A Modal Logic for Reasoning under the Good Faith Principle

Clara Smith[1], Ana María Monteiro[2]

Facultad de Informática, Universidad Nacional de La Plata.
Calle 50 esquina 115, primer piso. 1900 La Plata, Argentina.
csmith@info.unlp.edu.ar

Faculdade de Campo Limpo Paulista.
Campo Limpo, SP, Brasil.
anammont@ic.unicamp.br

Abstract. This paper describes issues encountered when trying to formalize the conception of *bona fide* as a legal interpretation method. We propose an objective trust delegation mechanism for reasoning about good faith in a multi agent environment which allows different accounts of agency.

Keywords. Legal reasoning, objective trust delegation mechanism, modal logics, agency.

1 Introduction

An act carried out in good faith is an act carried out honestly. From a juridical point of view notions such as good faith, complete trust or confidence are assumed to be fundamental legal principles. Good faith is not *a* norm, it is an *interpretation method* i.e. a criteria for rule application [1]. It may acquire several other abstract shapes e.g. *duty of loyalty, duty of cooperation, duty to inform.* It also limits another legal principles as the *freedom of contract principle* (see [2] for a legal analysis.) When the legislation needs to precise it in an indubitable and strict way, good faith explicitly raises under different manners in many norms. For example, it is usually concretely alluded in juridical institutions such as possession of profits, putative marriage, putative payment, *et cetera.* In these institutions good faith is present in the internal state of an agent, who presupposes that a given situation is real and it conforms the law. Good faith is also of crucial importance for the conduct of international relations in general and therefore recognized as an international principle according to the terms of the Vienna Convention.

In the automated legal reasoning area, different notions of trust have been defined. Linington et al. [3] conjectured the importance of the application of trust metrics to e.g. contracts and Shand and Bacon [4] also do, adopting a second order model of trustworthiness based on an extension of the Dempster-Shafer theory of evidence.

J. Padget, R. Neira, J. Díaz de León (Eds.). e-Government and e-Democracy: Progress and Challenges
© IPN, Mexico 2004.

Castelfranchi and Falcone [5] define a *degree of trust* as a combination of belief types agents might hold with regard to other agents. These three accounts to trust are subjective. Witkoswki et al. [6] propose an objective approach where agents select who they will trade with on the basis of a trust measure built on past experiences (although in our understanding the approach is *quasi-objective* as there is no authority or third party involved in the trust measurement i.e. there is no *pooling of trust.*) Throughout the paper we use the (legal) expression *good faith* and employ the term *trust* in its broadest sense i.e. as a synonym of *good faith.* Our "working definition" of good faith is thus given by the logics we construct.

We develop a theory for reasoning about presence of good faith in actions from an objective point of view, based on the idea that good faith in its wide-ranging sense is a *modality of behavior* which *modalises* any account of agency and involves a degree of interpretation. Our approach is mainly *juridical* although core guidelines are applicable to a context of a multi agent society (MAS).

We characterize in section two a modal logic simple enough for reasoning about good faith in acts. In section three we use as a basis the logic in section two to outline a more structured objective trust delegation mechanism in a multi agent logical framework [7]. Section four gives mathematical content to the principles given in section three. Section five outlines further considerations regarding good faith, collective actions and institutionalized power. Concluding remarks and lines of research are pointed out in section six.

2 Good Faith as a Modality

In this section we give a formal objective definition of the good faith principle keeping things as general as possible. We assume as a basis the basic modal language consisting in atomic propositional sentences A, B,... and complex expressions classically formed from these by means of $\sim, \wedge, \vee, \rightarrow, \leftrightarrow$. Actions are modeled barely, e.g. without paying attention to time or to who the agent is. We understand good faith as a modality of action, so we define an operator called G which focuses exclusively on the state of affairs around the good faith exhibited in sentences. Being A an *act description* we read the expression GA as 'A is carried out in (presence of) good faith' or 'A is behaved in good faith'. The logic of this G-modality can be described as an extension of propositional calculus (PC) with the following rule and axiom schemas:

$$(GR)\ A\,/\,GA \tag{1}$$

$$(GK)\ G(A \rightarrow B) \rightarrow (GA \rightarrow GB) \tag{2}$$

$$(G4)\ GGA \rightarrow GA\ . \tag{3}$$

The logics of G is classical, i.e. closed under logical equivalence:

$$\frac{A \leftrightarrow B}{GA \leftrightarrow GB}\ . \tag{4}$$

This *G-system* is a normal modal logics [8]. Rule GR (1) modalises provable sentences by stacking Gs in front. It formalizes good faith as an interpretation rule: good faith is presumed by the system i.e. if A is a theorem then GA is a theorem. This means there is a central trust authority who makes the presumption e.g. the juridical ordering. GK (2) states the logical implication closure that holds for any modal operator. It traduces a G-modalised formula in an implication, allowing classical reasoning inside the modal context. $G4$ (3) declares a kind of idempotence property of G. We avoid using any dual operator and read $\neg GA$ as ′ absence of good faith in A′ .

This is probably the most simple approach to an objective logic for good faith. Even more, it is solely in virtue of GR that we could claim that this system in a wide-ranging way models the good faith principle. Stated in this straightforward manner the semantics of the G-system collapses to a usual Kripke model semantics for modal operators [8]. Its soundness and completeness results fall into the S4 (KT4) class of reflexive transitive frames.

3 Good Faith, Obligations and Agency

From a juridical point of view, the notion of good faith as an interpretation method is clearly related to the notion of *obligation*. We are therefore mostly interested in connections among good faith, obligations and agency.

As far as it is presented, a G-system serves for reasoning solely about the presence or absence of good faith in acts. What follows is a richer characterization which falls into the class of logical frameworks for modeling norm governed systems and multi agent environments. We attempt to model an objective trust delegation mechanism *wrt* the logic of action described by Jones and Sergot (JS) in [7], called Φ. Φ is wide enough for us to show how to reason under the good faith principle in a complex environment. It has five modalities which we put together with G: we extend Φ with the core schemas (1)-(3) and provide bridging principles among G and Φ-operators. We on purpose put away the *counts as* operator ($\Rightarrow_i$) for institutionalized power and skip our comments on it until section 5.

3.1 Agency

Expressions of the type E_iA are read in Φ ′ agent i brings it about that A′ E_i is a well-known basic operator for agency (see [9] for background.) Sentences like C_iA mean ′i has the ability to produce A′ . Sentences of the form H_iA mean ′i attempts to see to it that A′ . The three modalities are closed under logical equivalence.

Agents have to act (e.g. do, bring about, have the ability to, *et cetera*) for their actions being G-modalised. GE_iA, GC_iA, GH_iA intuitively mean i behaves in good faith. We have then the success conditions:

$$\text{i) } G(E_i A) \rightarrow E_i A, \quad \text{ii) } G(C_i A) \rightarrow C_i A, \quad \text{iii) } G(H_i A) \rightarrow H_i A \; . \tag{5}$$

3.2 Belief

Expressions of the form B_iA are read agent *i* believes that A . For the interpretation of juridical institutions regarding putative (supposed) constituent features (e.g. *putative father, putative marriage*) good faith is present in the internal state of an agent, who presupposes that a given situation is real and it conforms the law. For this kind of analysis the following introspection schemas are useful:

$$\text{i) } G(E_i\text{A}) \rightarrow B_i\text{A} \quad \text{ii) } G(C_i\text{A}) \rightarrow B_i\text{A} \quad \text{iii) } G(H_i\text{A}) \rightarrow B_i\text{A} \tag{6}$$

which represent the expected juridical conjecture on the correspondence between agent s behavior and agent s beliefs: if he or she behaves in good faith is because he or she believes in A ((6)-like schemas for further agency operators are written the same way.) Notice that no modern juridical ordering can go *further inside* an agent than outlined in (6). Furthermore, we avoid converse schemas e.g. $B_i\text{A} \rightarrow G(E_i\text{A})$ as they would represent a strong presupposition on the correspondence between agent s subjectivity and agent s behavior. It is natural to assume that in spite of its beliefs the agent is (relatively) *free* to choose to behave in an untrustworthy manner or to act honestly.

3.3 Directive and Evaluative Modalities

The operator O is designated in Φ to specify obligations: what agents are obliged to do. I_i is a modality specifying that for a given agent *i*, something is ideal.

Next schemas link I and O with G. They model the idea of law presupposing good faith in acts.

$$\text{i) } I_iO\text{A} \wedge E_i\text{A} \rightarrow G(E_i\text{A}), \ \text{ii) } I_iO\text{A} \wedge C_i\text{A} \rightarrow G(C_i\text{A})$$
$$\text{iii) } I_iO\text{A} \wedge H_i\text{A} \rightarrow G(H_i\text{A}) \ . \tag{7}$$

i.e. if it is ideal for *i* that A is obligatory and *i* somehow does A, then such action conforms the good faith principle (e.g. *i* acted in good faith.) In turn we derive:

$$\text{i) } O\text{A} \wedge E_i\text{A} \rightarrow G(E_i\text{A}), \ \text{ii) } O\text{A} \wedge C_i\text{A} \rightarrow G(C_i\text{A})$$
$$\text{iii) } O\text{A} \wedge H_i\text{A} \rightarrow G(H_i\text{A}) \ . \tag{8}$$

which take as starting point directive sentences (we put the evaluative modality aside.) In a similar way we suggested for B_i, schema $I_i\text{A} \wedge E_i\text{A} \rightarrow G(E_i\text{A})$ is a strong assumption on i s subjectivity, so we avoid it. Finally, the idea that good faith is presupposed by law is captured in the following bridging principle:

$$O\text{A} \rightarrow G\text{A} \tag{9}$$

which is a general form of (8).

We can appreciate one of the advantages of an authorized and centralized good faith reasoning mechanism is that lack of trust can be improved. Undoubtedly, due to the objective approach, there is a poor degree of scalability, although e.g. previously unknown-to-each-other parties are free to enter into a precontractual bargaining process.

Schemas (7), (8) and (9) impact mainly on the formalization of e.g. (strictly speaking) legal systems because they provide a mean for interpreting agency and therefore for applying the law. Being general interpretation rules they appear to be too strong restrictions for systems governed by norms of various kinds where good faith is not objectively considered (e.g. it is modeled as complex internal states of agents.) Formal connections among objective and subjective trust mechanisms are to be investigated.

3.4 Necessity

In Φ the necessity operator (say N) has a usual S5 (KT5) semantics. We give:

$$NA \wedge OA \wedge E_i A \rightarrow G(E_i A) \qquad (10)$$

which is a variant of (8) (similarly for C_i and H_i.). Schema $NA \wedge E_i A \rightarrow G(E_i A)$ is clearly avoided.

4 Semantics for G, given ϕ

Next we give a mathematical content to axioms proposed in section two and three. Semantics for Φ-operators are detailed in [7]. We bring in the necessary details for making the semantics of G clear. We use as a basis JS´ s definitions (which in turn recall the ones in [9]):

· M = $(W, f_N, f_{Ci}, f_{Hi}, f_O, f_{Ii}, f_{Bi}, f_G, V)$ is a traditional model structure, with W a set of possible worlds, V a valuation function assigning to each sentence of Φ a set of possible worlds. The f members of M are functions employed in the specifications of the truth conditions for modal sentences. Function f_N is a unary function selecting for each world the set of propositions (i.e. a set of possible worlds) that are necessary relative to that world. Function f_{Ci} is a binary function that picks up the worlds where agent i realizes the ability he has in a world to bring about A. f_{Hi} selects for each world the set of propositions corresponding to the state of affairs i attempts to bring about at that world. Function f_O gets for each world the set of propositions obligatory at that world. Function f_{Ii} gets the set of propositions ideal for i at a given world. f_{Bi} picks up for each world the set of propositions believed by i at that world.

We introduce slightly variations on the original JS´ s model structure M: we put away $f_{\Rightarrow}$, and add the unary function f_G we define along the way.

· α, β are worlds, any members of W.

· $\|A\|^M$ is the truth set in M for A i.e. the set of worlds in M at which A is true, i.e. $\|A\|^M =_{def} \{\alpha: M,\alpha \models A\}$.

Function f_G picks up for each world the set of propositions that can be carried out in good faith in that world, written:

$$f_G(\alpha) \ . \tag{11}$$

Truth of a sentence of the form GA at a world α in a model M is specified as follows:

$$M, \alpha \models GA \text{ iff } \|A\|^M \in f_G(\alpha) \tag{12}$$

(12) is a definition of the presumption of good faith in acts: it means that in all worlds where A is true, it happens GA is also true.

Let X and Y be any subset of W. For the function f_G we adopt the followings constraints:

$$(cf_G) \text{ if } X \in f_G(\alpha) \text{ and } Y \in f_G(\alpha) \text{ then } X \cap Y \in f_G(\alpha) \ . \tag{13}$$

$$(tf_G) \text{ if } X \in f_G(\alpha) \text{ and } Y \in f_G(\alpha) \text{ then } \alpha \in X \ . \tag{14}$$

$$(4f_G) \text{ if } X \in f_G(\alpha) \text{ and } Y \in f_G(\alpha) \text{ then } \{\beta: \beta \in f_G(\beta)\} \in f_G(\alpha)\} \ . \tag{15}$$

From definition (12) and conditions (cf_G), (tf_G), $(4f_G)$, the validity of schemas (1)-(4) immediately follows.

The validity of the success condition (5i.) is secured by adopting the following constraint:

$$\text{if } \{\beta: \beta \in f_{Ci}(\beta, \|A\|^M)\} \in f_G(\alpha) \text{ then } \alpha \in f_{Ci}(\alpha, \|A\|^M) \tag{16}$$

given the constraint (for E_iA expressions): $M, \alpha \models E_iA$ iff $\alpha \in f_{Ci}(\alpha, \|A\|^M)$ [7].

The validity of the success conditions (5ii.) and (5iii.) are secured by adopting the following constraints:

$$\text{if } \{\beta: f_{Ci}(\beta, \|A\|^M) \neq \emptyset\} \in f_G(\alpha) \text{ then } f_{Ci}(\alpha, \|A\|^M) \neq \emptyset \text{ and} \tag{17}$$

$$\text{if } \{\beta: \|A\|^M \in f_{Hi}(\beta)\} \in f_G(\alpha) \text{ then } \|A\|^M \in f_{Hi}(\alpha),$$

provided the truth conditions for C_iA and H_iA are [7]:

$M, \alpha \models C_iA$ iff $f_{Ci}(\alpha, \|A\|^M) \neq \emptyset$ and $M, \alpha \models H_iA$ iff $\|A\|^M \in f_{Hi}(\alpha)$.

The introspection schema (6i.) is validated by building the constraint:

$$\text{if } \{\beta: \beta \in f_{Ci}(\beta, \|A\|^M)\} \in f_G(\alpha) \text{ then } \|A\|^M \in f_{Hi}(\alpha) \ . \tag{18}$$

Schema (6ii.) is validated by adopting the constraint:

$$\text{if } \{\beta: f_{Ci}(\beta, \|A\|^M) \neq \emptyset\} \in f_G(\alpha) \text{ then } \|A\|^M \in f_{Hi}(\alpha) \ . \tag{19}$$

And (6iii.) is secured by:

$$\text{if } \{\beta: \|A\|^M \in f_{H_i}(\beta,)\} \in f_G(\alpha) \text{ then } \|A\|^M \in f_{B_i}(\alpha) \ . \tag{20}$$

Schema (7i.), which links I and O with *G*, is validated by adopting the constraint:

$$\text{if } \{\beta: \|A\|^M \in f_O(\beta)\} \in f_{I_i}(\alpha) \text{ and } \alpha \in f_{C_i}(\alpha, \|A\|^M) \text{ then } \{\beta: \beta \in f_{C_i}(\beta, \|A\|^M)\} \in f_G(\alpha) \ . \tag{21}$$

Next, schema (7ii.) is validated by the condition:

$$\text{if } \{\beta: \|A\|^M \in f_O(\beta)\} \in f_{I_i}(\alpha) \text{ and } f_{C_i}(\alpha, \|A\|^M) \neq \emptyset \text{ then } \{\beta: f_{C_i}(\beta, \|A\|^M) \neq \emptyset\} \in f_G(\alpha) \ . \tag{22}$$

and (13iii.) is validated by adopting the constraint:

$$\text{if } \{\beta: \|A\|^M \in f_O(\beta)\} \in f_{I_i}(\alpha) \text{ and } \|A\|^M \in f_{h_i}(\alpha) \text{ then } \{\beta: \|A\|^M \in f_{H_i}(\beta)\} \in f_G(\alpha) \ . \tag{23}$$

The validity of (8.i) is secured by:

$$\text{if } \|A\|^M \in f_O(\alpha) \text{ and } \alpha \in f_{C_i}(\alpha, \|A\|^M) \text{ then } \{\beta: \|A\|^M \in f_{C_i}(\beta, \|A\|^M)\} \in f_G(\alpha) \ . \tag{24}$$

For (8.ii) and (8.iii) we adopt the following truth conditions:

$$\text{if } \|A\|^M \in f_O(\alpha) \text{ and } f_{C_i}(\alpha, \|A\|^M) \neq \emptyset \text{ then } \{\beta: f_{C_i}(\beta, \|A\|^M) \neq \emptyset\} \in f_G(\alpha) \tag{25}$$

and

$$\text{if } \|A\|^M \in f_O(\alpha) \text{ and } \|A\|^M \in f_{h_i}(\alpha) \text{ then } \{\beta: \|A\|^M \in f_{H_i}(\beta)\} \in f_G(\alpha). \tag{26}$$

Schema (9) is secured by:

$$\text{if } \|A\|^M \in f_O(\alpha) \text{ then } \|A\|^M \in f_G(\alpha) \ . \tag{27}$$

Finally, (10) is validated by:

$$\text{if } \|A\|^M \in f_M(\alpha) \text{ and } \|A\|^M \in f_O(\alpha) \text{ and } \alpha \in f_{C_i}(\alpha, \|A\|^M) \text{ then } \{\beta: \beta \in f_{C_i}(\beta, \|A\|^M)\} \in f_G(\alpha) \ . \tag{28}$$

5 Good Faith, Other Accounts of Agency, Institutionalized Power

The given formal approach to good faith is general enough to deal with several accounts of agency, as it is likely to be expected from a legal (juridical) perspective. It is not our purpose to investigate here different accounts of agency (specific bridging principles need to be built) but we can e.g. use G as a tool for reasoning about presence of good faith in coordinated actions. For example, as illustrated in [10] the expression $E_{\{i,j,k\}}$A means that agents i, j, k collectively bring about that A. If we assume individual behavior can not be identified, clearly the semantics of expressions of the form $G(E_{\{i,j,k\}}A)$ falls into the semantics already exposed.

A brief comment on the *counts as* ($\Rightarrow_s$) operator. A $\Rightarrow_s$ B sentences express the idea that, in institution s (e.g. a church), given A we have B (e.g. provided witnesses and special words said, a couple counts as married.) Therefore $\Rightarrow_s$ is an institution-sensitive good faith operator. Already argued by Gelati et al [10], we think that the type of reasoning involved in institutional and normative domains are essentially the same. We write thus an equivalent expression for A $\Rightarrow_s$ B i.e. GA $\rightarrow$ B.

6 Concluding Remarks and Lines of Research

We point out the following:

· Good faith as a legal interpretation method is objective and wide-ranging. We defined an objective (Φ-derived) trust delegation mechanism for making better the juridical understanding of different accounts and regulations of agency.

· Meaningful relations among good faith, other accounts of agency and institutionalized power are actually being addressed.

· We are also motivated by the hypothesis that a wide number of plausible state of affairs over which we can reason about presence of good faith are inconsistent. Reasoning objectively about good faith in acts should also handle reasoning with inconsistency: behaviors can be logically contradictory, however, opposite actions can be behaved in good faith.

· Connections between objective and subjective trust models for reasoning under the good faith principle are actually being adressed.

References

1. Borga, E.: Buena Fe. In: Enciclopedia Jurídica Omeba, Vol 2, pp 403-410. Editorial Bibliográfica Argentina, Buenos Aires.
2. Musy, A.: The Good Faith Principle in Contract Law and the Precontractual Duty to Disclose: Comparative Analysis of New Differences in Legal Cultures. In: Global Jurist Advances, Vol.1, Issue 1, 2001. The Berkeley Electronic Press (bepress) [Online]. Available: http://www.bepress.com/gj.

3. Linington, P.F., Milosevic, Z., Raymond, K.: Policies in Communities: Extending the ODP enterprise viewpoint. In: 2nd International Enterprise Distributed Object Computing Workshop, pp. 11-22, 1998.
4. Shand, B., Bacon, J.: Policies in Accountable Contracts. [Online]. Available: http://www.cl.cam.ac.uk/Research/SRG/opera/publications/Papers/S3_shand_b_distrib.pdf
5. Castelfranchi, C., Falcone, R.: Principles for Trust on MAS: Cognitive Anatomy, Social Importance, and Quantification. In: ICMAS ' 98, pp.7279.
6. Witkowski, M., Artikis, A., Pitt, J.: Trust and Cooperation in a Trading Society of Objective-Trust Based Agents. Proc. Autonomous Agents 2000 Workshop on Deception, Fraud and Trust in Agent Societies, pp. 127-136. [Online]. Available: http://www.iis.ee.ic.ac.uk/~m.witkowski/cogrob2/papers/aa00ws.pdf
7. Jones, A.J.I., Sergot, M.: Norms and Institutions. Tutorial given at ICAIL 2003, Edimburgh, Scotland, UK. Notes include part of the chapter "A Logical Framework", the final edited version in J. Pitt, ed. The Open Agent Society, John Wiley & Sons, Chichester, UK, 2003.
8. Blackburn, P., de Rijke, M., Venema, I.: Modal Logic. Cambridge Tracks in Theoretical Computer Science, 2001, Cambridge University Press.
9. Elgesem, D.: The Modal Logic of Agency. Nordic Journal of Philosophical Logic. Vol. 2, No. 2. pp. 1-46. Scandinavian University Press, 1997. [Online]. Available: http://www.hf.uio.no/filosofi/njpl/vol2no2/agency/agency.pdf
10. Gelati, J., Governatori, G., Rotolo, A., Sartor, G.: Declarative Power, Representation and Mandate: A Formal Analysis. In: JURIX' 02, pp. 131147. IOS Press, Amsterdam, 2002.