A conflict tolerant logic of explicit evidence

Thomas Studer
University of Bern,
Institute of Computer Science,
Neubrückstrasse 10, 3012 Bern, Switzerland.
E-mail: thomas.studer@inf.unibe.ch

Abstract: Standard epistemic modal logic is unable to adequately deal with the Frauchiger–Renner paradox in quantum physics. We introduce a novel justification logic CTJ, in which the paradox can be formalized without leading to an inconsistency. Still CTJ is strong enough to model traditional epistemic reasoning. Our logic tolerates two different pieces of evidence such that one piece justifies a proposition and the other piece justifies the negation of that proposition. However, our logic disallows one piece of evidence to justify both a proposition and its negation. We present syntax and semantics for CTJ and discuss its basic properties. Then we give an example of epistemic reasoning in CTJ that illustrates how the different principles of CTJ interact. We continue with the formalization of the Frauchiger–Renner thought experiment and discuss it in detail. Further, we add a trust axiom to CTJ and again discuss epistemic reasoning and the paradox in this extended setting.

Keywords: Conflicting evidence, justification logic, epistemic logic, Frauchiger–Renner paradox, quantum physics

1. Introduction

Nurgalieva and del Rio [Nurgalieva, del Rio, 2019] challenge the logic community to find a sound logical system to analyze agents' reasoning when quantum measurements are involved. They show that standard epistemic modal logic is inadequate in quantum settings. In particular, they investigate the Frauchiger–Renner paradox [Frauchiger, Renner, 2018] and establish that modal logics are unable to deal properly with this paradox. Moreover, they show that candidate workarounds like keeping track of the context of each statement, are unpractical, requiring exponentially large memories.

In this paper, we present an epistemic logic that can

1. adequately deal with the Frauchiger-Renner paradox, in particular without resulting in an inconsistency, and also

2. model classical epistemic reasoning.

In order to achieve this, we develop a novel justification logic, CTJ. Justification logic is a variant of modal logic where the $\Box$-modality is replaced with explicit evidence terms [Artemov, Fitting, 2019 ; Kuznets, Studer, 2019]. Thus, instead of of formulas $\Box_a A$ meaning *agent a believes A*, justification logic features formulas $[s]_a A$ meaning *agent a believes A for reason s*. In the first justification logic, the Logic of Proofs, the evidence terms represented formal proofs in say Peano arithmetic [Artemov, 2001; Kuznets, Studer, 2016]. Later, epistemic semantics for justification logic have been developed where evidence terms can represent general justifications for an agent's belief like direct observation or communication with other agents [Artemov, 2006; Artemov, 2008; Bucheli et al., 2011; Bucheli et al., 2014; Fitting, 2005; Kuznets, Studer, 2012; Studer, 2013].

The defining principle of our logic CTJ is the following: given some evidence s

it is not possible that

s justifies some proposition P and s justifies the negation of P.

We call this a principle of *no conflicts* as any given evidence cannot justifiy conflicting propositions.

However, our logic CTJ also is *conflict tolerant* in the sense that there may be two different pieces of evidence, s and t, such that s justifies P and t justifies $\neg P$. That is CTJ tolerates two contradicting pieces of evidence; but it disallows one piece of evidence to support two conflicting propositions. Thus CTJ maintains a fine balance between accepting all beliefs and banning all contradictory beliefs.

In the next section, we introduce an axiomatic system for CTJ and discuss its basic properties, in particular with respect to consistency statements of the form $[s]_a\bot$ and $[s]_a(A \wedge \neg A)$. Then we give semantics to CTJ using subset models. The main idea there is that evidence terms are interpreted as sets of possible worlds and a formula $[s]_a A$ is true if the interpretation of s is a subset of the truthset of A. Subset models have been introduced in [Lehmann, Studer, 2019] and turned out the be useful in many different contexts [Lehmann, Studer, 2020]. Section 4. deals with epistemic reasoning in CTJ. We present an example that shows how the principle of no conflicts interacts with positive introspection and also with the other axioms and rules of CTJ. Then we give a first formalization of the Frauchiger–Renner paradox in CTJ showing that it does not lead to an inconsistency. In Section 6. we discuss our formalization and compare it with formalizations in traditional modal logic. For Sections 7. and 8. we extend CTJ with a trust axiom and we discuss our epistemic reasoning example and the formalization of the Frauchiger–Renner paradox in this setting. The last section concludes the paper.

The justification logic principle of no conflicts has first been considered in a deontic setting [Faroldi et al., 2020] where it states that two obligations A and $\neg A$ cannot be mandatory for one and the same reason. In the present paper, we put this principle in the frame of epistemic justification logic.

The Frauchiger-Renner paradox has been presented as a no-go theorem stating that a particular situation is physically impossible. No-go theorems are known also in other areas where they also have been investigated by logical methods. In social choice theory, there is Arrow's theorem [Arrow, 1950] saying that no voting system is possible that meets certain fairness conditions. Arrow's theorem has been formalized in independence logic by Pacuit and Yang [Pacuit, Yang, 2016]. In data privacy, there are a no-go theorems stating that certain combinations of desirable privacy properites are impossible [Studer, Werner, 2014]. These results have been analyzed and generalized using modal logic [Studer, 2020].

2. Syntax

Justification terms are built from countably many constants c_i and variables x_i according to the following grammar:

$$t ::= c_i \mid x_i \mid t \cdot t \mid !t \ .$$

The set of constants is denoted by Cons. The set of terms is denoted by Tm. A term is called *ground* if it does not contain variables. We use a finite set of agents Ag.

Formulas are built from countably many atomic propositions P_i and the symbol $\bot$ according to the following grammar where $a \in \mathsf{Ag}$ and $t \in \mathsf{Tm}$:

$$F ::= P_i \mid \ \bot \ \mid F \to F \mid [t]_a F \ .$$

The set of atomic propositions is denoted by Prop and the set of all formulas is denoted by Fml. The other classical Boolean connectives $\neg, \top, \wedge, \vee, \leftrightarrow$ are defined as usual, in particular we have $\neg A := A \to \bot$ and $\top := \neg\bot$. Note that our language does not include the usual sum-operation of justification logic. This is on purpose, see Remark 1 later.

The axioms of conflict tolerant justification logic CTJ are the following:

cl all axioms of classical logic
j $[s]_a(A \to B) \to ([t]_a A \to [s \cdot t]_a B)$
noc $\neg([s]_a A \wedge [s]_a \neg A)$
j4 $[s]_a A \to [!s]_a [s]_a A$

Justification logics are parameterized by a so-called constant specification, which is a set of formulas

$$\mathsf{CS} \subseteq \{[c_n]_{a_n} \dots [c_1]_{a_1} A \mid \text{for } n \geq 1,\ c_i \in \mathsf{Cons},\ a_i \in \mathsf{Ag} \text{ and } A \text{ is an axiom of } \mathsf{CTJ}\}$$

that is downward closed, i.e. for $n > 1$

$$[c_n]_{a_n} \dots [c_1]_{a_1} A \in \mathsf{CS} \quad \text{implies} \quad [c_{n-1}]_{a_{n-1}} \dots [c_1]_{a_1} A \in \mathsf{CS}.$$

Our logic $\mathsf{CTJ}_{\mathsf{CS}}$ is now given by the axioms of CTJ and the rules *modus ponens*

$$\frac{A \qquad A \to B}{B}\ (\mathrm{MP})$$

and *axiom necessitation*

$$\frac{}{[c_n]_{a_n} \dots [c_1]_{a_1} A}\ (\mathrm{AN}) \quad \text{where } [c_n]_{a_n} \dots [c_1]_{a_1} A \in \mathsf{CS}\ .$$

We write $\Delta \vdash_{\mathsf{CS}} A$ to mean that a formula A is derivable in $\mathsf{CTJ}_{\mathsf{CS}}$ from a set of formulas Δ. As usual, we use Δ, A for $\Delta \cup \{A\}$.

Definition 1 (Axiomatically appropriate CS)**.** A constant specification CS is called *axiomatically appropriate* if for each axiom A and each sequence of agents $a_n, \dots, a_1$, there is a sequence of constants $c_n, \dots, c_1$ such that

$$[c_n]_{a_n} \dots [c_1]_{a_1} A \in \mathsf{CS}.$$

Axiomatically appropriate constant specifications are important as they provide a form of necessitation. For a proof of the following lemma see, e.g., [Artemov, 2001; Artemov, Fitting, 2019 ; Kuznets, Studer, 2019].

Lemma 1 (Internalization[1])**.** *Let CS be an axiomatically appropriate constant specification. For arbitrary formulas $A, B_1, \dots, B_n$, arbitrary terms $s_1, \dots, s_n$, and an arbitrary agent a, if*

$$B_1, \dots, B_n \vdash_{\mathsf{CS}} A,$$

then there is a term t such that

$$[s_1]_a B_1, \dots, [s_n]_a B_n \vdash_{\mathsf{CS}} [t]_a A.$$

[1] We follow the naming convention of [Kuznets, Studer, 2019]. Internalization means that a justification logic internalizes its own notion of derivation. In other places, e.g. [Artemov, Fitting, 2019], this result is called *lifting lemma*.

A consequence of internalization is that we can combine justifications in $\mathsf{CTJ}_{\mathsf{CS}}$ (for an axiomatically appropriate CS), i.e. we have the following lemma.

Lemma 2. *Let* CS *be an axiomatically appropriate constant specification. For all formulas A and B, there exists a term r such that for all terms s and t, the following is provable in* $\mathsf{CTJ}_{\mathsf{CS}}$

$$[s]_a A \wedge [t]_a B \to [r \cdot s \cdot t]_a (A \wedge B).$$

Proof. By internalization, there exists a term r with

$$\vdash_{\mathsf{CS}} [r]_a (A \to (B \to (A \wedge B))).$$

Thus from $[s]_a A$ and $[t]_a B$ and using axiom **j** and modus ponens we get

$$[r \cdot s \cdot t]_a (A \wedge B).$$

■

In $\mathsf{CTJ}_{\mathsf{CS}}$ we may have situations where there is a justification for A and (another) justification for $\neg A$, see Example 1 later. Neither does the logic $\mathsf{CTJ}_{\mathsf{CS}}$ exclude a justification for $\bot$; and there may be a justification for a formula of the form $A \wedge \neg A$. That means that the formulas

$$[s]_a \bot \quad \text{and} \quad [s]_a (A \wedge \neg A),$$

respectively, are satisfiable (if the constant specification is not too strong).

However, what is excluded in $\mathsf{CTJ}_{\mathsf{CS}}$ is the existence of a justification s such that s justifies A and s also justifies $\neg A$. That is the formula $[s]_a A \wedge [s]_a \neg A$ is not satisfiable as this directly contradicts axiom **noc**. Hence, in particular, there cannot be one justification for everything.

The situation is different when we consider schematic reasoning. We will give the definition of schematic constant specifications and then show that they are not compatible with conflicting evidence.

Definition 2 (Schematic constant specification)**.** A constant specification CS is called *schematic* if for each sequence of constants $c_n, \ldots, c_1$ and each sequence of agents $a_n, \ldots, a_1$, the set of axioms $\{A \mid [c_n]_{a_n} \ldots [c_1]_{a_1} A \in \mathsf{CS}\}$ consists of all instances of one or several (possibly zero) axioms schemes of CTJ.

Schematic constant specifications are often considered for justification logics as they support subsitutions in theorems.

Lemma 3. *Let CS be a schematic constant specification. Let σ be a substitution that in a given formula simultaneously replaces variables with terms and atomic propositions with formulas. We have*

$$\vdash_{\mathsf{CS}} A \quad \textit{implies} \quad \vdash_{\mathsf{CS}} A\sigma.$$

However, axiomatically appropriate and schematic constant specifications prohibit conflicting justifications. Therefore, we will not allow them for the rest of this paper.

Lemma 4. *Let CS be an axiomatically appropriate and schematic constant specification. It is inconsistent in $\mathsf{CTJ}_{\mathsf{CS}}$ to have a justification for A and (another) justification for $\neg A$. That is for all formulas $[s]_a A$ and $[t]_a \neg A$ we have*

$$\vdash_{\mathsf{CS}} ([s]_a A \land [t]_a \neg A) \to \bot.$$

Proof. Assume $[s]_a A$ and $[t]_a \neg A$. Using Lemma 2 we find a term r such that

$$[r \cdot s \cdot t]_a (A \land \neg A).$$

Since CS is axiomatically appropriate, we find by internalization a term k with

$$[k]_a (A \land \neg A \to P).$$

Since CS is schematic, we find by Lemma 3 that

$$[k]_a (A \land \neg A \to \neg P)$$

holds, too. Hence we find that both

$$[k \cdot (r \cdot s \cdot t)]_a P \quad \text{and} \quad [k \cdot (r \cdot s \cdot t)]_a \neg P,$$

which contradicts axiom **noc**. ■

Remark 1. Note that if our language included the usual sum-operation of justification logic, then

$$\vdash_{\mathsf{CS}} ([s]_a A \land [t]_a \neg A) \to \bot$$

would hold for arbitrary constant specifications. Indeed, the sum axiom is

$$[s]_a A \lor [t]_a A \to [s + t]_a A.$$

Hence, if this axiom is present, we immediately obtain that $[s]_a A \land [t]_a \neg A$ implies $[s + t]_a A \land [s + t]_a \neg A$, which contradicts axiom **noc**. One possibility to still include a sum-like principle could be to use an axiom like

$$([s]_a A \land \neg [t]_a \neg A) \to ([s + t]_a A \land [t + s]_a A).$$

3. Semantics

We base our semantics on subset models, which have recently been introduced in justification logic [Lehmann, Studer, 2019; Lehmann, Studer, 2020].

Definition 3 (Subset model)**.** Given some constant specification CS, then a CS-subset model $\mathcal{M} = (W, W_0, V, E)$ is defined by:

- W is a set of objects called worlds.
- $W_0 \subseteq W$ and $W_0 \neq \emptyset$.
- $V : W \times \mathsf{Fml} \to \{0, 1\}$ such that for all $\omega \in W_0$, $t \in \mathsf{Tm}$, $F, G \in \mathsf{Fml}$:
  - $V(\omega, \bot) = 0$;
  - $V(\omega, F \to G) = 1 \quad \text{iff} \quad V(\omega, F) = 0$ or $V(\omega, G) = 1$;
  - $V(\omega, [t]_a F) = 1 \quad \text{iff} \quad E_a(\omega, t) \subseteq \{v \in W \mid V(v, F) = 1\}$.
- $E : \mathsf{Ag} \to (W \times \mathsf{Tm} \to \mathcal{P}(W))$ that meets the following conditions where we write E_a for $E(a)$ and use the notation

  $$[A] := \{\omega \in W \mid V(\omega, A) = 1\}. \tag{1}$$

  For all $a \in \mathsf{Ag}$, all $\omega \in W_0$, and all $s, t \in \mathsf{Tm}$:

  - $E_a(\omega, s \cdot t) \subseteq \{v \in W \mid \forall F \in \mathsf{APP}_{a,\omega}(s, t)(v \in [F])\}$ where APP contains all formulas that can be justified by an application of s to t, see below;
  - $\exists v \in W_{\mathsf{nc}}$ with $v \in E_a(\omega, t)$ where

    $$W_{\mathsf{nc}} := \{\omega \in W \mid \\ \text{for all formulas A } (V(\omega, A) = 0 \text{ or } V(\omega, \neg A) = 0)\};$$

  - $E_a(\omega, !t) \subseteq$

    $$\{\, v \in W \mid \forall F \in \mathsf{Fml}\,(V(\omega, [t]_a F) = 1 \Rightarrow V(v, [t]_a F) = 1)\,\};$$

  - for all $[c_n]_{a_n} \dots [c_1]_{a_1} A \in \mathsf{CS}$:

    $$E_{a_n}(\omega, c_n) \subseteq [\,[c_{n-1}]_{a_{n-1}} \dots [c_1]_{a_1} A\,].$$

The set APP is formally defined as follows:

$$\mathsf{APP}_{a,\omega}(s, t) := \{F \in \mathsf{Fml} \mid \exists H \in \mathsf{Fml} \text{ s.t.} \\ E_a(\omega, s) \subseteq [H \to F] \text{ and } E_a(\omega, t) \subseteq [H]\}.$$

W_0 is the set of *normal* worlds. The conditions on V for normal worlds tell us, in particular, that the laws of classical logic hold in normal worlds. The set $W \setminus W_0$ consists of the *non-normal* worlds. Moreover, using the notation introduced by (1), we can read the condition on V for justification formulas $[t]_a F$ as:

$$V(\omega, [t]_a F) = 1 \quad \text{iff} \quad E_a(\omega, t) \subseteq [F].$$

Since the valuation function V is defined on worlds and formulas, the definition of truth is standard.

Definition 4 (Truth)**.** Given a subset model

$$\mathcal{M} = (W, W_0, V, E)$$

and a world $\omega \in W$ and a formula F we define the relation $\Vdash$ as follows:

$$\mathcal{M}, \omega \Vdash F \quad \text{iff} \quad V(\omega, F) = 1.$$

Validity is defined with respect to the normal worlds.

Definition 5 (Validity)**.** Let CS be a constant specification. We say that a formula F is CS-*valid* if for each CS-subset model

$$\mathcal{M} = (W, W_0, V, E)$$

and each $\omega \in W_0$, we have $\mathcal{M}, \omega \Vdash F$.

As expected, we have soundness and completeness. A completeness proof is easily obtained by combining the completeness proofs of [Lehmann, Studer, 2019] and [Faroldi et al., 2020].

Theorem 1. *Let* CS *be an arbitrary constant specification. For each formula* F *we have that*

$$\mathsf{CTJ}_{\mathsf{CS}} \vdash F \quad \textit{iff} \quad F \textit{ is } \mathsf{CS}\textit{-valid.}$$

Let us now present a small but instructive example of our semantics.

Example 1. There is a subset model $\mathcal{M}$ with a normal world ω such that

$$\mathcal{M}, \omega \Vdash [s]_a P \quad \text{and} \quad \mathcal{M}, \omega \Vdash [t]_a \neg P.$$

for some terms s and t, some agent a, and some atomic proposition P.

Indeed, let $\mathcal{M}$ be given by

1. $W := \{\omega, \mu, \nu\}$;

2. $W_0 := \{\omega, \mu\}$;

3. $V(\omega, P) := 0$, $V(\mu, P) := 1$, $V(\nu, P) := 1$, $V(\nu, \neg P) := 1$;

4. $E_a(\omega, s) := \{\mu, \nu\}$, $E_a(\omega, t) := \{\omega, \nu\}$.

By the definition of V, we find

$$[P] = \{\mu, \nu\} \quad \text{and} \quad [\neg P] = \{\omega, \nu\}.$$

Hence

$$E_a(\omega, s) \subseteq [P] \quad \text{and} \quad E_a(\omega, t) \subseteq [\neg P]$$

and thus (since $\omega \in W_0$)

$$V(\omega, [s]_a P) = 1 \quad \text{and} \quad V(\omega, [t]_a \neg P) = 1$$

as desired.

Note that the model $\mathcal{M}$ can never satisfy an axiomatically appropriate and schematic constant specification (see Lemma 4). However, this example *at least* implies $\nvdash_{\emptyset} ([s]_a A \wedge [t]_a \neg A) \rightarrow \bot$.

Remark 2. The logic $\mathsf{CTJ}_{\mathsf{CS}}$ with an axiomatically appropriate and non-schematic constant specification CS is not an explicit counterpart of any modal logic. We can map formulas of justification logic to formulas of modal logic as follows. The forgetful projection of a formula A of Fml is the result of replacing all occurences of $[t]_a$ in A with $\Box_a$, i.e. we forget the explicit justification for agent a to believe a proposition and only represent that a believes the proposition.

By the previous example, it is consistent in $\mathsf{CTJ}_{\mathsf{CS}}$ to have $[s]_a P$ and $[t]_a \neg P$ for two different terms s and t. Thus in a corresponding modal logic we have $\Box_a P$ and $\Box_a \neg P$. This, however, contradicts the forgetful projection of axiom **noc**, which is $\neg(\Box_a P \wedge \Box_a \neg P)$.

We finish this section with a remark on the notion of negation in $\mathsf{CTJ}_{\mathsf{CS}}$.

Remark 3. The justifcation operators of $\mathsf{CTJ}_{\mathsf{CS}}$ provide hyperintensional contexts. That is they make it possible to distinguish between logically equivalent formulas, which is necessary for a tolerant treatment of conflicts. Thus the question arises which notion of negation do we get in these hyperintensional contexts provided by $\mathsf{CTJ}_{\mathsf{CS}}$.

Let CS be an axiomatically appropriate but non-schematic constant specification. Then $\mathsf{CTJ}_{\mathsf{CS}}$ internalizes the rules of contraposition and double negation. Formally we can prove in $\mathsf{CTJ}_{\mathsf{CS}}$

$$[x]_a(A \to B) \to [r_1]_a(\neg B \to \neg A)$$
$$[x]_a A \to [r_2]_a \neg\neg A$$
$$[x]_a \neg\neg A \to [r_3]_a A$$

for suitable terms r_1, r_2, and r_3.

However, ex contradictione rules cannot be internalized in $\mathsf{CTJ}_{\mathsf{CS}}$, i.e. it is in general not provable that

$$[x]_a(A \to B) \wedge [y]_a(A \to \neg B) \;\to\; [r_4]_a(A \to C)$$

for any term r_4.

4. Epistemic Reasoning

In this section, we discuss an epistemic situation that illustrates the use and interplay of axiom **noc**, positive introspection and an axiomatically appropriate constant specification. Note, in particular, how axiom **noc** is used to state that if one observes that a hat is not red, then the same observation cannot lead to the result that the hat is red.

Before we present our example for reasoning in $\mathsf{CTJ}_{\mathsf{CS}}$, let us talk about terminology. Often we will read a formula $[t]_aF$ as *agent a knows F for reason t* or *t justifies agent a's knowledge of F*. However, we should emphasize that $\mathsf{CTJ}_{\mathsf{CS}}$ does not include a factivity (or truth) axiom of the form $[t]_aF \to F$.[2] The reason that we still talk of knowledge is that we want to stay as close as possisble to the presentation of the Frauchiger–Renner paradox given in[Nurgalieva, del Rio, 2019]. A more appropriate reading of $[t]_aF$ in the context of $\mathsf{CTJ}_{\mathsf{CS}}$ would be *t justifies agent a's belief in F* or *agent a accepts t as evidence for F*.

Consider the following scenario where we work with an axiomatically appropriate constant specification. There are two agents, a and b. Agent a wears a hat, which may be red or not. We use the propositional atom red to state whether the hat is red. Assume further that agent a cannot see the color of the hat. But a red hat will attract b's attention and b will observe (and thus know) that the hat is red. Formally, we express this by

$$\mathsf{red} \to [\mathsf{obs}]_b\mathsf{red}$$

[2] The T in $\mathsf{CTJ}_{\mathsf{CS}}$ stands for *tolerant* and not (as usual) for *truth*.

where obs is a term representing b's observation. We also assume that agent a knows that a red hat will attract b's attention and hence there is a term s_1 with

$$[s_1]_a(\mathsf{red} \to [\mathsf{obs}]_b\mathsf{red}).$$

From this and the axiomatically appropriate constant specification, we can construct a term s_2 with

$$[s_2]_a(\neg[\mathsf{obs}]_b\mathsf{red} \to \neg\mathsf{red}). \tag{2}$$

Now suppose that the color of the hat was not red but yet agent b noticed it and observed that the hat is not red. Hence we have

$$[\mathsf{obs}]_b\neg\mathsf{red}.$$

Agent b knows this by positive introspection (axiom **j4**), i.e. we have

$$[!\mathsf{obs}]_b[\mathsf{obs}]_b\neg\mathsf{red}. \tag{3}$$

Using axiom **noc** we find

$$[\mathsf{obs}]_b\neg\mathsf{red} \to \neg[\mathsf{obs}]_b\mathsf{red}.$$

Since we work with an axiomatically appropriate constant specification, we can use Lemma 1 to find a term t with

$$[t]_b([\mathsf{obs}]_b\neg\mathsf{red} \to \neg[\mathsf{obs}]_b\mathsf{red}).$$

This, together with axiom **j** and (3), leads to

$$[t\cdot!\mathsf{obs}]_b\neg[\mathsf{obs}]_b\mathsf{red}. \tag{4}$$

That means that agent b knows that

$$\neg[\mathsf{obs}]_b\mathsf{red}, \tag{5}$$

i.e. agent b knows that it is not the case that agent b observed that the hat is red. Hence agent b can announce (5) to agent a. Then we get

$$[\mathsf{ann}]_a\neg[\mathsf{obs}]_b\mathsf{red}, \tag{6}$$

where ann is a term representing the announcement.

Now agent a knows that it is not the case that agent b observed that the hat is red. Combining this with (2) yields

$$[s_2 \cdot \mathsf{ann}]_a\neg\mathsf{red},$$

which means that after b's announcement, agent a knows that the hat is not red.

Remark 4. We use announcements in a very informal way and we did not include any principles formalizing announcements in $\mathsf{CTJ_{CS}}$. In the above example, we have an announcement in the step from (4) to (6). We assume this to work as follows. Agent b has evidence r for F, i.e. $[r]_b F$. Thus agent b can announce F to agent a. Then this announcement, represented by ann, is a's evidence for F, i.e. $[\mathsf{ann}]_a F$.

5. Frauchiger–Renner Experiment

The Frauchiger–Renner thought experiment [Frauchiger, Renner, 2018] is used to formulate a no-go theorem in quantum physics. We follow the presentation of the thought experiment that is given in [Nurgalieva, del Rio, 2019]. We omit all details and give only a very rough description of the experiment where we omit the actual quantum physical construction and focus on the epistemic logic view on the experiment. Thus we do not dicuss the quantum physical assumptions of the paradox. Neither do we discuss whether the thought experiment really is paradoxical, for more on this, see, e.g. [Lazarovici, Hubert, 2019].

The original no-go theorem claims that no physical theory can simultaneously satisfy the assumptions:

(Q) compatibility with the Born rule of quantum mechanics;

(C) logical consistency among agents;

(S) experimenters having the subjective experience of seeing only one outcome.

A fourth, implicitly used, assumption is discussed in [Nurgalieva, del Rio, 2019]:

(U) all agents are considering the evolution of the other agents in their labs unitary.

This means that agents, when reasoning about statements other agents make, do it according to a specific assumption of time evolution. Roughly, assumption **(U)** states that time evolution is such that the probability of the quantum system is conserved.

The setup of the experiment consists of four participants, Alice, Bob, Ursula, and Wigner, where each of them is equipped with a quantum memory (A, B, U, and W, respectively). The procedure of the experiment is as follows.

1. Alice measures a qubit R in a basis $\{|0\rangle_R, |1\rangle_R\}$. She records the outcome in her memory A and, depending on the outcome, prepares a qubit S in a certain way and sends it to Bob.

2. Bob measures S in a basis $\{|0\rangle_S, |1\rangle_S\}$ and records the outcome in his memory B.

3. Ursula measures Alice's lab (consisting of R and A) in a basis $\{|\mathsf{ok}\rangle_{RA}, |\mathsf{fail}\rangle_{RA}\}$.

4. Wigner measures Bob's lab (consisting of S and B) in a basis $\{|\mathsf{ok}\rangle_{SB}, |\mathsf{fail}\rangle_{SB}\}$.

5. Ursula and Wigner compare the outcomes of their measurements. If they are both ok, they halt the experiment, otherwise they reset the experiment and repeat it.

It can be shown that this experiment will halt at some point and we postselect on this event. The setup of the experiment (i.e. the initial qubit R, the construction of qubit S, and the bases in which the measurements are performed) is carefully chosen such that the following hold:

$$\text{If Ursula observes outcome } \mathsf{ok}\text{, then Bob obtained outcome 1.} \tag{7}$$

$$\text{If Bob observes outcome 1, then Alice obtained outcome 1.} \tag{8}$$

$$\text{If Alice observes outcome 1, then Wigner will obtain outcome } \mathsf{fail}. \tag{9}$$

Since the setup of the experiment is common knowledge, Wigner knows the above implications. Hence by simple logical reasoning, Wigner knows that

$$\text{if Ursula observes outcome } \mathsf{ok}\text{, then Wigner will obtain outcome } \mathsf{fail}. \tag{10}$$

Since we consider the event when the experiment halts, Wigner knows that

$$\text{Wigner observes outcome } \mathsf{ok} \tag{11}$$

and

$$\text{Ursula observes outcome } \mathsf{ok}. \tag{12}$$

Taking (12) and (10) together, we obtain that Wigner knows that

$$\text{Wigner observes outcome } \mathsf{fail},$$

which contradicts (11) if agents experience only a single outcome of measurements.

Let us now formalize the experiment in the language of justification logic. We start with the fact that Wigner knows the implications (7)–(9). That means there exists a term r such that

$$[r]_W\big((\mathsf{u} = \mathsf{ok}) \to (\mathsf{b} = 1)\big)$$
$$[r]_W\big((\mathsf{b} = 1) \to (\mathsf{a} = 1)\big)$$
$$[r]_W\big((\mathsf{a} = 1) \to (\mathsf{w} = \mathsf{fail})\big),$$

where we treat $(\mathsf{u} = \mathsf{ok})$, $(\mathsf{b} = 1)$, $(\mathsf{a} = 1)$, and $(\mathsf{w} = \mathsf{fail})$ as propositional atoms. Further we let $(\mathsf{w} = \mathsf{ok})$ be an abbreviation for $\neg(\mathsf{w} = \mathsf{fail})$. Now we can reason in $\mathsf{CTJ}_{\mathsf{CS}}$ as follows. Using axiom **j** we find that for all terms x

$$[x]_W(\mathsf{u} = \mathsf{ok}) \to [r \cdot (r \cdot (r \cdot x))]_W(\mathsf{w} = \mathsf{fail}). \tag{13}$$

Again, since the setup is common knowledge and the experiment halts, Wigner knows that both Ursula and Wigner obtained outcome ok. Since $(\mathsf{w} = \mathsf{ok})$ is $\neg(\mathsf{w} = \mathsf{fail})$, we may assume that there is a term s such that

$$[s]_W(\mathsf{u} = \mathsf{ok}) \tag{14}$$
$$[s]_W\neg(\mathsf{w} = \mathsf{fail}). \tag{15}$$

From (14) and (13) we get

$$[r \cdot (r \cdot (r \cdot s))]_W(\mathsf{w} = \mathsf{fail}), \tag{16}$$

which, in $\mathsf{CTJ}_{\mathsf{CS}}$, does not contradict (15). A model for a similar case is provided in Example 1.

Note that there are sever restrictions on CS for (15) and (16) to not contradict each other. In particular, the constant specification cannot be both axiomatically appropriate and schematic (see Lemma 4). A good choice for CS would be an axiomatically appropriate CS that is not schematic because then we still have internalization (Lemma 1). This is important for epistemic reasoning in general (see Section 4.), and, in particular, in the context of assumption **(C)** of the Frauchiger–Renner paradox (see the discussion in the next section).

This formalization shows that justification logic is an adequate framework to represent the Frauchiger–Renner paradox. $\mathsf{CTJ}_{\mathsf{CS}}$ is strong enough to model complex epistemic situations (as shown in the previous section) yet formalizing the paradox does not lead to an inconsistency.

6. Discussion

Nurgalieva and del Rio [Nurgalieva, del Rio, 2019] discuss formalizations of the Frauchiger–Renner paradox in modal logic. They claim that modal logic is not adequate in quantum settings since formalizing the Frauchiger–Renner paradox in modal logic leads to inconsistencies.

Essentially, their formalization is along the same lines as the one we present in justification logic (actually we followed their model). The important difference, however, is that in modal logic one only has the $\Box$-modality at hand and thus cannot distinguish between different reasons for an agent's belief. So instead of our (15) and (16), one obtains in a modal logic setting

$$\Box_W \neg(\mathsf{w} = \mathsf{fail}) \quad \text{and} \quad \Box_W(\mathsf{w} = \mathsf{fail}), \tag{17}$$

respectively.

In modal logic D, where the axiom

$$\neg \Box_a \bot \tag{18}$$

is present for all agents a, the situation (17) is obviously inconsistent. An easy way out would be to drop axiom (18) and simply use modal logic K to avoid the contradiction. However, this is not an option since axiom (18) is necessary to adequately model the assumptions of the Frauchiger–Renner paradox in modal logic. In particular, we have assumption

(S) experimenters having the subjective experience of seeing only one outcome,

which is taken care of by (18) in the sense of *experimenters cannot have the subjective experience of contradicting outcomes.* The problem, of course, is that (18) does not talk about subjective experience but about belief of an agent; and in the language of modal logic, one cannot distinguish whether an agent's belief originates from subjective experience, communication with other agents, or logical reasoning.

In justification logic $\mathsf{CTJ}_{\mathsf{CS}}$, asssumption **(S)** is modelled by axiom **noc** saying that it is not possible that the same evidence justifies both a proposition and its negation. This matches better the idea behind **(S)** because now we can state that measurements have a unique outcome (from the perspective of the agent carrying out the experiment). Yet, communication with other agents and logical reasoning may lead to contradicting beliefs.

Now let us briefly look at assumption **(C)**. Nurgalieva and del Rio state that **(C)** is modelled by the distributivity axiom of modal logic, which corresponds to axiom j in $\mathsf{CTJ}_{\mathsf{CS}}$. If we work with an axiomatically appropriate constant specification we additionally have an analogue to the necessitation rule of modal logic (see Lemma 1). A more detailed discussion of assumption **(C)** is given in the next two sections.

Nurgalieva and del Rio [Nurgalieva, del Rio, 2019] discuss a formalization of the Frauchiger–Renner paradox in a modal logic with contexts [Schroeter, 2019]. There, the contradiction is avoided since the distributivity axiom can only be applied in matching contexts. However, this also means that even simple logical reasoning often cannot be performed. Another problem is that the contexts may grow exponentially. There is strong evidence that such an exponential blow-up does not happen in justification logic. Brezhnev and Kuznets [Brezhnev, Kuznets, 2006] present a realization procedure of the modal logic $\mathsf{S4}$ into the Logic of Proofs LP that produces justification terms of at most quadratic length. Although $\mathsf{CTJ}_{\mathsf{CS}}$ is not an explicit counterpart of a modal logic (see Remark 2) and thus we cannot establish a realization result, we take Brezhnev and Kuznets' result as a hint that also $\mathsf{CTJ}_{\mathsf{CS}}$ behaves well with respect to complexity.

7. Epistemic Reasoning with Trust

Assumption **(C)** is originally explained as follows (where we omit the references to time) [Frauchiger, Renner, 2018]: *A theory T that satisfies* **(C)** *allows any agent Alice to reason as follows. If Alice has established 'I am certain that agent B (upon reasoning using T) is certain that P', then Alice can conclude 'I am certain that P.'*

Therefore, Nurgalieva and del Rio [Nurgalieva, del Rio, 2019] suggest to use a trust axiom of the form

$$\Box_a \Box_b P \to \Box_a P$$

to model **(C)** properly. The above axiom expresses that agent a trusts agent b. Formally one could consider a system where all agents trust each other or a system with an explicit trust relation.

To extend $\mathsf{CTJ}_{\mathsf{CS}}$ with trust, we need a new operation on terms. Namely,

if s is a term, then $\downarrow s$ is a term, too.

Then we can state the trust axiom as

ju $[s]_a [t]_b A \to [\downarrow s]_a A)$

Using this axiom, we get a more accurate formalization of the epistemic situation given in Section 4.. This concerns the step when agent b makes the announcement to agent a.

We start with (4)

$$[t{\cdot}!\mathsf{obs}]_b\neg[\mathsf{obs}]_b\mathsf{red}.$$

Now agent b announces this to agent a. Then we have

$$[\mathsf{ann}]_a[t{\cdot}!\mathsf{obs}]_b\neg[\mathsf{obs}]_b\mathsf{red},$$

where ann is a term representing the announcement.

Now we use the Trust axiom **ju** to derive

$$[\downarrow\mathsf{ann}]_a\neg[\mathsf{obs}]_b\mathsf{red},$$

i.e. agent a knows that it is not the case that agent b observed that the hat is red. Combining this with (2) yields

$$[s_2{\cdot}\downarrow\mathsf{ann}]_a\neg\mathsf{red},$$

which means that after b's announcement, agent a knows that the hat is not red. Note that the evidence term for a's knowledge contains the $\downarrow$-operation meaning that the trust relation was used to obain that knowledge. One could also extend the language and index the $\downarrow$ with the agents to show who trusted whom, similar to Yavorskaya's evidence conversion operator $\uparrow^a_b$, see [Yavorskaya, 2007]. Actually, the combination of the announcement and the trust axiom employed in our example above has the same effect as the evidence conversion operation, which is axiomatized by

$$[t]_bA\rightarrow[\uparrow^a_b t]_aA.$$

Hence, in Yavorskaya's system we would apply the $\uparrow^b_a$-operation to (4) to obtain

$$[\uparrow^a_b(t{\cdot}!\mathsf{obs})]_a\neg[\mathsf{obs}]_b\mathsf{red}$$

and using (2) we could conclude

$$[s_2{\cdot}\uparrow^a_b(t{\cdot}!\mathsf{obs})]_a\neg\mathsf{red}.$$

Note that single agent versions of the trust axiom are discussed in the frame of deontic justification logic by Faroldi and Protopopescu [Faroldi, Protopopescu, 2019]. Also Fitting [Fitting, 2016] discusses them in the context of realizing Geach logics.

We want to finish this section with a remark showing how the trust principle and our (informal) announcements play together in the context of conflicting evidence.

Remark 5. Assume agent b has conflicting justifications $[s]_b F$ and $[t]_b \neg F$. Using **j4** we find that $[!s]_b[s]_b F$ and $[!t]_b[t]_b \neg F$. By Lemma 2 we find a term r such that we get $[r{\cdot}!s{\cdot}!t]_b([s]_b F \wedge [t]_b \neg F)$. Now agent b can announce $[s]_b F \wedge [t]_b \neg F$ to agent a, which gives us

$$[\mathsf{ann}]_a([s]_b F \wedge [t]_b \neg F).$$

Since CS is axiomatically appropriate there are terms p_1 and p_2 such that

$$[p_1]_a([s]_b F \wedge [t]_b \neg F \to [s]_b F) \quad \text{and} \quad [p_2]_a([s]_b F \wedge [t]_b \neg F \to [t]_b \neg F)$$

are provable in $\mathsf{CTJ}_{\mathsf{CS}}$. Thus we obtain

$$[p_1 \cdot \mathsf{ann}]_a[s]_b F \quad \text{and} \quad [p_2 \cdot \mathsf{ann}]_a[t]_b \neg F.$$

By **ju** we get

$$[{\downarrow}(p_1 \cdot \mathsf{ann})]_a F \quad \text{and} \quad [{\downarrow}(p_2 \cdot \mathsf{ann})]_a \neg F.$$

The last two formulas are in accordance with **noc**. But this requires p_1 and p_2 to be two different terms. Hence hyperintensionality, in particular distinguishing between $A \wedge B$ and $B \wedge A$, is essential for making our approach work.

8. Frauchiger–Renner with trust

In [Nurgalieva, del Rio, 2019], there is also a modal logic analysis of the Frauchiger–Renner paradox given that takes into account the trust relation between the agents where Bob trusts Alice, Ursula trusts Bob, Wigner trusts Ursula.

Again we closely follow the presentation of [Nurgalieva, del Rio, 2019] and adapt it to justification logic. Before the experiment begins, but after the agents talked to each other, we have the following statements about Wigner's beliefs:

$$\begin{gathered}
[r_1]_W[s_1]_U\big([\mathsf{obs}_U]_U(\mathsf{u} = \mathsf{ok}) \to [v_1(\mathsf{obs}_U)]_B(\mathsf{b} = 1)\big) \\
[r_2]_W[s_2]_U[v_2]_B\big([\mathsf{obs}_B]_B(\mathsf{b} = 1) \to [w(\mathsf{obs}_B)]_A(\mathsf{a} = 1)\big) \\
[r_3]_W[s_3]_U[v_3]_B[w_3]_A\big([\mathsf{obs}_A]_A(\mathsf{a} = 1) \to [r_4(\mathsf{obs}_A)]_W(\mathsf{w} = \mathsf{fail})\big).
\end{gathered}$$

Applying the trust axiom several times leads to

$$\begin{gathered}
[{\downarrow} r_1]_W\big([\mathsf{obs}_U]_U(\mathsf{u} = \mathsf{ok}) \to [v_1(\mathsf{obs}_U)]_B(\mathsf{b} = 1)\big) \\
[{\downarrow\downarrow} r_2]_W\big([\mathsf{obs}_B]_B(\mathsf{b} = 1) \to [w(\mathsf{obs}_B)]_A(\mathsf{a} = 1)\big) \\
[{\downarrow\downarrow\downarrow} r_3]_W\big([\mathsf{obs}_A]_A(\mathsf{a} = 1) \to [r_4(\mathsf{obs}_A)]_W(\mathsf{w} = \mathsf{fail})\big).
\end{gathered}$$

Thus we find a term r_5 such that

$$[r_5]_W\big([x]_U(\mathsf{u} = \mathsf{ok}) \to [r_4(w(v_1(x)))]_W(\mathsf{w} = \mathsf{fail})\big). \tag{19}$$

Now we run the experiment and consider the case when the experiment halts. Then Ursula and Wigner are both ok, i.e. there are terms obs_U and obs_W such that

$$[\mathsf{obs}_U]_U(\mathsf{u} = \mathsf{ok}) \quad \text{and} \quad [\mathsf{obs}_W]_W(\mathsf{w} = \mathsf{ok}).$$

Moreover they learn of each other's outcomes, in particular, there is a term r_6 such that

$$[r_6]_W[\mathsf{obs}_U]_U(\mathsf{u} = \mathsf{ok}).$$

Plugin this into (19) yields

$$[r_5 \cdot r_6]_W[r_4(w(v_1(\mathsf{obs}_U)))]_W(\mathsf{w} = \mathsf{fail}).$$

Since Wigner trusts himself, we conclude

$$[\downarrow(r_5 \cdot r_6)]_W(\mathsf{w} = \mathsf{fail}),$$

which again does not contradict $[\mathsf{obs}_W]_W(\mathsf{w} = \mathsf{ok})$ in $\mathsf{CTJ}_{\mathsf{CS}}$ where we again use an axiomatically appriopriate but non-schematic CS.

This is in contrast to the modal logic formalization given in [Nurgalieva, del Rio, 2019] where we obtain a contradiction in the modal logic D extended by the trust axioms.

9. Conclusion

We introduced $\mathsf{CTJ}_{\mathsf{CS}}$, a new epistemic justificaton logic. $\mathsf{CTJ}_{\mathsf{CS}}$ disallows one piece of evidence to justify both a proposition and its negation but still tolerates conflicting beliefs. We studied epistemic reasoning in $\mathsf{CTJ}_{\mathsf{CS}}$ and showed that it can adequately represent the Frauchiger–Renner paradox from quantum physics. Further, we investigated an extension of $\mathsf{CTJ}_{\mathsf{CS}}$ with trust axioms.

The price we had to pay for obtaining conflict tolerance was to drop the sum-operation from standard justification logic. So this paper can also be seen as a contribution to understanding the role of the $+$-operation. But sum also is one of the most intuitive operations for justification logic and it is crucial for obtaining normal realizations. Thus further investigations on the sum operation are definitely needed.

It might also be interesting to investigate the relationship of $\mathsf{CTJ}_{\mathsf{CS}}$ and conflict tolerant non-normal modal logics. For instance, in Chellas' minimal deontic logic $\neg\Box\bot$ does not imply $\neg(\Box A \wedge \Box\neg A)$, see [Chellas, 1980].

Acknowledgements. We are grateful to Lídia del Rio who introduced us to the physics of the Frauchiger–Renner paradox. We would also like to thank Michael Baur for proof reading this paper. Further we thank the anonymous reviewers for many very valuable comments. This work is supported by the Swiss National Science Foundation grant 200020_184625.

References

Arrow, 1950 – K. J. Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, 58(4):328–346, 1950.

Artemov, 2001 – S. Artemov. Explicit provability and constructive semantics. *Bulletin of Symbolic Logic*, 7(1):1–36, Mar. 2001.

Artemov, 2006 – S. N. Artemov. Justified common knowledge. *TCS*, 357(1–3):4–22, July 2006.

Artemov, 2008 – S. N. Artemov. The logic of justification. *RSL*, 1(4):477–513, Dec. 2008.

Artemov, Fitting, 2019 – S. Artemov and M. Fitting. *Justification Logic: Reasoning with Reasons.* Cambridge University Press, 2019.

Brezhnev, Kuznets, 2006 – V. N. Brezhnev and R. Kuznets. Making knowledge explicit: How hard it is. *TCS*, 357(1–3):23–34, July 2006.

Bucheli et al., 2011 – S. Bucheli, R. Kuznets, and T. Studer. Justifications for common knowledge. *Applied Non-Classical Logics*, 21(1):35–60, Jan.–Mar. 2011.

Bucheli et al., 2014 – S. Bucheli, R. Kuznets, and T. Studer. Realizing public announcements by justifications. *Journal of Computer and System Sciences*, 80(6):1046–1066, 2014.

Chellas, 1980 – B. F. Chellas. *Modal Logic. An Introduction.* Cambridge University Press, Cambridge, 1980.

Faroldi et al., 2020 – F. Faroldi, M. Ghari, E. Lehmann, and T. Studer. Impossible and conflicting obligations in justification logic. In A. Marra, F. Liu, P. Portner, and F. Van De Putte, editors, *Proceedings of DEON 2020*, 2020.

Faroldi, Protopopescu, 2019 – F. L. G. Faroldi and T. Protopopescu. A hyperintensional logical framework for deontic reasons. *Logic Journal of the IGPL*, 27:411–433, 2019.

Fitting, 2005 – M. Fitting. The logic of proofs, semantically. *APAL*, 132(1):1–25, Feb. 2005.

Fitting, 2016 – M. Fitting. Modal logics, justification logics, and realization. *Annals of Pure and Applied Logic*, 167(8):615–648, 2016.

Frauchiger, Renner, 2018 – D. Frauchiger and R. Renner. Quantum theory cannot consistently describe the use of itself. *Nature Communications*, 9, 2018.

Kuznets, Studer, 2012 – R. Kuznets and T. Studer. Justifications, ontology, and conservativity. In T. Bolander, T. Braüner, S. Ghilardi, and L. Moss, editors, *Advances in Modal Logic, Volume 9*, pages 437–458. College Publications, 2012.

Kuznets, Studer, 2016 – R. Kuznets and T. Studer. Weak arithmetical interpretations for the Logic of Proofs. *Logic Journal of IGPL*, 24(3):4243–440, 2016.

Kuznets, Studer, 2019 – R. Kuznets and T. Studer. *Logics of Proofs and Justifications.* College Publications, 2019.

Lazarovici, Hubert, 2019 – D. Lazarovici and M. Hubert. How quantum mechanics can consistently describe the use of itself. *Scientific Reports*, 9, 2019.

Lehmann, Studer, 2019 – E. Lehmann and T. Studer. Subset models for justification logic. In R. Iemhoff, M. Moortgat, and R. de Queiroz, editors, *Logic, Language, Information, and Computation - WoLLIC 2019*, pages 433–449. Springer, 2019.

Lehmann, Studer, 2020 – E. Lehmann and T. Studer. Belief expansion in subset models. In *Proceedings of Logical Foundations of Computer Science LFCS'20.* Springer, 2020.

Nurgalieva, del Rio, 2019 – N. Nurgalieva and L. del Rio. Inadequacy of modal logic in quantum settings. In P. Selinger and G. Chiribella, editors, *Proceedings 15th International Conference on Quantum Physics and Logic, QPL 2018, Halifax, Canada, 3-7th June 2018.*, volume 287 of *EPTCS*, pages 267–297, 2019.

Pacuit, Yang, 2016 – E. Pacuit and F. Yang. Dependence and independence in social choice: Arrow's theorem. In S. Abramsky, J. Kontinen, J. Väänänen, and H. Vollmer, editors, *Dependence Logic: Theory and Applications*, pages 235–260. Springer, 2016.

Schroeter, 2019 – L. Schroeter. Two-Dimensional Semantics. In E. N. Zalta, editors, *The Stanford Encyclopedia of Philosophy.* Metaphysics Research Lab, Stanford University, 2019.

Studer, 2013 – T. Studer. Decidability for some justification logics with negative introspection. *JSL*, 78(2):388–402, June 2013.

Studer, 2020 – T. Studer. No-go theorems for data privacy. In *Proceedings of the 7th International Cryptology and Information Security Conference 2020*, pages 74–84, 2020.

Studer, Werner, 2014 – T. Studer and J. Werner. Censors for boolean description logic. *Transactions on Data Privacy*, 7:223–252, 2014.

Yavorskaya, 2007 – T. Yavorskaya (Sidon). Interacting explicit evidence systems. *Theory of Computing Systems*, 43(2):272–293, Aug. 2008. Published online October 2007.