Theoretical Background for the Project

Epistemic Logic in the movie *Fight Club*

Huaitian Lu (S3339246), Bettina Soós (S3421554), Xingchi Su (P286498)

1 First-order Modal Logic and Formalization

A previous related work on our topic is the term-modal logic where terms are introduced into the language to divide agents and the names of agents. In philosophical sense, term-modal logic can express a non-rigid reference of the names of agents, which implies that over different possible worlds, different agents could have different names and one agent could even have one more names.

Unfortunately, term-modal logic cannot be applied in our case of *Fight Club*. The models that term-modal logic concerns about are basically first-order Kripke models where the accessibility relations are defined with agents, instead of the names of agents. A philosophical commitment should be noticed here where one agent only has one set of epistemic information, which implies that on each possible world, the undistinguished worlds for one agent are fixed. However, in *Fight Club*, Jack and Tyler have different information on the events happened in the story which means they are different agents epistemically, although they are the same person physically.

In order to deal with this conflict, we commit that different agents are the agents have different epistemic information, instead of different agents in physical sense. By this commitment, Jack and Tyler are no longer the same agent and the terms (or names) are not needed in our language of the logic any more. Moreover, Jack and Tyler are two independent agents in the domain of our first-order Kripke models. In term-modal logic, the identity of two terms is defined as follows:

$$M, w, \sigma \models t \approx t' \;\Leftrightarrow\; \sigma_w(t) = \sigma_w(t')$$

Intuitively, the above truth condition defines the identity of two terms t and t' by checking whether the agent that t refers to is the same as the agent that t' refers to. But we aim to formalize the process of Jack's awareness that Tyler is just himself. If we define the identity with the identity of agents, Jack and Tyler are not identified permanently due to the commitment that Jack and Tyler are two different agents. Thus, a new truth condition of identity should be given in our case:

$$M, w, \sigma \models x \approx y \;\Leftrightarrow \text{for any predicate } P,\ [\sigma_w(x) \in \rho(P, w) \Leftrightarrow \sigma_w(y) \in \rho(P, w)]$$

This truth condition is inspired by the Leibniz Principle[1]. We do not need to check whether they refer to the same agent. Instead, we check whether two agents agree on all the predicates.

[1] $\forall x \forall y \forall P(x = y \leftrightarrow (Px \leftrightarrow Py))$

Alongside the story line, Jack would eliminate the worlds where he and Tyler obey distinguished predicates step by step. Consequently, the remaining possible worlds in the end are the worlds where Jack and Tyler share the absolutely same predicates. The logic ELA (Epistemic Logic with Assignment) will be given in following subsections.

1.1 The Language of ELA

Definition 1. *Given a denumerable set of variables $\boldsymbol{X}$, and a denumerable set of unary predicate symbols, the language ELA is defined as:*

$$\phi ::= (x \approx y) \mid Px \mid \neg\phi \mid (\phi \wedge \phi) \mid K_x\phi$$

where $x \in \boldsymbol{X}$ and $P \in \boldsymbol{P}$.

1.2 Semantics

We define the semantics of ELA over first-order Kripke models.

Definition 2. *A first-order Kripke model M for ELA is a tuple $\langle W, A, R, \rho\rangle$ where:*

- *W is a non-empty set of possible worlds.*
- *A is a finite set of agents.*
- *$R : A \to 2^{W\times W}$ assign a binary relation $R(i)$ (also written R_i) between worlds, to each agent $i \in A$.*
- *$\rho : P \times W \to 2^A$ assigns an unary relation $\rho(P, w)$ between agents to each unary predicate P at each possible worlds w.*

It should be noticed that the interpretation of predicates varies over different possible worlds, which corresponds to our case where Jack thinks Tyler is a different person who doubtlessly has distinct predicates from him.

To interpret free variables, we need a variable assignment like first-order logic: $\sigma : \mathbf{X} \to A$, which means we interpret every variable x into an agent. Following σ, the truth conditions of ELA can be given as follows:

- $M, w, \sigma \models x \approx y \;\Leftrightarrow\;$ *for any predicate* P, $[\sigma_w(x) \in \rho(P, w) \Leftrightarrow \sigma_w(y) \in \rho(P, w)]$
- $M, w, \sigma \models Px \;\Leftrightarrow\; \sigma_w(x) \in \rho(P, w)$
- $M, w, \sigma\neg\phi \;\Leftrightarrow\; M, w, \sigma \not\models \phi$
- $M, w, \sigma \models (\phi \wedge \psi) \;\Leftrightarrow\; M, w, \sigma \models \phi$ *and* $M, w, \sigma \models \psi$
- $M, w, \sigma \models K_x\phi \Leftrightarrow\; M, w, \sigma \models \phi$ *for all* v *s.t.* $wR_{\sigma_w(x)}v$
- $M, w, \sigma \models [x := y]\phi \;\Leftrightarrow\; M, w, \sigma[x \mapsto \sigma_w(y)] \models \phi$