A Logic for Emotions:
a basis for reasoning about commonsense psychological knowledge

Kathryn E. Sanders
Brown University Department of Computer Science
Providence, RI 02912, U.S.A.

Technical Report No. CS-89-23, March 1989

A Logic for Emotions: a basis for reasoning about commonsense psychological knowledge

Kathryn E. Sanders
Department of Computer Science, Box 1910
Brown University, Providence, RI 02912
(401) 863-7672
ks@cs.brown.edu [1]

March 23, 1989

[1] This research has been supported by NSF grants IRI 8515005 and IRI 8801253. The author gratefully acknowledges the assistance of Leora Morgenstern, Tom Dean, Eugene Charniak, Robert McCartney, and Mary Harper.

Abstract

There is a body of commonsense knowledge about human psychology that we all draw upon in everyday life to interpret our own actions, and those of the people around us. In this paper, we define a logic in which this knowledge can be expressed. We focus on a cluster of emotions, including approval, disapproval, guilt, and anger, most of which involve some sort of ethical evaluation of the action that triggers them. As a result, we are able to draw on well-studied concepts from deontic logic, such as obligation, prohibition, and permission. We formalize a portion of commonsense psychology and show how some simple problems can be solved using our logic.

In order to handle concrete problems, since emotions do not occur in a vacuum, it is also necessary to formalize some commonsense knowledge about actions and the probable evaluation of those actions. Specifically, we focus on a cluster of actions having to do with ownership and possession of property – giving, lending, selling, and stealing. We demonstrate that our logic is sufficiently expressive to handle a variety of information about human actions and responses, in a way that is substantially more formal than previous work in this area.

1 Introduction

There is a body of commonsense knowledge about human psychology that we all draw upon in everyday life to interpret our own actions, and those of the people around us. This knowledge is brought to bear by the reader in interpreting texts involving human actions and reactions.

In this paper, we define a logic in which commonsense knowledge about human psychology can be expressed. We formalize a portion of this knowledge and show how some simple problems can be formulated and solved using this logic. We focus on a cluster of emotions, including approval, disapproval, guilt, and anger, most of which involve some sort of ethical evaluation of the action that triggers them. As a result, we are able to draw on well-studied concepts from deontic logic, such as obligation, prohibition, and permission, in formalizing this knowledge.

In order to handle concrete problems, since emotions do not occur in a vacuum, it is also necessary to formalize some commonsense knowledge about actions and the probable evaluation of those actions by the agents and others. Specifically, we focus on a cluster of actions having to do with ownership and possession of property – giving, lending, selling, and stealing – and the predictable responses to those actions. We demonstrate that our logic is sufficiently expressive to handle a variety of information about human actions and responses, in a way that is substantially more formal than previous work in this area.

We will focus on the following examples:

1. Jack went to the supermarket. He parked his car in a legal parking place. When he came out, it was gone.

   Infer that Jack will be angry.
   Infer that Jack will not be grateful.

2. Jack wanted to go to the supermarket. He needed a car. Wilma lent him her car.

   Infer that Jack will be grateful.
   Infer that Jack will not be angry.

3. Barney knew that Jack had donated groceries to a charity.

   Infer that Barney will approve.

4. Jack stole a candy bar from the supermarket. Betty saw him.

   Infer that Betty will disapprove.

5. Jack stole a candy bar from the supermarket. Betty saw him. Jack knows that she saw him and disapproves of what he did, and he cares about her opinion.

   Infer that Jack will be ashamed.

6. The cashier at the supermarket gave Jack too much change. Jack knew that the cashier had given him too much change. He took the change.

   Infer that Jack will feel guilty.

In Section 2, we describe the logic used in this paper. In Section 3, we prove the desired inferences. In Section 4 we contrast our theory with previous work, and in Section 5 we present our conclusions and discuss future work.

2 The logic

2.1 Syntax

In this paper, we use an extension of the temporal logic developed by Shoham in [16], modified to incorporate the three modal operators 'want,' 'know,' and 'believe.' We use an S5 axiom set for 'know,' weak S5 for 'believe,' T without veridicality for 'want,' and the inference rules modus ponens and universal instantiation (cf. [9]). Finally, we use an Axiom of Noncontradiction for 'want' and 'believe': $(Ox\phi) \rightarrow \neg(Ox\neg\phi)$ (where 'O' is either 'believe' or 'want'). In the case of 'know', it follows from the axiom of veridicality that if you know some proposition, you do not also know its negation. The Axiom of Noncontradiction provides a similar rule for 'want' and 'believe.' Unlike Shoham, we assume that the intervals over which an assertion is interpreted are closed. Shoham deliberately makes no commitment one way or the other; we have found that, for purposes of our proofs, it is useful to make a choice, and in general, closed intervals seem more intuitive.

The syntax of our language is the same as in [16], with the following additions:

1. The symbols in the language include the three modal operators, 'want,' 'know,' and 'believe.'

2. The set of *modal formulas* is defined as follows:

   (a) If O is one of the three modal operators, x is a nontemporal term denoting some person, r is an n-ary relation symbol, and $trm_1, ..., trm_n$

are nontemporal terms, then $(O\ x\ (r\ trm_1...trm_n))$ is a modal formula.

(b) If O is one of the three modal operators, x is a nontemporal term denoting some person, and ϕ is a modal formula, then $(O\ x\ \phi)$ is a modal formula.

3. If trm_a and trm_b are temporal terms and ϕ is a modal formula, then TRUE(trm_a, trm_b, ϕ) is a formula.

2.2 Semantics

2.2.1 Informal semantics

Intuitively, the semantics of our language can be understood as follows. We are given a set that includes all of the possible worlds. Possible worlds have a temporal dimension. That is, each possible world is a complete possible history of the world, like a timeline extending infinitely far into the past and the future.

In all the worlds that are knowledge-accessible to an agent x, the propositions that x knows about the past, present, and future all hold. All other propositions vary from world to world. Similarly, in all the worlds that are belief-accessible to x (which may not include the 'real' world, if x has beliefs that are inconsistent with reality) all of x's beliefs hold. Propositions about which x has no particular opinion vary from world to world. The knowledge-accessible worlds are a subset of the belief-accessible worlds, since knowledge implies belief. Finally, the propositions that x wants to be true are true in all the 'want-accessible' worlds, the propositions x wants to be false are false, and propositions about which x is indifferent vary from world to world.

2.2.2 Formal semantics

Formally, the semantics of our language are as follows:

Let D be a domain of individuals, and let $P \subset D$ be a (nonempty) subset of D consisting of all of the persons in D. Let PW be the set of all possible worlds. With each $x \in P$ we associate three relations on PW, B_x, W_x, and K_x, corresponding to the modal operators 'believe', 'want', and 'know', respectively. Let O represent any one of the three modal operators, and let O_x represent any of the three relations. Each of these relations is serial, i.e., it has the property that from any given world at any time, at least one other world is accessible: $\forall w_i \in PW, \forall t_i (\exists w_j \in PW (O_x\ w_i\ w_j\ t_i))$. An interpretation I is a function that maps the nonlogical symbols in the language to some element of D.

Given these definitions, a sentence ϕ is true in a world $w_i \in PW$ under an interpretation I and a variable assignment VA if and only if one of the following is true:

1. ϕ has the form $trm_1 = trm_2$ and $I(w_i, trm_1) = I(w_i, trm_2)$.
2. ϕ has the form $trm_1 \preceq trm_2$ and $I(w_i, trm_1) \leq I(w_i, trm_2)$.
3. ϕ has the form TRUE$(trm_a, trm_b, (r\ trm_1 ... trm_n))$ and the relation $I(w_i, I(trm_a), I(trm_b), r)$ holds on $I(w_i, I(trm_a), I(trm_b), trm_1)$ through $I(w_i, I(trm_a), I(trm_b), trm_n)$.
4. ϕ has the form TRUE(trm_a, trm_b, ψ), where ψ is a modal formula, in the form $\psi = (O\ x\ \zeta)$, O_x is the relation corresponding to O, and TRUE(trm_a, trm_b, ζ) holds in all worlds w_j such that $\forall t_i, trm_a \leq t_i \leq trm_b, (O_x\ w_i\ w_j\ t_i))$.
5. ϕ has the form $\phi_1 \wedge \phi_2$, and both ϕ_1 and ϕ_2 are true.
6. ϕ has the form $\neg\phi_1$, and ϕ_1 is false.
7. ϕ has the form $\forall z \phi_1$, and ϕ is true under all variable assignments VA$'$ that agree with VA everywhere except possibly on z.

2.2.3 Basic concepts

As stated above, we use some key concepts from deontic logic: 'permitted', 'prohibited', and 'obligated.' [19] 'Permitted' is a predicate on actions. Thus, the proposition '(permitted rules (do x a))' is true in a given world if and only if the relation corresponding to the symbol 'permitted' holds on the elements corresponding to 'rules' and '(do x a)' in that world, that is, if x is in fact permitted by the body of rules in question (e.g., law or ethics) to perform that action. For example, we might have ¬(permitted Law (do Antigone (bury Polynices))) and ¬(permitted Religion (do Antigone ¬(bury Polynices))). Permission might also be granted by an individual, for example, (permitted Wilma (do Jack (take car))).

'Obligated' and 'prohibited' are defined in terms of 'permitted' in the usual way. An action is prohibited if performing it is not permitted. An action is obligated if *not* performing it is not permitted.

Where the source of the rules is religion or ethics, these predicates might be expressed in terms such as 'right' and 'wrong', or 'should' and 'should not.'

The more general formulation allows us to express a variety of types of rules and prohibitions using a single set of predicates.

We represent actions as sets of intervals. Thus, the action (do Janet (sell Jack bicycle)) is the set of intervals during which Janet sells Jack a bicycle, in all possible worlds; and the proposition 'TRUE(t_i, t_j, (occurs (do Janet (sell Jack bicycle))))' holds in a given world if the interval (t_i, t_j) in that world belongs to that set. The negation of an action is the complement of the set of intervals represented by the action. Thus, '($\neg$(do Janet (sell Jack bicycle)))' is the set of all intervals during which Janet does *not* sell Jack a bicycle.

In addition, we define the predicate, 'occurs-in,' to handle the situation where the exact boundaries of the interval during which an action occurs are unknown. The proposition 'TRUE(t_i, t_j, (occurs-in (do x a)))' holds if and only if some subinterval of the interval (t_i, t_j) belongs to the set denoted by (do x a). This allows us to reason using information such as 'Jack went to the supermarket on Friday,' without knowing the exact time at which the trip was made.

An action is considered to 'benefit' someone if it is a step in a plan to achieve one of his or her goals. We represent plans as sets of actions, using a notation based on [12]. For example, we can represent a plan for Jack going to the supermarket as ((do Wilma (lend Jack car)) (do Jack (drive car supermarket))). '(plan-step action plan)' holds at any time in any world if 'action' is an element of the set 'plan.' Thus, in the above example, '((do Wilma (lend Jack car))' is a step in the plan for Jack going to the supermarket. Finally, we can express the relationship between the above plan and the goal it is intended to achieve as TRUE(t_1, t_2, (to-do (achieve TRUE(t_3, t_4, (occurs (do Jack6 (go supermarket4))))) ((do Wilma22 (lend Jack6 car8)) (do Jack6 (drive car8 supermarket4))))). Intuitively, the action '(achieve goal)' means causing the goal to become true, and 'TRUE(t_i, t_j, (to-do (achieve goal) plan))' holds in a given world iff the execution of the steps of the plan would cause the goal to become true.

We take 'own' and 'possess' as primitives. Intuitively, ownership is a collection of legal rights, including the right to possess, use, give, rent, or sell an object. By contrast, you possess an object if you physically control it. Thus, if you borrow a car, you possess the car, even though it is still owned by whoever you borrowed it from.

A gift is defined as an unconditional transfer of ownership; a sale is a transfer conditioned on the performance of some action, typically a payment. Similarly, a loan is defined as an unconditional (temporary) transfer of possession, and a rental is a temporary transfer of possession in exchange for the performance of some action, usually a rental payment. (Bank loans are an exception to this general usage, in that the transfer of money is in exchange for an interest payment. For our purposes, this transfer will be considered a 'rental.')

Using these concepts, we define anger, gratitude, shame, guilt, approval, and disapproval. Informally, the definitions are as follows. You become angry at someone if you think they did something wrong, you didn't want them to do it, and you think they knew it was wrong. You are grateful to someone if you think they did something that benefited you that you wanted them to do – such as giving you something or lending you something – and the action was not conditioned on receiving anything in return. You approve of someone if you believe they have done something they should and disapprove if you believe they have done something they shouldn't. You feel ashamed if you believe that you have done something that someone else thinks is wrong, you think they know what you've done, and you care what they think. You feel guilty if you believe that you have done something wrong, and you knew that it was wrong at the time you did it. Formal definitions are given in Figure 1.

3 Proofs

In Section 3.1, we give the proper axioms used in the proofs, in Section 3.2, some useful lemmas, and in Sections 3.3 - 3.8, formal proofs of each of the benchmark examples.

3.1 Axioms

- axioms about time

Axiom 1 *A proposition is true after a given time point iff it is true in some nonpoint interval, open on the left-hand side, starting at that point. In our notation, if a proposition p holds over some interval, TRUE(t_1, t_2, p), then it is true at both endpoints of the interval.*

$\forall$ p, t_1,
[true-after(t_1, p)
$\leftrightarrow$
$\exists t_2, t_1 < t_2$,
$\forall t_3, t_1 < t_3 \leq t_2$,
[TRUE(t_3, t_2, p)]]

anger $\forall x, t_1 \leq t_2, t_3 \leq t_4, y, a,$

[TRUE(t_1, t_2, (believe x TRUE(t_3, t_4, (occurs (do y a))))) $\wedge$
TRUE(t_1, t_2, (believe x TRUE(t_3, t_4, (obligated Ethics $\neg$(do y a))))) $\wedge$
TRUE(t_1, t_2, (want x TRUE($t_3, t_4, \neg$ (occurs (do y a))))) $\wedge$
TRUE(t_1, t_2, (believe x TRUE(t_3, t_4, (know y TRUE(t_3, t_4, (obligated Ethics $\neg$(do y a)))))))
$\leftrightarrow$
TRUE(t_1, t_2, (angry x y (do y a) t_3 t_4))]

You become angry at someone if you think they did something wrong, you didn't want them to do it, and you think they knew it was wrong.

gratitude $\forall x, a, t_3 \leq t_1 \leq t_2, t_3 \leq t_4 \leq t_2, y \neq x,$

[TRUE(t_1, t_2, (believe x TRUE(t_3, t_4, (occurs (do y a))))) $\wedge$
TRUE(t_1, t_2, (want x TRUE(t_3, t_4, (occurs (do y a))))) $\wedge$
TRUE(t_1, t_2, (believe x TRUE(t_3, t_4, (benefit x (do y a))))) $\wedge$
TRUE(t_1, t_2, (believe x $\neg$ TRUE(t_3, t_4, (conditional (do y a)))))
$\leftrightarrow$
TRUE(t_1, t_2, (grateful x y (do y a) t_3 t_4))]

You are grateful to someone if you think they did something that you wanted them to do that benefited you, and their action was not conditioned on receiving anything in return.

approval $\forall x, y, t_1 \leq t_2, t_3 \leq t_4, a,$

[TRUE(t_1, t_2, (believe x TRUE(t_3, t_4, (occurs (do y a))))) $\wedge$
TRUE(t_1, t_2, (believe x TRUE(t_3, t_4, (obligated Ethics (do y a)))))
$\leftrightarrow$
TRUE(t_1, t_2, (approve x y (do y a) t_3 t_4))]

You approve of someone if you believe that they have done something they should.

disapproval $\forall x, y, t_1 \leq t_2, t_3 \leq t_4, a,$

[TRUE(t_1, t_2, (believe x TRUE(t_3, t_4, (occurs (do y a))))) $\wedge$
TRUE(t_1, t_2, (believe x TRUE(t_3, t_4, (obligated Ethics $\neg$(do y a)))))
$\leftrightarrow$
TRUE(t_1, t_2, (disapprove x y (do y a) t_3 t_4))]

You disapprove of someone if you believe that they have done something they shouldn't.

shame $\forall x, t_1 \leq t_2, t_3 \leq t_4 \leq t_2, a,$

[TRUE(t_1, t_2, (believe x TRUE(t_3, t_4, (occurs (do x a))))) $\wedge$
$\exists y \neq x,$
[TRUE(t_1, t_2, (believe x TRUE(t_1, t_2, (disapprove y x (do x a) t_3 t_4)))) $\wedge$
TRUE(t_1, t_2, (want x TRUE(t_1, t_2, $\neg$(disapprove y x (do x a) t_3 t_4))))]
$\leftrightarrow$
TRUE(t_1, t_2, (ashamed x (do x a) t_3 t_4))]

You feel ashamed if you believe that you have done something that someone else thinks is wrong, you think they know what you've done, and you care what they think.

guilt $\forall x, t_1 \leq t_2, t_3 \leq t_4 \leq t_2, a,$

[TRUE(t_1, t_2, (believe x TRUE(t_3, t_4, (occurs (do x a))))) $\wedge$
TRUE(t_1, t_2, (believe x TRUE(t_3, t_4, (obligated Ethics $\neg$(do x a))))) $\wedge$
TRUE(t_1, t_2, (believe x TRUE(t_3, t_4, (know x TRUE(t_3, t_4, (obligated Ethics $\neg$(do x a)))))))
$\leftrightarrow$
TRUE(t_1, t_2, (guilty x (do x a) t_3 t_4))]

You feel guilty if you believe that you have done something that you think is wrong, and you believe that you knew it was wrong at the time you did it.

Figure 1: Definitions.

Axiom 2 *A proposition is true until a given time point iff it is true in some nonpoint interval, ending at that point. This predicate allows us to reason using information such as 'Wilma owned the car when she sold it to the dealer,' where the time at which she began to own the car is unknown or irrelevant.*

$\forall$ p, t_2,
[true-until(t_2, p)
$\leftrightarrow$
$\exists t_1, t_1 < t_2$,
[TRUE(t_1, t_2, p)]]

- axioms concerning 'occurs' and 'occurs-in'

Axiom 3 *An action 'occurs-in' a given interval if and only if there exists some subinterval of that interval in which the action 'occurs.'*

$\forall$ x, a, $t_1 \leq t_4$,
[TRUE(t_1, t_4 (occurs-in (do x a)))
$\leftrightarrow$
$\exists t_2, t_3, t_1 \leq t_2 \leq t_3 \leq t_4$,
[TRUE(t_2, t_3 (occurs (do x a)))]]

Axiom 4 *If the predicate 'occurs-in' holds in an interval, it holds in every interval containing that interval.*

$\forall$ x, a, $t_1 \leq t_2 \leq t_3 \leq t_4$,
[TRUE(t_2, t_3 (occurs-in (do x a)))
$\rightarrow$
TRUE(t_1, t_4 (occurs-in (do x a)))]

- axioms concerning knowledge and belief

Axiom 5 *Knowledge implies belief.*

$\forall$ x, $\phi, t_1 \leq t_2$,
[TRUE(t_1, t_2 (know x ϕ))
$\rightarrow$
TRUE(t_1, t_2 (believe x ϕ))]

Axiom 6 *Belief is downward-hereditary in the sense of [16], that is, whenever it holds over an interval, it holds over all of its subintervals (for our purposes, it is unnecessary to distinguish between point and nonpoint subintervals).*

$\forall$ x, ϕ, $t_1 \leq t_4$,
[TRUE(t_1, t_4 (believe x ϕ))
$\longrightarrow$
$\forall t_2, t_3, t_1 \leq t_2 \leq t_3 \leq t_4$,
[TRUE(t_2, t_3 (believe x ϕ))]]

Axiom 7 *Knowledge is downward-hereditary.*

$\forall$ x, ϕ, $t_1 \leq t_4$,
[TRUE(t_1, t_4 (know x ϕ))
$\longrightarrow$
$\forall t_2, t_3, t_1 \leq t_2 \leq t_3 \leq t_4$,
[TRUE(t_2, t_3 (know x ϕ))]]

- axioms concerning permission, obligation, and prohibition

Axiom 8 *An action is prohibited if performing it is not permitted.*

$\forall$ x, y, a, $t_1 \leq t_2$,
[TRUE(t_1, t_2 (prohibited x (do y a))
$\leftrightarrow$
TRUE($t_1, t_2 \neg$(permitted x (do y a)))]

Axiom 9 *An action is obligated if* not *performing it is not permitted.*

$\forall$ x, y, a, $t_1 \leq t_2$,
[TRUE(t_1, t_2 (obligated x (do y a))
$\leftrightarrow$
TRUE($t_1, t_2 \neg$(permitted x $\neg$(do y a)))]

Axiom 10 *Permission is downward-hereditary.*

$\forall$ x, y, a, $t_1 \leq t_4$,
[TRUE(t_1, t_4 (permitted x (do y a)))
$\rightarrow$
$\forall t_2, t_3, t_1 \leq t_2 \leq t_3 \leq t_4$,
[TRUE(t_2, t_3 (permitted x (do y a)))]]

- axioms concerning 'want'

Axiom 11 *Want is downward-hereditary.*

$\forall$ x, ϕ, $t_1 \leq t_4$,
[TRUE(t_1, t_4 (want x ϕ))
$\rightarrow$
$\forall t_2, t_3, t_1 \leq t_2 \leq t_3 \leq t_4$,
[TRUE(t_2, t_3 (want x ϕ))]]

Axiom 12 *If you own something and you haven't given anyone permission to take it, then you don't want anyone to take it.*

$\forall$ x, y $\neq$ x, object, $t_1 \leq t_2 \leq t_3 \leq t_4$,
[TRUE(t_1, t_4, (owns x object))
$\wedge$
TRUE(t_1, t_4, $\neg$(permitted x (do y (take object x))))
$\rightarrow$
TRUE(t_1, t_4, (want x TRUE($t_2, t_3, \neg$occurs(do y (take object x)))))]

- axioms about ethical rules

Axiom 13 *Contingent agreements*

Informally, the idea is that if x and y agree that x will do a if and only if y does b, and x performs his part of the deal, y should then perform his or her part.

Note: We take 'agree' as a primitive; intuitively, x and y agree if they have a deal, explicit or implicit, that x will do some action a in exchange for y's doing b. Thus, for example, if x sells y a car, they agree that x will transfer ownership of the car to y in exchange for y's paying the price of the car. On the other hand, if x lends y a car, y is obligated to return it at some future point, but the loan is not in exchange for the return – returning the car does not constitute payment for the loan – so the predicate 'agree' does not apply.

$\forall$ x, y, a, b, $t_1 \le t_2 \le t_3 \le t_4$,
[TRUE(t_1, t_4 (agree x y (do x a) (do y b))) $\wedge$
TRUE(t_2, t_3 (occurs (do x a))) $\wedge$
$\neg$TRUE(t_1, t_4 (occurs-in (do y b)))
$\rightarrow$
TRUE(t_3, t_4 (obligated Ethics (do y b)))]

Axiom 14 *If you do not own something, you are not permitted to sell it to someone else. (It might seem more appropriate to cast rules about ownership as legal rules; however, in this area, ethical and legal rules more or less coincide, and it is the ethical rules that we need for purposes of these proofs.)*

$\forall$ x, y, object, $t_1 \le t_2$,
[TRUE(t_1, t_2, (owns x object))
$\rightarrow$
TRUE(t_1, t_2, (prohibited Ethics (do x (sell y object))))]

Axiom 15 *It is prohibited to take something that doesn't belong to you without permission.*

$\forall$ x, y, object, $t_1 \le t_2$,
[TRUE(t_1, t_2, $\neg$(owns x object))
$\wedge$
TRUE(t_1, t_2, (owns y object))
$\wedge$
TRUE(t_1, t_2, $\neg$(permitted y (do x (take object y))))
$\wedge$
TRUE(t_1, t_2, $\neg$(permitted Law (do x (take object y))))
$\rightarrow$
TRUE(t_1, t_2, (prohibited Ethics (do x (take object y))))]

Axiom 16 *Ethics requires that you give to those in need. We take 'need' as a primitive; intuitively, it refers to basic needs such as food, clothing, shelter, and health care.*

$\forall$ x, y $\neq$ x, object, $t_1 \leq t_2$,
[TRUE(t_1, t_2, (need x object))
$\longrightarrow$
TRUE(t_1, t_2, (obligated Ethics (do y (give object x))))

- axioms about ownership, possession, and transfers

  If you own an object, you can possess it, sell it, give it away, use it, or transfer to someone else the right to possess it or use it. These concepts are partially axiomatized here, to the extent necessary for our proofs. A more complete formalization of these concepts is outside the scope of this paper. The basic primitives are 'owns,' which implies the bundle of rights listed above, and 'possess.'

Axiom 17 *Ownership is downward-hereditary.*

$\forall$ x, object, $t_1 \leq t_4$,
[TRUE(t_1, t_4 (owns x object))
$\longrightarrow$
$\forall t_2, t_3, t_1 \leq t_2 \leq t_3 \leq t_4$,
[TRUE(t_2, t_3 (owns x object))]]

Axiom 18 *Possession is downward-hereditary.*

$\forall$ x, object, $t_1 \leq t_4$,
[TRUE(t_1, t_4 (possess x object))
$\longrightarrow$
$\forall t_2, t_3, t_1 \leq t_2 \leq t_3 \leq t_4$,
[TRUE(t_2, t_3 (possess x object))]]

Axiom 19 *One way of taking something away from someone is by moving it out of their possession.*

$\forall$ x, y $\neq$ x, object, ptA, ptB, $t_1 \leq t_2$,
[TRUE(t_1, t_2, (occurs (do x (move object ptA ptB))))
$\wedge$
TRUE(t_1, t_2, (owns y object))
$\wedge$
TRUE(t_1, t_2, $\neg$(possess y ptB))
$\longrightarrow$
TRUE(t_1, t_2, (occurs (do x (take object y))))]

Axiom 20 *A point is in someone's possession if they are at that point, or if it is in a location that they possess (e.g., home, office, etc.)*

$\forall$ x, ptA, $t_1 \le t_2$,
[TRUE(t_1, t_2, (possess x ptA))
$\leftrightarrow$
TRUE(t_1, t_2, (at x ptA))
$\vee$
$\exists$ location,
[TRUE(t_1, t_2, (in ptA location))
$\wedge$
TRUE(t_1, t_2, (possess x location))]]

Axiom 21 *A store possesses a point if it is in a location that the store possesses.*

This axiom could be extended to apply to all legal entities such as businesses or corporations.

$\forall$ store, ptA, $t_1 \le t_2$,
[TRUE(t_1, t_2, (possess store ptA))
$\leftrightarrow$
$\exists$ location,
[TRUE(t_1, t_2, (in ptA location))
$\wedge$
TRUE(t_1, t_2, (possess store location))

Axiom 22 *Store merchandise is never free.*

$\forall$ store, object, $t_1 \le t_2$,
[TRUE(t_1, t_2, (merchandise-of store object))
$\rightarrow$
TRUE(t_1, t_2, (> (price-of object) 0))]

Axiom 23 *Stores do not permit you to take their merchandise unless you buy it.*

$\forall$ store, customer, object, $t_1 \le t_2$,
[TRUE(t_1, t_2, (owns store object))
$\rightarrow$
TRUE(t_1, t_2, $\neg$(permitted store (do customer (take object store))))]

Axiom 24 *The law does not permit you to take a store's merchandise unless you buy it.*

∀ store, customer, object, $t_1 \leq t_2$,
[TRUE(t_1, t_2, (owns store object))
$\rightarrow$
TRUE(t_1, t_2, ¬(permitted Law (do customer (take object store))))]

Axiom 25 *The law does not permit anyone to take a car unless it is parked in an illegal location, and then only if the person is an official.*

∀ x, y ≠ x, car, $t_1 \leq t_2$,
[TRUE(t_1, t_2, (permitted Law (do y (take car x))))
$\leftrightarrow$
TRUE(t_1, t_2, (official y))
∧
∃ loc,
[TRUE(t_1, t_2, (prohibited Law (do x (put car loc))))
∧
TRUE(t_1, t_2, (at car loc))]]

Axiom 26 *Payment*

∀ x, y, amount, $t_1 \leq t_3$,
[TRUE(t_1, t_3 (occurs (do x (pay y amount))))
$\leftrightarrow$
$\exists t_2, t_1 \leq t_2 \leq t_3$,
[TRUE(t_2, t_2 (occurs (do x (transfer-ownership y amount))))]
∨
∃ amount1, amount2, $t_2, t_1 \leq t_2 < t_3$,
[TRUE(t_1, t_2 (occurs (do x (pay y amount1)))) ∧
TRUE(t_3, t_3 (occurs (do x (transfer-ownership y amount2)))) ∧
TRUE(t_1, t_3 (= (amount (+ amount1 amount2))))]]

Axiom 27 *Total payment*

Informally, a payment is a transfer of an amount of money, or the sum of a sequence of transfers. The total payment made during a given interval of time is the sum of all the payments made during that interval. (This assumes that only a finite number of payments can be made in a given interval.)

$\forall$ x, y, amount, number > amount, $t_1 \leq t_2$,
[TRUE(t_1, t_2 (occurs (do x (total-pay y amount))))
$\leftrightarrow$
TRUE(t_1, t_2 (occurs (do x (pay y amount)))) $\wedge$
$\neg$TRUE(t_1, t_2 (occurs (do x (pay y number))))]

Axiom 28 *Net payment*

The net payment made by x to y in a given interval is the sum of all amounts transferred by x to y during the interval, less the sum of all amounts returned.

$\forall$ x, y, amount, amount1, amount2, $t_1 \leq t_2$,
[TRUE(t_1, t_2 (occurs (do x (net-pay y amount))))
$\leftarrow$
TRUE(t_1, t_2 (occurs (do x (total-pay y amount1)))) $\wedge$
TRUE(t_1, t_2 (occurs (do y (total-pay x amount2)))) $\wedge$
TRUE(t_1, t_2 (= (amount (− amount1 amount2))))]

Axiom 29 *Maximum value of net payment*

If x has net-paid a certain amount, he or she has not net-paid any amount greater than that.

Strictly speaking, x has not net-paid any amount less than 'amount' either. This meaning could be captured by simply making axiom 28 an equivalence rather than an implication, and deleting axiom 29. However, this would allow the inference that x has failed to pay for a purchase if he or she makes an overpayment. Axioms 28 and 29 are designed to block this inference.

$\forall$ x, y, amount, number > amount, $t_1 \leq t_2$,
[TRUE(t_1, t_2 (occurs (do x (net-pay y amount))))
$\rightarrow$
$\neg$TRUE(t_1, t_2 (occurs (do x (net-pay y number))))]

Axiom 30 *Transfers of ownership.*

$\forall$ x, y, object, t_1,
[TRUE(t_1, t_1 (occurs (do x (transfer-ownership y object))))
$\leftrightarrow$
true-until(t_1, (owns x object))
$\wedge$
true-until($t_1, \neg$(owns y object))
$\wedge$
true-after(t_1, (owns y object))
$\wedge$
true-after($t_1, \neg$(owns x object))]

Axiom 31 *Transfers of possession.*

$\forall$ x, y, object, t_1,
[TRUE(t_1, t_1 (occurs (do x (transfer-possession y object))))
$\leftrightarrow$
true-until(t_1, (possess x object))
$\wedge$
true-until($t_1, \neg$(possess y object))
$\wedge$
true-after(t_1, (possess y object))
$\wedge$
true-after($t_1, \neg$(possess x object))]

Axiom 32 *If you own something, you go on owning it until you transfer it to someone else.*

$\forall$ x, y, object, $t_1 \leq t_2 \leq t_3$,
[TRUE(t_1, t_2 (owns x object))
$\wedge$
TRUE(t_1, t_3 $\neg$(occurs-in (do x (transfer-ownership y object))))
$\rightarrow$
TRUE(t_2, t_3 (owns x object))]

Axiom 33 *Similarly, if you do not own something, you continue not owning it until someone transfers it to you.*

$\forall$ x, y, object, $t_1 \leq t_2 \leq t_3$,
[TRUE(t_1, t_2 $\neg$(owns x object))

$\wedge$
TRUE(t_1, t_3 $\neg$(occurs-in (do y (transfer-ownership x object))))
$\rightarrow$
TRUE(t_2, t_3 $\neg$(owns x object))]

Axiom 34 *Sales*

A sale is defined as a transfer of ownership of an object, payment of the price of the object, and an agreement, implicit or explicit, that the payment is in exchange for the transfer.

$\forall$ x, y, object, $t_1 \leq t_3$,
[TRUE(t_1, t_3 (occurs (do x (sell y object))))
$\leftrightarrow$
$\exists t_2, t_1 \leq t_2 \leq t_3$,
[TRUE(t_2, t_2 (occurs (do x (transfer-ownership y object))))
$\wedge$
TRUE(t_1, t_3 (occurs (do y (net-pay x (price-of object)))))] $\wedge$
TRUE(t_1, t_3 (agree x y (do x (transfer-ownership y object)) (do y (net-pay x (price-of object)))))]

Axiom 35 *Gifts*

A gift is defined as a transfer of ownership with no payment in return.

$\forall$ x, y, object, $t_1 \leq t_3$,
[TRUE(t_1, t_3 (occurs (do x (give y object))))
$\leftrightarrow$
TRUE(t_1, t_1 (occurs (do x (transfer-ownership y object))))] $\wedge$
$\neg\exists z, b$,
[TRUE(t_1, t_3 (agree x z (do x (transfer-ownership y object)) (do z b)))]

Axiom 36 *Rentals*

A rental is defined as a temporary transfer of possession in exchange for a rent payment. (Generally, this distinction between lending and renting corresponds to the normal English usage; however, lending money at a positive rate of interest would be considered a rental.)

$\forall$ x, y, object, $t_1 \leq t_4$,
[TRUE(t_1, t_4 (occurs (do x (rent y object))))

$\leftrightarrow$
$\exists t_2, t_3, t_1 \leq t_2 \leq t_3 \leq t_4,$
[TRUE(t_2, t_2 (occurs (do x (transfer-possession y object))))
$\wedge$
TRUE(t_1, t_4 (owns x object))
$\wedge$
TRUE(t_1, t_3 (occurs (do y (net-pay x (rental-of object)))))] $\wedge$
TRUE(t_2, t_4 (obligated Ethics (do y (transfer-possession x object)))) $\wedge$
TRUE(t_1, t_4 (agree x y (do x (transfer-possession y object)) (do y (net-pay x (rental-of object)))))]

Axiom 37 *Loans*

A loan is defined as a transfer of possession with no payment in exchange for the transfer, but with an obligation to return the object lent.

$\forall$ x, y, object, $t_1 \leq t_3$,
[TRUE(t_1, t_3 (occurs (do x (lend y object))))
$\leftrightarrow$
TRUE(t_1, t_1 (occurs (do x (transfer-possession y object)))) $\wedge$
TRUE(t_1, t_3 (owns x object)) $\wedge$
TRUE(t_1, t_3 (obligated Ethics (do y (transfer-possession x object)))) $\wedge$
$\forall$ z, b,
[$\neg$TRUE(t_1, t_3 (agree x z (do x (transfer-possession y object)) (do z b)))]]

Axiom 38 *Paying cashiers*

Paying the cashier is the same as paying the store. (This might be true in general of agents, but the rules of agency are complicated, and outside the scope of this paper. For purposes of these proofs, the following two axioms will suffice.)

$\forall$ customer, store, amount, $t_1 \leq t_2$
[TRUE(t_1, t_2 (occurs (do customer (pay (cashier-of store) amount))))
$\leftrightarrow$
TRUE(t_1, t_2 (occurs (do customer (pay store amount))))]

Axiom 39 *Receiving change*

Receiving change from the cashier is the same as receiving it from the store.

$\forall$ customer, store, amount, $t_1 \leq t_2$
[TRUE(t_1, t_2 (occurs (do (cashier-of store) (pay customer amount))))
$\leftrightarrow$
TRUE(t_1, t_2 (occurs (do store (pay customer amount))))]

- axioms about sales agreements

Axiom 40 *Stores offer their merchandise for sale.*

$\forall$ store, customer, $t_1 \leq t_2$, object,
[TRUE(t_1, t_2 (merchandise-of store object))
$\leftrightarrow$
TRUE(t_1, t_2 (owns store object))
$\wedge$
TRUE(t_1, t_2 (offer store (do store (sell object customer))))]

Axiom 41 *If an object is merchandise of a store, it is not owned by anyone but the store.*

$\forall$ store, x $\neq$ store, $t_1 \leq t_2$, object,
[TRUE(t_1, t_2 (merchandise-of store object))
$\rightarrow$
TRUE(t_1, t_2 $\neg$(owns x object))]

Axiom 42 *A sales agreement is made up of an offer and an acceptance.*

$\forall$ store, customer, $t_1 \leq t_2 \leq t_3$, object,
[TRUE(t_1, t_2 (offer store (do store (sell object customer))))
$\wedge$
TRUE(t_1, t_2 (accept customer (do store (sell object customer))))
$\leftrightarrow$
TRUE(t_1, t_3 (agree store customer (do store (transfer-ownership customer object)) (do customer (net-pay store (price-of object)))))]

Axiom 43 *An offer to pay constitutes an acceptance of the offer to sell.*

$\forall$ store, customer, $t_1 \leq t_2$, object,
[TRUE(t_1, t_2 (accept customer (do store (sell object customer))))
$\leftrightarrow$
TRUE(t_1, t_2 (offer customer (do customer (net-pay store (price-of object)))))]

Axiom 44 *An actual payment of any amount constitutes an offer to pay in full.*

$\forall$ store, customer, $t_1 \leq t_2$, amount,
[TRUE(t_1, t_2 (occurs (do customer (pay store amount)))) $\wedge$
TRUE(t_1, t_2 (> amount 0))
$\rightarrow$
$\exists$ object,
[TRUE(t_1, t_2 (offer customer (do customer (net-pay store (price-of object)))))]]

- Axiom about plans

Axiom 45 *An action is a step in a plan if it is an element of the set represented by 'plan.'*

$\forall$ plan, action, $t_1 \leq t_2$,
[TRUE(t_1, t_2, (plan-step action plan))
$\leftrightarrow$
TRUE(t_1, t_2, action $\in$ plan)]

- Axioms about motion and locations

Axiom 46 *Every person is in exactly one place.*

$\forall$ person, loc1, $t_1 \leq t_2$
[TRUE(t_1, t_2 $\neg$(at person loc1))
$\leftrightarrow$
$\exists$ loc2 $\neq$ loc1,
TRUE(t_1, t_2 (at person loc2))]

Axiom 47 *Things don't move by themselves. (Note: For simplicity, we assume that human agency is required. Ideally, we should incorporate a theory of physics and causation, but for purposes of our proofs, this axiom will suffice.)*

$\forall$ ptA, object, $t_1 \leq t_2$,
[TRUE(t_1, t_1, (at ptA object))
$\wedge$
TRUE(t_2, t_2, $\neg$(at ptA object))
$\rightarrow$
$\exists$ person, ptB $\neq$ ptA,
TRUE(t_1, t_2, (occurs-in (do person (move object ptA ptB))))]

Axiom 48 *If you move something from point A to point B, it will then be at point B.*

$\forall$ person, ptA, ptB, object, $t_1 \leq t_2$,
[TRUE(t_1, t_2, (occurs (do person (move object ptA ptB))))
$\rightarrow$
TRUE(t_2, t_2, (at ptB object))]

Axiom 49 *'Put' is defined in terms of 'move': if you put something at point A, that is the same as moving it from some unspecified location to point A.*

$\forall$ person, ptA, object, $t_1 \leq t_2$,
[TRUE(t_1, t_2, (occurs (do person (put object ptA))))
$\leftrightarrow$
$\exists$ ptB,
[TRUE(t_1, t_2, (occurs (do person (move object ptB ptA))))]]

- definition of 'benefit'

Axiom 50 *An action benefits you iff it helps you to achieve one of your goals.*

$\forall$ x, y, action, $t_1 \leq t_2$,
[TRUE(t_1, t_2, (benefit x (do y action)))
$\leftrightarrow$
$\exists$ goal, plan,
[TRUE(t_1, t_2, (want x goal)) $\wedge$
TRUE(t_1, t_2, (to-do (achieve goal) plan)) $\wedge$
TRUE(t_1, t_2, (plan-step action plan))]]

- definition of 'conditional' action

  Axiom 51 *A 'conditional' action is one the agent performs pursuant to an agreement.*

  $\forall$ x, a, $t_1 \leq t_2$,
  [TRUE(t_1, t_2 (conditional (do x a)))
  $\leftrightarrow$
  $\exists$ y, b,
  [TRUE(t_1, t_2 (agree x y (do x a) (do y b)))]]

- miscellaneous axioms

  Axiom 52 *The laws of arithmetic and set theory.*

  Axiom 53 *All agents know all of the logical axioms.*

  Axiom 54 *All agents know all of the proper axioms.*

3.2 Lemmas

Lemma 1 *If a customer pays the cashier at a store, the cashier returns change, and no other amounts are exchanged, the net amount paid by the customer to the store is the amount paid, less the change.*

Given:
TRUE(t_1, t_2, (occurs (do customerA (pay (cashier-of storeB) initial-paymentA))))
$\forall$ number > initial-paymentA,
$\neg$TRUE(t_1, t_2, (occurs (do customerA (pay (cashier-of storeB) number))))
TRUE(t_1, t_2, (occurs (do (cashier-of storeB) (pay customerA changeB))))
$\forall$ number > changeB,
$\neg$TRUE(t_1, t_2, (occurs (do (cashier-of storeB) (pay customerA number))))

Show:
TRUE(t_1, t_2 (occurs (do customerA (net-pay storeB (− initial-paymentA changeB)))))

Proof:
1. $\forall$ amount1, amount2,
[TRUE(t_1, t_2 (occurs (do customerA (net-pay storeB (− initial-paymentA changeB)))))

$\leftarrow$

TRUE(t_1, t_2 (occurs (do customerA (total-pay storeB amount1)))) $\wedge$

TRUE(t_1, t_2 (occurs (do storeB (total-pay customerA amount2)))) $\wedge$

TRUE(t_1, t_2 (= (– initial-paymentA changeB) (– amount1 amount2)))]

– Axiom 28; universal instantiation

2. $\forall$ amount, number $>$ amount,

[TRUE(t_1, t_2 (occurs (do customerA (total-pay storeB amount))))

$\leftrightarrow$

TRUE(t_1, t_2 (occurs (do customerA (pay storeB amount)))) $\wedge$

$\neg$TRUE(t_1, t_2 (occurs (do customerA (pay storeB number))))]

– Axiom 27; universal instantiation

3. TRUE(t_1, t_2 (occurs (do customerA (pay (cashier-of storeB) initial-paymentA))))

$\leftrightarrow$

TRUE(t_1, t_2 (occurs (do customerA (pay storeB initial-paymentA))))

– Axiom 38; universal instantiation

4. TRUE(t_1, t_2 (occurs (do customerA (pay (cashier-of storeB) initial-paymentA))))

– Given.

5. TRUE(t_1, t_2 (occurs (do customerA (pay storeB initial-paymentA))))

– Steps 3, 4, and modus ponens.

6. $\forall$ number,

[$\neg$TRUE(t_1, t_2 (occurs (do customerA (pay (cashier-of storeB) number))))

$\leftrightarrow$

$\neg$TRUE(t_1, t_2 (occurs (do customerA (pay storeB number))))]

– Axiom 38; universal instantiation; definition of equivalence.

7. $\forall$ number $>$ initial-paymentA,

$\neg$TRUE(t_1, t_2, (occurs (do customerA (pay (cashier-of storeB) number))))

– Given.

8. $\forall$ number $>$ initial-paymentA,

$\neg$TRUE(t_1, t_2, (occurs (do customerA (pay storeB number))))

– Steps 6, 7, and modus ponens.

9. TRUE(t_1, t_2 (occurs (do customerA (total-pay storeB initial-paymentA))))

– Steps 2, 5, 8, and modus ponens.

10. $\forall$ amount, number $>$ amount,

[TRUE(t_1, t_2 (occurs (do storeB (total-pay customerA amount))))

$\leftrightarrow$

TRUE(t_1, t_2 (occurs (do storeB (pay customerA amount)))) $\wedge$

$\neg$TRUE(t_1, t_2 (occurs (do storeB (pay customerA number))))]
– Axiom 27; universal instantiation

11. TRUE(t_1, t_2 (occurs (do (cashier-of storeB) (pay customerA changeB))))

$\leftrightarrow$
TRUE(t_1, t_2 (occurs (do storeB (pay customerA changeB))))
– Axiom 39; universal instantiation

12. TRUE(t_1, t_2 (occurs (do (cashier-of storeB) (pay customerA changeB))))

– Given.

13. TRUE(t_1, t_2 (occurs (do storeB (pay customerA changeB))))
– Steps 11, 12, and modus ponens.

14. $\forall$ number,
[$\neg$TRUE(t_1, t_2 (occurs (do (cashier-of storeB) (pay customerA number))))
$\leftrightarrow$
$\neg$TRUE(t_1, t_2 (occurs (do storeB (pay customerA number))))]
– Axiom 39; universal instantiation; definition of equivalence.

15. $\forall$ number $>$ changeB,
$\neg$TRUE(t_1, t_2, (occurs (do (cashier-of storeB) (pay customerA number))))
– Given.

16. $\forall$ number $>$ changeB,
$\neg$TRUE(t_1, t_2, (occurs (do storeB (pay customerA number))))
– Steps 14, 15, and modus ponens.

17. TRUE(t_1, t_2 (occurs (do storeB (total-pay customerA changeB))))
– Steps 10, 13, 16, and modus ponens.

18. TRUE(t_1, t_2 (occurs (do customerA (net-pay storeB (– initial-paymentA changeB)))))
– Steps 1, 9, 17, Axiom 52, and modus ponens. *Q.E.D.*

Lemma 2 *Shoplifting is wrong.*

Given:
TRUE(t_1, t_2, (merchandise-of storeA objectB))

Show:
TRUE(t_1, t_2, (prohibited Ethics (do customerC (take objectB storeA))))

Proof:
1. TRUE(t_1, t_2 (merchandise-of storeA objectB))
$\leftrightarrow$
TRUE(t_1, t_2 (owns storeA objectB))

$\wedge$
TRUE(t_1, t_2 (offer storeA (do storeA (sell objectB customerC))))
– Axiom 40; universal instantiation.

2. TRUE(t_1, t_2 (merchandise-of storeA objectB))

– Given.

3. TRUE(t_1, t_2 (owns storeA objectB))

– Steps 1-2 and modus ponens.

4. [TRUE(t_1, t_2 (merchandise-of storeA objectB))

$\longrightarrow$
TRUE(t_1, t_2 $\neg$(owns customerC objectB))]
– Axiom 41; universal instantiation.

TRUE(t_1, t_2 $\neg$(owns customerC objectB))

– Steps 2, 4, and modus ponens.

6. [TRUE(t_1, t_2, (owns storeA objectB))

$\longrightarrow$
TRUE(t_1, t_2, $\neg$(permitted storeA (do customerC (take objectB storeA))))]
– Axiom 23; universal instantiation.

7. TRUE(t_1, t_2, $\neg$(permitted storeA (do customerC (take objectB storeA))))]

– Steps 3, 6, and modus ponens.

8. [TRUE(t_1, t_2, (owns storeA objectB))

$\longrightarrow$
TRUE(t_1, t_2, $\neg$(permitted Law (do customerC (take objectB storeA))))]
– Axiom 24; universal instantiation.

9. TRUE(t_1, t_2, $\neg$(permitted Law (do customerC (take objectB storeA))))

– Steps 3, 8, and modus ponens.

10. [TRUE(t_1, t_2, $\neg$(owns customerC objectB))

$\wedge$
TRUE(t_1, t_2, (owns storeA objectB))
$\wedge$
TRUE(t_1, t_2, $\neg$(permitted storeA (do customerC (take objectB storeA))))
$\wedge$
TRUE(t_1, t_2, $\neg$(permitted Law (do customerC (take objectB storeA))))
$\longrightarrow$
TRUE(t_1, t_2, (prohibited Ethics (do customerC (take objectB storeA))))]
– Axiom 15; universal instantiation.

11. TRUE(t_1, t_2, (prohibited Ethics (do customerC (take objectB storeA))))]

– Steps 3, 5, 7, 9, and modus ponens. *Q.E.D.*

Lemma 3 *If you put something at a given point, it will then be at that point.*

Given:
TRUE(t_1, t_2, (occurs (do person (put object ptA))))
Show:
TRUE(t_2, t_2, (at ptA object))
Proof:
1. TRUE(t_1, t_2, (occurs (do person (put object ptA))))
– Given.
2. $\exists$ ptB,
[TRUE(t_1, t_2, (occurs (do person (move object ptB ptA))))]
– Axiom 49; universal instantiation; step 1; modus ponens.
3. TRUE(t_2, t_2, (at ptA object))
– Axiom 48; universal instantiation; step 2; modus ponens. *Q.E.D.*

Lemma 4 *All agents believe all of the logical axioms.*

Proof:
Follows from axioms 53 and 5.

Lemma 5 *All agents believe all of the proper axioms.*

Proof:
Follows from axioms 54 and 5.

3.3 Proof 1

Given:

- Fact 1:
  $t_1 \leq t_2 \leq t_4 \leq t_5 \leq t_3$.
- Fact 2:
  TRUE(t_1, t_2, (occurs (do Jack6 (put car3 pkg-space2))))
- Fact 3:
  TRUE(t_3, t_3, $\neg$(at car3 pkg-space2))
- Fact 4:
  TRUE(t_3, t_3, (at Jack6 pkg-space2))

- Fact 5:
  $\forall$ loc,
  [TRUE(t_4, t_5, (at car3 loc)) $\rightarrow$
  TRUE(t_2, t_3, $\neg$(possess Jack6 loc))]

- Fact 6:
  $\forall$ x,
  TRUE(t_1, t_3, $\neg$(permitted Jack6 (do x (take car3 Jack6))))

- Fact 7:
  TRUE(t_1, t_3, (permitted Law (do Jack6 (put car3 pkg-space2))))

- Fact 8:
  TRUE(t_1, t_3, (owns Jack6 car3))

- Fact 9:
  $\forall$ x $\neq$ Jack6,
  TRUE(t_4, t_5, $\neg$(owns x car3))

- Fact 10:
  $\forall$ person,
  TRUE(t_1, t_3, (know person ϕ)), where ϕ is any of the above givens.

Show:
TRUE(t_3, t_3, (angry Jack6 personX (do personX (take car3 Jack6)) t_4 t_5))
Proof:

- Steps 1-9: Jack believes that someone took his car.

  1. TRUE(t_1, t_2, (occurs (do Jack6 (put car3 pkg-space2))))
     – Given.
  2. TRUE(t_2, t_2, (at car3 pkg-space2))
     – Lemma 3; universal instantiation; step 1; modus ponens.
  3. TRUE(t_3, t_3, $\neg$(at car3 pkg-space2))
     – Given.
  4. $\exists$ personX, locX $\neq$ pkg-space2,
     [TRUE(t_2, t_3, (occurs-in (do personX (move car3 pkg-space2 locX))))]
     – Axiom 47; universal instantiation; steps 2-3; modus ponens.
  5. $\exists t_4, t_5, t_2 \leq t_4 \leq t_5 \leq t_3$,
     [TRUE(t_4, t_5 (occurs (do personX (move car3 pkg-space2 locX))))]
     – Axiom 3; universal instantiation; step 4; modus ponens.

6. TRUE(t_4, t_5, (owns Jack6 car3)).
– Fact 8; axiom 17; universal instantiation; modus ponens.

7. TRUE(t_4, t_5, $\neg$(possess Jack6 locX)).
– Fact 5; universal instantiation.

8. TRUE(t_4, t_5, (occurs (do personX (take car3 Jack6))))
– Axiom 19; universal instantiation; steps 5-7; modus ponens.

9. TRUE(t_3, t_3, (believe Jack6 TRUE(t_4, t_5, (occurs (do personX (take car3 Jack6))))
– Lemmas 4 and 5; Fact 10; axiom 5; consequential closure.

- Steps 10-15: Jack believes that it was wrong for someone to take his car.

10. TRUE(t_4, t_5, $\neg$(permitted Jack6 (do personX (take car3 Jack6))))
– Fact 6; universal instantiation; axiom 10; universal instantiation; modus ponens.

11. TRUE(t_4, t_5, (permitted Law (do Jack6 (put car3 pkg-space2))))
– Fact 7; axiom 10; universal instantiation; modus ponens.

12. TRUE(t_4, t_5, $\neg$(permitted Law (do personX (take car3 Jack6))))
– Axiom 25; universal instantiation; step 11; modus ponens.

13. TRUE(t_4, t_5, $\neg$(owns personX car3))
Fact 9.

14. TRUE(t_4, t_5, (prohibited Ethics (do personX (take car3 Jack6)))
– Axiom 15; universal instantiation; steps 6, 10, 12-13; modus ponens.

15. TRUE(t_3, t_3, (believe Jack6 TRUE(t_4, t_5, (prohibited Ethics (do personX (take car3 Jack6)))
– Lemmas 4 and 5; Fact 10; axiom 5; consequential closure.

- Steps 16-19: Jack didn't want anyone to take his car.

16. TRUE(t_1, t_3, (owns Jack6 car3)).
– Fact 8.

17. TRUE(t_1, t_3, $\neg$(permitted Jack6 (do personX (take car3 Jack6))))
– Fact 6; universal instantiation.

18. TRUE(t_1, t_3, (want Jack6 TRUE(t_4, t_5, $\neg$(occurs (do personX (take car3 Jack6))))
– Axiom 12; universal instantiation; steps 16-17; modus ponens.

19. TRUE(t_3, t_3, (want Jack6 TRUE(t_4, t_5, $\neg$(occurs (do personX (take car3 Jack6))))
– Axiom 11; universal instantiation; step 18; modus ponens.

Note: For this reason, it follows that Jack will *not* be grateful that his car has been stolen: it is required as part of the definition of gratitude that the action be one that you wanted to occur.

- Steps 20-21: Jack believes that the person who took his car knew it was wrong.

  20. TRUE(t_4, t_5, (know personX TRUE(t_4, t_5, (prohibited Ethics (do personX (take car3 Jack6)))
  – Axioms 54 and 53; universal instantiation; Fact 10; consequential closure; axiom 7; universal instantiation; modus ponens.

  21. TRUE(t_3, t_3, (believe Jack6 TRUE(t_4, t_5, (know personX TRUE(t_4, t_5, (prohibited Ethics (do personX (take car3 Jack6)))
  – Lemmas 4 and 5; Fact 10; axiom 5; consequential closure.

- Step 22: As a result, Jack was angry.

  22. TRUE(t_3, t_3, (angry Jack6 x (do x (take car3 Jack6)) t_4 t_5))
  – Steps 14, 15, 19, 21; definition of anger; universal instantiation; modus ponens. *Q.E.D.*

3.4 Proof 2

Given:

- Fact 1:
  $t_1 \leq t_2 \leq t_3 \leq t_4$.

- Fact 2:
  TRUE(t_1, t_2, (occurs (do Wilma22 (lend Jack6 car8))))

- Fact 3:
  TRUE(t_1, t_4, (want Jack6 TRUE(t_1, t_2, (occurs (do Wilma22 (lend Jack6 car8))))))

- Fact 4:
  TRUE(t_1, t_4, (want Jack6 TRUE(t_3, t_4, (occurs (do Jack6 (go supermarket4))))))

- Fact 5:
  TRUE(t_1, t_2, (to-do (achieve TRUE(t_3, t_4, (occurs (do Jack6 (go supermarket4)))))
  ((do Wilma22 (lend Jack6 car8)) (do Jack6 (drive car8 supermarket4)))))

- Fact 6:
  $\forall$ person,
  TRUE(t_1, t_3, (know person ϕ)), where ϕ is any of the above givens.

Show:
TRUE(t_1, t_2, (grateful Jack6 Wilma22 (do Wilma22 (lend Jack6 car8) t_1 t_2))
Proof:

- Steps 1-3: Jack wants Wilma to lend him the car and believes that she has done so.

  1. TRUE(t_1, t_2, (occurs (do Wilma22 (lend Jack6 car8))))
  – Given.

  2 . TRUE(t_1, t_2, (believe Jack6 TRUE(t_1, t_2, (occurs (do Wilma22 (lend Jack6 car8))))
  – Fact 6; Fact 2; axiom 5.

  3. TRUE(t_1, t_2, (want Jack6 TRUE(t_1, t_2, (occurs (do Wilma22 (lend Jack6 car8))))))
  – Fact 3; axiom 11.

  Note: For this reason, it follows that Jack will *not* be angry that Wilma lent him her car: it is required as part of the definition of anger that the action be one that you did not want to occur.

- Steps 4-8: Jack believes that the loan will benefit him.

  4. TRUE(t_1, t_2, (want Jack6 TRUE(t_3, t_4, (occurs (do Jack6 (go supermarket4))))))
  – Fact 4; axiom 11.

  5. TRUE(t_1, t_2, (to-do (achieve TRUE(t_3, t_4, (occurs (do Jack6 (go supermarket4)))))
  ((do Wilma22 (lend Jack6 car8)) (do Jack6 (drive car8 supermarket4)))))
  – Given.

  6. TRUE(t_1, t_2, (plan-step (do Wilma22 (lend Jack6 car8)) ((do Wilma22 (lend Jack6 car8)) (do Jack6 (drive car8 supermarket4)))))
  – Axiom 45; universal instantiation; Step 5; modus ponens.

  7. TRUE(t_1, t_2, (benefit Jack6 (do Wilma22 (lend Jack6 car8))))
  – Axiom 50; universal instantiation; Steps 4-6; modus ponens.

  8. TRUE(t_1, t_2, (believe Jack6 TRUE(t_1, t_2, (benefit Jack6 (do Wilma22 (lend Jack6 car8))))
  – Lemmas 4 and 5; Fact 6; axiom 5; consequential closure.

$\exists$ loc,
[TRUE(t_1, t_3, (in pt loc))
$\wedge$
TRUE(t_1, t_3, $\neg$(possess supermarket22 loc))]]

- Fact 8:
  TRUE(t_1, t_3, (know Betty8 ϕ)), where ϕ is any of the above givens.

Show:
TRUE(t_1, t_3, (disapprove Betty8 Jack6 (do Jack6 (take candy-bar6 supermarket22)) t_1 t_3))

Proof:
1. TRUE(t_1, t_3, (prohibited Ethics (do Jack6 (take candy-bar6 supermarket22))))
– Lemma 2; universal instantiation.
2. TRUE(t_1, t_3, (merchandise-of supermarket22 candy-bar6))
– Given.
3. TRUE(t_1, t_3, (owns supermarket22 candy-bar6))
– Axiom 40; universal instantiation; Step 2; modus ponens.
4. TRUE(t_1, t_3, (occurs (do Jack6 (move candy-bar6 ptA ptB))))
– Given.
5. TRUE(t_1, t_3, $\neg$(in ptB buildingX))
– Given.
6. $\exists$ loc,
[TRUE(t_1, t_3, (in ptB loc))
$\wedge$
TRUE(t_1, t_3, $\neg$(possess supermarket22 loc))]]
– Facts 6-7; universal instantiation; modus ponens.
7. TRUE($t_1, t_3, \neg$(possess supermarket22 ptB))
– Axiom 21; universal instantiation; Step 6; modus ponens.
8. TRUE(t_1, t_3, (occurs (do Jack6 (take candy-bar6 supermarket22))))
– Steps 3, 4, 7; axiom 19; universal instantiation; modus ponens.
9. TRUE(t_1, t_3, (believe Betty8 TRUE(t_1, t_3, (prohibited Ethics (do Jack6 (take candy-bar6 supermarket22))))))
– Lemmas 4 and 5; step 1; Fact 8; axiom 5; consequential closure.
10. TRUE(t_1, t_3, (believe Betty8 TRUE(t_1, t_3, (occurs (do Jack6 (take candy-bar6 supermarket22))))
– Lemmas 4 and 5; steps 2-7; Fact 8; axiom 5; consequential closure.
11. TRUE(t_1, t_3, (disapprove Betty8 Jack6 (do Jack6 (take candy-bar6 supermarket22)) $t_1 t_3$))

2. TRUE(t_1, t_2, (obligated Ethics (do Jack6 (give charity22 food))))
– Step 1; axiom 16; universal instantiation; modus ponens.

3. TRUE(t_1, t_2, (believe Barney3 TRUE(t_1, t_2, (obligated Ethics (do Jack6 (give charity22 food))))
– Lemmas 4 and 5; steps 1-2; Fact 4; axiom 5; consequential closure.

4. TRUE(t_1, t_2, (occurs (do Jack6 (give charity22 food))))
– Given.

5. TRUE(t_1, t_2, (believe Barney3 TRUE(t_1, t_2, (occurs (do Jack6 (give charity22 food))))))
– Lemmas 4 and 5; step 4; Fact 4; axiom 5; consequential closure.

6. TRUE(t_1, t_2, (approve Barney3 Jack6 (do Jack6 (give charity22 food)) t_1 t_2))
– Steps 3, 5; definition of approval; universal instantiation; modus ponens. *Q.E.D.*

3.6 Proof 4

Given:

- Fact 1:
  $t_1 \leq t_2 \leq t_3$.
- Fact 2:
  TRUE(t_1, t_3, (merchandise-of supermarket22 candy-bar6))
- Fact 3:
  TRUE(t_1, t_3, (possess supermarket22 buildingX))
- Fact 4:
  TRUE(t_1, t_3, (occurs (do Jack6 (move candy-bar6 ptA ptB))))
- Fact 5:
  TRUE(t_1, t_3, (in ptA buildingX))
- Fact 6:
  TRUE(t_1, t_3, $\neg$(in ptB buildingX))
- Fact 7:
  $\forall$ pt,
  [TRUE(t_1, t_3, $\neg$(in pt buildingX))
  $\rightarrow$

– definition of disapproval; universal instantiation; steps 9-10; modus ponens. *Q.E.D.*

3.7 Proof 5

Given:

- Fact 1:
  $t_1 \leq t_2 \leq t_3$.
- Fact 2:
  TRUE(t_1, t_3, (merchandise-of supermarket22 candy-bar6))
- Fact 3:
  TRUE(t_1, t_3, (possess supermarket22 buildingX))
- Fact 4:
  TRUE(t_1, t_3, (occurs (do Jack6 (move candy-bar6 ptA ptB))))
- Fact 5:
  TRUE(t_1, t_3, (in ptA buildingX))
- Fact 6:
  TRUE(t_1, t_3, $\neg$(in ptB buildingX))
- Fact 7:
  $\forall$ pt,
  [TRUE(t_1, t_3, $\neg$(in pt buildingX))
  $\longrightarrow$
  $\exists$ loc,
  [TRUE(t_1, t_3, (in pt loc))
  $\wedge$
  TRUE(t_1, t_3, $\neg$(possess supermarket22 loc))]]
- Fact 8:
  $\forall$ a, t_4, t_5,
  TRUE(t_1, t_3, (want Jack6 TRUE(t_1, t_3, $\neg$(disapprove Betty8 (do Jack6 a) t_4 t_5)))
- Fact 9:
  TRUE(t_1, t_3, (know Betty8 ϕ)), where ϕ is any of the above givens.

- Fact 10:
  TRUE(t_1, t_3, (know Jack6 ϕ)), where ϕ is any of the above givens.

- Fact 11:
  $\forall\, t_4, t_5$,
  TRUE(t_4, t_5, Betty8 $\neq$ Jack6).

Show:
TRUE(t_1, t_3, (ashamed Jack6 (do Jack6 (take candy-bar6 supermarket22)) t_1 t_3))

Proof:

1. TRUE(t_1, t_3, (occurs (do Jack6 (take candy-bar6 supermarket22))))
– See Proof 4.

2. TRUE(t_1, t_3, (disapprove Betty8 (do Jack6 (take candy-bar6 supermarket22)) t_1 t_3))
– See Proof 4.

3. TRUE(t_1, t_3, (believe Jack6 TRUE(t_1, t_3, (occurs (do Jack6 (take candy-bar6 supermarket22))))
– Lemmas 4 and 5; step 1; Fact 10; axiom 5; consequential closure.

4. TRUE(t_1, t_3, (believe Jack6 TRUE(t_1, t_3, (disapprove Betty8 (do Jack6 (take candy-bar6 supermarket22)) t_1 t_3))
– Lemmas 4 and 5; step 2; Fact 10; axiom 5; consequential closure.

5. TRUE(t_1, t_3, (want Jack6 TRUE(t_1, t_3, $\neg$(disapprove Betty8 (do Jack6 (take candy-bar6 supermarket22)) t_1 t_3))
– Fact 8; universal instantiation.

6. $\exists y \neq Jack6$,
[TRUE(t_1, t_3, (believe Jack6 TRUE(t_1, t_3, (disapprove y Jack6 (do Jack6 (take candy-bar6 supermarket22)) t_1 t_3))))
$\wedge$
TRUE(t_1, t_3, (want Jack6 TRUE(t_1, t_3, $\neg$(disapprove y (do Jack6 (take candy-bar6 supermarket22)) t_1 t_3))))]
– Fact 11; Steps 4-5; existential abstraction.

7. TRUE(t_1, t_3, (ashamed Jack6 (do Jack6 (take candy-bar6 supermarket22)) t_1 t_3))
– definition of shame; universal instantiation; steps 3, 6; modus ponens.

3.8 Proof 6

Given:

- Fact 1:
  $t_1 \leq t_2 \leq t_3$.

- Fact 2:
  TRUE(t_1, t_3, (occurs (do Jack6 (pay (cashier-of supermarket7) initial-payment8))))

- Fact 3:
  $\forall$ number > initial-payment8,
  $\neg$TRUE(t_1, t_3, (occurs (do Jack6 (pay (cashier-of supermarket7) number))))

- Fact 4:
  TRUE(t_1, t_3, (occurs (do (cashier-of supermarket7) (pay Jack6 change2))))

- Fact 5:
  $\forall$ number > change2
  $\neg$TRUE(t_1, t_3, (occurs (do (cashier-of supermarket7) (pay Jack6 number))))

- Fact 6:
  TRUE(t_1, t_3, (> (price-of object3) (− initial-payment8 change2)))

- Fact 7:
  TRUE(t_1, t_2, (merchandise-of supermarket7 object3))

- Fact 8:
  TRUE(t_2, t_2, (occurs (do supermarket7 (transfer-ownership Jack6 object3))))

- Fact 9:
  TRUE(t_1, t_3, (> initial-payment8 0)).

- Fact 10:
  TRUE(t_1, t_3, (know Jack6 ϕ)), where ϕ is any of the above givens.

Show:

TRUE(t_2, t_3, (guilty Jack6 $\neg$(net-pay supermarket7 object3) t_1 t_3))

Proof:

- Steps 1-7: Jack did not pay the full price of his purchases.

  1. TRUE(t_1, t_3, (occurs (do Jack6 (pay (cashier-of supermarket7) initial-payment8))))
  – Given.

  2. $\forall$ number $>$ initial-payment8,
  $\neg$TRUE(t_1, t_3, (occurs (do Jack6 (pay (cashier-of supermarket7) number))))
  – Given.

  3. TRUE(t_1, t_3, (occurs (do (cashier-of supermarket7) (pay Jack6 change2))))
  – Given.

  4. $\forall$ number $>$ change2
  $\neg$TRUE(t_1, t_3, (occurs (do (cashier-of supermarket7) (pay Jack6 number))))
  – Given.

  5. TRUE(t_1, t_3 (occurs (do Jack6 (net-pay supermarket7 ($-$ initial-payment8 change2)))))
  – Steps 1-4 and Lemma 1.

  6. TRUE(t_1, t_3, ($>$ (price-of object3) ($-$ initial-payment8 change2)))
  – Given.

  7. TRUE(t_1, t_3 (occurs $\neg$(do Jack6 (net-pay supermarket7 (price-of object3)))))
  – Axiom 29; Steps 5-6; universal instantiation; modus ponens.

- Steps 8-16: Jack should have paid for his purchases.

  8. TRUE(t_1, t_2 (offer supermarket7 (do supermarket7 (sell object3 Jack6))))
  – Axiom 40; universal instantiation; Fact 7; modus ponens.

  9. TRUE(t_1, t_3, (occurs (do Jack6 (pay (cashier-of supermarket7) initial-payment8))))
  – Given.

  10. TRUE(t_1, t_3, (occurs (do Jack6 (pay supermarket7) initial-payment8)))
  – Axiom 38; Step 9; modus ponens.

  11. TRUE(t_1, t_3, ($>$ initial-payment8 0)).
  – Given.

12. [TRUE(t_1, t_3 (offer Jack6 (do Jack6 (net-pay supermarket7 (price-of object3)))))]
– Axiom 44; universal instantiation; Steps 10-11; modus ponens.

13. [TRUE(t_1, t_3 (accept Jack6 (do supermarket7 (sell object3 Jack6))))]
– Axiom 43; universal instantiation; Step 12; modus ponens.

14. TRUE(t_1, t_3, (agree supermarket7 Jack6
(do supermarket7 (transfer-ownership Jack6 object3))
(do Jack6 (net-pay supermarket7 (price-of object3)))))
– Axiom 42; universal instantiation; Steps 8 and 13; modus ponens.

15. TRUE(t_2, t_2, (occurs (do supermarket7 (transfer-ownership Jack6 object3))))
– Given.

16. TRUE(t_2, t_3 (obligated Ethics (do Jack6 (net-pay supermarket7 (price-of object3)))))
– Axiom 13; universal instantiation; steps 7, 14-15, and modus ponens.

- Steps 17-18: Jack believes that he didn't pay and that he should have paid.

  17. TRUE(t_2, t_3, (believe Jack6
  TRUE(t_1, t_3 (occurs (do Jack6 $\neg$(net-pay supermarket7 (price-of object3)))))))
  – Lemmas 4 and 5; step 7; Fact 10; axiom 5; consequential closure; axiom 6.

  18.TRUE(t_2, t_3 (believe Jack6
  TRUE(t_2, t_3 (obligated Ethics (do Jack6 (net-pay supermarket7 (price-of object3)))))))
  – Lemmas 4 and 5; step 16; Fact 10; axiom 5; consequential closure; axiom 6.

- Steps 19-20: Jack believes that he knew at the time that it was wrong to keep the extra change.

  19. TRUE(t_2, t_3 (know Jack6
  TRUE(t_2, t_3 (obligated Ethics (do Jack6 (net-pay supermarket7 (price-of

object3)))))))
– Step 16; Fact 10; axioms 53 and 54; consequential closure.

20. TRUE(t_2, t_3 (believe Jack6
TRUE(t_2, t_3, (know Jack6
TRUE(t_2, t_3 (obligated Ethics (do Jack6 (net-pay supermarket7 (price-of object3)))))))))
– Lemmas 4 and 5; step 19; Fact 10; axiom 5; consequential closure.

- Step 21: As a result, Jack feels guilty.

  21. TRUE(t_2, t_3, (guilty Jack6 (do Jack6 $\neg$(net-pay supermarket7 (price-of object3))) t_2 t_3))
  – Definition of guilt; universal instantiation; steps 17-18,20; modus ponens. *Q.E.D.*

4 Previous Work

Little previous work has been done in this area. Kube encodes a portion of commonsense psychology in [10]. That paper is restricted to the ways in which people acquire knowledge and beliefs, however, and explicitly excludes any consideration of agents' emotions or intent.

Dyer attempted to incorporate psychological knowledge in his program BORIS [5]. BORIS's approach is substantially less formal than ours. It processes one basic story using some fairly simple definitions for the emotions involved. According to these definitions, for example, you are happy if you achieve a goal; if one of your goals fails or is suspended, you will be unhappy; and if someone else causes this to happen, you will become angry. In general, these goal-oriented definitions are too broad. For example, if you go to the bank and there is a long line at the teller machine, the people in front of you are causing one of your goals to fail – the goal of obtaining cash quickly. BORIS's definition would sanction the inference that you are angry at all of those people. Our definition would allow this inference only if you believe that their actions are wrong. You might become angry at someone who cut in front of you in line, but not at someone who merely arrived a few minutes before you.

There is a substantial psychological literature on the emotions [1, 7, 14, 15, 17, 18]. In general, however, this research disregards the kind of commonsense knowledge we are trying to encode. Instead, it focuses on such issues as the possible neurological causes of emotion. Rather than using a commonsense theory as a basis, even when addressing some of the same issues, this work usually takes an independent approach [7].

For our purposes, the most interesting of the psychological theories of emotion is the recent work by Ortony, Clore, and Collins [14]. They attempt to characterize the range of possible emotions, rather than the emotional reactions which would be likely to occur in a given culture. They provide a broad framework which is generally consistent with our theory. Like our theory, theirs assumes that emotions are largely caused by people's beliefs about the world. They divide these beliefs into three categories – beliefs about actions, events, and objects – and divide emotions into three basic types, according to the kind of beliefs by which they are triggered. Thus, for example, joy is a positive response to an event, approval is a positive response to an action, and liking is a positive response to an object. All of the emotions considered in this paper would be classified as responses to actions; however, our theory could be extended to handle the other areas as well.

Unlike our theory, Ortony, Clore, and Collins's work is very informal. Their terminology has no precise semantics. The reader must rely on intuition for the meaning of terms such as "joy-intensity" and "fear-potential." Our theory introduces very few new primitives; most predicates are defined in terms of a small group of well-studied logical concepts. O'Rorke and Cain have begun working on a story-comprehension system based on Ortony, Clore, and Collins's theory [2]. Because of the breadth of the underlying theory, their system promises to be more general than BORIS; however, like BORIS, this work should be considerably less formal than ours.

5 Conclusions and Future Work

In this paper, we define a logic in which commonsense knowledge about human psychology can be expressed. We demonstrate, by formulating and solving a set of benchmark problems, that our logic is sufficiently expressive to handle a variety of information about human actions and responses, in a way that is substantially more formal than previous work in the area.

Obvious extensions to this work include defining further emotions, such as hope, fear, surprise, and impatience, along with their causes and results. Because our theory is compositional, some extensions fall easily out of the definitions. For example, we could define an emotion that might be labelled 'remorse' – the response when you realize that you have done something wrong, although you didn't know it was wrong at the time – by modifying the definition of guilt only slightly. Other extensions could be obtained by incorporating additional predicates in the theory, perhaps following the directions suggested by [14].

In addition, we would like to integrate this work with a theory of causation. Note that our rules purport to define the 'causes' of states such as anger, gratitude, shame, and guilt. For the purposes of this paper, we have treated the causal inferences as though they were equivalent to logical inference. Technically, however, causation is both stronger and weaker than logical inference (see discussion in [16], pp. 166ff). Ideally, a theory of emotion should incorporate a more precise understanding of causation.

Another area for future work is the semantics of 'want.' As we have defined 'want' here, there are only three possibilities: x either wants a proposition to be true, wants it to be false, or is indifferent. Unfortunately, this semantics does not capture the idea of preference. Traditional possible-world semantics does not allow us to express degrees of want, belief, or knowledge. In the case of knowledge, this is intuitively reasonable, and even in the case of belief, the loss due to this abstraction is often acceptable. In the case of want, however, we would like to be able to reason about the ways in which alternative desires and goals are weighed against each other. This might be achieved by defining a preference criterion of the kind suggested in [16].

Finally, we would like to explore an issue that is implicit in the definitions of approval and disapproval. Note that these definitions imply a symmetry between approval and disapproval that does not in fact exist. You might disapprove of someone for committing murder (say), but you do not constantly approve of everyone who is refraining from murder. Similarly, you might approve of someone who makes a large donation to charity, while not disapproving of those who do not.

This asymmetry results from a fact which holds for the other emotions as well: typically, we only react to things which we do not take for granted. You do not approve of everyone who fails to commit murder, unless the temptation is particularly severe, because you take that for granted. Similarly, children are not always grateful for the food, clothing, and shelter provided by their parents, because they take these things for granted. We only react to actions that we notice, for one reason or another. Possibly this could be tied in with notions of awareness and implicit/explicit knowledge from epistemic logic. [6] Developing a general theory of what kinds of things we take for granted is outside the scope of this paper, but remains an interesting area for future work.

References

[1] Averill, James R. *Anger and aggression: an essay on emotion.* NY: Springer-Verlag (1982).

[2] Cain, Timothy, and Paul O'Rorke. Explanations involving emotions. *Proceedings of the AAAI-88 Workshop on Plan Recognition* (1988).

[3] Davis, Ernest. Inferring Ignorance from the Locality of Visual Perception. *AAAI-88* (1988), 786-790.

[4] Davis, Ernest. *Representations of Knowledge for Commonsense Reasoning.* (draft, May 6, 1987).

[5] Dyer, Michael G. *In-Depth Understanding.* Cambridge, MA: MIT Press (1983).

[6] Fagin, Ronald and Joseph Y. Halpern. Belief, Awareness, and Limited Reasoning: Preliminary Report. *IJCAI-85* (1985), 491-501.

[7] Harre, Rom, David Clarke, and Nicola DeCarlo. *Motives and Mechanisms: an introduction to the psychology of action.* London: Methuen (1985).

[8] Hayes, Patrick. Naive Physics I: ontology for liquids. J. Hobbs and R. Moore, editors, *Formal Theories of the Commonsense World,* Ablex 1985.

[9] Hughes, G.E. and M.J. Cresswell. *An Introduction to Modal Logic.* London: Methuen (1968).

[10] Kube, Paul. Cognitive propositional attitudes. In *Commonsense Summer: Final Report.* Stanford University: Center for the Study of Language and Information, Report No. CSLI-85-35 (October 1985).

[11] McCarty, L. Thorne. Permissions and Obligations. *IJCAI-83.*

[12] McDermott, Drew. A Temporal Logic for Reasoning About Processes and Plans. *Cognitive Science 6:* 101-155 (1982).

[13] Mendelson, Elliott. *Introduction to Mathematical Logic.* NY: Van Nostrand (2d ed. 1979).

[14] Ortony, Andrew, Gerald L. Clore, and Allan Collins. *The Cognitive Structure of Emotions.* Cambridge: Cambridge University Press (1988).

[15] Plutchik, Robert and Henry Kellerman, edd. *Emotion: theory, research, and experience.* NY: Academic Press (1980).

[16] Shoham, Yoav. *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence.* Cambridge, MA: MIT Press (1988).

[17] Strongman, K. T. *The Psychology of Emotion.* Chichester, U.K.: John Wiley & Sons (3d ed. 1987).

[18] Tavris, Carol. *Anger: the misunderstood emotion.* NY: Simon & Schuster (1982).

[19] Wright, G. H. von. *An Essay in Modal Logic.* North-Holland (1951).

- Steps 9-11: Jack believes that the loan was made unconditionally.

  9. $\neg\exists z, b,$
  [TRUE(t_1, t_2 (agree Wilma22 z (do Wilma22 (transfer-possession Jack6 car8)) (do z b)))]
  – Axiom 37; universal instantiation; Step 1; modus ponens.

  10. $\neg$TRUE(t_1, t_2 (conditional (do Wilma22 (lend Jack6 car8))))
  – Axiom 51; universal instantiation; Step 9; modus ponens.

  11. TRUE(t_1, t_2 (believe Jack6 $\neg$TRUE(t_1, t_2 (conditional (do Wilma22 (lend Jack6 car8))))))
  – Lemmas 4 and 5; Fact 6; axiom 5; consequential closure.

- Step 12: And as a result, Jack is grateful to Wilma for lending him the car.

  12. TRUE(t_1, t_2, (grateful Jack6 Wilma22 (do Wilma22 (lend Jack6 car8) t_1 t_2))
  – Definition of gratitude; universal instantiation; Steps 2, 3, 8, 11; modus ponens.

3.5 Proof 3

Given:

- Fact 1:
  $t_1 \leq t_2$.

- Fact 2:
  TRUE(t_1, t_2, (need charity22 food))

- Fact 3:
  TRUE(t_1, t_2, (occurs (do Jack6 (give charity22 food))))

- Fact 4:
  TRUE(t_1, t_2, (know Barney3 ϕ)), where ϕ is any of the above givens.

Show:
TRUE(t_1, t_2, (approve Barney3 Jack6 (do Jack6 (give charity22 food)) t_1 t_2))
Proof:

1. TRUE(t_1, t_2, (need charity22 food))
– Given.