

TRUST & ROLES
Andrew J I Jones
Dept of Computer Science
King's College London
ajijones@dcs.kcl.ac.uk

Background to this presentation:

- **The development of a multi-modal logical language within the the European IST project ALFEBIITE (IST-1999-10298).**
- **The language is capable of representing a number of concepts central to the characterisation of norm-governed interaction in multi-agent systems.**
- **See 'A Logical Framework', by Andrew J I Jones, ch. 3 of Jeremy Pitt, ed., *The Open Agent Society*, John Wiley & Sons, forthcoming.**

- **The principal modalities include:**
 - **Modalities for action, attempted action and ability**
 - **The belief modality**
 - **Directive deontic modalities, to represent permissions and obligations**
 - **The modal connective \Rightarrow_s (the 'counts as' connective, Jones & Sergot 1996)**

- **In terms of these basic modalities, a number of more complex notions have been defined, including *rights* (see, e.g., Jones & Sergot, 1992) and *institutionalised power*.**

- **Within ALFEBIITE:**
 - **Communication: includes a new formal theory of ACLs (see also Jones & Parent, 2003)**
 - **Trust (see also Jones, 2002)**
 - **Roles**

- **Note: a principal aim is to place these compound ('molecular') concepts within a general framework for formally describing agent interaction, and to characterise them in terms of a small group of 'atomic' modalities.**

- **Turning now to the concept *role*, consider the following:**

- **It is a characteristic feature of [...] normative systems that they employ the attributes defined by [...] qualifying or constitutive norms [...] as conditions in norms of conduct. For instance, a normative system which has defined the attributes of being a child may go on to require that a child shall not take part in a performance or rehearsal the duration of which exceeds three and a half hours, that a child shall not be present at a place of performance or rehearsal before ten in the morning unless the child lives or receives education in that place, etc. If so, the normative system clearly uses an attribute defined in the system as the condition of some of its norms. Because of their prevalence in normative systems clusters of norms organized in this way deserve a name of their own. We shall call them role structures because in terms of them it is possible to define the sociological notion of a role. (Pörn, 1977, p.62.)**

- **The basic form of a role condition:**

$$F_i \Rightarrow_s G_i$$

which says that, according to institution s , agent i 's having property F counts as i 's having property G .

- **F might denote a rather simple property (e.g., 'being on s 's payroll may count, for s , as being an employee of s), but it might in other cases consist of a complex package of properties (consider, e.g., the definition of British nationality).**
- **In a number of cases, the properties included in F may include a set of proven practical skills that i must have if he is to be deemed by s to have property G . (Consider, e.g., what counts as a medical doctor, what counts as a car mechanic.)**

- **The essential characteristic of a role structure is that any given role condition has associated with it a package of norms, specifying the obligations, permissions, rights and powers of any agent who fulfils the role condition.**
- **So expressions of the following type will ordinarily be associated with any given role condition:**
 - **$OE_j A$**
 - **$PE_j A$**
 - **$E_j A \Rightarrow_s E_k A$**
- **The approach here taken facilitates the systematic investigation of various types of *role conflict*.**

- **'On the Concept of Trust', *DECISION SUPPORT SYSTEMS* 33 (2002), 225-232**

Some examples

In each of the following situations it is true to say that *x* trusts *y*:

S1- (the regularity scenario)

***X* believes that there exists a regularity in *y*'s behaviour, so that under particular kinds of circumstances *y* exhibits a particular kind of behaviour (he does *Z*). In addition, *x* believes that this regularity will also be instantiated on some future occasion(s); that is to say, *x* believes that the future occasion(s) will not prove to be an exception.**

S2- (the obligation scenario)

***X* believes that there is a rule requiring *y* to do *Z*, and that *y*'s behaviour will in fact comply with this rule. For instance, *x* believes that *y* is under an obligation to repay a debt, and that *y* will indeed make the repayment.**

S3 – (*the role scenario*)

X believes that y occupies some particular role, and that y will perform the acts associated with that role in a competent manner. This is what is meant when it is said, for instance, that x trusts his doctor, or x trusts his car mechanic.

S4 – (*the informing scenario*)

X believes that y is transmitting some information to him, and that the content of y's message, or signal, is reliable. For instance, y says to x "Norwegians eat rotten fish", and x believes what he says.

S5 – (*the intention/interests scenario*)

X intends to see to it that some state of affairs Z is realised, and/or x believes that it is in his interests that Z obtains; furthermore, x believes that y will do nothing which will make it less likely that Z obtains.

- In each of these scenarios, the core of x 's trusting attitude lies in two beliefs, which will be called the 'rule-belief' and the 'conformity-belief', respectively.

S1- (the regularity scenario)

In S1, the rule-belief is x 's belief that there exists a regularity in y 's behaviour.

x 's rule-belief: $B_x(A \rightsquigarrow E_y Z)$

x 's conformity-belief: a belief to the effect that exceptional circumstances will not in fact arise on the occasion or occasions concerned, and that the regularity will then again be instantiated: $B_x(A \rightarrow E_y Z)$

♦ **In distinguishing cases of type S1 from cases of type S2, it should be noted that there is in the former, as here understood, no assumption of an agreement between *x* and *y*, or of the existence of an obligation, according to which *y* is *required* to do *Z*. This feature of cases of type S1 might be described by saying that *x*'s expectation vis-à-vis *y* is a purely factual - rather than normative - expectation. If *y* does not do *Z*, and thus fails to act in accordance with *x*'s expectation, *x* will see this as an *exception* to the believed regularity in *y*'s behaviour and not as an act of *violation* of some obligation.**

S2- (the obligation scenario)

- In S2, the rule-belief component of *x*'s trusting attitude is *x*'s belief that *y* is under an obligation to do Z. And the conformity-belief component is *x*'s belief that *y*'s behaviour will be of a kind which fulfils this obligation. Here, *x* may be said to have a normative expectation vis-à-vis *y* in the sense that *x* believes that there is a requirement that *y* is to do Z. This expectation - *x*'s rule-belief - is in itself compatible with a belief, or suspicion, on *x*'s part that *y* will violate his obligation; however, *x*'s conformity-belief is that what in fact will happen is that *y* will meet his obligation, i.e., that *y* will do what he is supposed to do. In cases of type S2, then, trust amounts to belief in *de facto* conformity to normative requirement.

- **Rule-belief:** $B_x O E_y Z$

- **Conformity-belief:** $B_x (O E_y Z \rightarrow E_y Z)$

S3 – (*the role scenario*)

- ◆ **Scenarios of type S3 are intended to cover such uses of 'trust' as are exemplified by "x trusts his doctor", "x trusts his car-mechanic", and so on. The assumption is that what is said to be trusted in these instances is behaviour associated with some particular role(s): x trusts his doctor/car-mechanic to perform competently the roles associated with being a doctor/being a car mechanic.**

- ◆ **The rule-belief/conformity-belief model again applies, provided that one is prepared to accept that a central feature of any given role is that it has associated with it a set of normative standards. It is expected (required) of a doctor, for instance, that he exercise particular skills in ways which meet certain standards of competence. To say that *y* occupies the role of doctor, is not just to say that *y* is recognised as having particular kinds of skills; it is also to say that he is required to exercise these skills in a proficient manner. Thus the rule-belief component of *x*'s trust in his doctor *y*, is *x*'s belief that there are standards that the actions of an agent occupying the role of doctor are required to meet. The conformity-belief is *x*'s belief that *y*'s actions will satisfy these standards.**
- ◆ **Thus, on this approach, scenarios of type S3 turn out to be particular instances of scenarios of type S2.**

