Social Institutions, Norms, and Practices[1]

Wolfgang Balzer, Institut PLW, Universität München
Raimo Tuomela, Academy of Finland

Introduction

The need for clarifying the interplay of actions and norms within social institutions is keenly felt among social scientists and in the multi-agent community. In sociology, the mainstream approach to institutions is in game theoretic terms, e.g. (Schotter, 1981), but there also are approaches using a power structure (Balzer, 1990), (Coleman, 1974), and stressing the cognitive level (Conte & Castelfranchi, 1995). In game theory the representation of actions and expectations is very idealized and far away from application to comprehensive real-life institutions. In the power centered approach so far the intentional, normative part has remained at an informal level. In AI, the study of cooperation has included organizational features (Durfee et al., 1987), (Prietula et al., 1998) and norms ('prohibitions', 'social laws') (Moses & Tennenholtz, 1995), and has led to formal accounts of institutionalized power, norms, rights, and obligations (Jones & Sergot,1997). One main restriction of these accounts is their lack of reference to the mental sphere of attitudes which prevents the exploitation of attitudes as a means of governing action. On the other hand, philosophical accounts of institutions focus on the normative aspects, neglecting explicit treatment of the corresponding systems of actions (Tuomela, 1995), (Ullmann-Margalit, 1977).

A comprehensive theory of institutions is still missing which makes explicit the overall macro structure, the norms, and the systems of actions as well as the interplay between these components. These features have to be formalized so that a comprehensive model may guide further fine grained studies which can lead to implementations. We submit a model of social institutions which captures the normative and the action component. It binds together a) a 'behavioral' system of social practices as repeated patterns of collective intentional actions and b) the normative Überbau consisting of a task-right system which on the one hand is influenced and in basic cases even induced by the 'underlying' practices and on the other hand serves to stabilize them. The model is not fully general in that we leave corporate actors and some aspects of jointness out of consideration.

An explicit connection in terms of sanctions is drawn between actions which are obligatory or permitted by special positions on the one hand and the 'ordinary' course of actions which occurs in social practices within an institution on the other hand. Though this connection has been discussed for quite some time (e.g. Pörn,1970), it has not received the manageable formalization needed for computer applications. The new feature of our model is that obligations and

[1] An extended version of the paper will be found under
http//www.lrz-muenchen.de/~W.Balzer/BALZER.html

rights are not simply bound to actions, but to *systems* of actions given in the form of systems of social practices. This adds an essential component which has been neglected so far (but see (Balzer, 1990)). The inclusion of social practices yields a rich structure in which the emergence and maintenance of norms can be tackled in a realistic way. The model offers a fresh start by making explicit the interplay between actions, the attitudes which are at work in triggering them, and the system of rights and obligations which stabilizes the system of social practices (actions) in an institution. We believe that the model yields a realistic basis for detailed case studies,[2] and also for subsequent studies of the emergence of task right systems.

1 States and Actions

Our model is a state space model in which the states are sets of sentences, indexed by a time variable. The states are relativized to individuals or groups, so that we can describe different states in which different persons or groups find themselves at the same time. States need not be closed under implication and no consistency requirements are made.

We do not aim at a philosophically satisfactory representation of actions (see e.g. (Tuomela, 1977,1995) for details). Actions are modelled as changes of state. Any pair (C, E) of sets of sentences of a given language L describes a potential transition from a 'previous' state C to a subsequent, expected state E. The sentences occurring in C and E must be such that under the right conditions they could describe some real action. In this case C describes a state in which the conditions for the action are satisfied, and E describes a state in which the desired effect of the action obtains. We distinguish between a) action *types* (C, E) for which the elements of C and E are formulas of L possibly containing variables, b) *potential actions* for which members of C and E must be sentences (closed formulas) and c) *actions* which really are performed. The latter are represented by $perf(t, i, (C, E))$, reading 'at time t, individual or group i performs the action described by (C, E)'. An action (C, E) at t may fail to produce the desired effect E (see below).

For a set C of formulas we write $C[t, i]$ and $C[t, i_1, ..., i_n]$ to denote the set of sentences obtained from C by replacing all variables by the names t, i, resp. $t, i_1, ..., i_n$ for instants and persons. To economize on notation we also write $C[t, i]$ if i denotes a group, $i = \{i_1, ..., i_n\}$. For the actual performance of an action we assume that the names occurring in the sets $C[t, i], E[t, i]$ are the same that are used in the $perf$ predicate: $perf(t, i, (C[t, i], E[t, i]))$, and we simply write $perf(t, i, (C, E))$.

A primitive A is used in order to pick out those pairs (C, E) representing action types from the set of all pairs (C, E) of sets of formulas. From A, a set A^* of (descriptions of) potential actions can be defined in terms of closure. A^*

[2] Though even a simple example is beyond the space available here.

contains all pairs (C^*, E^*) such that, for some $(C, E) \in A$, C^*, E^* are the sets of closures of formulas in C and E. If $S(L)$ and $F(L)$ denote the sets of *sentences* and *formulas* of a language L, we thus distinguish between a) transition types $(C, E) \in \mathbf{po}(F(L)) \times \mathbf{po}(F(L))$, b) action types $(C, E) \in A$, c) potential actions $(C, E) \in A^*$, and d) actions $perf(t, i, C[t, i], E[t, i])$.

The sentences in $S(L)$ will also be used in order to express the content of some mutual belief held among the persons considered which is central and constitutive for a social institution. Roughly, this content expresses that all members in the institution behave according to the tasks and rights assigned to them by their respective positions in the institution. As this content comprises a major part of the structure of an institution, the sentences in $S(L)$ must be rich enough to express this structure. In the extended version of the paper, see note 1), the construction of such a language is described in detail.

2 Frames

The conceptual arena in which we will talk about actions, rights, obligations, social practices and institutions we call a *frame*.

A frame is built up from

- a non-empty, finite set J of individuals or persons
- a finite, non-empty set G of groups such that $G \subseteq \mathbf{po}(J)$ and each $g \in G$ has at least two elements (we use I as an abbreviation for $J \cup G$)
- a non-empty, finite set ATT of attitude kinds containing at least *belief*, *intention* and *goal*
- a finite, linear order $(T, <)$, representing time
- a finite set O of 'ordinary objects'
- a language L with sets $S(L)$ and $F(L)$ of sentences and formulas
- a set $A \subseteq \mathbf{po}(F(L)) \times \mathbf{po}(F(L))$, of descriptions of *action types*
- a function $x : T \times I \to \mathbf{po}(S(L))$, the *state function*
- a function $caus : T \times \mathbf{po}(S(L)) \times T \to \mathbf{po}(S(L))$, the *causal function*
- a relation $perf \subseteq T \times I \times A^*$, the relation of *actual performance*
- a relation $catt \subseteq T \times G \times ATT \times A^*$ expressing *collective attitudes*
- a relation $incom \subseteq A^* \times A^*$ of *incompatibility* of potential actions
- a relation $ex \subseteq T \times J$, 'existence'
- a relation $sanc \subseteq \{+, -\} \times A \times A$ ('sanctions').

A frame basically consists of a state space, the states of which are described by sets of sentences (members of $S(L)$). The development of states over time is represented by the state function x which is relativized to individuals or groups. The sentences in $x(t, i)$ describe the state in which individual or group i is at time t. For each non-maximal instant t, the 'next' instant is denoted by $t + 1$. $caus(t, X, t')$ denotes the effect at time t' caused by the presence of X at t. The 'cause' here is described by the sentences in X. If these sentences are satisfied at t, then the cause X is present at t. At t' the ensuing effect is $caus(t, X, t')$, $caus(t, X, t') \subseteq \cup_{i \in I} x(t', i)$.

$perf(t, i, (C[t, i], E[t, i]))$ reads: at t, i performs (or the members of i collectively perform) action $(C[t, i], E[t, i])$. For proper individuals $i \in J$ this comprises individual action and for groups $i \in G$ collective action. An action may fail in the sense that for all subsequent t', $E[t, i] \not\subseteq caus(t, C[t, i], t')$.

$incom(a, b)$ expresses that the potential actions a and b are incompatible. This may be much weaker than inconsistency, incompatibility may simply be due to practical reasons.

$ex(t, j)$ means that at t, individual j exists as an active member. For each $g \in G$ and each t, we denote by g_t the set of members of g existing at t, $g_t = \{i \in J/ex(t, i)\}$. We assume that for each $j \in J$ there exist t^l_j, t^u_j such that $t^l_j < t^u_j$ and for all t with $t^l_j \leq t < t^u_j$, $ex(t, j)$, and t^l_j and t^u_j are the 'smallest' and 'largest' such instants.

$sanc$ is used to express that an action b is a sanction for another action a. We distinguish between sanctions of the form $(+, a, b)$ representing a sanction b following the performance of action a, and sanctions of the form $(-, a, b)$ in which b is a sanction for action a not having been performed. We say that *i's action a at t is sanctioned* iff $\exists b \exists j \exists t'(t < t' \wedge (+, a, b) \in sanc \wedge perf(t, i, a) \wedge perf(t', j, b))$. Similarly *i's not doing a at t is sanctioned* iff not $perf(t, i, a)$ and there are b, j and $t' > t$ such that $(-, a, b) \in sanc$ and $perf(t', j, b)$. Sanctions here are always understood in the negative sense.

In A we may distinguish between action types (and potential actions) involving one or more individuals. Action types (C, E) satisfying $\forall t, i : perf(t, i, C[t, i], E[t, i]) \rightarrow \exists j \in J(i = \{j\})$ are called *individual*, those which do not satisfy this condition being called *collective* action types. By CA and IA we denote the sets of collective and individual action types.

A *frame* y thus has the form $y = (J, T, ATT, O, G, <, L, A, x, caus, perf, catt, incom, ex, sanc)$.

3 Social Practices

A social institution consists of two central parts, an 'underlying' system of social practices and a (weakly) normative Überbau. We analyzed single social practices in (Balzer & Tuomela, 1999).

A social practice roughly is a repeated pattern of collective action in which a collective attitude of kind att (usually belief or intention) with content B is formed in a group, and an action of a corresponding action type (C, E) is then performed. For example, the collective intention of actors i, j to perform some action a means that the two intend to do their respective parts of a, believe that the respective Other intends to do his part, believe that the respective other belives that 'I' intend to do 'my' part, and so on, see e.g. (Balzer & Tuomela, 1997), (Wooldridge & Jennings, 1997) for accounts of collective attitudes. In general, the relation between content B and action type (C, E) may be opaque, but in the present first analysis we assume that both are identical, i.e. $B = (C, E)$. For example, if the attitude kind is *intention*, the group may repeatedly form the collective intention "we have sauna together next Saturday" and perform the collective action of having sauna together each 'next' Saturday. Both the content "we have sauna together next Saturday" and the corresponding action are represented in the format (C, E) of an action type where C contains sentences like "the sauna is operative", "most persons in the group are healthy" etc., and E contains sentences like "sufficiently many persons meet at 10 a.m. in the lobby", "the persons enter the sauna and bath" etc.[3]

Slightly modifying the account in (Balzer & Tuomela, 1999), the core of a social practise is given by three items:
- a kind *att* of attitude
- a content (C, E) of that attitude such that
- (C, E) is a collective action type.

By a collective action type we mean a type which is realized by a 'collective' of several persons, in contrast to individual action types, the actions of which can be performed by one person. In a frame $y = (J, T, ATT, O, G, <, L, A, x, caus, perf, catt, incom, ex, sanc)$ we assume that $att \in ATT$ and $(C, E) \in A$.

To these core items we add functions describing trigger conditions for attitudes (*trigatt*) and actions (*trigact*) which are specific for the particular action type (C, E) under consideration and are represented by sets of formulas. If all the trigger conditions in these sets are instantiated and true this will lead to the formation of the collective attitude, and to the sbsequent performance of the collective action (C, E). In the sauna case, a trigger condition for the attitude might be, for example, that the persons call each other to see whether they will have company, and a trigger condition for action will be that it is Saturday, 10 a.m.

Moreover, we use numerical functions *suc* for the *success* of a collective action, and *thr* to specify a threshold. The value of *suc* is increased or decreased depending on the success of the performance of the action, and the constant *thr* gives a threshold. If the success function drops below the threshold for several successive repetitions of the practice, the practice is likely to terminate.

Each formation of the collective attitude followed by a corresponding action and the latter's causal effects takes place in one *period* $z = (t_1, ..., t_4)$ in which four points of time are distinguished. At the first point t_1 the trigger conditions for the attitude are present, at t_2 the collective attitude is formed, at t_3 the corresponding action is executed, and at t_4 the causal effects of that action are noted. In a social practice such a four step pattern is repeated over and over, so we consider a sequence of periods $(z^i)_{i=1,2,3,...}$. By P^* we denote the set of all periods z^i pertaining to a given social practice.

In a frame y a *social practice with core* (g, att, a) now can be defined as a system $(g, att, (C, E), (z^i)_{i=1,2,3,...}, trigatt, trigact, suc, thr)$, where $g \in G$ is a group, *att* a kind of attitude, (C, E) a collective action type, $(z^i)_{i=1,2,3,...}$ a sequence of periods and

[3] See (Balzer & Tuomela, 1999) for a detailed analysis and more elaborate examples.

- $trigatt : T \times \{g\} \times \{att\} \times A \rightarrow \mathbf{po}(S(L))$,
- $trigact : T \times \{g\} \times \{att\} \times A \rightarrow \mathbf{po}(S(L))$,
- $thr : \{g\} \times \{att\} \times A \rightarrow \mathbf{N}$,
- $suc : P^* \times \{g\} \times \{att\} \times A \rightarrow \mathbf{N}$.

Moreover, some axioms have to assure that the four step schema described above is repeated over a sufficiently large number of periods. In particular, we assume
1) for all $i = 1, 2, 3, ...$ there exist $t_1^i, t_2^i, t_3^i, t_4^i$ such that $z^i = (t_1^i, ..., t_4^i), t_j^i \in T$ and $t_1^i < ...t_4^i < t_1^{i+1}$,
2) for all $i = 1, 2, 3, ...$ and all components t of $z^i = (t_1^i, ..., t_4^i)$, if n is the number of variables for different persons in $C \cup E$, then g_t has at least n members.
3) $\forall t \in T(catt(t, g_t, att, (C, E)) \in x(t, g_t) \rightarrow (trigact(t, g_t, att, (C, E)) \subseteq x(t, g_t)$
$\leftrightarrow perf(t + 1, g_{t+1}, (C, E)) \in caus(t, \{catt(t, g_t, att, C, E)\}, t + 1)))$.
4) $\forall z^i, i = 1, 2, 3, ..., z^i = (z_1^i, ..., z_4^i)$:
$trigatt(z_1^i, g_{z_1^i}, att, (C, E)) \subseteq x(z_1^i, g_{z_1^i})$
$\wedge\, thr(g_{z_1^i}, att, (C, E)) \leq suc(z, g_{z_1^i}, att, (C, E)))$
$\leftrightarrow catt(z_2^i, g_{z_2^i}, att, (C, E)) \in caus(z_1^i, x(z_1^i, g_{z_1^i}), z_2^i) \cap x(z_2^i, g_{z_2^i})$.
5) $\forall z^{i+1} \forall (C, E)$: if $perf(z_3^i, g_{z_3^i}, (C, E)) \in caus(z_2^i, \{catt(z_2^i, g_{z_2^i}, att, (C, E))\}, z_3^i)$
then $suc(z^{i+1}, g_{z_1^{i+1}}, att, (C, E)) = suc(z^i, g_{z_4^i}, att, (C, E))+1$ if $E \subseteq caus(z_3^i, C, z_4^i)$
and $= suc(z^i, g_{z_4^i}, att, (C, E)) - 1$ if $E \not\subseteq caus(z_3^i, C, z_4^i)$.

In 1) the sequence (z^i) of periods is embedded into the overall time structure $(T, <)$ such that the periods 'follow' each other. At the different points of time t the active members of group g are those found in g_t. 2) assures that in each period and at each specified instant t of that period, g_t contains 'sufficiently many' members so that the characteristic action type (C, E) can be performed.

3) says that if the collective attitude with content (C, E) is present in the group at t (among the active members g_t) then the trigger conditions for action will lead at the next instant $t+1$ to the action's (C, E) being performed 'because of' that attitude, 'and conversely'.[4] According to 4), if in the first instant of a period the trigger conditions for the attitude with content (C, E) obtain for the active members of group g and the success level for actions of the kind (C, E) is above the threshold, then the collective attitude will be formed and be present at the second point of time in that period, 'and conversely'. 5) is a simple rule for updating the success function at the end of period z^i. If the action performed in that period was a success, the function value is increased by one, otherwise it is decreased by one. 'Success' is expressed by reference to the action description (C, E). The action is successful if its expected effects, E, in fact, are among the causal consequences of its conditions C ($E \subseteq caus(z_3^i, C, z_4^i)$).[5]

[4] $t + 1$ need not be chosen according to the pattern of instants in the periods. We assume that attitudes persist in the sense of (Cohen & Levesque, 1990).

[5] Using slightly different formulations of these axioms, in (Balzer & Tuomela, 1999) necessary and sufficient

In order to define a *system* of several different social practices we use a set SP of *names* for social practices, and a function f which to each (name of a) social practice assigns a value $(g, att, (C, E))$ specifying the group g, the kind of attitude att and the action type (C, E) specific for that practice (its *core*).

D1 s is a *system of social practises* iff $s = (J, T, ATT, O, SP, G, <, L, A, x, caus, perf, catt, incom, ex, sanc, f)$ and
1) $y = (J, T, ATT, O, G, <, L, A, x, caus, perf, catt, incom, ex, sanc)$ is a frame.
2) SP is a finite, non-empty set (of labels of social practices).
3) $f : SP \rightarrow G \times ATT \times CA$ and $\cup\{\pi_1(f(sp))/sp \in SP\} = J$.
4) for all $sp \in SP$ and all g, att, a, if $f(sp) = (g, att, a)$ then in y there exists a social practice with core (g, att, a).

We do not require that different practices in a system of practices be compatible though this assumption makes good sense in most institutions, and in particular in organizations whose task right system is officially specified.

4 Obligations and Rights

In an institution, obligations and rights are attached to the positions pos which the persons occupy in it. Each person *holds* a specific position pos which we identify with two sets of action types, $pos = (OB_{pos}, RI_{pos})$, $OB_{pos} = \{o_1, ..., o_m\}, RI_{pos} = \{r_1, ..., r_n\}$ such that holders of pos are obliged to performed actions of types $o_1, ..., o_m$ and have the right to perform actions of types $r_1, ..., r_n$. Obligations and rights thus are represented in the following way. Person i in position pos is *obliged* to do a iff a is one of the action types occurring in OB_{pos} and the conditions for executing a obtain. Briefly, an obligation to do a is represented by '$a \in OB_{pos}$' for some position pos in the institution. Similarly, a right to do a in position pos is represented by '$a \in RI_{pos}$'.

Using the format (C, E) for action types, with conditions C and expected effects E, and the state function x and performance relation $perf$ described earlier, this representation of rights and obligations may easily be connected with actions. Consider some person i in position pos, and some action type $o = (C, E)$ obligatory for pos, i.e. $o \in OB_{pos}$. If i is in a state $x(t, i)$ in which the conditions for o are satisfied ($C[t, i] \subseteq x(t, i)$) then i should perform o. At the non-normative level 'i should perform o' corresponds to 'if i does not perform o then i gets sanctioned': $\neg perf(t, i, C, E) \rightarrow \exists j \exists t' \exists b(t < t' \wedge perf(t', j, b) \wedge sanc(-, o, b))$. In the case of rights the connection is a bit more complicated. If $r = (C, E)$ is covered by a right of i ($r \in RI_{pos}$ and $holds(t, i, pos)$) and i is in a state in which she could perform r ($C[t, i] \subseteq x(t, i)$) then no other person j should perform any action b interfering with r. That is, for any other person j and action $b = (C', E')$ which j could perform at time t ($C'[t, j] \subseteq x(t, j)$), and which is incompatible with r ($incom(r[t, i], b[t, j])$), j should not not perform b. Again, 'j should not perform b' at the non-normative level corresponds to

conditions are stated for the 'survival' of a practice over time.

‘if j would perform b then j would get sanctioned: $perf(t, j, b) \rightarrow \exists k \exists t' \exists c(t < t' \wedge perf(t', k, c) \wedge sanc(+, b, c))$.

This account provides a relatively simple connection between the normative level, the normative force of obligations and rights, and the level of actions and sanctions. It thus might serve as a basis for further investigations of how and why obligations and rights emerge and are upheld.

In order to anchor the action types attached to rights and obligations in a system of social practices we make the global assumption that each such action type comes from one of the practices in an ‘underlying’ system of practices, i.e. the action type is ‘part of’ the core of such a practice. This assures that no contrived actions figure in the rights and obligations. Rights and obligations are concerned only with socially entrenched action types. We cannot assume, however, that an action type expressing, say, an obligation, is simply identical with the action type of a social practice, for the latter describes a collective action while the former describes an individual one. In order to bridge this gap we use a relation *part* between collective actions (or action types) and their individual *parts*. We write $part((C, E), i, (C^i, E^i))$ to express that (C^i, E^i) is an individual action (type) which forms person i's part of the collective action (type) (C, E). A part (C^i, E^i) need not be unique; a person i may have several parts to perform in the collective action (C, E).[6]

D2 *norm* is a *task right system* for the system of social practises $s = (J, T, ATT, O, SP, G, <, L, A, x, caus, perf, catt, incom, ex, sanc, f)$ iff there exist $POS, part$ and $holds$ such that $norm = (POS, part, holds)$ and 1) for all pos, $pos \in POS$ iff there exist $o_1, ..., o_n, r_1, ..., r_m$ such that $pos = (OB_{pos}, RIpos)$, where $OB_{pos} = \{o_1, ..., o_n\} \subseteq IA$ and $RI_{pos} = \{r_1, ..., r_m\} \subseteq IA$. 2) $part \subseteq CA \times J \times IA$. 3) $holds \subseteq T \times J \times POS$. 4) for all pos, t, i, if $holds(t, i, pos)$ then $ex(t, i)$. 5) for all $pos = (OB_{pos}, RI_{pos}) \in POS$, all $(C, E) \in OB_{pos} \cup RI_{pos}$, all $i \in J$ and all $t \in T$, if $holds(t, i, pos)$ then there exist (C^*, E^*) and $sp \in SP$ such that 5.1) $f(sp) = (g, att, (C^*, E^*))$, 5.2) $part((C^*, E^*), i, (C, E))$.

The action types $(C, E) \in OB_{pos}$ are those which holders of position pos are *obliged* to perform (under the right conditions). Whenever the conditions C are satisfied for a person i holding position pos (i.e. $C \subseteq x(t, i)$) then i is obliged to perform an action of type (C, E). Action types a in RI_{pos} specify the *rights* of persons holding position pos. Person i has the right to perform actions of type a (within the frame considered) iff every other person j is obliged to refrain from performing any potential action b which is incompatible with an action a^* of type a $(incom(a^*, b))$.

Using a weak negation of action (‘it is not the case that i performs a’),

[6] Of course, this covers up all the problems of spelling out the individual parts of a collective action, and of constructing collective actions out of individual ones. However, for practical purposes it can be assumed that a collective action in fact is constituted by individual, ‘basic’ actions in the way of dynamic logic, i.e. by recursively forming bigger actions of the form $a \mid\mid b$ and $a;b$ out of simpler ones, see (Harel, 1984), (Sandu & Tuomela, 1996).

inflating the number of obligations, and assuming some kind of consistency of the task right system we can express the usual connection between rights and obligations as follows. If $a \in RI_{pos}$ and $holds(t,i,pos)$ then for all a^* of type a, all b and all j: if $incom(a^*,b)$ and $holds(t,j,pos')$ then among the obligations of pos' there is one obliging j not to perform b ('if i has the right to do a then every j ought to refrain from actions incompatible with a'). Conversely, if $a \in OB_{pos}$ then there is no right (in the system) of performing an action incompatible with a.

5 Social Institutions

A social institution now consists of a system of social practices plus a task right system for it. The system of tasks and rights on the one hand normatively mirrors certain combinations of collective action as found in the system of social practices. On the other hand, the normative task right system by its obligations and rights provides external reasons of institutional action. We submit three axioms. The first, D3-3, is a central, analytic condition. It states that among the members of an institution there is a common belief ($mubel$)[7] that everybody behaves according to the obligations and rights attached to his position. The other two hypotheses are of a contingent, empirical nature, and aim at explaining the role of the normative system. D3-4 and 5 say that people 'usually' perform the actions they are obliged to perform, and 'usually' refrain from actions conflicting with the rights of other members. 'Usually' has to be understood in a statistical way, refering to the numbers of performances and the weights of the different actions and types.[8]

In order to formulate these regularities, let us define, for $a = (C,E) \in A$, and $pos \in POS$, the numbers

- $exopp(a,pos)$, the number of *execution opportunities* of a in pos, as the number of $(t,i) \in T \times J$ such that $holds(t,i,pos) \wedge C[t,i] \subseteq x(t,i)$,
- $exec(a,pos)$, the number of *executions* of a in pos as the number of $(t,i) \in T \times J$ such that $holds(t,i,pos) \wedge C[t,i] \subseteq x(t,i) \wedge perf(t,i,(C,E))$,
- $freq(a,pos)$, the *frequency of executions* of a in pos, by $exec(a,pos)/exopp(a,pos)$,
- $vio(a/pos)$, the number of actions *conflicting* with a in pos as the number of $(t,i,j,b) \in T \times J \times J \times A$ such that $holds(t,i,pos)$ and $incom(a[t,i],b[t,j])$ and $perf(t,i,a[t,i])$ and $perf(t,j,b[t,j])$.

Note that in $exopp$, $C[t,i] \subseteq x(t,i)$ need not lead to action, the trigger conditions also must occur.

[7] See (Balzer & Tuomela, 1997), (Colombetti, 1993) or (Wooldridge & Jennings, 1997) for accounts of mutual belief.

[8] In order to avoid the mutual beliefs in D3-3 to be irrational, given the probabilistic formulations of D3-4 and 5, we should better use an approximate version of D3-3, too. However, as this would involve substantial additional formalism, we prefer to stick to the simpler, somewhat problematic formulation.

D3 x is a *social institution* iff there exist y and $norm$ such that $x = (y, norm)$ and[9] 1) y is a system of social practices. 2) $norm$ is a task right system for y. 3) for all $t \in T$: $mubel(t, J, p)$ where $p = p_1 \wedge p_2$ is the following sentence
$p_1 \equiv \forall j \in J \forall pos \in POS \forall t \in T \forall (C, E)$
if $pos \in POS \wedge (C, E) \in OB_{pos} \wedge C[t, j] \subseteq x(t, j) \wedge holds(t, j, pos)$
then $perf(t, j, C, E)$, and
$p_2 \equiv \forall i, j \in J \forall pos \in POS \forall (C, E) \in RI_{pos} \forall t \in T \forall (C^*, E^*) \in A$,
if $holds(t, j, pos) \wedge C[t, j] \subseteq x(t, j) \wedge C^*[t, i] \subseteq x(t, i) \wedge perf(t, i, (C^*, E^*))$
$\wedge\, incom((C[t, j], E[t, j]), (C^*[t, i], E^*[t, i]))$ then i gets sanctioned.
4) for all $pos = (OB_{pos}, RI_{pos}) \in POS$ and all $a \in OB_{pos}$, $freq(a, pos)$ is close to 1.
5) for all $pos = (OB_{pos}, RI_{pos}) \in POS$ and all $a \in RI_{pos}$, $vio(a/pos)$ is close to 0.

Sentence p expresses that all members behave (in the social practices) according to their positions (tasks and rights). p_1 says that whenever the conditions of an action type to which i is obliged in her position obtain then i will perform an action of that type. p_2 expresses that all persons can act according to their rights. If another person i performs some action incompatible with j's potential action $(C[t, j], E[t, j])$ to which j is entitled $((C, E) \in RI_{pos} \wedge holds(t, j, pos))$ then i gets sanctioned. These are of course the ideal versions of proxy formulations.

The hypotheses D3-4 and 5 alternatively may be read as criteria indicating the extent to which people behave according to the normative frame,[10] or the extent to which the institution is 'in force'. We prefer the regularity reading because if these statements are true only for large degrees of approximation - $freq(a, pos)$ being near 0, and $vio(a/pos)$ being a large number - one cannot say that the system modelled is an institution even if D3-3 is satisfied.

Finally, we want to point out a difficulty that arises when we reformulate the model keeping syntax and semantics separate in the usual way. In such a setting the sentence p in D3 expressing the mutual belief contains variables for the items J, A etc. making up the institution which have to be interpreted when mutual belief is expressed in sentence p. But then the mutual belief that p in an institution x presupposes the 'right' interpretation, namely an interpretation in the very system x. In order to express that people have the mutual belief about 'their' institution we therefore have to refer to this interpretation in the sentence p. This creates a rather strange, circular situation. The present, set-theoretic approach avoids this at the cost of loosing the explicit, deductive part. At least in the beginning however, this loss seems to be bearable in view of the cost of having syntax separated.

[9] The present definition is still a bit general insofar as positional action may be *joint* action, a case which is not included in the mutual beliefs in D3-3. We will address this problem in future work.

[10] Note that our notion of norms given by rights and obligations is rather weak, and does not require their being officially stated or being upheld by official prodecures.

References

W.Balzer, 1990: A Basic Model of Social Institutions, Journal of Mathematical Sociology 17, 1-29.
W.Balzer & R.Tuomela, 1997: A Fixed Point Approach to Collective Attitudes, in (Holmström-Hintikka & Tuomela, 1997), 115-42.
W.Balzer & R.Tuomela, 1999: Social Practices and Attitudes, manuscript.
J.S.Coleman, 1974: Power and the Structure of Society, New York: Norton.
M.Colombetti, 1993: Formal Semantics for Mutual Belief, Artificial Intelligence 62, 341-53.
R.Conte & C.Castelfranchi, 1995: Cognitive and Social Action, London: UCL.
E.H.Durfee, V.R.Lesser, D.D.Corkill, 1987: Coherent Cooperation Among Communicating Problem Solvers, IEEE Transactions on Computers, 36, 1275-91.
D.Harel, 1984: Dynamic Logic, in D.Gabbay & F.Günthner (eds.), Handbook of Philosophical Logic, Vol.II, Dordrecht: Reidel, 497-604.
G.Holmström-Hintikka & R.Tuomela (eds.), 1997: Contemporary Action Theory, Vol.2, Dordrecht: Kluwer.
A.J.I.Jones & M.Sergot, 1997: A Formal Characterization of Institutionalized Power, in E.G.Valdéz et al. (eds.), Normative Systems in Legal and Moral Theory, Berlin: Duncker & Humblodt, 349-67.
Y.Moses & M.Tennenholtz, 1995: Artificial Social Systems, Computers and Artificial Intelligence 14, 533-62.
I. Pörn, 1970: The Logic of Power, Oxford: Blackwell.
M.Prietula, K. Carley, L.Gasser, (eds.), 1998: Simulating Organizations: Computational Models of Institutions and Groups, Cambridge MA: MIT Press.
G.Sandu & R.Tuomela, 1996: Joint Action and Group Action Made Precise, Synthese 105, 319-45.
A.Schotter, 1981: The Economic Theory of Social Institutions, Cambridge: UP.
E.Ullmann-Margalit, 1977: The Emergence of Norms, Oxford: UP.
R.Tuomela, 1977: Human Action and its Explanation, Dordrecht: Reidel.
R.Tuomela, 1995: The Importance of Us, Stanford, Stanford University Press.
M.Wooldridge & N.R.Jennings, 1997: Formalizing the Cooperative Problem Solving Process, in (Holmström-Hintikka & Tuomela, 1997), 143-61.

Wolfgang Balzer
Institute for Philosophy, Logic and Philosophy of Science
University of Munich
Ludwigstr.31
D-80539 München
Germany

Fax: +49-89-2180-2902 balzer@lrz.uni-muenchen.de

Raimo Tuomela
Department of Philosophy
University of Helsinki
Unioninkatu 40B
SF-00014 Helsinki
Finland

Title: Social Institutions, Norms, and Practices

Abstract:

We submit a model of social institutions which binds together the two central components of institutions, a) a 'behavioral' system of social practices as repeated patterns of collective intentional actions and b) the normative 'Ueberbau' consisting of a task-right system which on the one hand is influenced and in basic cases even induced by the 'underlying' practices and on the other hand serves to stabilize them.

An explicit and relatively simple connection in terms of sanctions is drawn between actions which are obligatory or permitted by special positions on the one hand and the 'ordinary' course of actions which occurs in social practices within an institution on the other hand. Obligations and rights are not simply bound to actions, but to systems of actions given in the form of systems of social practices. This adds an essential component which has been neglected in formal treatments so far. The inclusion of social practices yields a rich structure in which the emergence and maintenance of of norms can be tackled in a realistic way.

Keywords:
Institution, Organization, Norm, Joint Attitude, Social Practice