a certain extent, this can be associated to *default negation*.

Using this pattern, given any position, we can construct a triple of positions. Only one of the three may hold at a certain time. The three symbols +, −, 0, can be interpreted both in *state* and *transition* terms: i.e. as identifying a certain local state of this positional triple, or as identifying the event bringing about that state. Exploiting this two-fold interpretation, and the focus on local states, we ground our language to the computational model given by the Petri net notation, as in Fig. 1b (see e.g. [14] for an introduction to Petri nets).

3 Cognitive Components

3.1 Commitment

Commitment identifies a general motivational component, i.e. an internal cognitive mechanism which eventually converges or plays a role in driving towards

[3] The triangle of contrariety (with nodes A, E, Y) can be derived from the Aristotelian square (A, E, O, I), with $Y = O \wedge I$, see e.g. [12,13]. In relation to Hohfeld, see [1].

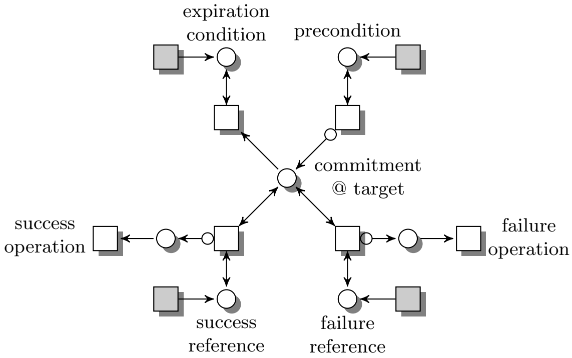


Fig. 2. Operational structure of commitment

action or action avoidance. Following the triangle of contrariety, commitment can be declined into the positions of *positive commitment*, *negative commitment* and *no-commitment*.

Positive Commitment. A positive commitment position can be specified via the elements illustrated in Fig. 2.[4]

The *target* identifies what the agent is committed to. The *precondition* represents the situation in which the commitment is instantiated. If the precondition is present, the resulting element is a *conditional commitment*, as in 'If I listen to the Sirens' voice, I want to follow them'. The *success reference* corresponds to the specification on how to recognize its satisfaction (usually a proxy of the target). The *success operation* describes what to perform after the commitment is recognized as satisfied. The *failure reference/operation* identify how to recognize/what to do when the commitment fails. These may be used to instantiate *backup* commitments. The *expiration condition* expresses situations that remove the commitment, independent from the ones defined in success or failure.

The structure allows to distinguish easily *achievement goals* from *maintenance goals*, depending on the presence in success operation of the removal the commitment (cf. [15]). Similarly, a commitment is called *non-persistent* instead of *persistent* if it nullifies itself after that the failure is recognized (cf. [16] for analogous institutional notions). For instance, Ulysses' intent to reach the sirens is persistent, as he continues to strive even when he acknowledges of not being able to be freed from the rope.

Negative Commitment. A negative commitment reflects a negative position of the agent towards a reference. In practice, the agent is committed to *avoid* the situation expressed in the target. The structure is the same as the previous one, apart that the recognition of the target situation (what the agent does *not* want) is encoded in the *failure* field this time. On the other hand, the *success* field reference is a situation in which this "negative" desire has been respected.

[4] The gray transitions are not part of this module, but highlight that *this structure is operational only if some mechanism evaluates the associated expressions.*

The positive and negative specifications of commitment offer two different frames to the modeler. For instance, in 'I want to listen to the sirens before we arrive at destination', reaching the destination can be interpreted as an event making explicit the failure of the commitment. If we consider a rephrasing the previous statement 'I do not want to arrive at destination before listening to the sirens', the success and failure fields are the same of the previous case. As a human reader, however, we recognize that the two phrases transport a different *pragmatic* meaning. The first case is clearly a matter of direct planning; in the second, there is an implicit reference to something that is blocking the path towards the desired outcome. See § 4.1.

No Commitment. A no commitment position corresponds to the absence of commitment towards the reference. Consequently, there are no failures, no successes to be accounted.

3.2 Expectation

If commitments are essential for the definition of the *subject*, expectations reflect the *situatedness* of the subject in the world. What the agent expects from the world is what he believes the world is, *actually* and *potentially.*

Positive Expectation. The structure of expectation is the same as that of commitment, and, therefore, a similar analysis of the components applies. If there is no precondition, the expectation is a *belief* about what is currently holding. The target defines the propositional content of the belief. The precondition specifies how the expectation can be formed. The success/failure referents are used to specify the means to verify/defeat the expectation, and they are usually built upon primitive perceptions or on other expectations. Differently, the expiration condition can be used to put a limit to such expectations, e.g. 'after the rain, the wind flows for a couple of hours'.

Negative Expectation. A negative expectation specifies what the agent thinks it is not the case or impossible. For instance 'Sirens do not exist'. It can be used to include constraints in the knowledge base of the agent.

No Expectation. This position states that the agent has not constructed any belief about the matter: it is an *agnostic* position. For instance, 'I do not know whether sirens exist', or 'I don't know whether people follow the sirens when they hear their voice'.

3.3 Affordance

In its traditional form, an *affordance* reifies the possibility of the agent to adopt a certain behaviour in certain conditions to achieve a certain result [17]. Affordances interact with commitments to define which behaviour the agent will perform.

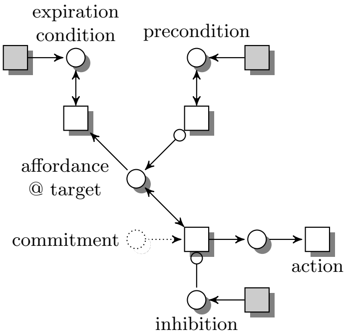


Fig. 3. Operational structure of affordance.

Positive Affordance. The structure to specify affordances is illustrated in Fig. 3. Its components can be compared to those of action languages: the precondition is basically a *pre* element, while the target identifies what the agent would achieve with that action, a subset of a traditional *post* element. Precondition and expiration elements are like in the previous structures. The *action* element corresponds to a plan of actions. The *inhibition* field identifies a situation in which the affordance, although available, cannot be used. This can be exploited to solve conflicts raised at motivational level. See § 4.1.

Negative Affordance. Negative affordance, or negative power, reflects on the ability of the agent to adopt a certain plan to *prevent* to reach a certain state of affairs. For instance, Ulysses perceives that his plan provides him with a negative affordance about falling for the sirens. In principle, the situation expressed by the target represents what is going to occur if the action is not performed.

Disability. The absence of any type of association between the behaviour of the agent to a certain target corresponds to *disability*. In general, given a certain action, infinite disabilities can be expressed (all of that cannot be achieved via that action). This would be redundant and uninteresting information to be maintained. When disability is expressed explicitly, usually it is because it settles domain limitations to an existing ability.

3.4 Susceptibility

The positional triple correlative to affordance is susceptibility. The agent is susceptible to a certain event if he has some reaction to its occurrence, at least at epistemic level.

Positive Susceptibility. Positive susceptibility describes the attention of the agent around a potential situation, identified by the target, whose occurrence is associated to a certain *reaction*. The structure is similar to that of affordance, apart from the renaming of action with reaction. Interestingly, all conditional positions can be transformed using susceptibility. A more general mechanism is then unveiled: all situations accounted in the structures have to be aligned with adequate susceptibilities. See § 4.2.

Negative Susceptibility. At the opposite polarity, we find negative susceptibility. If the stimulus described by the *target* occurs, then the agent generates a negative commitment towards what is specified in the reaction field. The sailors on Ulysses' boat receives orders from their captain to untie him. While usually they would have followed those, now they avoid doing what he asks.

No Susceptibility. The absence of susceptibility is *unresponsiveness* or no susceptibility. For instance, because of the wax in the ears, sailors are unresponsive to the voice of the Sirens. As with no-affordance, this serves to define the boundaries of an existing domain of susceptibility.

4 Operationalization

4.1 From Commitment to Action

To operationalize the connection between commitment and action, we relate these four cognitive components using the *prevent-acquire-cure-keep* (PACK) psychological framework [18]: the presence/absence of a positive/negative condition guides the agent to select a certain behaviour, in order to promote/demote such condition. This can be translated with the following rules:

Acquire (A). If you have a commitment towards a certain target, not holding at the moment, and an associated affordance is available, then use it.

Keep (K). If you have a commitment towards a certain target, which is holding at the moment, and you have a negative affordance associated with its negation, then use it. Furthermore, if there are available affordances that may produce this outcome as expected side-effect, inhibit them.[5]

Prevent (P). If you have a negative commitment towards a certain target, which is not holding at the moment, and you have a negative affordance towards it, then use such affordance. Similarly to the *keep* case, we also have to consider to inhibit affordances with undesired side-effects.

Cure (C). If you have a negative commitment towards a certain target, which is holding at the moment, and you have the affordance associated with its negation, then use such affordance.

Current agent platforms focus mostly on the first reasoning pattern. Integrating the others, however, we are able to explain the pragmatic difference between the two rephrasing of commitment given in § 3.1. The positive characterization triggers a mechanism A targeting the goal; the negative one a mechanism P in order to avoid not to bring about the goal. The two frames activate and interact with different distributions of affordances and expectations.

[5] This solution is simplistic: mutually excluding commitments would inhibit both affordances. A natural correction would be to introduce priorities between commitments.

4.2 From Commitment to Monitoring

Not all what the agent may perceive or infer from his knowledge is *relevant* to his commitments. The relevance relation can be extracted from the commitment specifications, considering two directions: *forward*, i.e. identifying potential situations enabling changes, because of preconditions, expiration conditions related to current positions and potential positions addressed by the PACK; *backward*, i.e. circumscribing success and failure references, necessary to readdress the current configuration. Both forward and backward components are required for the viability of the system. The first allows to respond adequately to changes in the environment. The second provides the means for steering, enabling repair, and for reifying errors in expectations, useful for adaptation purposes.

Identifying the primitive expectations relevant to a certain commitment configuration is however not sufficient. Expectations may for instance expire, the agent still requiring such knowledge. Agents have typically to start a specific course of action to retrieve missing information. This evidence-oriented focus is particularly relevant for our intended institutional applications (see § 1.1).

5 Discussion and Further Developments

The paper traces an outline of an agent architecture based on commitments, expectations, affordances, and susceptibilities, distinguished in positive, negative and null positions. For reasons of space, it overlooks technical details, preferring to give a wide overview of the system. A preliminary proof-of-concept of the architecture is being developed at the moment, and we are evaluating further elements (for instance, the suspension of commitments, cf. [19]).

Our aim is to fill a representational gap experienced while modeling complex institutional scenarios with current agent-based platforms. Such models are intended to be used for simulation, and for model-based diagnosis or similar abductive processes. Interestingly, the resulting architecture may be used as well to model characters in narratives, as the proposed Ulysses' story.

The agents we target are *non-reflective*: they cannot modify their own scripts. However, contrary to what the Petri net notation may suggest, they may show proactivity, as a consequence of maintenance goals.

Another important issue we are confronted with is of a computational nature. Kowalski et al. have convincingly argued that there are widespread confusions about the different natures of production and declarative rules [20], which we think can be aligned with some of the problems observable in e.g. deontic logic with contrary-to-duty obligations, in analytic philosophy with dispositions, etc. Our hypothesis is that, focusing on a computational model like Petri nets, constructed upon the notion of causation (and therefore, on local states), we are able to put aside the problem, still exploiting the advantage of formal grounding.

In the past, other authors worked already on connecting agents with Petri nets [21–23], especially for model checking reasons, but the models they proposed do not primarily focus on agent cognition. This work essentially aims to start filling this gap.

References

1. Sileno, G., Boer, A., van Engers, T.: On the interactional meaning of fundamental legal concepts. In: Proc. Int. Conf. Legal Knowledge and Information Systems - JURIX 2014, pp. 39–48 (2014)
2. Hohfeld, W.N.: Fundamental legal conceptions as applied in judicial reasoning. The Yale Law Journal **26**(8), 710–770 (1917)
3. Boer, A., van Engers, T.: An agent-based legal knowledge acquisition methodology for agile public administration. In: Proc. Int. Conf. on Artificial Intelligence and Law - ICAIL 2011, pp. 171–180 (2011)
4. Lehmann, J., Breuker, J.A., Brouwer, P.W.: Causation in AI & Law. Artificial Intelligence and Law **12**(4), 279–315 (2004)
5. Pennington, N., Hastie, R.: Reasoning in explanation-based decision making. Cognition **49**, 123–163 (1993)
6. Bordini, R.H., Hübner, J.F., Wooldridge, M.: Programming multi-agent systems in AgentSpeak using Jason. John Wiley & Sons Ltd. (2007)
7. Rao, A.S.: AgentSpeak (L): BDI agents speak out in a logical computable language. In: Proc. Workshop on Modelling Autonomous Agents in a Multi-Agent World (1996)
8. Dastani, M.: 2APL: a practical agent programming language. Autonomous Agents and Multi-Agent Systems **16**(3), 214–248 (2008)
9. Hindriks, K.V.: Programming rational agents in GOAL. In: Multi-agent Programming: Languages, Platforms and Applications, pp. 119–157 (2009)
10. Kowalski, R., Sadri, F.: From Logic Programming towards Multi-agent systems. Annals of Mathematics and Artificial Intelligence **25**, 391–419 (1999)
11. Costantini, S., Tocchio, A.: DALI: An Architecture for Intelligent Logical Agents. In: Proc. Workshop on Architectures for Intelligent Theory-Based Agents (AITA) (2008)
12. Blanché, R.: Sur l'opposition des concepts. Theoria **19**(3), 89–130 (1953)
13. Béziau, J.Y.: The Power of the Hexagon. Logica Universalis **6**(1–2), 1–43 (2012)
14. Bobbio, A.: System modelling with Petri nets. In: Systems Reliability Assessment, pp. 102–143 (1990)
15. Hindriks, K.V., van Riemsdijk, M.B.: Satisfying maintenance goals. In: Baldoni, M., Son, T.C., van Riemsdijk, M.B., Winikoff, M. (eds.) DALT 2007. LNCS (LNAI), vol. 4897, pp. 86–103. Springer, Heidelberg (2008)
16. Governatori, G., Rotolo, A.: Norm compliance in business process modeling. In: Dean, M., Hall, J., Rotolo, A., Tabet, S. (eds.) RuleML 2010. LNCS, vol. 6403, pp. 194–209. Springer, Heidelberg (2010)
17. Gibson, J.: The ecological approach to visual perception. Houghton Mifflin, Boston (1979)
18. Ogilvie, D.M., Rose, K.M.: Self-with-other representations and a taxonomy of motives: two approaches to studying persons. Journal of Personality **63**(3), 643–679 (1995)
19. Meneguzzi, F., Telang, P., Singh, M.: A first-order formalization of commitments and goals for planning. In: Proc. 27th AAAI Conference on Artificial Intelligence, pp. 697–703 (2013)
20. Kowalski, R., Sadri, F.: Integrating logic programming and production systems in abductive logic programming agents. In: Polleres, A., Swift, T. (eds.) RR 2009. LNCS, vol. 5837, pp. 1–23. Springer, Heidelberg (2009)

21. Behrens, T., Dix, J.: Model checking with logic based petri nets. In: Proc. of CLIMA (2007)
22. Celaya, J.R., Desrochers, A.A., Graves, R.J.: Modeling and analysis of multi-agent systems using petri nets. Journal of Computers **4**(10), 981–996 (2009)
23. Purvis, M., Cranefield, S.: Agent modelling with petri nets. In: Proc. Computational Engineering in Systems Applications - CESA 1996, pp. 602–607 (1996)