On the Attitude of Trust: A Formal Characterization of *Trust*, *Distrust*, and Associated Notions

Andrew J.I. Jones

Abstract Using modal logics to represent an agent's *beliefs*, *knowledge* and *wants*, an analysis is given of *trust* in terms of an agent's certainty that a particular, desired state-of-affairs will be realized. Similarly, a corresponding analysis of *distrust* is given. Placing these formal representations of *trust* and *distrust* at each of the ends of a spectrum, four intermediary structures may be identified, representing *hope*, two species of *anxiety*, and *fear*. In this way the relationships between the attitudes of trust/distrust and some basic types of emotional state may be precisely articulated. Some suggestions are also made regarding the analysis of some more complex types, including *regretting* that one trusted, and *being ashamed* that one trusted.

The paper employs modalities of type KD and KT for, respectively, the logics of *belief* and *knowledge*. It is shown that use of stronger doxastic and epistemic logics – of the kind often favoured in Artificial Intelligence – containing the positive and negative introspection axioms, would make three of the spectrum's four intermediary positions logically inconsistent. It is suggested that this result provides good reason for rejecting the stronger logics, and that their adoption in AI has often been motivated primarily by considerations of computational convenience, rather than by considerations of conceptual accuracy.

Keywords Trust • Distrust • Hope • Anxiety • Fear • Regret • Shame • Formal taxonomy of the emotions • Applied modal logic

Dedicated to the memory of my former mentor and research colleague Ingmar Pörn, who died in February 2014.

A.J.I. Jones (✉)
Department of Informatics, King's College London, London, UK

Imperial College London, London, UK
e-mail: andrewji.jones@kcl.ac.uk

A. Herzig, E. Lorini (eds.), *The Cognitive Foundations of Group Attitudes and Social Interaction*, Studies in the Philosophy of Sociality 5,
DOI 10.1007/978-3-319-21732-1_6

1 Introduction

Consider the following examples in which one agent trusts another:

- *x* trusts *y* to fulfil a contractual obligation;
- *x* trusts *y* to fulfil properly a role;
- *x* trusts what *y* says.

In Jones (2002) it was argued that, in each of these examples, the *content*, or *object*, of *x*'s trusting attitude concerns trustee compliance: *y*'s conformity to some governing norms or conventions. For the first example, the case is obvious; the contractual obligation is specified by some norm or other, and what *x* trusts is that *y* will comply with that norm. In the second example, the case turns on the assumption that roles are characterized, in part, in terms of a set of norms that apply to the role-holder when he is acting in that role – *cf.* (Pörn 1977, pp. 61–63). Trusting one's physician, for instance, amounts to trusting that he acts in ways that conform to the standards governing members of the medical profession. For the third example, the case turns on an assumption to the effect that indicative signalling, verbal or non-verbal, exploits conventions that correlate signalling act-types to types of states of affairs; when an instance of a given signalling act-type is performed, the conventionally correlated state of affairs *ought* then to hold. (For a detailed development of this approach to indicative signalling, see Jones and Kimbrough (2008); Jones and Parent (2007). The origin of the approach lies in Stenius (1967).)

In what follows, this 'trustee-compliance' view of the object of the trusting attitude will be presupposed.[1] The focus here will be not on *the object of* the trusting attitude, but rather on the *trusting attitude itself*. And in regard to the characterization of that attitude, Jones (2002) fell short in at least two respects:

- it described the *cognitive* aspect of the truster's attitude in terms of mere belief; but the fully trusting agent feels sure, certain, secure that trustee compliance will occur;
- it overlooked the *volitional* component.

The second of these two points reflects the fact that, ordinarily, it *matters* to truster *x* that compliance is forthcoming; compliance is not an issue on which *x* is indifferent: compliance is something that he wants. The presence of the volitional component in the trusting attitude explains, at least in part, why *trust* is so often linked to the notion of *risk*.

In order to develop an improved account of the trusting attitude, capable of repairing the shortcomings of the earlier approach, the point of departure here will be Pörn's modal-logical taxonomy of types of emotions (Pörn 1986), in which modal logics are used to represent the cognitive and volitional components alluded

[1] In my opinion most, if not all, other typical examples of situations in which one agent trusts another can also be understood in terms of this 'trustee-compliance' view of the object of the trusting attitude; but I shall not here argue that case.

to above.[2] One consequence will be that it becomes possible to get a clearer picture of the relationship between *trust* and *distrust*, on the one hand, and the cognitive and volitional aspects of *hope*, *anxiety* and *fear*, on the other.

2 Cognitive and Volitional Positions

Pörn (1986) applied the combinatory method of maxi-conjunctions, developed by Kanger for classifying types of Hohfeldian rights-relations.[3]

For the logic of *belief*, a modality of type KD is used, with the operator relativized to individual agents. The system KD of modal logic is formed by adding to the smallest normal system K – as defined in Chellas (1980) – the schema D:

D $B_x p \rightarrow \neg B_x \neg p$

which says that if an agent x believes that p, where p is any proposition, then he does not believe not-p. For the logic of *knowledge*, a modality of type KT is used, with the operator again relativized to individual agents. The system KT of modal logic is formed by adding to the smallest normal system K the schema T:

T $K_x p \rightarrow p$

A central conjecture in Pörn (1986) is that an agent's *certainty that p* may be represented as the agent's *believing that he knows that p*. Accordingly, the following two *certainty* positions may be identified:

$B_x K_x p$: x is certain that p
$B_x K_x \neg p$: x is certain that $\neg p$

In virtue of the logical properties of the two modalities B and K, as modalities of type KD and KT, respectively, the following relations of logical implication may be shown to hold between the two certainty positions and other, weaker doxastic-epistemic positions:

$B_x K_x p \rightarrow \neg B_x \neg K_x p \rightarrow \neg B_x K_x \neg p$
$B_x K_x p \rightarrow B_x \neg K_x \neg p \rightarrow \neg B_x K_x \neg p$
$B_x K_x \neg p \rightarrow \neg B_x \neg K_x \neg p \rightarrow \neg B_x K_x p$
$B_x K_x \neg p \rightarrow B_x \neg K_x p \rightarrow \neg B_x K_x p$

The class of doxastic-epistemic 'positions' may now be generated as follows: first take the four positive expressions $B_x K_x p$, $B_x K_x \neg p$, $B_x \neg K_x p$, $B_x \neg K_x \neg p$, and then

[2] Not long after the publication of Jones (2002), Ingmar Pörn suggested to me in conversation that the account there put forward had completely overlooked the *affective* aspect of trust. The present paper aims, in part, to remedy that oversight.

[3] For an overview of the method of generating 'normative positions', and for references to the work of Kanger and Hohfeld, see Jones and Sergot (1993).

the corresponding negative expressions $\neg B_x K_x p$, $\neg B_x K_x \neg p$, $\neg B_x \neg K_x p$, $\neg B_x \neg K_x \neg p$. These eight expressions can be arranged as four truth-functional tautologies:

1. $B_x K_x p \vee \neg B_x K_x p$
2. $B_x K_x \neg p \vee \neg B_x K_x \neg p$
3. $B_x \neg K_x p \vee \neg B_x \neg K_x p$
4. $B_x \neg K_x \neg p \vee \neg B_x \neg K_x \neg p$

Obviously, for any given agent, and for any proposition p, precisely one of the disjuncts in each of (1)–(4) must hold. There are 16 ways of selecting precisely one disjunct from each of (1)–(4), to form 16 conjunctions of four conjuncts each. Of these 16 conjunctions, just 6 are logically consistent, given the logical properties adopted for the two modal operators. The 6 logically consistent conjunctions are:

(DE1) $B_x K_x p \,\&\, \neg B_x \neg K_x p \,\&\, \neg B_x K_x \neg p \,\&\, B_x \neg K_x \neg p$
(DE2) $\neg B_x K_x p \,\&\, B_x \neg K_x p \,\&\, B_x K_x \neg p \,\&\, \neg B_x \neg K_x \neg p$
(DE3) $\neg B_x K_x p \,\&\, B_x \neg K_x p \,\&\, \neg B_x K_x \neg p \,\&\, B_x \neg K_x \neg p$
(DE4) $\neg B_x K_x p \,\&\, B_x \neg K_x p \,\&\, \neg B_x K_x \neg p \,\&\, \neg B_x \neg K_x \neg p$
(DE5) $\neg B_x K_x p \,\&\, \neg B_x \neg K_x p \,\&\, \neg B_x K_x \neg p \,\&\, B_x \neg K_x \neg p$
(DE6) $\neg B_x K_x p \,\&\, \neg B_x \neg K_x p \,\&\, \neg B_x K_x \neg p \,\&\, \neg B_x \neg K_x \neg p$

It may be shown that these six positions are mutually exclusive, and their disjunction is a logical truth. So precisely one of (DE1)–(DE6) must hold for any given proposition p.

Concerning (DE6), Pörn said (Pörn 1986), p. 208 that it "... is the epistemic null-position; it is doubtful whether it is relevant for the analysis of emotions since in this position the subject has no belief at all concerning p. (An epistemic null-position may of course be the object of an emotion, but that is another matter.)" And in the development of his analysis of atomic emotional types he chose to disregard (DE6). In what follows, however, the possibility will be left open that (DE6) might be of relevance, particularly in the context of comparing *trust* with the (doxastic-epistemic components of) emotions. So (DE6) will be retained.

Each of the six (DE) positions may be simplified by removing any conjuncts that are themselves logically implied by one or more other conjunct. The result of that simplification is as follows:

(SDE1) $B_x K_x p$
(SDE2) $B_x K_x \neg p$
(SDE3) $B_x \neg K_x p \,\&\, B_x \neg K_x \neg p$
(SDE4) $B_x \neg K_x p \,\&\, \neg B_x K_x \neg p \,\&\, \neg B_x \neg K_x \neg p$
(SDE5) $B_x \neg K_x \neg p \,\&\, \neg B_x K_x p \,\&\, \neg B_x \neg K_x p$
(SDE6) $\neg B_x \neg K_x p \,\&\, \neg B_x \neg K_x \neg p$

As regards the logic of volition, let expressions of the form $D_x p$ be read 'x desires/wants that p', where D_x is a (relativized) normal modality of type KD. It may then readily be shown that there are just three basic volitional positions for any given agent x and any proposition p. They are:

(V1) $D_x p$
(V2) $D_x \neg p$
(V3) $\neg D_x p \ \& \ \neg D_x \neg p$

Following Pörn, (V3) may be said to be the position of 'volitional indifference'. He was inclined to the view that it is irrelevant to the analysis of the emotions; given the present interest in the analysis of *trust*, and given what was said in the introductory remarks to the effect that it ordinarily *matters to* truster x that trustee y acts in a way that fulfils the trust bestowed upon him, the focus here will also be exclusively on volitional positions (V1) and (V2); indifference will be disregarded.

3 Cognitive and Volitional Positions Combined

The result of conjoining (V1) and (V2), respectively, to each of (SDE1)–(SDE6) is given in the following list of 12 doxastic-epistemic/volitional positions:

(DEV1) $B_x K_x p \ \& \ D_x p$
(DEV2) $B_x K_x p \ \& \ D_x \neg p$
(DEV3) $B_x K_x \neg p \ \& \ D_x p$
(DEV4) $B_x K_x \neg p \ \& \ D_x \neg p$
(DEV5) $B_x \neg K_x p \ \& \ B_x \neg K_x \neg p \ \& \ D_x p$
(DEV6) $B_x \neg K_x p \ \& \ B_x \neg K_x \neg p \ \& \ D_x \neg p$
(DEV7) $B_x \neg K_x p \ \& \ \neg B_x K_x \neg p \ \& \ \neg B_x \neg K_x \neg p \ \& \ D_x p$
(DEV8) $B_x \neg K_x p \ \& \ \neg B_x K_x \neg p \ \& \ \neg B_x \neg K_x \neg p \ \& \ D_x \neg p$
(DEV9) $B_x \neg K_x \neg p \ \& \ \neg B_x K_x p \ \& \ \neg B_x \neg K_x p \ \& \ D_x p$
(DEV10) $B_x \neg K_x \neg p \ \& \ \neg B_x K_x p \ \& \ \neg B_x \neg K_x p \ \& \ D_x \neg p$
(DEV11) $\neg B_x \neg K_x p \ \& \ \neg B_x \neg K_x \neg p \ \& \ D_x p$
(DEV12) $\neg B_x \neg K_x p \ \& \ \neg B_x \neg K_x \neg p \ \& \ D_x \neg p$

(DEV1)–(DEV10) are Pörn's ten 'atomic emotional types'. It is interesting to consider the labels he gave to them. (DEV1) and (DEV4) are both types of *security*, in the sense that, in each case, what the agent is certain of *matches* what he desires; by contrast, (DEV2) and (DEV3) represent *despair* (Pörn's label), or *hopelessness*, since in each case what the agent is certain of is the *opposite* of that which he desires.[4] (DEV5) and (DEV6) both represent a form of *anxiety*, in as much as the agent believes that he does not know whether p holds – and in the one case he wants p, whereas in the other he doesn't. Consider next (DEV7): the agent desires p and (first conjunct) believes that he does not know that p; although (second conjunct) he does not believe that he knows not-p, his knowing not-p is compatible with all

[4]Consider the renowned cartoon-style picture by Roy Lichtenstein of the face of a young woman, resting on a pillow, tears flowing, thinking to herself 'That's the way it *should* have *begun*! But it's hopeless!' The situation is essentially that described by (DEV3): she is certain that it (the affair??) did *not* begin in that way, and she wishes that it *had*.

that he believes (third conjunct). (That description of the third conjunct follows (Hintikka 1962) by interpreting '$\neg B_x \neg$' as 'compatibility with all that x believes'.) So (DEV7), and its counterpart (DEV10), represent *fear*. Parallel considerations lead to the conclusion that (DEV8) and (DEV9) represent *hope*.

(DEV11) and (DEV12) of course do not figure among Pörn's atomic types, since they are based on the epistemic null-position. But perhaps a case can be made for maintaining that they represent *another* type of *anxiety*. If it is compatible with all that an agent believes that he knows that p, but also compatible with all that he believes that he knows that not-p, then it would seem that he totally lacks any information that would enable him to decide the question of p's truth/falsity. But then if he is not indifferent, either wanting p to be the case or wanting not-p to be the case, he has grounds for anxiety – albeit grounds of a cognitive type different from that expressed in (DEV5) and (DEV6). This point will be considered further below, in the discussion of *trust* and *distrust*.

It is important to note Pörn's emphasis that – as he put it – the atomic types are 'unrestricted', in as much as their characterization is independent of any particular specification of the kind of state-of-affairs p describes. He then considers ((Pörn 1986), pp. 209–210), by way of contrast, some examples of emotions, such as *anger*, that *are* 'restricted to objects of a certain kind'. It is at this point that the above account of the doxastic-epistemic/volitional positions can be linked to the introductory discussion of the object of the trusting attitude.

4 A Spectrum of Cases

Assume now that the proposition p in (DEV1)–(DEV12) is restricted to *trustee compliance* in the sense described in the first paragraph of this paper, and elaborated in (Jones 2002). It is appropriate then to confine attention to those six cases in which $D_x p$, rather than $D_x \neg p$, appears, since the assumption is that truster x desires compliance on the part of the trustee. The contracted list is this:

(DEV1) $B_x K_x p \ \& \ D_x p$
(DEV3) $B_x K_x \neg p \ \& \ D_x p$
(DEV5) $B_x \neg K_x p \ \& \ B_x \neg K_x \neg p \ \& \ D_x p$
(DEV7) $B_x \neg K_x p \ \& \ \neg B_x K_x \neg p \ \& \ \neg B_x \neg K_x \neg p \ \& \ D_x p$
(DEV9) $B_x \neg K_x \neg p \ \& \ \neg B_x K_x p \ \& \ \neg B_x \neg K_x p \ \& \ D_x p$
(DEV11) $\neg B_x \neg K_x p \ \& \ \neg B_x \neg K_x \neg p \ \& \ D_x p$

A suitable label for (DEV1) is *complete trust*: x is certain that compliance, which he desires, will be forthcoming. Thus *complete trust* is a specific instance of what Pörn calls *security*. By contrast (DEV3) represents *complete distrust*: x is certain that compliance, which he desires, will *not* be forthcoming. Thus, plausibly enough, *complete distrust* is a specific instance of *despair* or *hopelessness*.

The six cases might in fact be considered to constitute a spectrum, with *complete trust* at the left-hand end and *complete distrust* at the right-hand end. Next to *complete trust* comes (DEV9), which represents *hope-of-compliance*, and

immediately preceding *complete distrust* comes (DEV7), which represents *fear-of-non-compliance*. The middle of the spectrum is occupied by the positions (DEV5) and (DEV11), which represent two species of *anxiety-about-whether-compliance-will-occur*. So the spectrum looks like this:

(DEV1) – (DEV9) – [(DEV5), (DEV11)] – (DEV7) – (DEV3)

One of the contexts in which matters of *trust* have lately been given a great deal of attention is the field of e-commerce. In that context, a distinction is often drawn between commercial interactions in which the traders have some previous experience of each other, and the so-called 'first-trade scenario', where the parties may be completely unfamiliar with one another. That distinction may perhaps be used to illustrate the differences between the two types of *anxiety* in the spectrum, (DEV5) and (DEV11). The latter fits, it seems, the kind of situation that would arise in a 'first-trade scenario' if the one party, x, totally lacks information relevant to assessing the trustworthiness of the other party, y, whereas (DEV5) would be a more appropriate description of the situation in which, on the basis of previous experience of y, x has come to the conclusion that he just doesn't know whether or not y can be trusted.

Another way of highlighting the difference between (DEV5) and (DEV11) is as follows: in virtue of its first conjunct, (DEV11) logically implies $\neg B_x \neg p$, and in virtue of its second conjunct it logically implies $\neg B_x p$. However, $B_x p$ may be consistently conjoined with (DEV5), and $B_x \neg p$ may be consistently conjoined with (DEV5) – but obviously not both, because of the D schema. (DEV11) is characterized by the agent's lack of relevant information; only when that lack is remedied can he move to a position that would be compatible either with the belief that p, or with the belief that not-p.

Some may object to the description of the spectrum offered above, and indeed more generally to Pörn's approach to the characterization of the emotions, on the grounds that there is more to an emotional state than the mere combination of epistemic-doxastic and volitional elements, making it inappropriate to use such terms as *hope*, *fear, anxiety* as labels. But nothing essential hinges on the use of those terms; the six positions, and their ordering on the spectrum, are clearly characterized by means of the component logics, and the entire account could thus be re-formulated without appeal to the emotion-terms. The key point is that a small set of modal building-blocks have been used to describe precisely and formally the attitudes of *complete trust* and *complete distrust*, and to exhibit their respective relationships to, and differences from, a set of intermediary attitudes.

5 Strengthening the Logics of Belief and Knowledge

It is commonly accepted that knowledge implies belief. So now add to the logics described above the schema:

KB $\quad K_x p \rightarrow B_x p$

Furthermore, it has been usual in Artificial Intelligence to adopt KD45 for the logic of belief and KT5 for the logic of knowledge. Essentially, this amount to adding to the logic KD (for belief), and the logic KT (for knowledge), the so-called *positive* and *negative introspection* schemas*:*

B4 $B_x p \rightarrow B_x B_x p$ *(positive introspection)*
B5 $\neg B_x p \rightarrow B_x \neg B_x p$ *(negative introspection)*
K4 $K_x p \rightarrow K_x K_x p$ *(positive introspection)*
K5 $\neg K_x p \rightarrow K_x \neg K_x p$ *(negative introspection)*

From the semantical point of view, this strengthened logic of knowledge and belief can be characterized by means of a standard model (in the sense of Chellas (1980)) in which there are two binary accessibility relations $R^K{}_x$ and $R^B{}_x$ satisfying the following properties:

$R^K{}_x$ is both reflexive and Euclidean
$R^B{}_x$ is serial, transitive and Euclidean
$R^B{}_x$ is a sub-relation of $R^K{}_x$.

The basic truth condition for sentences of the form $K_x p$ is given as follows:

(TCK) At any world w in any standard model M, $K_x p$ is true at w iff p itself is true at every world w_1 such that $\langle w, w_1 \rangle \in R^K{}_x$

Similarly, the basic truth condition for sentences of the form $B_x p$ is given by:

(TCB) At any world w in any standard model M, $B_x p$ is true at w iff p itself is true at every world w_1 such that $\langle w, w_1 \rangle \in R^B{}_x$.

Adoption of this strengthened logic of knowledge and belief would have significant consequences for the 'trust-distrust' spectrum described above, since *three* of the six component positions – (DEV7), (DEV9) and (DEV11) – would become *inconsistent*. In terms of the semantics, the inconsistencies may be demonstrated in the following way, starting from (DEV11).

Suppose that each of the first two conjuncts of (DEV11), $\neg B_x \neg K_x p$ and $\neg B_x \neg K_x \neg p$, is true at some world w in a model M of the kind just outlined above. Since $\neg B_x \neg K_x p$ holds at w, it follows by (TCB) that there must be some world w_1 such that $\langle w, w_1 \rangle \in R^B{}_x$ and such that $K_x p$ holds at w_1. Similarly, since $\neg B_x \neg K_x \neg p$ holds at w, it follows by (TCB) that there must be some world w_2 such that $\langle w, w_2 \rangle \in R^B{}_x$ and such that $K_x \neg p$ holds at w_2. Since $R^B{}_x$ is a sub-relation of $R^K{}_x$, it now follows that $\langle w, w_1 \rangle \in R^K{}_x$ and that $\langle w, w_2 \rangle \in R^K{}_x$. But the relation $R^K{}_x$ is Euclidean, so it now follows that $\langle w_1, w_2 \rangle \in R^K{}_x$. But then, since $K_x p$ holds at w_1, it follows by (TCK) that p itself must hold at w_2. However, since $R^K{}_x$ is also reflexive, it follows that $\langle w_2, w_2 \rangle \in R^K{}_x$ and thus, since $K_x \neg p$ holds at w_2, it follows by (TCK) that $\neg p$ also holds at w_2. This reduces to absurdity the initial assumption that each of $\neg B_x \neg K_x p$ and $\neg B_x \neg K_x \neg p$ holds at w.

Consider next (DEV9), and suppose that its second and third conjuncts, $\neg B_x K_x p$ and $\neg B_x \neg K_x p$, are both true at some world w in a model M of the kind under

consideration. Since $\neg B_x \neg K_x p$ holds at w, it follows by (TCB) that there must be some world w_1 such that $<w, w_1> \in R^B_x$ and such that $K_x p$ holds at w_1. Similarly, since $\neg B_x K_x p$ holds at w, it follows by (TCB) that there must be some world w_2 such that $<w, w_2> \in R^B_x$ and such that $\neg K_x p$ holds at w_2. Since R^B_x is a sub-relation of R^K_x, it now follows that $<w, w_1> \in R^K_x$ and that $<w, w_2> \in R^K_x$. But the relation R^K_x is Euclidean, so it now follows that $<w_2, w_1> \in R^K_x$. Now, since $\neg K_x p$ holds at w_2, it follows by (TCK) that there must be some world w_3 such that $<w_2, w_3> \in R^K_x$ and such that $\neg p$ holds at w_3. Since it has been established that $<w_2, w_1> \in R^K_x$ and that $<w_2, w_3> \in R^K_x$, it now follows from a further application of the Euclidean property of R^K_x that $<w_1, w_3> \in R^K_x$. But $K_x p$ holds at w_1; so it follows by (TCK) that p itself must hold at w_3. This reduces to absurdity the initial assumption that each of $\neg B_x K_x p$ and $\neg B_x \neg K_x p$ holds at w. (By means of the same pattern of argument it may also be demonstrated that (DEV7) is inconsistent.)

Apart from the basic truth conditions for the knowledge and belief modalities, these proofs of inconsistency turn on three properties: that R^K_x is Euclidean; that R^K_x is reflexive; and that R^B_x is a sub-relation of R^K_x. The second and third properties are unproblematic: reflexivity is the key to guaranteeing validity of the T schema, and thus that knowledge implies truth, and the sub-relation property guarantees the KB schema, and thus that knowledge implies belief. So the problem lies in the adoption of the Euclidean property for the epistemic accessibility relation – the very property that plays the key role in validating the positive and negative introspection schemas for the knowledge modality. Given the intuitive plausibility of (DEV7), (DEV9) and (DEV11) as representations of, respectively, the *fear-of-non-compliance* position, the *hope-of-compliance* position, and one of the *anxiety* positions, the conclusion to be drawn is that the Euclidean property, and the corresponding introspection schemas, should be rejected.[5]

Why was KT5 often the epistemic logic of choice in AI ? Part of the answer to that question, perhaps, lies in the fact that KT5 has properties that are attractive from the point of view of *computational* tractability. The practice of allowing *computational* considerations to play a significant role in determining choice of *conceptual* model has been quite widespread in AI and in the closely allied research field of Agents and Multi-agent Systems. The problematic nature of that practice is discussed in some detail in (Jones et al. 2013), which outlines an approach to the design of intelligent socio-technical systems in which conceptual and computational models are properly *integrated*.

[5]In some ongoing work on the application of the modal logic of belief to the characterization of *self-deception*, I have reached the same type of conclusions regarding KD45; that work identifies a class of intuitively plausible 'self-deception positions', each of which can be consistently represented if the belief logic is of type KD, and all of which become logically inconsistent if the logic used is KD45. See Jones (2013).

6 Non-atomic Types

Pörn (1986, pp. 210–213), discusses ways in which atomic types of emotions can be combined to form complex types. Pörn considers, for instance, *envy*; suppose

(i) x is envious of y because y got the job.

Here, he suggests, we have a situation in which two instances of *despair* are combined, in that

(ii) x is certain that y got the job, but wishes that he (y) had not; and
(iii) x is certain that he himself did not get the job, but wishes that he had.

So, where $p =$ 'y got the job' and $q =$ 'x got the job', the logical form of (i) becomes:

(iv) $B_xK_xp \;\&\; D_x\neg p \;\&\; B_xK_x\neg q \;\&\; D_xq$

Another way in which complex emotional types can be formed, Pörn suggests, is when the object of an emotion is itself also an emotion. Consider this in relation to the case of *complete trust*, as analysed above[6]:

(v) x is certain that p (i.e., that y will comply), and x desires that p.

And now consider how to interpret

(vi) x *regrets* putting complete trust in y.

What, according to (vi), is x's attitude (doxastic-epistemic and volitional) towards the object of his regret, as expressed by (v) ? A natural answer is that x is certain that (v) and desires that it had not been the case that (v). Expressed formally:

(vii) $B_xK_x(B_xK_xp \;\&\; D_xp) \;\&\; D_x\neg(B_xK_xp \;\&\; D_xp)$

Consider a specific example: x trusted Nick Clegg and the Liberal Democrats at the 2010 UK General Election; x was certain (believed that he knew) that they would deliver on their (manifold) promises, and desired that they should do so. x now regrets trusting: he is quite sure that he had that trust, and he wishes that he hadn't !

What then would be the difference between *that* situation and one in which x is *ashamed of* having trusted Clegg and the Liberal Democrats ? One suggestion would be that x's shame combines his regret with a conviction that he *ought not* to have trusted in the first place: he should have known better, should have been able to see through the pretence . . . . . If a suggestion of that sort is accepted, then the modal-logical language needs to be supplemented with an appropriate normative modality to represent 'ought'. When that component is supplied, the sentence

[6]Although the account that now follows bears some similarities to Pörn's, it differs substantially from it in points of detail. In particular, I do not make use of the notion of the *appropriateness* or *inappropriateness* of an emotion.

(viii) x is ashamed of putting complete trust in y

may be rendered as

(ix) $B_xK_x(B_xK_xp \;\&\; D_xp) \;\&\; D_x\neg(B_xK_xp \;\&\; D_xp) \;\&\; B_xK_x\; Ought\neg(B_xK_xp \;\&\; D_xp)$

7 Concluding Remark

The paper has indicated a way of placing the concepts of *trust* and *distrust* in relation to a broader class of attitudes, in which doxastic-epistemic and volitional components are combined. It is evident that much work remains to be done on exploring the cognitive, volitional and perhaps normative aspects of the structure of complex types of emotions. Hopefully, however, the discussion presented here provides grounds for thinking that, in the spirit of Pörn's work, these phenomena are amenable to rigorous and systematic analysis by means of application of the tools of modal logic.

Acknowledgements 1. A second paper by Ingmar Pörn on the emotions, written in Swedish under the title 'Jagvärderande emotioner' ('Emotions of self-assessment'), also proved useful in writing the present piece. I am most grateful to my daughter, Linnea Erasmie-Jones, for producing an English translation of it. The Swedish copy that I have – a photocopied typescript – is dated 1994. Although I do not have specific details, it was apparently intended for a collection of papers on the emotions to be published by a philosophical society in Sweden or Finland. 2. My thanks to Robert Demolombe for helpful comments on an earlier version of this paper.

References

Chellas, Brian F. 1980. *Modal logic – An introduction*. Cambridge: Cambridge University Press.

Hintikka, Jaakko. 1962. *Knowledge and belief – An introduction to the logic of the two notions*. Ithaca/London: Cornell University Press.

Jones, Andrew J.I. 2002. On the concept of trust. In *Formal modeling and electronic commerce* (Special issue of Decision Support Systems, vol. 33, no. 3), ed. Steven Kimbrough et al., 225–232.

Jones, Andrew J.I. 2013. *On the formal-logical characterisation of self-deception.* Invited talk at ArgMAS 2013, a workshop at AAMAS 2013, St. Paul, Minnesota, May 2013. Draft paper nearing completion.

Jones, Andrew J.I., and Steven O. Kimbrough. 2008. The normative aspect of signalling and the distinction between performative and constative. *Journal of Applied Logic* 6(2): 218–228.

Jones, Andrew J.I., and Xavier Parent. 2007. A convention-based approach to agent communication languages. *Group Decision and Negotiation* 16: 101–141.

Jones, Andrew J.I., and Marek J. Sergot. 1993. On the characterisation of law and computer systems: The normative systems perspective, chapter 12. In *Deontic logic in computer science: Normative system specification* (Wiley professional computing series), ed. J.-J.Ch. Meyer, and R.J. Wieringa, 275–307. Wiley.

Jones, Andrew J.I., Alexander Artikis, and Jeremy V. Pitt. 2013. The design of intelligent socio-technical systems. *Artificial Intelligence Review* 39(1): 5–20.

Pörn, Ingmar. 1977. *Action theory and social science – Some formal models*, Synthese library, vol. 120. Dordrecht/Holland: Reidel.

Pörn, Ingmar. 1986. On the nature of emotions. In *Changing positions*, Philosophical Studies published by the University of Uppsala, Sweden: vol. 38, 205–214.

Stenius, Erik. 1967. Mood and language game. *Synthese* 17: 254–274.