



ELSEVIER

Journal of Cognitive Systems Research 2 (2001) 39–54

**Cognitive Systems**  
RESEARCH

www.elsevier.com/locate/cogsys

# Socio-cognitive mechanisms of belief change

## Applications of generalized game theory to belief revision, social fabrication, and self-fulfilling prophesy

Action editor: Ron Sun

Tom R. Burns<sup>a,\*</sup>, Anna Gomolińska<sup>b</sup>

<sup>a</sup>*Uppsala Theory Circle, Department of Sociology, University of Uppsala, Box 821, 75108 Uppsala, Sweden*

<sup>b</sup>*Department of Mathematics, University of Białystok, Akademicka 2, 15267 Białystok, Poland*

Received 15 October 2000; accepted 31 October 2000

---

### Abstract

In this article, generalized game theory (GGT) is used to conceptualize and explain key socio-cognitive processes in multi-agent interaction, in particular belief revision. GGT is based on the mathematics of rules and rule complexes (drawing on developments at the interface of mathematics, logic, and computer science). Rule concepts are used to formalize game, social relationships, and role as well as a major component of role, namely model or belief structure. This is an agent's 'situational view,' providing a perspective on and a basis for understanding and analyzing interaction situations with others. GGT conceptualizes the way that actors, when confronted with new information or candidates for belief, integrate them into their models, or reject them. This occurs through *rules of composition*. Several social factors can be identified as key variables incorporated or expressed in composition rules and judgments which regulate belief revision and learning processes: (1) degree of trust in a source of belief or message; (2) the social status (professional expertise, ethnicity, gender, age, etc.) of the source relative to the recipient; (3) the strength of commitment with respect to a belief structure; and (4) the strength of collective sanctioning. The theory is applied to multi-agent games, where the social relationships among actors, status and authority differences, the level of trust and expected honesty affect belief change — in large part by affecting the composition rules which are applied to 'candidates for belief'. The article shows that in some cases of belief revision falsehood is produced — indeed, deception and fabrication are part and parcel of many multi-agent interaction systems. However, in social life, even false beliefs — produced through acceptance of expert or authoritative judgements or beliefs — may become true through self-fulfilling processes. © 2001 Elsevier Science B.V. All rights reserved.

*Keywords:* Generalized game theory; Rule complex; Social role; Model; Composition rules; Belief revision; Collective ignorance; Fabrication; Self-fulfilling prophesy

---

---

\*Corresponding author. Tel.: +46-18-471-1203; fax: +46-18-471-1170.

E-mail addresses: tom.burns@soc.uu.se (T.R. Burns), anna.gom@math.uw.bialystok.pl (A. Gomolińska).

## 1. Introduction

This article applies the generalized game theory (GGT) to individual and collective learning processes, focusing, in particular on changes in beliefs: (1) it shows how belief change may be facilitated or blocked by the character of the relationships among agents, the level of trust and openness, and status and authority differences — in a word, social context; (2) it also points up the role of deception and fabrication in game situations — that is, the production of false information and beliefs (which, nevertheless, may become true through self-fulfilling prophecies). In a certain sense, one may speak not only about ‘learning’ in its positive meaning but also about the construction of ignorance in multi-agent games.

## 2. Presentation of the theory of games

In their classic work, Von Neumann and Morgenstern (1972: 49) defined a game as simply the totality of the pre-determined rules which describe it and to which players must and do conform. They did not, however, elaborate a theory of rules, or deal with rules as mathematical objects.<sup>1</sup> Systems of rules guiding and regulating social actors in their activities and interactions may be formalized in a uniform and general way by means of *rule complexes*. According to this approach, games are rule complexes where the rules may be imprecise, possibly inconsistent, and open to a greater or lesser extent to modification and transformation by the participants.

The mathematical theory of rules (Burns & Gomolinska, 1998, 2000a,b; Burns, Gomolinska & Meeker, 2001; Burns, Gomolinska, Meeker & De-

Ville, 1998; Gomolinska, 1999) provides a basis for extending and generalizing game theory.<sup>2</sup> (The mathematics is based on contemporary developments at the interface of mathematics, logic, and computer science). A major generalized game theory (GGT hereafter) concept is rule complex.<sup>3</sup> The motivation behind development of this concept has been to consider repertoires of rules in all their complexity with complex interdependencies among the rules and, hence, to not merely consider them as sets of rules. The organization of rules in rule complexes provides us with a powerful tool to investigate and describe various sorts of rules with respect to their

<sup>2</sup>Rules are a type of knowledge, which is related to action and the processes which feed into and organize and regulate action (Burns & Flam, 1987; Burns, 1990). An abstract conception of a rule may be expressed as follows (Burns & Gomolinska, 1998; Gomolinska, 1999):

$$r: \frac{X:Y}{\gamma}$$

where  $X$  is the set of premises or conditions and  $Y$  is the set of justifications (default provisions or exception conditions), and  $\gamma$  is the ‘conclusion’. The latter either provides information, evaluation, or a directive or requirement for action (in this case the actor is supposed to perform it).  $(X, Y, \gamma) \in r$  informally reads as follows: if all premises of  $X$  hold and all justifications of  $Y$  may hold, then conclude  $\gamma$  according to  $r$ .  $\alpha$  may hold (that is in any given situation, an actor may not know that  $\alpha$  does not hold but acts presuming that it does hold. When  $\alpha$  is a justification of a rule  $r$  and  $\alpha$  does not hold, then an ‘exception’ obtains, that is the rule  $r$  is not applied).

If the set of justifications is non-empty, then the rule is a kind of proper default rule (Reiter’s default logic (Reiter, 1980)). If the set of premises and justifications are both empty, then the rule is ‘a fact’ or axiomatic. Axiomatic rules can be viewed as equivalent with their conclusions. Since all formulas can be rewritten in the form of axiomatic rules, the basic objects (points) of our conceptual space are just rules. Letting  $X = \{\alpha_1, \alpha_2, \dots, \alpha_m\}$  and  $Y = \{\beta_1, \beta_n\}$ , rule  $r$  can be written as:

$$r: \frac{\alpha_1, \alpha_2, \dots, \alpha_m; \beta_1, \dots, \beta_n}{\gamma}$$

One can distinguish several types of rules, e.g., declarative, prescriptive, proscriptive, evaluative, decision rules, etc.

<sup>3</sup>A *rule complex* is a set obtained according to the following formation rules: (1) any finite set of rules is a rule complex; (2) if  $C_1$  and  $C_2$  are rule complexes, then their union  $C_1 \cup C_2$  and the power set of  $C_1$ ,  $P(C_1)$ , are rule complexes; (3) if  $C_1 \subseteq C_2$  and  $C_2$  is a rule complex, then  $C_1$  is a rule complex. For any rule complexes  $C_1$  and  $C_2$ , their intersection  $C_1 \cap C_2$ , and their difference  $C_1 - C_2$  are rule complexes as well.

<sup>1</sup>The ‘various rules’ that Von Neumann and Morgenstern (1972) and others refer to may be of a very different character: laws of nature and physical and ecological constraints (nature); norms, laws, and constitutions (culture and institutions); conventions and traditions, interpersonal understandings and sentiments between the particular actors; resources and technologies. This mixing or conflation of radically different types of ‘rules’ is a major flaw in the classical theory. For instance, social rules can and should be analytically distinguished from physical, technical, and other material constraints. What is called for is a systematic theory of rules and rule complexes, to which our work is a contribution.

functions such as values, norms, judgment rules, prescriptive rules, and meta-rules as well as more complex objects consisting of rules such as roles, routines, algorithms, action modalities, models of reality as well as social relationships and games. Informally speaking, a rule complex is a set of rules and/or other rule complexes.

A well-specified game at time  $t$ ,  $G(t)$ , is a particular interaction situation where the participating actors have defined roles and role relationships (although not all games are necessarily well-defined with, for instance, clearly specified and consistent roles and role relationship(s)). In general, actors are involved in a number of social relationships and institutional domains and play a number of different roles. Suppose that actor  $i$  has several roles at  $t$  such as citizen, supervisor, wife, and mother. The notion of a *situation* is a primitive. Situations are denoted by  $S$  with subscripts, if needed. We use the lowercase  $t$ , possibly with subscripts, to denote points of time (or reference). Thus,  $S_t$  denotes a situation at time  $t$ .<sup>4</sup>

Given a concrete situation  $S_t$ , a *general game* is represented as a particular rule complex  $G(t)$ . This complex includes roles as rule subcomplexes along with norms and other rules. Suppose that a group or collective  $I = \{1, \dots, m\}$  of actors is involved in a game  $G(t)$ .  $\text{ROLE}(i, t, G)$  denotes actor  $i$ 's *role complex* in the game  $G(t)$ .  $\text{ROLE}(I, t, G)$  denotes a *role configuration* of all actors in  $I$  engaged in  $G(t)$ .

The individual role,  $\text{ROLE}(i, t, G)$  for every actor  $i = 1, \dots, m$ , is a subcomplex of the role configuration  $\text{ROLE}(I, t, G)$  and the latter complex is a subcomplex of  $G(t)$  in the formalism of rule theory:<sup>5</sup>

$$\text{ROLE}(i, t, G) \subseteq_g \text{ROLE}(I, t, G) \subseteq_g G(t) \quad (1)$$

<sup>4</sup>Given a situation  $S_t$ , we can associate with each actor  $i$  a rule complex  $\text{ACTOR}(i, t)$ , containing all rules the actor  $i$  has in  $S_t$ . We refer to this rule complex as  $i$ 's actor complex at  $t$ . Some rules of  $\text{ACTOR}(i, t)$  relate to  $i$ 's roles at  $t$ . If we neglect or delete the rules that are irrelevant for the topic 'role', we obtain a subcomplex  $\text{ROLE}(i, t)$  of  $\text{ACTOR}(i, t)$ , written  $\text{ROLE}(i, t) \subseteq_g \text{ACTOR}(i, t)$ , referred to as the role complex of  $i$  at  $t$ .

<sup>5</sup>Eq. (1) may be also written as follows (where  $X, Y$  are rule complexes, the notation  $X[Y]$  is to indicate that  $Y \subseteq_g X$ , that is, that  $Y$  is a subcomplex of  $X$ ):

$$G(t)[\text{ROLE}(I, t, G)[\text{ROLE}(1, t, G), \dots, \text{ROLE}(m, t, G)]] \quad (1A)$$

It should be emphasized that role, role configuration, and the game complex  $G(t)$  also contain other rules (or rule complexes) which describe and regulate the game such as the 'rules of the game', general norms, and practical rules such as initiation and stop rules as well as meta-rules,<sup>6</sup> indicating, for instance, how seriously or strict the roles and rules of the game are to be implemented, and possibly the ways to adapt or to adjust the rule complexes to particular situations.

Any game defined as the complex  $G(t)$  involves a set of actors  $I$  who (1) are conscious of being involved in an interaction with others and (2) operate more or less according to a common rule complex, that is shared or common knowledge of the game complex including their respective roles vis-à-vis one another as well as other relevant norms and rules of the game. For our purposes here, a simple, well-defined game is one in which each player occupies a single specified role.

Role relationships provide contextualizing frames of appropriate rules including values and norms, particular ways in which actions are classified and judged, and 'internal' interpretations and meanings (Burns & Flam, 1987). 'Non-cooperation' in, for instance, a prisoners' dilemma (PD for short) situation will not be merely 'defection' in the case that the actors are friends or relatives in a solitary relationship, but a form of 'disloyalty' or 'betrayal' and subject to harsh social judgment and sanction. In the case of enemies, 'defection' in the PD game would be fully expected and considered 'natural' — neither shameful nor contemptible, but right and proper rejection of or damage to the other, and, hence, not a matter of 'defection' at all. Such a perspective on games enables us to systematically identify and analyze their symbolic and moral as-

<sup>6</sup>To enhance the expressive power of the concept of rule complex, we do not separate the object level from meta-levels. In practice, people carry on both the operative and the meta-levels; in part, they switch back and forth or operate simultaneously on both levels. Nevertheless, we are aware of the circularities which may occur. As a consequence, we can uniformly and relatively easily investigate various sorts of rules (e.g., evaluative rules, norms, judgment rules, and action rules) in relation to one another as well as complex objects consisting of rules (e.g., roles, procedures and social algorithms, game complexes, and the models of reality and of actors and their relationships).

pects associated with established social relationships.

An actor's role is specified in GGT in terms of a few basic cognitive and normative components (formalized as mathematical objects in (Burns & Gomolinska, 1998, 2000a,b; Burns et al., 1998, 2001; Gomolinska, 1999). The role complex includes, among other things: particular beliefs or rules that frame and define the reality of relevant interaction situations (the concept of a socio-cognitive model is introduced later); norms and values relating, respectively, to what to do and what not to do and what is good or bad; repertoires of strategies, programs, and routines; and modalities to organize the determination of decisions and actions in relation to particular other agents.

*Judgment* is a central concept in GGT. The major basis of judgment is matching, that is a process of comparing and determining similarity (or dissimilarity) (Burns & Gomolinska, 2000a,b; see also Sun, 1995). Let  $J_t$  be the rule complex or, possibly an algorithm, with which an actor makes the comparison and judgment of the degree of similarity or goodness of fit (or dissimilarity) of two objects of consideration in situation  $S_t$  at time  $t$ . The result of application of  $J_t$  to a pair of objects  $(x_1, x_2)$  —  $J_t(x_1, x_2)$  — is an expression describing in qualitative or quantitative terms whether or not  $x_1, x_2$  are similar and to what extent. It can take the form: 'similar', 'dissimilar', 'not decided', 'almost similar', 'sufficiently similar', 'highly similar', or 'dissimilar to the degree  $d$ ', etc. Associated with the judgment complex  $J_t$ , there is typically a threshold for similarity of prescribed consequences.<sup>7</sup>

The capacity of actors to judge similarity or likeness (that is, up to some threshold, specified by a meta-rule or norm of stringency), plays a major part in belief change and action processes generally. This is also the foundation for rule-following or rule-

application activity (Burns & Gomolinska, 2000a; Gomolinska, 1999).<sup>8</sup>

In an interaction process or game, each actor may plan and construct or find an action  $a$  and its consequences which she believes realizes an appropriate norm or value  $r$  in situation  $S$ . Note that a may be a complex plan involving multiple tasks and subtasks or performances. By the realization of  $r$ , we mean that implementation of a results in *consequences* that match or correspond to those specified or prescribed by the norm or value  $r$ . Realization occurs when the finite set of expected or predicted consequences of  $a$ ,  $\text{Con}(a, t)$  in situation  $S$  at time  $t$ , is judged *sufficiently similar* to the set of those consequences which the norm or value  $r$  prescribes  $\text{Con}(r)$  (for purposes of simplification, we drop  $t$  that contextualizes  $J$ ; it will be understood that it applies for a specific actor or group of actors in a particular situation and time  $t$ ):<sup>9</sup>

### 2.1. The principle of action determination

Construct, or find and select, an action  $a$  in situation  $S$  at time  $t$  which satisfies the following equation:

<sup>8</sup>This relates to the Wittgensteinian problem of 'following a rule'. From the GGT perspective, 'following a rule' (or rule complex) entails *several phases and a sequence* of judgments: in particular, activation and application together with relevant judgments such as judgments of appropriateness for a given situation or judgments of applicability. To apply a rule (or rule complex), one has to know (1) the conditions under which the application is possible or allowed and (2) the particular conditions of execution or application of the rule (in part, whether other rules may have to be applied earlier). The application of a rule (or rule complex) is not then simply a straightforward matter of 'following' and 'implementing' it: the conditions of execution may be problematic; the situation (or situational data) may not fit or be fully coherent with respect to the rule (or rule complexes); actors may reject or refuse to seriously implement a rule (or rule complex); a rule (or rule complex) may be incompatible or inconsistent with another rule that is to be applied in the situation. In general, actors may experience ambiguity, contradiction, dilemmas, and predicaments in connection with 'following a rule' making for a problematic situation and possibly the unwillingness or inability to 'follow a particular rule'.

<sup>9</sup>This entails following or applying a rule in one of its senses (see later). The basic process is one of comparison-test, universal in constructing, selecting, and assessing action.

<sup>7</sup>Consider a membership function  $M_{\{x\}}: U \rightarrow [0,1]$ .  $U$  is the universe of all considered objects. The degree of membership of an object  $y$  in the class of objects with respect to similarity to  $x$  is expressed by  $M_{\{x\}}(y)$ , that is by the value of  $M_{\{x\}}$  on  $y$ .  $M_{\{x\}}(x) = 1$ . Now assume that there is some threshold given by a function  $f$ , that is a number  $z = f(x, y)$  in  $[0, 1]$ . We define that  $y \sim x$  iff  $z \leq M_{\{x\}}(y)$ , that is the degree  $M_{\{x\}}(y)$  is at least equal to  $z$ . Of course, if  $M_{\{x\}}(y) = 1$ , it would not mean in general that  $y = x$  but rather that  $y$  and  $x$  are indiscernible.

$J(\text{Con}(a, t), \text{Con}(r)) = \text{sufficiently similar}$  (2)

This states that the expected or predicted consequences in situation  $S_t$  of action  $a$ ,  $\text{Con}(a, t)$ , are judged to satisfy or match consequences prescribed by the norm or value  $r$ ,  $\text{Con}(r)$ .<sup>10</sup> Action determination is, of course, an *ex ante* process, that is prior to the performance or implementation of the action.

The principle of action determination — corresponding to the principle of maximizing utility in rational choice theory — subsumes several distinct modalities of action determination, each with its own ‘logic’ (Burns & Gomolinska, 2000a,b; Burns et al., 2001). GGT distinguishes such modalities as normative, instrumental, dramaturgical-communicative, and play as well as other modalities of action determination. These are distinguished by  $\text{Con}(r)$  which orient the actor(s) to attending to and trying to regulate corresponding consequences in the actions they construct or consider for choice. In an instrumental modality, for instance, the value of acts derives from assessments or evaluative judgments of action outcomes, whereas the value of action in the case of normative modality derives from judgments of the qualities of the action itself (including possibly the intentionality of the actor).

<sup>10</sup>Consider the case of realization of normative type rules. (1) First, the case of realizing a value. The values of actor  $i$  at  $t$  are represented in the form of evaluative rules  $r_0, \dots, r_m$  in  $\text{VALUE}(i, t)$ . Suppose that  $\gamma_0^*, \gamma_1^*, \dots, \gamma_k^*$  are actual consequences of the rule governed acts/activities performed by the actor. That is, actor  $i$  has already performed some actions when applying a rule or rule complex. Now suppose  $r_j$  has  $\gamma$  as the conclusion. For instance,  $\gamma$  may say:  $\delta$  is good. If there is  $\gamma_n^*$  such that  $\gamma_n^*$  is sufficiently similar to  $\delta$ , then we say that actor  $i$  has realized  $r_j$  at  $t$ . It is because some of the actual consequences are sufficiently similar to value indications about what is considered as good or valuable to the actor  $i$  at  $t$ . (2) Now consider the case of following a norm(s). Suppose that norms of actor  $i$  at  $t$  are represented in the form of normative rules  $r'_0, \dots, r'_m$ . Suppose that  $\gamma_0^*, \gamma_1^*, \dots, \gamma_k^*$  are actual consequences of the rule governed acts/activities performed by the actor, as earlier. And  $i$  judges the actions, that is, their consequences in terms of the relevant norms. Now suppose  $r'_j$  has  $\gamma$  as the conclusion. For instance,  $\gamma$  may say: (a) it ought to be the case that  $\delta$ ; or (b)  $\delta$  is forbidden. If there is  $\gamma_n^*$  such that  $\gamma_n^*$  is sufficiently similar to  $\delta$ , then we say that actor  $i$  has followed (realized)  $r'_j$  at  $t$  in the case (a). In the case (b), we say that actor  $i$  does not followed  $r'_j$  at  $t$ . In (a) (and (b)) it is because some of the actual consequences is sufficiently similar to something which is considered as obligatory (forbidden) to the actor  $i$  at  $t$ .

### 3. Socially embedded belief processes

Actors operate with models in their interactions. The models are made up of beliefs and knowledge structures. Actor  $i$ 's *model* or belief structure of reality in situation  $S_t$ ,  $\text{MODEL}(i, t)$ , is the actor's ‘situational view’, providing a perspective on, and a basis for understanding and analyzing the reality of the situation  $S_t$ . In this perspective, beliefs are rules (Burns & Gomolinska, 2000a).  $\text{MODEL}(i, t)$  consists then of beliefs that the actor  $i$  has about herself and her environment, the interaction conditions, opportunities and constraints, and relevant actors in the situation. The model typically may include beliefs about other actors' expectations or strategic predispositions (Burns et al., 2001). If  $i$  is fixed, we may drop the parameter in  $\text{MODEL}(i, t)$  and write  $\text{MODEL}(t)$  instead.

Through their interactions, actors are confronted with new information which may or may not fit their beliefs. We focus here on the processes of belief dissonance and change, the conditions under which change occurs, and the way in which it takes place. In maintaining or changing their beliefs, actors make judgments, for instance, judgments about the relevance or the reliability of the new information.

In the course of, and as a result of, deliberation, observation or interaction with other actors, the actor  $i$  may acquire new beliefs (or knowledge) and integrate them into her  $\text{MODEL}(t)$ , that is, she introduces new beliefs, revises old ones, removes them, etc. The process of change of beliefs of  $i$  may be seen as a *composition* of her actual model,  $\text{MODEL}(t)$ , with a rule complex consisting of candidates for ‘new’ beliefs, say  $C$ , under a composition rule complex  $D$ . The latter is a judgment complex, in some cases an algorithm, for composing rule complexes. For simplicity, we may assume that the result of such a composition is a rule complex as well:  $\text{MODEL}(t + 1)$  in a resulting situation  $S_{t+1}$  (1 represents simply the next point of time or referent).<sup>11</sup> Thus, for actor  $i$

$$\text{MODEL}(t + 1) = \text{Compose}(\text{MODEL}(t), C, D). \quad (3)$$

<sup>11</sup>One can also imagine the case that the composition yields a set (or class) of models.

To make the picture even simpler we may assume that  $\text{MODEL}(t)$  is composed with a rule complex consisting of a single rule  $r$ , i.e.,  $C = \{r\}$ . In this case we may write:

$$\text{MODEL}(t + n) = \text{Compose}(\text{MODEL}(t), r, D). \quad (4)$$

It should be emphasized that every kind of belief or knowledge change may be viewed as a form of composition of rule complexes, where the application of  $D$  to a new piece of information articulated as a single rule  $r$  results in judgments that maintain or change the model,  $\text{MODEL}(t)$ . As we show later, composition rules are a function of the type of social relationship, the level of solidarity and trust, and the distribution of status and authority among actors.

Consider a  $\text{MODEL}(t)$  representing the belief state of actor  $i$  in a situation  $S_i$ , a rule  $r$  which is a candidate for a new belief in  $\text{MODEL}(t)$ , and a composition rule complex  $D$ . According to  $D$ , the actor  $i$  decides whether or not  $r$  should become a 'new' belief (or piece of knowledge). Suppose  $r$  is accepted.<sup>12</sup> The rule  $r$  may be an ordinary object rule like: 'All humans are untrustworthy'. The actor incorporates it into  $\text{MODEL}(t)$  according to rules of  $D$ . These rules specify subcomplexes of  $\text{MODEL}(t)$ , say  $C_1, \dots, C_k$ , in which to insert  $r$ . That is,  $D$  contains rules that identify or indicate relevant subcomplexes in  $\text{MODEL}(t)$ , for instance, sub-complexes where there are beliefs in relation to which the new belief is relevant. If the resulting rule complex  $\text{MODEL}(t)(C_1 \cup \{r\}/C_1, \dots, C_k \cup \{r\}/C_k)$  is inconsistent (for instance, Der-inconsistent (Burns et al., 1998)), appropriate rules of  $D$  are used to restore its consistency.  $\text{MODEL}(t)(C_1 \cup \{r\}/C_1, \dots, C_k \cup \{r\}/C_k)$  denotes the rule complex obtained from  $\text{MODEL}(t)$  by replacing  $C_1 \cup \{r\}$  for  $C_1, \dots, C_k \cup \{r\}$  for  $C_k$ . According to the rules of  $D$ , priority is given to  $r$  (as well as or to other beliefs). Hence the kind of change (i.e., revision/updating or arbitration (see Alchourrón, Gärdenfors & Makinson, 1985;

Gärdenfors, 1992; Hansson, 1991; Katsuno & Mendelzon, 1992; Revesz, 1997, among others)) depends on the situation and/or the rule complex  $D$ , among other things. In a given game or social relational context,  $D$  will contain rules such as: 'accept new beliefs that come from scientific or expert sources' or 'give priority to expert beliefs over traditional or everyday beliefs'; or, 'new beliefs from untrustworthy or alien sources should be rejected'.

### 3.1. Principles of belief judgment

In retaining or changing a belief, an actor makes comparisons and judges similarity (or dissimilarity) between new information articulated as a single rule  $r$  and an existing belief  $r'$  (or, in the general case, beliefs in several subcomplexes of her model) in the situation  $S_i$ . This is represented in (5) (for purposes of simplification, we drop  $t$  that contextualizes  $J_i$ ; it will be understood that it applies in a particular situation and at time  $t$ ):<sup>13</sup>

$$J(r, r') = \text{sufficiently similar} \quad (5)$$

Typically, judgment of similarity is merely one type of judgment performed in processes of belief change. There are usually multiple rules and meta-judgments about the reliability, validity, trustworthiness of the established belief  $r'$  (or, in general, beliefs in several subcomplexes of the actor's model) versus the new information (observation, communication) articulated in  $r$ . The latter is *judged* in terms of its validity, based on the likely reliability or trustworthiness of the source. Several *social factors* can be identified as key variables affecting the rules and meta-judgments that operate in composition processes, and, therefore regulate belief and learning processes: (1) *Degree of trust* in a source (or in the strength of norms constraining deception and harming). For instance, actor  $i$  believes (or does not believe) that the source — let us say actor  $j$  — is not deceptive or acting with malice toward  $i$ . (2) The *social status of the source  $j$*  relative to that of the

<sup>12</sup>The rule  $r$  may be a meta-rule w.r.t.  $\text{MODEL}(t)$  saying, for instance, (1) that another belief  $r'$  being a generalized member of  $\text{MODEL}(t)$ ,  $r' \in_g \text{MODEL}(t)$ , should be removed from  $\text{MODEL}(t)$ . In other words, that  $\text{MODEL}(t)$  should be contracted by  $r'$ . (2) Or, that a belief  $r'$  should be replaced by a complex  $C$  of beliefs in a subcomplex of  $\text{MODEL}(t)$ .

<sup>13</sup>This entails following or applying a rule in one of its senses (see later). The basic process is one of comparison-test, universal in constructing, selecting, and assessing action.

recipient  $i$  where such status may be based on professional expertise, legitimate authority, or general status characteristics such as gender, age, and ethnicity, which, as we shall see, play a key role in belief processes. (3) The *strength of collective commitment* with respect to an established belief or belief structure. In a community with a high degree of commitment to a particular belief or belief structure, members experience dissonance if they express themselves or act contrary to the belief (Kuran, 1998; Machado, 1998). (4) The *strength of collective sanctioning*. Members of the collective expect punishment or a strong feeling of shame or guilt if they express themselves or act contrary to core group beliefs. Some groups and communities are subject to high levels of internal control, while other groups are much less restrictive.

In this perspective the mechanisms of cognition, judgment, and belief change are socially embedded or contextualized. By social context we mean in particular the relationship(s) and institutions organizing and regulating actors agents in concrete situations and affecting, among other things, cognitive processes and learning. These processes take place in interaction situations or games where actors have certain roles and are oriented to particular norms (or particular role conceptions and norms are absent in the situation). These become the basis of their judgments underlying belief processes. In the literature several different though interrelated types of belief change are distinguished. (Alchourrón et al., 1985; Fuhrmann, 1991; Fuhrmann & Morreau, 1991; Gomolinska, 1998; Gomolinska & Pearce, 1999; Grahne, Mendelzon & Revesz, 1992; Gärdenfors, 1992; Gärdenfors & Makinson, 1988; Hansson, 1991; Hansson & Rabinowicz, 1995; Katsuno & Mendelzon, 1992; Lindström & Rabinowicz, 1991; Nebel, 1994; Revesz, 1997; Ryan, 1991, among others). The major types of belief change are expansion, contraction, revision, updating, and arbitration. In the literature, an actor's beliefs are typically represented by formulas and her belief state is represented by a set of formulas of a given language. Operators of change of belief state are characterized by sets of 'rationality' postulates that should be satisfied; or, they are defined directly, usually in semantic terms. Formulated in our terms, types of belief change may take the following forms:

(1) *Expansion* (see, for instance, Alchourrón et al., 1985) of  $MODEL(t)$  with a new belief  $r$  consists in the insertion of  $r$  into particular relevant subcomplexes  $C_1, \dots, C_k$  of the rule complex  $MODEL(t)$ . Then

$$MODEL(t+1) = MODEL(t)(C_1 \cup \{r\}/C_1, \dots, C_k \cup \{r\}/C_k). \quad (6)$$

Such belief change occurs when, on a meta-level, the actor judges the new piece of information or communication articulated as  $r$  to be valid or reliable, and no contradiction between  $r$  and  $MODEL(t)$  is detected (or a contradiction has been detected but is totally ignored by the actor (see below)).

(2) *Revision* (Alchourrón et al., 1985) of  $MODEL(t)$  takes place in the case the actor decides to add  $r$  among her beliefs. However, simple expansion of  $MODEL(t)$  with  $r$  leads to contradiction with, for instance, an established belief  $r'$  in  $MODEL(t)$  (this may be also expressed as the case where a simple expansion of  $MODEL(t)$  with  $r$  results in an inconsistent belief state  $MODEL(t+1)$ ). That is,  $r$  and  $r'$  are dissonant or contradictory with respect to one another given  $MODEL(t)$ , and the contradiction is resolved in favor of  $r$  (a more general case of revision is where the belief  $r$  is contradictory with some subcomplexes  $C_1, \dots, C_s$  of  $MODEL(t)$ ). Here then, not only is  $r$  added but  $r'$  as well as possible other beliefs are removed from relevant subcomplexes of  $MODEL(t)$  in order to avoid or to resolve contradiction.  $r$  replaces  $r'$ , because it is judged more valid or trustworthy. Such judgments may be based on the association of  $r$  with science while  $r'$  is associated with tradition or everyday beliefs. Or, it derives from more 'trustworthy sources.' In revolutions (Burns & Carson, 2001), many of the beliefs and ideas of the old regime are judged as invalid or untrustworthy. Revision is motivated for reasons internal to the actors involved. In many cases, the external world to which  $r$  refers has not changed.

Updating (see Katsuno & Mendelzon, 1992) for the original concept of updating) of  $MODEL(t)$  by  $r$  is similar to revision but in this case it is assumed that change of the external world is the reason for the change of beliefs in  $MODEL(t)$ . In both revision and

updating the ‘new’ belief  $r$  is given priority to any other belief(s) such as  $r'$  from  $\text{MODEL}(t)$ .

(3) *Retention or Maintenance*. Such belief maintenance takes place when the actor judges that (a)  $r$  roughly matches a belief  $r'$  already in  $\text{MODEL}(t)$ , that is,  $r$  corresponds to  $r'$ , or (b) the belief candidate  $r$  is rejected as a new belief. Then  $\text{MODEL}(t)$  remains the same. In the case (a), continuation of  $r'$  occurs when the new piece of information  $r$  in situation  $S$  at time  $t$ , is judged *sufficiently similar* to a relevant belief  $r'$  in  $\text{MODEL}(t)$  (that is, Eq. (5) is satisfied).

In case (b), the candidate for belief  $r$  is simply viewed as either irrelevant or not reliable or trustworthy and, therefore, can be ignored or rejected.  $\text{MODEL}(t)$  is maintained unchanged. Such judgments of reliability or trustworthiness relate, above all, to the status or reliability of the source (or the method or procedure whereby the information was obtained and  $r$  was constructed). A composition process takes place which results in no change in the model:  $\text{MODEL}(t+1)=\text{MODEL}(t)$ .

(4) *Contraction*. Contraction (see Alchourrón et al., 1985) of  $\text{MODEL}(t)$  by  $r$  takes place if the actor decides to remove an existing belief  $r$  from among her beliefs in  $\text{MODEL}(t)$ , for instance to stop accepting it as a belief. Then  $r$  is removed from relevant subcomplexes  $C_1, \dots, C_s$  of  $\text{MODEL}(t)$ , usually together with other beliefs, in such a way that one cannot again conclude  $r$  from the resultant belief state represented by  $\text{MODEL}(t+1)$  (in the sense of derivability (see Gomolinska, 1999; Burns & Gomolinska, 2000a)).

(5) *Arbitration* (see Revesz, 1997) for original concept of arbitration) is a form of belief change, where a candidate for belief  $r$  is incorporated but not given priority to the old beliefs, that is, to members of  $\text{MODEL}(t)$ . In this case  $r$  is inserted into belief state  $\text{MODEL}(t)$ , as in the case of expansion, because  $r$  is judged by the actor  $i$  as sufficiently valid or trustworthy. In the next step, that of ‘consolidation’ (Hansson, 1991), consistency is restored whenever necessary. In general, restoration of consistency of a rule complex may consist in removing less

entrenched beliefs which contribute to inconsistency or in inserting meta-rules which block usage of contradictory rules or rule complexes in given situations. Such a meta-rule may say: In such and such case never use both  $r$  and  $r'$ . Potentially, a composition process takes place. In some cases, actors may not be concerned about consistency of their models or belief states until they face a situation that evokes the contradictory beliefs.

(6) *Blocking (or conditional usage) of rules*. This form of belief change is a version of expansion, revision or updating. In our opinion, however, it deserves special mention. If in the situation  $S_i$  the actor  $i$  stops believing  $r$  which is a (generalized) member of  $\text{MODEL}(t)$ , she may decide not to contract  $\text{MODEL}(t)$  by  $r$ , but to limit or condition its usage. Often it suffices to expand (or to revise/update) the model with a meta-rule blocking the usage of  $r$  in a particular class of situations. That is, such a rule or rules prevent the actor from using  $r$  in specified situations.

## 4. Applications

The applications of GGT presented in this section concern multi-agent game or interaction situations, where the social relationships among actors, the status and authority differences, the level of trust and honesty affect belief change — in large part by affecting the composition rules which are applied to candidates for belief. In some cases of belief revision, falsehood is produced. But, as we shall show, even false beliefs may become true through self-fulfilling processes. The key principle in these applications is that social relational factors are articulated in composition rules which regulate actors’ responses to information and communication from others; that is, whether a new belief expressed or articulated in a rule is likely to be accepted or rejected.

### 4.1. Belief change as a function of solidarity and trust

Actors’ social or role relationships entail different

levels of solidarity and trust. High solidarity implies a normative predisposition not to deceive others in harmful ways. It also implies a predisposition or likelihood of accepting a communication or piece of information as ‘honest’ or ‘non-deceptive.’ These conditions make for optimization of ‘truthful’ communication and belief revision based on such communication. The more or less free communication of ideas and ‘discoveries’ facilitates selective change in belief — expansion, revision, updating, contraction, arbitration, etc.<sup>14</sup> The social relationships provide a *normative basis* effective collective problem-solving and learning. In general, group members operate with composition rules that incorporate beliefs communicated by other members, particularly high status members.

Social relationships where there is mutual distrust or dislike make for strategies of deception as well as distrust of one another’s communications. This of course blocks effective information exchange, and collective learning and problem-solving.<sup>15</sup> Each actor

would be disposed to reject or ignore proposals coming from other(s). Information, presentations, and proposals, etc., from ‘other’ will be distrusted or discounted — unless it is information or communication that fit the image of the other as deceptive, dangerous, etc. New information, which is dissonant with respect to the beliefs the actors have about one another, will be rejected and the given beliefs maintained. Relationships of distrust, including the extreme case of enmity, and the particular rules and expectations associated with these relationships, make for greater or lesser communicative deception, distortion of beliefs, and likelihood of unsuccessful bargaining and agreement. The difficulties — and transaction costs — of reaching a settlement are greatest, the more deception there is, that is, the more agents distort their true positions (for instance, what terms of exchange they would accept), which, in general, they are disposed to do. This explains why collective learning and problem-solving are so difficult when actors distrust or dislike one another — even when they have the possibilities to communicate with one another and there are positive incentives to cooperate. In the case one or another makes genuine efforts to communicate honestly, there would be a general distrust of the messages and offers, thus blocking the emergence and development of new beliefs. In general, actors in such social relationships operate with composition rules that reject new beliefs or proposals from one another, accepting only those messages that fit or reinforce their stereotypical images of one another.

In sum, qualities of social relationships such as level of trust and solidarity affect belief revision processes, in particular, the composition rules with which they operate. Actors in a solidary group operate with composition rules for group interaction that dispose members to incorporate beliefs and information communicated by other group members, particularly high status actors. In the case of communication between distrustful or hostile actors, they operate with composition rules that reject communications and beliefs expressed by one another. Other social factors that play a role in belief changes are the relative status and authority of the actors. Composition rules tend to be incorporative or inclusive when new beliefs are expressed by experts as well as

<sup>14</sup>For instance, market social relationships vary considerably, from long-term enduring cooperative relationships to more momentary, ‘spot market’ relationships. Of particular interest are market relationships characterized by highly stable and enduring relationships between particular buyers and sellers on industrial markets. These are shown to involve a high degree of open communication, information and ‘discovery’ sharing, and minimal deception (Forsgren & Johanson, 1992; Haakansson, 1982; Johanson, 1988). Their research shows that market actors may establish enduring relationships, and these are the basis for long-term cooperation in research, production and technical development, custom-design and adaptation of products. While price remains important, the participants are predisposed to refer to, and be governed by, norms of fairness, sharing of risks and gains. And, what is important for our purposes here, they share information and knowledge, participating in collective belief revision processes. Johanson and associates in their comparative research on industrial networks, found that such patterns appear to be relatively common in continental Europe and Japan, but not so much in the USA where a more pure market ethos (and relationships) prevail — with high price sensitivity and less commitment to establishing or maintaining solidary or cooperative relationships.

<sup>15</sup>It increases ‘transaction costs’. In general, the belief or experience of the other as deceptive, or bluffing leads to actors discounting one another’s statements and claims. Under such conditions it is difficult to convince the other that the truth is being told.

other high status actors.<sup>16</sup> And the rules are exclusive or rejecting when the sources are agents defined as incompetent, or untrustworthy.

#### 4.2. The communication of 'false' beliefs and self-fulfilling processes

In his characterization of the phenomenon of self-

---

<sup>16</sup>A number of small group and organizational studies (Anderson, Berger, Cohen & Zelditch, 1966; Berger, Zelditch & Anderson 1966a; Berger, Cohen & Zelditch, 1966b, 1972) have shown that status relationships affect the readiness to accept the opinion(s) of others. In group situations where members are committed to successful performance of a task or to the achievement of an instrumental goal, they are strongly disposed to defer to the opinions or judgments of high status persons. While performance expectations — and differential response to the judgments of others — arise from direct experience and proven success, they are also embodied in stereotypical beliefs about people and categories of people who are characterized by certain status values. If actors *i* and *j* differ in status (with respect to occupation, social class, sex, ethnicity, age, education, etc.), then under conditions where there is no other basis on which to attribute competence and performance capabilities, then beliefs relating to status characteristics become the major basis for performance expectations and differential readiness to accept or reject the judgments of others. This is found to be true even in situations where the abilities associated with the group task are not directly related to the stereotyped conceptions that *i* and *j* have of one another. In general, prior or external status differences becomes the basis for differential influence in new interaction situations (Berger, Zelditch & Anderson, 1966, 1972; Berger, Cohen & Zelditch, 1966). This type of influence process — with its impact on judgment — based on preexisting status and authority relationships operates most powerfully when there is no objective basis, or alternative or competing authority, on which to make independent judgments. (In some social systems, the leadership is in a position to control information about performance failings and, in general, to fabricate significant parts of reality). In such group processes, there is a characteristic patterning — highly asymmetric — of belief change among group members, with high status members bringing about significantly more change in the beliefs of low status members than vice versa. Only under certain restrictive conditions is it possible to make a group or organizational status structure — as a basis for patterning belief change — independent of those status relationships found in the wider community or society: the group can withdraw or segregate itself from the larger social system so as to neutralize or restructure the external status relations (Zelditch, Berger & Cohen, 1966: 292). A status characteristic (such as gender, race, ethnicity, profession or occupation) may be defined as irrelevant or may be suppressed through group controls.

fulfilling prophesy, Merton (1968: 421) refers to an initial false definition of the situation evoking a new behavior which makes the originally false definition come true. He refers to the bankruptcy of a major US bank in 1932 (many banks were forced to close their doors in this period). This was precipitated by rumors that the bank was in the process of going bankrupt. Depositors rushed to remove their money before the bank went under.<sup>17</sup> As a result of their run on the bank, the bank collapsed. According to Merton (1968: 422):

*The stable financial structure of the bank had depended upon one set of definitions of the situation: belief in the validity of the interlocking system of economic promises men live by. Once depositors had defined the situation otherwise, once they questioned the possibility of having these promises fulfilled, the consequences of this unreal definition were real enough.*

The shared belief among depositors in the liquidity of the bank — its capacity to pay depositors their claims and to assure the security of their depositions — is a strategic, constitutive factor in generating or constructing the social fact: the liquidity of the bank.<sup>18</sup> In Merton's view, the belief that the bank could not meet its obligations was false in that the

---

<sup>17</sup>A bank run is technically defined as a sudden and unexpected increase in deposit withdrawals from a bank. Such sudden surges in net deposit withdrawals risks triggering a bank run that would eventually put a bank into insolvency (Saunders, 1994: 302). such panics also cause a flight to safe assets such as government bills and real estate or gold.

<sup>18</sup>The liquidity was validated in part by the belief itself! The belief enters into judgments, actions, and conditions and influences potentialities, events, and developments in the situation, above all the viability of the bank, since a bank by lending out money is typically in a state where it is incapable of collecting all of its loans and returning on short notice all of the deposits to its clients. One cannot, therefore, simply speak of 'true' or 'false' or 'real' or 'unreal', at least in the sense of statements made independent of the situation referred to, when these statements (and the underlying beliefs) enter into actions and interaction and maintain or breakdown a social definition of the situation. One should refer instead to potentialities, possibilities, or opportunities which agents may or may not exploit.

bank was in a reasonably liquid state.<sup>19</sup> However, as a definition of the situation, it became true. In such terms, he evokes the ‘Thomas theorem’ that, if human actors define a situation as true, it may become true in its consequences.

Examples of self-fulfilling predictions, prophecies, and other statements are numerous and widespread in social life (Henshel, 1976). In 1981, the Wall Street analyst Joseph Granville ‘predicted’ that the New York Stock Market’s advance was ending and advised the 3000 subscribers to his Granville Market Letter to sell everything. Enough believed him to send the Dow Jones industrial average tumbling down 23.9 points on 7 January 1981. Gerald Tsai, another prominent stock market analyst, acquired an influential reputation as a forecaster of stock market trends. Like Granville, Tsai had a large following of true believers who trusted his forecasts and followed his advice on stock trading. Tsai concentrated his forecasts on thinly traded stocks — stocks which experienced little turnover on a given day. Because of low rates of buying and selling, Tsai’s pronouncements on these stocks tended to have substantial impacts on their subsequent value. If Tsai predicted

that the stocks would rise in value and many of his followers — along with others alert to apparent opportunities — purchased the stock, their value increased, at least temporarily. Predictions about heavily traded stocks, such as American Telephone and Telegraph, IBM, and General Motors, would not, in general, have the same impact.

Self-fulfilling processes can be conceptualized as multi-agent interaction systems where the communication of ostensibly false beliefs results in behavior which tends to validate the beliefs (Burns, 1998). Let us briefly consider one type of social order that makes for such processes. Actors are differentiated into at least two role categories: (1)  $i$ ’s who are experts, scientists, priests, or others providing knowledge, predictions, forecasts, prophecies, etc.; (2)  $j$ ’s who are subjects, consumers, followers, responding directly or indirectly to  $i$ ’s communicated statements. The  $i$ – $j$  relationship is related to the social situation or phenomena  $S_t$ . The  $i$ ’s make certain statements about future states or social phenomena of the situation  $S_t$ , which persuade  $j$ ’s engaged in  $S_t$  to change their beliefs and to act on the basis of their beliefs in such a way as to bring about  $S_{t+n}^*$ , which  $i$  predicted.<sup>20</sup>

The  $i$ – $j$ – $S$  represents a minimum social system with differentiated agents, a social relationship with influence potentialities, resource control or power with respect to events and developments in  $S_t$ , and concrete decisions and action processes that make a difference with respect to the state of affairs of  $S_t$ . The actors,  $i$ ’s and  $j$ ’s, are differentiated in their cognitive capabilities and interests. Between them, there are certain communication links as well as authority and influence relations. The  $j$ ’s have capabilities or resource controls sufficient to affect the phenomena or state of  $S_t$ . The  $i$ – $j$ – $S$  model entails the following key processes (where we assume a single expert or authority  $i$  for our purposes here):

1. *The production of statements about social phenomena*: the social agent,  $i$ , produces statements, including predictions, about social reality, in

<sup>19</sup>The contrary belief — that the bank could not meet its obligations — is, of course, a potential threat to the liquidity of the bank and, thus, a threat to its capacity to pay depositors on demand, particularly in the case where large numbers of depositors — came to have this belief. When such a belief spreads among a substantial number of people who act on the belief, the liquidity of the bank is, indeed, undermined. The rumor — the spreading belief — that the bank might fail to meet the claims of its depositors was not entirely invalid (contrary to Merton’s argument). Certainly not under conditions where there is a significant potentiality for the bank failing to meet obligations. Under circumstances of widespread diffusion among bank depositors of the belief about pending bank illiquidity, the belief was and became, indeed, a valid one. A shared belief or public trust in the liquidity of a bank — in its capability to pay all its depositors their claims on demand — contributes to sustaining a bank, even a bank having serious liquidity problems. The public trust is a major constitutive component of the socially defined and organized order, the order of the bank. The bank is viable as long as the shared belief — public trust — is maintained and other events and developments do not undermine such trust in the bank’s status. An undermining factor would be that other banks, or the Central Bank, refuse — or possibly only hesitate — to express trust or to act in a supportive manner toward the bank. Or they fail to take stabilizing measures vis-à-vis the bank in question.

<sup>20</sup> $S_{t+n}^*$  means a situation at some time  $t+n$  later than  $t$ . In the simplest case  $n=1$  as in our presentation of selected types of belief change.

particular concerning the phenomena or state  $S_t$ ,  $i$  has an interest in doing this and believes herself capable of doing so. Using certain methods, models, procedures and the organization of knowledge production, she produces statements about future phenomena or the state of  $S_t$ , predicting or prophesizing the state of affairs  $S_{t+n}^*$  but where the most likely outcome or development would actually be  $S_{t+n}$ ,  $S_{t+n} \neq S_t$ . Examples of institutionalized production of statements about the future are: the Greek Oracle at Delphi, market analysts such as Granville and Tsai, expert groups making technology assessments, economists and econometricians (Baumgartner & Midttun, 1986; Burns & Flam, 1987).

2. *The infrastructure of communication*: there are technological means and social networks or organizations, such as mass media, that enable  $i$ 's messages to be disseminated to a population of  $j$ 's. In the absence of such communication infrastructure, most large-scale or societal level self-mediating processes could not take place.
3. *The process of communication*:  $i$ 's communication about the social phenomena  $S_t$  is motivated by certain interests in or incentives to communicate publicly, possibly to the  $j$ 's in particular. The  $j$ 's need not, however, have been the targets of the communication. In some cases, other groups are intended targets, but the information communicated reaches the  $j$ 's. For example, labor unions develop their wage bargaining objectives and strategies after receiving government wage development forecasts. The communication is organized in specific ways and involves the use of various technologies and channels. There is a language as well as a 'medium'.
4. *The process of interpretation and composition*: the  $j$ 's have cognitive structures and skills which enable them to understand and interpret  $i$ 's statements so that they can be related to their models and decisions. In other words, the communications have meaning. The interpretations which the  $j$ 's give to  $i$ 's statements need not, however, correspond to  $i$ 's intentions. The  $j$ 's are predisposed to incorporate the  $i$ 's beliefs or statements into their models (and ultimately to use the models in planning and determining their actions). The  $j$ 's incorporate  $i$ 's communications about  $S_t$

into a common model, MODEL( $t$ ), assuming for our purposes here a shared model relating to  $S_t$ . That is, the  $j$ 's composition rules are inclusive or incorporative when the source of a new belief or message is an actor  $i$  judged to be a trustworthy expert or authority.

5. *The  $j$ 's judgment/decision process*: this entails: (a) a certain interest or motivation on  $j$ 's part to use the new beliefs about  $S_t$  and, in particular, (b) a readiness to relate these beliefs in MODEL( $j$ ,  $t+n$ ) to their judgments and actions with respect to  $S_t$ .
6.  *$j$ 's implementation and action processes*: the  $j$ 's have the interest and capabilities, including the necessary resource control, to act in relation to and to affect the situation or phenomena  $S_t$  to which  $i$ 's statements refer, ultimately transforming the latter into a situation  $S_{t+n}^*$ . The latter differs from the most likely state  $S_{t+n}$  which would occur if the  $j$ 's had not revised their beliefs and acted on them.  $S_{t+n}^*$  corresponds to  $i$ 's prediction.

This model stresses the importance of correspondence between  $i$ 's statements referring to  $S_t$ , and  $j$ 's interests and action capabilities with respect to  $S_t$ . Instances of non-correspondence define, in part, the limits of self-fulfilling predictions.  $i$ 's statements might refer to a state of affairs or phenomenon, for instance at a national or international level phenomena, whereas  $j$ 's action capabilities are limited to purely local level situations. The  $j$ 's would be unable to influence the state of affairs to which  $j$ 's statements refer. No self-fulfilling (or self-defeating) process could occur. In general, the social production of reality on the basis of self-fulfilling processes can take only take place when the predictions about reality are somehow connected — e.g., through communication processes — with agents who are involved in the relevant processes and who have the necessary interest and action capabilities to realize the new beliefs in practice.

Although self-fulfilling (and also self-defeating) predictions are universal features of human social life, their occurrence in any concrete situation depends on a number of structural and action conditions. They are knowledge-mediated social processes taking place in the context of a multi-agent

structure with particular communication linkages and particular interest and influence potentialities.

#### 4.3. Authoritarian belief control: Coercive equilibrium and normative disequilibrium

Consider a social context where there is a power or authority agent  $P$  who enforces a belief  $r$  through selective sanctions. The actors in the organization or community  $I$  are forced to communicate and act publicly consistent with  $r$  as demanded by  $P$ , at the same time that members of the organization or community disagree with the belief, rather adhering privately to their own common belief  $r'$ , which may be, for instance, traditional, religious, or professional in character. Specialized agents (supervisors, police, courts, etc.) maintain and reinforce the pattern of public compliance. Members of  $I$  are ‘forced’ to act as if they adhere to the belief  $r$ .

(1)  $P$  applies selective sanctions in order to guarantee that people demonstrate adherence to or compliance with  $r$ . Through  $P$ ’s intervention in the interaction situation  $S_i$ , the likely (and more or less predictable) patterns of publicly expressed belief and action are transformed. Consider that some action  $b_i$  may be such that it expresses or is consistent with  $r$ . Assume that  $a_i$  is an action expressing or embodying the belief  $r'$  where  $a_i \neq b_i$ .  $P$ ’s selective sanctions alter the consequences of actions  $a_i$  and  $b_i$  in the situation  $S_i$  (more precisely, the new context defined by  $P$ ’s impositions  $S_{i+1}$ ). In the formulation below, we use  $\text{Con}(a_i, t)^-$  to express that the overall value of the expected consequences of  $a_i$  has decreased relative to  $\text{Con}(a_i, t)$ , and, similarly, we write  $\text{Con}(b_i, t)^+$  to say that the expected consequences of  $b_i$  are valued as more valuable or attractive than it was before  $P$ ’s intervention (see Eq. (2)).

$$r': \text{Con}(a_i, t) \rightarrow \text{Con}(a_i, t)^-$$

$$r: \text{Con}(b_i, t) \rightarrow \text{Con}(b_i, t)^+$$

In this way, the expected consequences of the actions  $a_i$  and  $b_i$  in  $S_i$  — and, therefore, the consequences of expressing or putting into practice the authority’s belief  $r$  or that of the group or community  $r'$  — are transformed.

(2)  $P$ ’s selective sanctions not only alter the consequences but also the actors’ values which are salient

in the situation  $S_i$ . She introduces into the situation other stakes, such as matters of welfare, health, or life and death.<sup>21</sup> Thus,  $P$ ’s sanctions — or threat of sanctions — activate in the situation  $S$  an *additional value or values*, for instance, a ‘survival value’  $v_s$ . The latter might refer to ‘survival’ concerning self, family, business, or community.

(3)  $P$ ’s impositions transform actors’ judgment processes and orientations. The actors in the community  $I$  reorient from  $r'$  to  $r$ , and act to construct or select an action  $b_i$  that complies with or follows  $P$ ’s prescription. The performance or implementation of  $b_i$  is a type of *social equilibrium*, individual or collective, to the degree it realizes  $r$ , to  $P$ ’s satisfaction (Burns & Gomolinska, 2000b; Burns et al., 2001). Those conforming in this way expect to assure the realization of their survival value  $v_s$  (Eq. (2) is satisfied with respect to value  $v_s$ ). In the case some actors choose to deviate so that  $r'$  is expressed or realized and  $r$  violated, there would be a substantial price to pay. They would risk survival under these conditions, that is, failing to satisfy value  $v_s$  in Eq. (2).

The exercise of  $P$ ’s power activates existential value commitments among the population — symbolized here by  $v_s$  — which are not or would not normally be a part of the situation  $S_i$  and whose salience in the situation dominates commitment to the core community belief  $r'$  (there is a meta-value defining such an ordering). Obviously, the meaning of adherence or obedience to  $r$  through action  $b_i$  is to realize or satisfy  $i$ ’s value of survival,  $v_s$ . Compliance with  $r$  implies the expected realization of  $v_s$ .

(4) If in the context of  $P$ ’s power the selective sanctions are effectively executed, then the preference or evaluative ordering  $J(a_i) \gg J(b_i)$  is transformed into  $J(a_i) \ll J(b_i)$ . The members of  $I$  do  $b_i$ . Then a social but non-normative equilibrium with

<sup>21</sup>A dictator  $P$  not only affects certain consequence orders (incentive systems), but (1) may impose constraints on the actors’ repertoires in the game and (2) may propagate and reinforce values that result in certain action patterns and predictability. But then democracies use these approaches as well. The difference is that democracies try to legitimize these methods, while dictatorships may ignore this matter, or try to impose legitimacy (which is ultimately based on actors judging real conditions themselves).

respect to  $r$  obtains, based of course on  $P$ 's system of sanctions and incentives.<sup>22</sup> This pattern of activity would make for a degree of social order and predictability.  $P$ 's sanctions are sufficient to overcome or supersede what each and every actor would sacrifice or pay to realize  $r'$ , at least in the context of  $P$ 's power. If the community belief  $r'$  is a powerful one, however, then  $P$  will be compelled to apply considerable force in an effective way (which is highly resource demanding) in order to assure compliance. Indeed, in the case that  $r'$  is a sacred belief in the group or community  $I$ , the cost of maintaining conformity to  $r$  is not only high but also becomes very uncertain. People may be prepared to die or to make other heroic sacrifices in refusing or opposing conformity to  $P$ 's demands.

By changing the consequences of their actions and their evaluations of their options,  $P$  induces some level of public compliance to her beliefs in the transformed situation. Compliance with  $r$  under these circumstances is for purely instrumental reasons, i.e., because of the punishment or losses people avoid or the rewards they expect to achieve for such conformity. The members of  $I$  are likely to judge such compliance, locally as well as globally, as normatively wrong. There is a *normative disequilibrium* with respect to the core belief  $r$ , applying to the situation  $S$ . The coerced equilibrium, the  $b_i$  pattern (violating  $r'$ ), is unstable; it not only lacks normative force or commitment but clashes with an important community belief. Actors in  $I$  will deviate from this 'equilibrium pattern', if able to avoid  $P$ 's controls.

Examples of such situational compliance are plentiful: above all, situations where an authoritarian agent imposes what are considered by members of a community as harsh, unjust controls. Under such controls, members of the community or society conform in ways which clash with community beliefs and inclinations. At the same time, the actors pretend or fabricate behavior which is incompatible with what they truly feel or judge right and proper.

This is typical 'pragmatic behavior' observable under authoritarian rule.<sup>23</sup> In general, such situations make for dissonance between relevant, meaningful beliefs and actual behavior (Festinger, 1957; Kuran, 1998; Machado, 1998). Actors are disposed to break out of the patterns. Opportunities to bring about change are readily exploited and cause the games to be played differently with new (and less dissonant) interactions and outcomes.

#### 4.4. Collective fabrication and ignorance

Consider a case when actors in  $I$  experience diffuse community pressure to conform to a belief  $r$  and at the same time believe that other actors are truly committed to the belief, although this is in fact not the case. Indeed, for many there are strong incentives for local, individual deviation. When not observed, members of  $I$  reject and deviate from  $r'$ . But publicly, there is an apparent conformity and normative equilibrium with respect to the belief. It is a coerced equilibrium, as in the previous case.

This conformity in word and deed is predicated on beliefs in, and perceptions of, community support for the belief  $r$  with potential sanctions for violations. The beliefs orient each actor and structure her perception of consequences and judgments. These beliefs underlie conformity in word and deed. There are parallels to the previous situation with a powerful agent imposing law and order. But such structuring is maintained on the basis of real or imagined normative constraints at the same time that there are barriers to interpersonal and collective communication about the degree of individual commitment to, and private compliance with, the norm. There is, therefore, stable collective ignorance about one another's feelings and the fact that are all fabricating,

<sup>22</sup>If this applies for each and every actor  $i$  in  $I$  ( $i \in I$ ), then a potential Nash equilibrium obtains, that is in this case (Burns & Gomolińska, 2000a):

$$J(b_1, b_2, \dots, a_i, \dots, b_m) \leq J(b_1, b_2, \dots, b_i, \dots, b_m).$$

<sup>23</sup>If through propaganda or effective demonstrations, a powerful agent  $P$  manages to convince the community  $I$ , to consider the belief  $r$  as right and proper for situation  $S$ , then  $b_i$  would become a legitimate and normative equilibrium competing with  $a_i$  (which is an expression or embodiment of the belief  $r'$ ) as a normative equilibrium. This makes for a clash of beliefs and judgments, and entails internal dilemmas, rather than a tension between an internalized norm and an external imposition.

concealing their true beliefs which, in fact, are shared.<sup>24</sup>

## 5. Conclusions

In the applications in this article, we have tried to show how GGT incorporating social science concepts such as status and authority, trust, self-fulfilling prophesy, authoritarian control, fabrication, and collective ignorance is useful in describing and explaining belief processes in multi-agent systems.

Among our empirical considerations, we considered two cases where members of a group or community, because of powerful sanctions, conform publicly in accordance with prescribed beliefs even if they disagree strongly with the beliefs. They fabricate false beliefs in their communications and behavior, thus constructing a particular social reality. The latter to a greater or lesser extent reinforce or help maintain the belief structures. Of particular importance were the communication patterns among group or community members. When they fear or distrust one another, there is not only a high level of deception but barriers to the communication of true beliefs. This blocks belief revision processes which could result in cognitive structures corresponding more closely to some states of reality (self-fulfilling processes may, of course, still go on). Distrust and barriers to communication are increased in authoritarian regimes due to the extensive use of undercover agents and spies. While the latter might be considered ‘enemies of the people,’ they often are difficult to identify, as the history of authoritarian systems demonstrate. Under such circumstances,

<sup>24</sup>Given that each actor has a private belief  $r'$ , which is incompatible with  $r$ , one of the costs of conforming is the experience of dissonance between one's own belief,  $r'$ , and the belief  $r$  with which one conforms (of course, such dissonance is usually experienced through action, the action of conforming with  $r$  and violating  $r'$  (Kuran, 1998; Machado, 1998)) The act of conformity  $b_i$  with respect to  $r$  does not fit or is dissonant with respect to  $r'$ , as in the previous case. Such dissonance gives rise to tension and efforts to avoid it or to resolve it (Festinger, 1957; Machado, 1998; Kuran, 1998). There are limits to the intensity of dissonance that an actor may tolerate, depending of course on her value complex and, in particular, her commitment to or identification with the belief  $r'$ .

actors are afraid and reluctant to share information or ‘discoveries’ except with closest and most trustworthy friends. They conform in ways that do not fit what they believe or know, contributing to dissonance, on the one hand, and to the experience that most people in the community agree or accept the prescribed beliefs, on the other. Collective ignorance about the extent of opposition tends to stabilize ‘false’ beliefs (as well as odious laws and policies). This social control mechanism is a double-edged sword, however. It maintains a modicum of order with respect to alien beliefs as well as norms. But this is at the price of social patterns of behavior where deception is the natural order, information and common beliefs are falsified, problems are not problems, solutions are not solutions. The barriers to learning are apparent. In a more open, democratic society, where there are high levels of trust and ‘truth-telling,’ the capacities for identifying and maintaining valid truths, effective belief revision, and collective learning and are considerably greater (Burns & Engdahl, 1998).

## References

- Alchourrón, C. E., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic* 50, 510–530.
- Anderson, B., Berger, J., Cohen, B. P., & Zelditch, M. (1966). Status classes in organizations. *Administrative Science Quarterly* 11, 264–283.
- Baumgartner, T., & Midttun, A. (Eds.), (1986). The politics of energy forecasting: a comparative study of energy forecasting in western industrialized nations, Oxford University Press, Oxford.
- Berger, J., Zelditch, M., & Anderson, B. (Eds.), (1966a). Sociological theories in progress, Vol. I, Houghton-Mifflin, Boston, MA.
- Berger, J., Cohen, B. P., & Zelditch, M. (1966b). Status characteristics and expectation states. In: Berger, J., Zelditch, M., & Anderson, B. (Eds.), Sociological theories in progress, Vol. 1, Houghton-Mifflin, Boston, MA.
- Berger, J., Cohen, B. P., & Zelditch, M. (1972). Status characteristics and social interaction. *American Sociological Review* 37, 241–255.
- Burns, T. R. (1998). The three faces of the coin: money as symbol, institution, and technology. In: Mongardini, C. (Ed.), *Il denaro nella cultura moderna (Money in Modern Culture)*, Bulzoni Editore, Rome.
- Burns, T. R. (1990). Models of social and market exchange: toward a sociological theory of games and human interaction. In: Calhoun, C., Meyer, M. W., & Scott, W. R. (Eds.), *Structures*

- of power and constraint: essays in honor of Peter M. Blau, Cambridge University Press, New York.
- Burns, T.R. & Carson, M. (2001). Actors, paradigms, and institutional dynamics: the theory of social rule systems applied to radical reforms. In: Hollingsworth, R. (Ed.), *Social Actors and The Embeddedness of Institutions*, Rowman and Littlefield, Oxford.
- Burns, T. R., & Engdahl, E. (1998). The social construction of consciousness: collective consciousness and its socio-cultural foundations; individual selves. Self-awareness, and reflectivity. Parts I and II. *Journal of Consciousness Studies* 5((1, 2)), 67–85.
- Burns, T. R., & Flam, H. (1987). *The shaping of social organization: social rule system theory with applications*, Sage Publications, (1987, reprinted 1990).
- Burns, T. R., & Gomolinska, A. (1998). Modelling social game systems by rule complexes. In: Polkowski, L., & Skowron, A. (Eds.), *Rough sets and current trends in computing*, Springer, Berlin/Heidelberg.
- Burns, T. R., & Gomolinska, A. (2000a). The theory of socially embedded games: the mathematics of social relationships, rule complexes, and action modalities. *Quality and Quantity: International Journal of Methodology* 34, 379–406.
- Burns, T. R., & Gomolinska, A. (2000b). The theory of social embedded games: norms, human judgment, and social equilibria. Paper presented at the Joint Conference of the American Sociological Association-International Sociological Association on Rational Choice Theory, Washington, DC, August 2000.
- Burns, T. R., Gomolinska, A., & Meeker, L. D. (2001). The theory of socially embedded games: applications and extensions to open and closed games. *Quality and Quantity: International Journal of Methodology*, in press.
- Burns, T. R., Gomolinska, A., Meeker, D., & DeVille, P. (1998). *The general theory of games: rule complexes, action modalities, and transformations*, Uppsala Theory Circle Report, Uppsala, Sweden.
- Festinger, L. (1957). *A theory of cognitive dissonance*, Stanford University Press, Stanford.
- Forsgren, M., & Johanson, J. (Eds.), (1992). *Managing networks in international business*, Gordon and Breach, Amsterdam.
- Fuhrmann, A. (1991). Theory contraction through base contraction. *Journal of Philosophical Logic* 20, 175–203.
- Fuhrmann, A., & Morreau, M. (Eds.), (1991). *The logic of theory change*, LNAI, 465, Springer, Berlin/Heidelberg.
- Gomolinska, A. (1998). Credibility of information for modelling belief state and its change. *Fundamenta Informaticae* 34, 33–51.
- Gomolinska, A. (1999). Rule complexes for representing social actors and interactions. *Studies in Logic, Grammar and Rhetoric* 3(16), 95–108.
- Gomolinska, A., & Pearce, D. (1999). Disbelief change. In *Spinning Ideas. Electronic Essays Dedicated to Peter Gärdenfors on His Fiftieth Birthday*. Lund: University of Lund. <http://www.lucs.lu.se/spinning>.
- Grahne, G., Mendelzon, A. O., & Revesz, P. Z. (1992). Knowledge base transformations. In *Proceedings of the 11th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, pp. 246–260.
- Gärdenfors, P. (Ed.), (1992). *Belief revision*, Cambridge University Press, Cambridge.
- Gärdenfors, P., & Makinson, D. (1988). Revisions of knowledge systems using epistemic entrenchment. In: Vardi, M. (Ed.), *Theoretical aspects of reasoning about knowledge*, Morgan Kaufmann, pp. 83–95.
- Haakansson, H. (Ed.), (1982). *Internal marketing and purchasing of industrial goods: an international approach*, Wiley, Chichester.
- Hansson, S. O. (1991). *Belief base dynamics*. Doctoral Dissertation. Uppsala: Uppsala University.
- Hansson, S. O. & Rabinowicz, W. (eds.) (1995). *Logic for a change. Essays dedicated to Sten Lindström on the occasion of his fiftieth birthday*. Uppsala: Uppsala Prints and Preprints in Philosophy 9.
- Henshel, R. L. (1976). *On the future of social prediction*, Bobbs-Merrill, Indianapolis, IN.
- Johanson, J. (1988). *Business relationships and industrial networks: observations from international business research*, Department of Business Administration, Uppsala.
- Katsuno, H., & Mendelzon, A. O. (1992). On the difference between updating a knowledge base and revising it. In: Gärdenfors, P. (Ed.), *Belief revision*, Cambridge University Press, Cambridge, pp. 183–203.
- Kuran, T. (1998). Social mechanisms of dissonance reduction. In: Hedström, P., & Swedberg, R. (Eds.), *Social mechanisms: an analytical approach to social theory*, Cambridge University Press, New York, pp. 147–171.
- Lindström, S., & Rabinowicz, W. (1991). In: *Epistemic entrenchment with incomparabilities and relational belief revision*. LNAI 465, Springer, Berlin/Heidelberg, pp. 93–126.
- Machado, N. (1998). *Using the bodies of the dead: legal, ethical, and organisational dimensions of organ transplantation*, Ashgate, Aldershot, UK.
- Merton, R. K. (1968). *Social theory and social structure*, Free Press, Glencoe, IL.
- Nebel, B. (1994). Base revision operations and schemes: semantics, representation, and complexity. In: Cohn, A. (Ed.), *Proceedings of the 11th European Conference on Artificial Intelligence (ECAI 94)*, Wiley, Chichester, pp. 341–345.
- Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence* 13, 81–132.
- Revesz, P. Z. (1997). On the semantics of arbitration. *International Journal of Algebra and Computation* 7(2), 133–160.
- Ryan, M. (1991). Defaults and revision in structured theories. In: *Proceedings of the 6th IEEE Symposium on Logic in Computer Science (LICS)*, pp. 362–373.
- Saunders, A. (1994). *Financial institutions management: a modern perspective*, Irwin, Illinois.
- Sun, R. (1995). Robust reasoning: integrated rule-based and similarity-based reasoning. *Artificial Intelligence* 75(2), 241–295.
- Von Neumann, J., & Morgenstern, O. (1972). *Theory of games and economic behavior*, Princeton University Press, Princeton.
- Zelditch, M., Berger, J., & Cohen, B. P. (1966). Stability of organizational status structures. In: Berger, J., Zelditch, M., & Anderson, B. (Eds.), *Sociological theories in progress*, Vol. 1, Houghton-Mifflin, Boston, MA.