Paraconsistent Justification Logic: a Starting Point

Che-Ping Su [1]

School of Historical and Philosophical Studies
The University of Melbourne

Abstract

In his 2003 paper, *Ten philosophical problems in belief revision*, Sven Ove Hansson argues that sometimes belief revision might essentially involve inconsistent epistemic states, and that to better model belief revision requires well modeling inconsistent epistemic states. In this paper, we are going to develop a type of justification logic, which is intended to model the agent's justification structure, when she is in an inconsistent epistemic state. We call the logic to be developed *paraconsistent justification logic*. The hope is that this logic could help us better model belief revision.
More specifically, we will construct a three-valued justification logic system, which serves as the starting point of the research. Roughly speaking, the system reflects the idea that committing to a contradiction does not imply committing to everything. This idea is taken to be our basic assumption about inconsistent epistemic states. In addition, the main technical result of the paper is that quasi-realization theorem – which holds for standard two-valued justification logic systems – also holds for this three-valued system.

Keywords: justification logic, paraconsistent logic, inconsistent epistemic state, quasi-realization, belief revision.

1 Introduction

The classic justification logic system $\mathsf{JT4}$ (the *Logic of Proofs*) [1] is the justification counterpart of the classic epistemic logic system $\mathsf{S4}$. Similarly, *paraconsistent justification logic* is intended to be the justification counterpart of *paraconsistent epistemic logic*. To introduce the former, let us start with the latter.

By 'paraconsistent epistemic logic', I mean the family of epistemic logic systems that have this property: $(\Box\phi \wedge \Box\neg\phi) \rightarrow \Box\psi$ is not a valid scheme. System K and all extensions of K do not have this property. In [5,8,10,13], systems with the property are introduced. And, in the literature, some works interpret $\Box$ occurring in the property as the belief operator. Hence, the property then

[1] sucheping@gmail.com is my e-mail. And, I would especially like to thank Graham Priest, Sergei Artemov, Melvin Fitting, Tudor Protopopescu and three anonymous referees.

expresses: having inconsistent beliefs does not imply believing everything. In the context of this paper, I would like to interpret $\Box\phi$ as 'the agent (explicitly) commits to ϕ'. In general, paraconsistent epistemic logic is intended to better model inconsistent epistemic states.

Paraconsistent justification logic is to be developed in this paper.[2] It shall refer to the family of justification logic systems that satisfy the property: it is not the case that for all justification terms s_1, s_2, there exists a justification term t such that $(s_1 : \phi \wedge s_2 : \neg\phi) \to t : \psi$ is a valid scheme.[3] $s_1 : \phi$ will be intuitively interpreted as 'the agent (explicitly) commits to ϕ for evidence s_1'. And, $s_2 : \neg\phi$ and $t : \psi$ will be interpreted in a similar way. Then, roughly speaking, the property expresses: having pieces of evidence that force us to (explicitly) commit to a contradiction does not imply that for each statement ψ, we have corresponding evidence forcing us to commit to ψ. In general, paraconsistent justification logic is intended to model the agent's justification structure, when she is in an inconsistent epistemic state.

How to construct paraconsistent justification logic is suggested by Fitting models [6] for justification logic. In standard justification logics such as J, $\mathsf{JT4}$ and $\mathsf{JT45}$, formula $t : \phi$ intuitively means that 'the agent believes/knows ϕ for evidence t'. And, informally, under Fitting models, the truth condition of $t : \phi$ is analyzed as the conjunction of:

- t is admissible evidence for ϕ;
- the agent believes/knows ϕ.

In Fitting models, the first condition is formally handled by evidence functions, which are syntactic functions that map worlds and justification terms to sets of formulas. And, Fitting models formalize the second condition by appealing to some epistemic logic. If we use, for example, K to handle the second condition and put corresponding constraints on evidence functions, roughly speaking we get J. Similarly, if we use $\mathsf{S4}$, then we get $\mathsf{JT4}$. Hence, the above suggests that using a paraconsistent epistemic logic system to handle the second condition, we might get a corresponding paraconsistent justification logic. This is how in this paper we are going to construct a paraconsistent justification logic.

There are three main tasks in this paper. First of all, we will try to persuade readers that paraconsistent justification logic is worth developing. Secondly, based on a specific paraconsistent epistemic logic, called $\mathsf{PE_b}$, we are going to construct a paraconsistent justification logic system, called $\mathsf{PJ_b}$. $\mathsf{PJ_b}$ serves as a starting point of the project of developing paraconsistent justification logic. Finally, we will prove: as modal logics, for example, K, $\mathsf{S4}$, $\mathsf{S5}$ can be embedded into justification logics J, $\mathsf{JT4}$, $\mathsf{JT45}$, respectively, $\mathsf{PE_b}$ can be embedded into $\mathsf{PJ_b}$, too. More specifically, we will follow Fitting's non-constructive way [6] to

[2] Joseph Lurie also independently comes up with the idea of merging justification logic with paraconsistent logic. Readers could go to [11] to see more details about how Lurie develops the idea.

[3] Actually, the paraconsistent justification logic system $\mathsf{PJ_b}$, which we are going to construct, satisfies stronger properties, that is, Fact 4.4's (iii) and (iv).

show that quasi-realization theorem holds for $\langle \mathsf{PE_b}, \mathsf{PJ_b} \rangle$.

2 Why Paraconsistent Justification Logic

In this section, I hope to persuade readers that paraconsistent justification logic is worth developing.

It was advertised in Introduction that in the context of this paper, paraconsistent epistemic logic is intended to model the epistemic state of 'explicitly committing to a contradiction'. And, paraconsistent justification logic is intended to model the agent's justification structure, when she is in this kind of epistemic state. Hence, to well motivate paraconsistent justification logic, answers to the following six questions should be provided:

(i) what the epistemic state – explicitly committing to a contradiction – is;

(ii) whether such sort of epistemic states really exist;

(iii) what the basic properties of the epistemic state are;

(iv) why talk about explicit commitments, rather than beliefs;

(v) why care about such sort of epistemic states;

(vi) why care about the agent's justification structure, when she is in this kind of epistemic state.

To the first question, let us start with distinguishing *explicit beliefs*, *implicit commitments* and *explicit commitments*. If an agent explicitly believes some statement, then the agent accepts this statement. Examples of explicit beliefs are my belief that $2 + 2 = 4$ and my belief that I am a human being. In contrast, implicit commitments are the agent's explicit beliefs' consequences that the agent is not aware of. For example, let ϕ be a deep consequence of Peano arithmetic that has not been proved by any mathematician. Then, all people believing PA implicitly commit to ϕ. On the other hand, explicit commitments are the agent's explicit beliefs consequences that the agent is aware of, but the agent might/might not accept.

Explicitly believing a contradiction is not usual. Implicitly committing to a contradiction sometimes happens. For instance, before knowing Russel's paradox of his set theory, Frege implicitly committed to a contradiction. Explicitly committing to a contradiction is the sort of epistemic state that paraconsistent justification logic intends to model.

Secondly, I will argue by example that the epistemic state of explicitly committing to a contradiction does occur. In 1859, Urbain Le Verrier discovered that the actual movement of the planet Mercury is not like what the traditional theory of Newtonian gravity predicts. In other words, the Newtonian gravity plus Le Verrier's obervation leads to a contradiction. The successful revision of the theory of gravity had to wait till Einstein's theory of general relativity (1915). Between 1859 and 1915, physicists could not simply give up/'freeze' Newtonian gravity to keep consistency, because it was the best available theory of the time. Physicists of this period were aware of the contradiction. The contradiction is a consequence of their explicit beliefs in Newtonian gravity and

Le Verrier's obervation. Hence, physicists of this period explicitly committed to the contradiction. However, they did not accept the contradiction to be true.

To the third question, in this paper, we assume that two properties hold for the epistemic state of explicitly committing to a contradiction. One is that explicitly committing to a contradiction does not imply explicitly committing to everything. The other property assumed is: having pieces of evidence that force us to explicitly commit to a contradiction does not imply that for each ψ, we have corresponding evidence forcing us to commit to ψ.

Fourth, in the context of this paper, to simplify things we assume that the agent we model is ideal in the sense that her awareness is logical omniscient. Hence, there is no implicit commitment for the agent; she can only either explicitly believe a contradiction or explicitly commit to a contradiction. Since it is rare that an agent explicitly believes a contradiction, we really need to argue for the category of explicit commitments to contradictions.

The fifth question is why we need to care about modeling the epistemic state of explicitly committing to a contradiction. In the following, I would like to answer the question from the angle of belief revision.

In Sven Ove Hansson's paper [9], it is argued that sometimes belief revision might essentially involve inconsistent epistemic states. Here is a summery of Hansson's point. In the literature of belief revision, there is one approach called *belief base belief revision*, where the belief set is not required to be closed under a consequence relation. In belief base belief revision, there are two ways to define the revision operator:

- revision = expansion + contraction
- revision = contraction + expansion

At least formally, these two ways of defining the revision operator do not collapse into the same operator. When the agent must accept the new information but it is unclear to give up which piece of old information, the first way – revision = expansion + contraction – seems to fit real psychology more. In the first way, there is an intermediate inconsistent epistemic state that occurs after expansion.

To explain Hansson's point, let us recall the example of Le Verrier's observation about Mercury. The physicists of that period ought to accept the new information, that is, Le Verrier's observation. However, it was unclear which part of the Newtonian gravity (or background postulates) we should give up/change. Before the successful revision, physicists explicitly committed to a contradiction. We might take the history as: first expanding with the new information (Le Verrier's observation); secondly, reaching the inconsistent epistemic state of explicitly committing to a contradiction; third, changing the original theory. Therefore, from the angle of belief revision, it seems that this sort of inconsistent epistemic states – explicitly committing to a contradiction

– should be paid attention to [4] .

To the final question, I would like to answer it also from the angle of belief revision. My answer begins with an intuition: the same belief/commitment content with different justification structures might lead to different belief revision results. Let us compare two example cases to illustrate the intuition. First, assume that John believes all european swans are white. He forms this belief based on observing most part of Northern Europe and randomly checking the rest of Europe. However, one day, he sees a black swan in Germany. The other case goes as follows. Here, Miles also believes all european swans are white. However, he forms the belief because he watches one program in the Discovery channel and the program says so. And, one day Miles also sees a black swan in Germany. In the first case, John probably can still keep his belief that all swans in Finland are white, because John has done careful survey in Northern Europe. But, it is less clear that in the second case, Miles can keep the same belief. In both cases, each agent starts with the same belief content. But the results of revision are different. It is probably because agents' justification structures are different.

Hence, putting the following two ideas together:

- the epistemic state of explicitly committing to a contradiction sometimes plays an essential role in the process of belief revision;
- the agent's justification structure is a factor to determine the revision result,

we might conclude that we should pay attention to the agent's justification structure, when she explicitly commits to a contradiction.

3 Paraconsistent Epistemic Logic $\mathsf{PE_b}$

In this section, we are going to construct a three-valued epistemic logic system such that $(\Box\phi\wedge\Box\neg\phi) \to \Box\psi$ is not a valid scheme. We call this system $\mathsf{PE_b}$: the subscript b stands for the third truth value b. Our strategy to construct such a system is to take as the propositional part of $\mathsf{PE_b}$ a non-classical propositional logic where from $\phi\wedge\neg\phi$ we can not derive everything. The specific non-classical propositional logic we pick here has been studied in the paraconsistent logic literature [2,3].

Here is the syntax of $\mathsf{PE_b}$. $\mathsf{PE_b}$ *formulas* are inductively defined as:

$$\phi \quad ::= \quad p \mid \neg\phi \mid (\phi\wedge\phi) \mid (\phi\to\phi) \mid \Box\phi$$

Intuitively, $\Box\phi$ means that 'the agent explicitly commits to ϕ'.

Here is the semantics of $\mathsf{PE_b}$. A *frame* is a pair $\langle W, R\rangle$, where W is a non-empty set of possible worlds and R is an accessibility relation of type $W\times W$. At the current stage, to simplify things, we do not put any constraint on R. A $\mathsf{PE_b}$ *model*, $\mathcal{M}$ is a three-tuple $\langle W, R, v\rangle$, where $\langle W, R\rangle$ is a frame, and v is

[4] Note that up to this point, we might have provided sufficient reasons to motivate developing *paraconsistent* dynamic epistemic logic for belief revision. [8] is along this line.

a function from worlds and propositional variables to $\{1, b, 0\}$. In addition, 1 and b are designated values. b intuitively means 'being both true and false'.

[12] provides one way to make sense of value b. Each world with some propositional variables assigned value b is an *impossible world*. "Impossible worlds are just more than one possible world taken together"[12]. An impossible world represents a way that the world can not be. Different impossible worlds represent different ways of 'clashes' (in commitment contents).

We extend v in the following way.

Definition 3.1 Let $\mathcal{M} = \langle W, R, v\rangle$ be a $\mathsf{PE_b}$ model. $v^{\mathcal{M}}$ extends v in the following way. Let u be a world in $\mathcal{M}$.

(i)

$$v_u^{\mathcal{M}}(p) = v_u(p)$$

(ii)

$$v_u^{\mathcal{M}}(\neg\phi) = \begin{cases} 1 & \text{if } v_u^{\mathcal{M}}(\phi) = 0; \\ b & \text{if } v_u^{\mathcal{M}}(\phi) = b; \\ 0 & \text{otherwise.} \end{cases}$$

(iii)

$$v_u^{\mathcal{M}}(\phi \wedge \psi) = \begin{cases} 1 & \text{if } v_u^{\mathcal{M}}(\phi) = 1 \text{ and } v_u^{\mathcal{M}}(\psi) = 1; \\ b & \text{if } v_u^{\mathcal{M}}(\phi) = b \text{ and } v_u^{\mathcal{M}}(\psi) \neq 0, \\ & \text{or } v_u^{\mathcal{M}}(\phi) \neq 0 \text{ and } v_u^{\mathcal{M}}(\psi) = b; \\ 0 & \text{otherwise.} \end{cases}$$

(iv)

$$v_u^{\mathcal{M}}(\phi \to \psi) = \begin{cases} v_u^{\mathcal{M}}(\psi) & \text{if } v_u^{\mathcal{M}}(\phi) \in \{1, b\}; \\ 1 & \text{otherwise.} \end{cases}$$

(v)

$$v_u^{\mathcal{M}}(\Box\phi) = \begin{cases} 1 & \text{if for all } u' \text{ with } uRu',\ v_{u'}^{\mathcal{M}}(\phi) \in \{1, b\}; \\ 0 & \text{otherwise.} \end{cases}$$

Define $\phi \vee \psi$ as the abbreviation of $\neg(\neg\phi \wedge \neg\psi)$ [5]; $\phi \leftrightarrow \psi$ as the abbreviation of $(\phi \to \psi) \wedge (\psi \to \phi)$. In addition, note that the conditional of $\mathsf{PE_b}$ can not be defined by $\neg$, $\wedge$ [4]. Furthermore, $\Box\phi$ never gets value b. The rationale is that although the agent's commitment content could be inconsistent, it is never the case that the agent *both* explicitly commits to some statement and does not explicitly commit to the same statement [6].

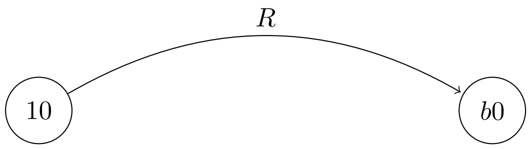
Let us look at an example model of the system.

---

[5] Assume this order on the truth values: $0 < b < 1$. Then, $\vee$ corresponds to the operation of taking the maximum value. This justifies the abbreviation.

[6] Note that not committing to ϕ is different from committing to not-ϕ.

Example 3.2 Let us consider a restricted language, which only has two propositional variables, p, q. Let 10 stand for a world where p is true (only) and q is false (only). Let $b0$ stand for a world where p is *both* true and false, but q is false (only) . (So, $b0$ is an impossible world.) Then, the following is an example $\mathsf{PE_b}$ model. Let us call it $\mathcal{N}$.



Accurately speaking, $\mathcal{N} = \langle W, R, v\rangle$, where $W = \{10, b0\}$; $R = \{\langle 10, b0\rangle\}$; $v_{10}(p) = 1$, $v_{10}(q) = 0$, $v_{b0}(p) = b$, $v_{b0}(q) = 0$. Further, note that $v^{\mathcal{N}}_{10}(\Box p) = v^{\mathcal{N}}_{10}(\Box\neg p) = 1$, but $v^{\mathcal{N}}_{10}(\Box q) = 0$. Hence, the world 10 in $\mathcal{N}$ shows that $(\Box\phi \wedge \Box\neg\phi) \to \Box\psi$ is not a valid scheme in $\mathsf{PE_b}$.

Here are some basic facts about $\mathsf{PE_b}$.

Fact 3.3 (i) $(\Box p \wedge \Box\neg p)$ *is satisfiable.*

(ii) $(\Box\phi \wedge \Box\neg\phi) \to \Box\psi$ *is not a valid scheme. That is, for some formulas* ϕ *and* ψ, $(\Box\phi \wedge \Box\neg\phi) \to \Box\psi$ *is invalid.*

(iii) *For all formulas* ϕ, *if* ϕ *contains no* $\Box$, *then there exist some formula* ψ *such that* $(\Box\phi \wedge \Box\neg\phi) \to \Box\psi$ *is invalid.*

(iv) *For all formulas* ϕ *and* ψ, *if both* ϕ, ψ *contain no* $\Box$; ϕ *and* ψ *share no propositional variable; and* ψ *is not valid, then* $(\Box\phi \wedge \Box\neg\phi) \to \Box\psi$ *is invalid.*

(v) *Modus Ponens preserves designated values.*

(vi) $\Sigma \cup \{\phi\} \Vdash \psi$, *iff* $\Sigma \Vdash \phi \to \psi$.[7]

(vii) $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$ *is a valid scheme.*

Proof. Please see the first half of Appendix A. □

The first four of Fact 3.3 form a group. It is about the paraconsistency of $\mathsf{PE_b}$. (i) says that an agent might explicitly commits to a contradiction. We explain (ii) and (iii) together. Maybe, the slogan for paraconsistent epistemic logic is that 'explicitly committing to a contradiction does not imply explicitly committing to everything'. However, the slogan is vague. There are two possible accurate readings of the slogan:

- for **some** contradiction $\phi \wedge \neg\phi$, explicitly committing to $\phi \wedge \neg\phi$ does not imply explicitly committing to everything;
- for **all** contradictions $\phi \wedge \neg\phi$, explicitly committing to $\phi \wedge \neg\phi$ does not imply explicitly committing to everything.

The first is weaker than the second. (ii) expresses the weaker reading of the

[7] We define the semantical consequence relation $\Vdash$ for $\mathsf{PE_b}$ as follows: $\Gamma \Vdash \tau$, iff for all $\mathsf{PE_b}$ models $\mathcal{M}$ and all worlds u in $\mathcal{M}$, if $v^{\mathcal{M}}_u(\tau') \in \{1, b\}$ for all $\tau' \in \Gamma$, then $v^{\mathcal{M}}_u(\tau) \in \{1, b\}$.

slogan. What (iii) expresses is closer to the stronger reading. However, (iii) does not quantify over unrestrictedly all contradictions, but just all contradictions that do not involve $\Box$. Now, we explain (iv). First of all, if two formulas share no propositional variable, intuitively we could take these two formulas to be irrelevant to each other. Secondly, based on the previous point, (iv) intuitively means:

- for all contradictions $\phi \wedge \neg\phi$ (with some restriction), for all statements ψ (with some restriction), if $\phi \wedge \neg\phi$ and ψ are irrelevant to each other in the sense that $\phi \wedge \neg\phi$ and ψ share no propositional variable, then explicitly committing to $\phi \wedge \neg\phi$ does not imply explicitly committing to ψ.

In addition, note that (iv) implies (iii), but not vice versa; (iii) implies (ii), but not vice versa.

The last three items of Fact 3.3 form another group. This group tells us that $\mathsf{PE_b}$ has Modus Ponens, deduction theorem and the validity of K-axiom. The way we define the conditional gives us these three. If we define $\phi \to \psi$ as $\neg\phi \vee \psi$, the resulting system will not have these three. As readers will see in the next section, making our system have these three will be helpful for letting the corresponding justification logic behave nicely in the technical aspect.

4 Paraconsistent Justification Logic $\mathsf{PJ_b}$

Based on $\mathsf{PE_b}$, we are going to construct a justification logic system $\mathsf{PJ_b}$ by giving its Fitting models. $\mathsf{PJ_b}$ reflects the idea: explicitly committing to a contradiction for two conflicting pieces of evidence does not imply that for all statement ψ, we have corresponding evidence that forces us to commit to ψ. This idea is accurately formulated in (ii) of Fact 4.4. Actually, $\mathsf{PJ_b}$ satisfies stronger paraconsistent properties, that is, (iii) and (iv) of Fact 4.4.

4.1 Syntax of $\mathsf{PJ_b}$

$\mathsf{PJ_b}$ has the following symbols.

(i) propositional variables, $p, q, r, \ldots$

(ii) connectives, $\neg$, $\wedge$, $\to$

(iii) justification variables, $x, y, \ldots$

(iv) justification constants, $c, d \ldots$

(v) function symbol, $\cdot$, $+$

(vi) operator symbol of the type $\langle term \rangle : \langle formula \rangle$

Justification terms are inductively defined as:

$$t \quad ::= \quad x \mid c \mid (t \cdot t) \mid (t + t)$$

$\mathsf{PJ_b}$ *formulas* are inductively defined as:

$$\phi \quad ::= \quad p \mid \neg\phi \mid (\phi \wedge \phi) \mid (\phi \to \phi) \mid t : \phi$$

Intuitively, $t : \phi$ is interpreted as ‘the agent explicitly commits to ϕ based on evidence t’.

4.2 Semantics of $\mathsf{PJ_b}$

We specify $\mathsf{PJ_b}$'s semantics by giving its Fitting models.

An *evidence function* on the frame $\langle W, R\rangle$ is a function $\mathcal{E}$ from worlds and justification terms to sets of formulas. In $\mathsf{PJ_b}$, we put the following constrains on the evidence function:

- **Application** $(\phi \to \psi) \in \mathcal{E}(u, s)$ and $\phi \in \mathcal{E}(u, t)$ implies $\psi \in \mathcal{E}(u, s \cdot t)$.
- **Sum** $\mathcal{E}(u, s) \cup \mathcal{E}(u, t) \subseteq \mathcal{E}(u, s + t)$.

A $\mathsf{PJ_b}$ *model*, $\mathcal{M}$ is a four-tuple $\langle W, R, \mathcal{E}, v\rangle$, where $\langle W, R\rangle$ is a frame, $\mathcal{E}$ is an *evidence function* on $\langle W, R\rangle$, and v is a function from worlds and propositional variables to $\{1, b, 0\}$. And, 1 and b are designated values.

We extend v in the following way.

Definition 4.1 Let $\mathcal{M} = \langle W, R, \mathcal{E}, v\rangle$ be a $\mathsf{PJ_b}$ model. $v^{\mathcal{M}}$ extends v in the following way. Let u be a world in $\mathcal{M}$.

(i) propositional variables, $\neg$, $\wedge$ and $\to$ are interpreted as in $\mathsf{PE_b}$

(ii)

$$v_u^{\mathcal{M}}(t : \phi) = \begin{cases} 1 & \text{if } \phi \in \mathcal{E}(u, t) \text{ and for all } u' \text{ with } uRu',\ v_{u'}^{\mathcal{M}}(\phi) \in \{1, b\}; \\ 0 & \text{otherwise.} \end{cases}$$

Definition 4.2 [constant specification] Let $X_0 = \{\phi \mid v_u^{\mathcal{M}}(\phi) = 1 \text{ or } b$ for every world u at every $\mathsf{PJ_b}$ models $\mathcal{M}\}$. Let $X_{n+1} = \{c : \phi \mid c$ is a justification constant and $\phi \in X_n\}$. Let $X = \bigcup_{i\in\mathbb{N}} X_i$. A *constant specification* $\mathcal{C}$ is a function from justification constants to subsets of X such that $\mathcal{C}$ satisfies the following two conditions:

- if $c_n : \cdots : c_1 : \phi \in \mathcal{C}(c_{n+1})$, then $c_{n-1} : \cdots : c_1 : \phi \in \mathcal{C}(c_n)$, where $n \geqslant 2$ and $\phi \in X_0$.
- if $c_1 : \phi \in \mathcal{C}(c_2)$, then $\phi \in \mathcal{C}(c_1)$, where $\phi \in X_0$.

A model $\mathcal{M} = \langle W, R, \mathcal{E}, v\rangle$ is said to *meet a constant specification* $\mathcal{C}$, iff for all $u \in W$, for all justification constants c, $\mathcal{C}(c) \subseteq \mathcal{E}(u, c)$.

Here is an example $\mathsf{PJ_b}$ model.

Example 4.3 Let $\mathcal{C}$ be a constant specification. Let $x, y, z_1, z_2, \ldots$ be an enumeration of all justification variables. We construct a $\mathsf{PJ_b}$ model $\mathcal{M}$ meeting $\mathcal{C}$, based on the $\mathsf{PE_b}$ model $\mathcal{N}$ defined in Example 3.2. Define $\mathcal{M} = \langle W, R, \mathcal{E}, v\rangle$, where $\langle W, R, v\rangle = \mathcal{N}$ and $\mathcal{E}$ is defined as follows:

- for all justification constants c, define $\mathcal{E}(u, c) = \mathcal{C}(c)$, where $u \in \{10, b0\}$;
- define $\mathcal{E}(10, x) = \{p\}$, $\mathcal{E}(10, y) = \{\neg p\}$ and $\mathcal{E}(10, z_i) = \emptyset$, where $i \in \mathbb{N}$;
- define $\mathcal{E}(b0, x) = \mathcal{E}(b0, y) = \mathcal{E}(10, z_i) = \emptyset$, where $i \in \mathbb{N}$;
- define $\mathcal{E}(u, s \cdot t) = \{\psi \mid \phi \to \psi \in \mathcal{E}(u, s) \text{ and } \phi \in \mathcal{E}(u, t)\}$, where $u \in \{10, b0\}$;
- define $\mathcal{E}(u, s + t) = \mathcal{E}(u, s) \cup \mathcal{E}(u, t)$, where $u \in \{10, b0\}$.

The fourth and fifth item guarantee that $\mathcal{E}$ satisfies **Application** and **Sum** condition on evidence function, respectively. The first item guarantees that $\mathcal{M}$ meets $\mathcal{C}$. Further, note that $v_{10}^{\mathcal{M}}(x : p) = v_{10}^{\mathcal{M}}(y : \neg p) = 1$, but $v_{10}^{\mathcal{M}}(t : q) = 0$ for all justification terms t.

4.3 Notations and Terminologies

Let $\mathcal{C}$ be a constant specification. A set Σ of formulas is said to be $\mathcal{C}$*-satisfiable*, if there are some model $\mathcal{M}$ that meets $\mathcal{C}$ and some world u in $\mathcal{M}$ such that $v_u^{\mathcal{M}}(\psi) \in \{1, b\}$ for all $\psi \in \Sigma$. A formula ϕ is *valid in a model* $\mathcal{M} = \langle W, R, \mathcal{E}, v\rangle$, if for all worlds u in $\mathcal{M}$, we have $v_u^{\mathcal{M}}(\phi) \in \{1, b\}$. A formula is $\mathcal{C}$*-valid*, if it is valid in every model that meets $\mathcal{C}$. Now, we define the semantical consequence relation. $\Sigma \Vdash_{\mathcal{C}} \phi$ iff for all models $\mathcal{M}$ that meet $\mathcal{C}$ and all worlds u in $\mathcal{M}$, if $v_u^{\mathcal{M}}(\psi) \in \{1, b\}$ for all $\psi \in \Sigma$, then $v_u^{\mathcal{M}}(\phi) \in \{1, b\}$.

Let Σ be a set of formulas. Define $v_u^{\mathcal{M}}[\Sigma] = \{v_u^{\mathcal{M}}(\phi) \mid \phi \in \Sigma\}$.

4.4 Basic Facts about $\mathsf{PJ_b}$

Here are some basic facts about $\mathsf{PJ_b}$. We divide them into two groups.

The following is the first group, which is about the paraconsistency of $\mathsf{PJ_b}$. To simplify the formulation, we use $\forall$, $\exists$ as the abbreviations of the meta-expressions, 'for all' and 'for some', respectively. s_1, s_2, t range over justification terms; x, y range over justification variables.

Fact 4.4 *Let $\mathcal{C}$ be a constant specification.*

(i) $(x : p \wedge y : \neg p)$ *is* $\mathcal{C}$*-satisfiable;*

(ii) $\exists s_1 \exists s_2 \exists \phi \exists \psi \forall t$, $\{s_1 : \phi, s_2 : \neg\phi\} \nVdash_{\mathcal{C}} t : \psi$*;*

(iii) $\forall s_1 \forall s_2 \forall \phi$*, if* ϕ *contains no justification term, then* $\exists \psi \forall t$, $\{s_1 : \phi, s_2 : \neg\phi\} \nVdash_{\mathcal{C}} t : \psi$*;*

(iv) $\forall s_1 \forall s_2 \forall \phi \forall \psi \forall t$*, if* ϕ *and* ψ *contain no justification term;* ϕ *and* ψ *share no propositional variable; and* ψ *is not* $\mathcal{C}$*-valid, then* $\{s_1 : \phi, s_2 : \neg\phi\} \nVdash_{\mathcal{C}} t : \psi$.

Proof. Please see the second half of Appendix A. □

The item (i) of Fact 4.4 says that an agent might explicitly commits to a contradiction based on two different pieces of evidence. The item (ii) intuitively means:

- for **some** two conflicting pieces s_1, s_2 of evidence, for **some** contradiction $\phi \wedge \neg\phi$, explicitly committing to $\phi \wedge \neg\phi$ based on s_1, s_2 does not imply that for all statement ψ, there is evidence t which forces us to explicitly commit to ψ.

The item (iii) expresses something stronger:

- for **every** two conflicting pieces s_1, s_2 of evidence, for **every** contradiction $\phi \wedge \neg\phi$ (with some restriction), explicitly committing to $\phi \wedge \neg\phi$ based on s_1, s_2 does not imply that for all statement ψ, there is evidence t which forces us to explicitly commit to ψ.

The item (iv) intuitively says:

- Given any two pieces s_1, s_2 of evidence, any contradiction $\phi \wedge \neg\phi$ (with some restriction) and any statement ψ (with some restriction), if $\phi\wedge\neg\phi$ is irrelevant to ψ (in the sense that $\phi \wedge \neg\phi$ and ψ share no propositional variable), then explicitly committing to $\phi \wedge \neg\phi$ based on s_1, s_2 does not imply that there is evidence t that forces us to explicitly commit to ψ.

And, note two things. First, (iv) implies (iii), but not vice versa; (iii) implies (ii), but not vice versa. Secondly, typically, in justification logic, we consider constant specifications that *entail internalization* in the sense that for all formulas ϕ, if ϕ is $\mathcal{C}$-valid, then $t : \phi$ is $\mathcal{C}$-valid for some justification term t. In standard two-valued justification logics, if the constant specification considered entails internalization, then all of (ii) – (iv) fail. However, in $\mathsf{PJ_b}$, (ii) – (iv) hold, even if the constant specification considered entails internalization.

Now, we move to the second group of facts about $\mathsf{PJ_b}$. They are good properties as a justification logic. Proofs for them are skipped.

Fact 4.5 *Let $\mathcal{C}$ be a constant specification.*

(i) *Modus Ponens preserves designated values;*

(ii) $\Sigma \cup \{\phi\} \Vdash_{\mathcal{C}} \psi$, *iff* $\Sigma \Vdash_{\mathcal{C}} \phi \to \psi$;

(iii) $s : (\phi \to \psi) \to (t : \phi \to (s \cdot t) : \psi)$ *is $\mathcal{C}$-valid;*

(iv) $s : \phi \to (s + t) : \phi$ *and* $t : \phi \to (s + t) : \phi$ *are $\mathcal{C}$-valid.*

The way we define the conditional gives us (i), (ii) and (iii). These three play important roles in proving quasi-realization theorem for $\langle \mathsf{PE_b}, \mathsf{PJ_b} \rangle$ (Theorem 6.3).

5 Axiomatic Soundness and Completeness of $\mathsf{PJ_b}$

In this section, we give an axiomatization of $\mathsf{PJ_b}$ and show its soundness and completeness. The machinery that we will use to prove completeness of $\mathsf{PJ_b}$ will help us prove quasi-realization theorem in the next section. [8]

5.1 An Axiomatic Proof System

Before giving the axiomatization, we list some setups. First, recall that $\phi \vee \psi$ is the abbreviation of $\neg(\neg\phi\wedge\neg\psi)$; $\phi \leftrightarrow \psi$ is the abbreviation of $(\phi \to \psi)\wedge(\psi \to \phi)$. Secondly, pick a specific propositional variable p. Define $\dot\neg\phi$ as the abbreviation of $\phi \to (\Box p \wedge \neg\Box p)$. Note that based on $\mathsf{PJ_b}$'s semantics, $\Box p \wedge \neg\Box p$ always gets value 0. Hence, $\dot\neg$ actually behaves like the classical negation. [9]

Let $\mathcal{C}$ be a constant specification. The $\mathsf{PJ_b}$ *axiomatic proof system w.r.t.* $\mathcal{C}$ is defined by the following axiom schemes and inference rules.

[8] Accurately speaking, it is the truth lemma for $\mathsf{PJ_b}$ – the key lemma to completeness – that will help us prove quasi-realization theorem.

[9] Roughly speaking, we can take $\mathsf{PJ_b}$ as having two negations: $\neg$ and $\dot\neg$. The former does not lead to explosion; the latter does. That is, $\phi \to (\neg\phi \to \psi)$ is not a valid scheme; however, $\phi \to (\dot\neg\phi \to \psi)$ is. $\neg$ helps us model paraconsistency (Fact 4.4). And, $\dot\neg$ helps us give a complete axiomatization.

A1. $\phi \to (\psi \to \phi)$	**A2.** $(\tau \to (\phi \to \psi)) \to ((\tau \to \phi) \to (\tau \to \psi))$
A3. $(\phi \vee \psi) \to (\psi \vee \phi)$	**A4.** $(\phi \vee \psi) \to ((\phi \to \tau) \to (\tau \vee \psi))$
A5. $\phi \vee \neg\phi$	**A6.** $\phi \vee \dot{\neg}\phi$
A7. $\phi \to (\dot{\neg}\phi \to \psi)$	**A8.** $\phi \to (\psi \to (\phi \wedge \psi))$
A9. $(\phi \wedge \psi) \to \phi$	**A10.** $(\phi \wedge \psi) \to (\psi \wedge \phi)$
A11. $\neg\phi \to \neg(\phi \wedge \psi)$	**A12.** $(\dot{\neg}\neg\phi \wedge \dot{\neg}\neg\psi) \leftrightarrow \dot{\neg}\neg(\phi \wedge \psi)$
A13. $\neg(\phi \to \psi) \leftrightarrow (\phi \wedge \neg\psi)$	**A14.** $\neg\neg\phi \leftrightarrow \phi$
A15. $\dot{\neg}\neg\neg\phi \leftrightarrow \dot{\neg}\phi$	**A16.** $s : (\phi \to \psi) \to (t : \phi \to (s \cdot t) : \psi)$
A17. $s : \phi \to (s + t) : \phi$	**A18.** $s : \phi \to (t + s) : \phi$
A19. $t : \phi \to (\neg t : \phi \to \psi)$	

R1.	*Modus Ponens*	$\phi, \phi \to \psi \Rightarrow \psi$
R2.	*$\mathcal{C}$ Axiom Necessitation*	$\Rightarrow c : \phi$ where $\phi \in \mathcal{C}(c)$ and either ϕ is an axiom **A1** – **A19** or ϕ is inferable using **R2**.

We use the notation $\vdash_{\mathcal{C}}$ to denote the proof-theoretic consequence relation with respect to a constant specification $\mathcal{C}$.

Following [6], when giving the axiomatization, actually we do not consider all constant specifications, but put further constraints. In this paper, we put the following constraint on constant specifications. A constant specification is said to be *strongly appropriate*, if every axiom has a justification constant; $c : \phi$ has a justification constant, whenever c is a justification constant for ϕ; and apart from these conditions, no other formulas have justification constants. This constraint helps, when we prove the axiomatic proof system is complete (Theorem 5.10) and satisfies internalization (Theorem 5.2).

We finish this subsections with two basic theorems. First, with the help of axiom schemes **A1** and **A2**, we can show the deduction theorem.

Theorem 5.1 *Let $\mathcal{C}$ be a constant specification. $\Sigma \cup \{\phi\} \vdash_{\mathcal{C}} \psi$, iff $\Sigma \vdash_{\mathcal{C}} \phi \to \psi$.*

Secondly, with the help of **A16**, **R2** and the strong appropriateness, we have the following theorem.

Theorem 5.2 (internalization) *Let $\mathcal{C}$ be a strongly appropriate constant specification. Then, if $\vdash_{\mathcal{C}} \phi$, then $\vdash_{\mathcal{C}} t : \phi$, for some justification term t.*

5.2 Soundness and Completeness

Proving the soundness of the axiomatic system is relatively simple, so we focus on completeness.

We prove completeness by the canonical model construction. One small difference from the classical case is that the notion of maximal consistent set is replaced by the notion of *$\mathcal{C}$-partition*. Before defining this notion, we need to introduce a notation: $\Gamma \vdash_{\mathcal{C}} \Gamma'$ holds, iff there are some formulas $\phi_1, \ldots, \phi_n$ in Γ' such that $\Gamma \vdash_{\mathcal{C}} \phi_1 \vee \cdots \vee \phi_n$. Now, we are ready to define the notion. Given a strongly appropriate constant specification $\mathcal{C}$ and two sets Γ, Γ' of formulas, $\langle \Gamma, \Gamma' \rangle$ is said to be a *$\mathcal{C}$-partition*, if the following three conditions hold:

(1) $\Gamma' \neq \emptyset$; (2) $\Gamma \nvdash_{\mathcal{C}} \Gamma'$;

(3) for all pairs $\langle \Omega, \Omega' \rangle$ with $\Gamma \subseteq \Omega$ and $\Gamma' \subseteq \Omega'$, if $\Gamma \subsetneq \Omega$ or $\Gamma' \subsetneq \Omega'$, then $\Omega \vdash_{\mathcal{C}} \Omega'$ holds.

By Zorn's lemma, we are able to prove the following lemma.

Lemma 5.3 *Let $\mathcal{C}$ be a strongly appropriate constant specification. Let $\langle \Sigma, \Sigma' \rangle$ be such that $\Sigma \nvdash_{\mathcal{C}} \Sigma'$ and $\Sigma' \neq \emptyset$. Then, there exists a $\mathcal{C}$-partition $\langle \Gamma, \Gamma' \rangle$ such that $\Sigma \subseteq \Gamma$ and $\Sigma' \subseteq \Gamma'$.*

The axiom scheme **A4** helps us prove the first item of the following fact. In addition, **A5** is for the second item; **A6** and **A7** are for the third item. Furthermore, the fourth item can be shown by using the first item.

Fact 5.4 *Let $\mathcal{C}$ be a strongly appropriate constant specification. Let $\langle \Gamma, \Gamma' \rangle$ be a $\mathcal{C}$-partition. Then, the following hold:*

(i) *for all formulas ϕ, either $\phi \in \Gamma$ or $\phi \in \Gamma'$.*

(ii) *for all formulas ϕ, either $\phi \in \Gamma$ and $\neg\phi \in \Gamma'$, or $\phi \in \Gamma'$ and $\neg\phi \in \Gamma$, or $\phi \in \Gamma$ and $\neg\phi \in \Gamma$.*

(iii) *for all formulas ϕ, either $\phi \in \Gamma$ and $\dot{\neg}\phi \in \Gamma'$, or $\dot{\neg}\phi \in \Gamma$ and $\phi \in \Gamma'$;*

(iv) *if $\Gamma \vdash_{\mathcal{C}} \phi$, then $\phi \in \Gamma$.*

Fact 5.4 is used, when we prove each item of the following fact. In addition, **A8**, **A9**, **A10** are for handling the first item of the following fact; **A10**, **A11**, **A12** are for the second item. Furthermore, **R1** is used in dealing with the third item; **A13** is for the fourth item.

Fact 5.5 *Let $\mathcal{C}$ be a strongly appropriate constant specification. Let $\langle \Gamma, \Gamma' \rangle$ be a $\mathcal{C}$-partition. Then, for all formulas ϕ and ψ, the following hold:*

(i) *$\phi \wedge \psi \in \Gamma$, iff $\phi \in \Gamma$ and $\psi \in \Gamma$.*

(ii) *$\neg(\phi \wedge \psi) \in \Gamma$, iff $\neg\phi \in \Gamma$ or $\neg\psi \in \Gamma$.*

(iii) *$\phi \to \psi \in \Gamma$, iff $\phi \in \Gamma'$ or $\psi \in \Gamma$.*

(iv) *$\neg(\phi \to \psi) \in \Gamma$, iff $\phi \in \Gamma$ and $\neg\psi \in \Gamma$.*

Definition 5.6 [canonical model] Let $\mathcal{C}$ be a strongly appropriate constant specification. The *canonical model in* $\mathsf{PJ_b}$ *w.r.t.* $\mathcal{C}$ is a four-tuple $\langle W, R, \mathcal{E}, v \rangle$, where

(i) W is the set of all $\mathcal{C}$-partitions.

(ii) for all $\langle \Gamma_1, \Gamma_1' \rangle$, $\langle \Gamma_2, \Gamma_2' \rangle \in W$, $\langle \Gamma_1, \Gamma_1' \rangle R \langle \Gamma_2, \Gamma_2' \rangle$ iff $\Gamma_1^\sharp \subseteq \Gamma_2$, where $\Gamma_1^\sharp$ is defined to be the following set:

$$\{\phi \mid t : \phi \in \Gamma_1, \text{ for some } t\}$$

(iii) for all $\langle \Gamma, \Gamma' \rangle \in W$, for all justification terms t, for all formulas ϕ, $\phi \in \mathcal{E}(\langle \Gamma, \Gamma' \rangle, t)$ iff $t : \phi \in \Gamma$.

(iv) for all $\langle \Gamma, \Gamma' \rangle \in W$ and all propositional variables p,

$$v_{\langle\Gamma,\Gamma'\rangle}(p) = \begin{cases} 1 & \text{if } p \in \Gamma \text{ and } \neg p \in \Gamma'; \\ b & \text{if } p \in \Gamma \text{ and } \neg p \in \Gamma; \\ 0 & \text{otherwise.} \end{cases}$$

Axiom schemes **A16**, **A17**, **A18** help us prove (i) of the following lemma. In addition, the inference rule **R2** helps us prove (ii).

Lemma 5.7 *Let $\mathcal{C}$ be a strongly appropriate constant specification.*

(i) *The canonical model in $\mathsf{PJ_b}$ w.r.t. $\mathcal{C}$ is a model in $\mathsf{PJ_b}$.*

(ii) *The canonical model in $\mathsf{PJ_b}$ w.r.t. $\mathcal{C}$ meets $\mathcal{C}$.*

The way we define the accessibility relation for the canonical model yields the following fact.

Fact 5.8 *Let $\mathcal{C}$ be a strongly appropriate constant specification. Let $\mathcal{M} = \langle W, R, \mathcal{E}, v\rangle$ be the canonical model in $\mathsf{PJ_b}$ w.r.t. $\mathcal{C}$. Let $\langle\Gamma,\Gamma'\rangle \in W$. Then, for all justification terms t and for all formulas ϕ, if $t : \phi \in \Gamma$, then for all $\langle\Omega,\Omega'\rangle \in W$ with $\langle\Gamma,\Gamma'\rangle R \langle\Omega,\Omega'\rangle$, it holds that $\phi \in \Omega$.*

With the help of Fact 5.5, Fact 5.8 and **A14**, **A15** and **A19**, we can show the truth lemma.

Lemma 5.9 (truth lemma for $\mathsf{PJ_b}$) *Let $\mathcal{C}$ be a strongly appropriate constant specification and $\mathcal{M} = \langle W, R, \mathcal{E}, v\rangle$ be the canonical model in $\mathsf{PJ_b}$ w.r.t. $\mathcal{C}$. Then, for all formulas ϕ, for all $\langle\Gamma,\Gamma'\rangle \in W$, the following hold:*

(i) *$\phi \in \Gamma$ and $\neg\phi \in \Gamma'$, iff $v^{\mathcal{M}}_{\langle\Gamma,\Gamma'\rangle}(\phi) = 1$.*

(ii) *$\phi \in \Gamma$ and $\neg\phi \in \Gamma$, iff $v^{\mathcal{M}}_{\langle\Gamma,\Gamma'\rangle}(\phi) = b$.*

(iii) *$\phi \in \Gamma'$ and $\neg\phi \in \Gamma$, iff $v^{\mathcal{M}}_{\langle\Gamma,\Gamma'\rangle}(\phi) = 0$.*

With the help of Lemma 5.3 and Lemma 5.9, we can prove completeness.

Theorem 5.10 (completeness of $\mathsf{PJ_b}$) *Let $\mathcal{C}$ be a strongly appropriate constant specification. If $\Sigma \Vdash_{\mathcal{C}} \phi$, then $\Sigma \vdash_{\mathcal{C}} \phi$.*

6 Quasi-Realization Theorem for $\langle\mathsf{PE_b}, \mathsf{PJ_b}\rangle$

We follow Fitting's non-constructive way [6] to prove quasi-realization theorem for $\langle\mathsf{PE_b}, \mathsf{PJ_b}\rangle$. This theorem establishes an embedding from $\mathsf{PE_b}$ into $\mathsf{PJ_b}$. In proving quasi-realization theorem for $\langle\mathsf{PE_b}, \mathsf{PJ_b}\rangle$, the only part that is not totally obviously suggested by Fitting's proof in [6] is how to handle the negation.

Before going to proofs, here are some setups. Let ϕ be a formula in $\mathsf{PE_b}$. Assume that ϕ is fixed for the rest of the section. In the following, for the sake of simplicity, when we talk about some *subformula* of ϕ, we actually mean some *occurrence* of this subformula of ϕ. In addition, *positive* subformulas and *negative* subformulas are defined as usual. More specifically, ϕ itself is a positive subformula of ϕ. Given a subformula α of ϕ, if α is of the form $\Box\beta$ or $\beta \wedge \gamma$,then the polarity of β and the polarity of γ are the same as α. If α is of the form $\beta \to \gamma$, then the polarity of β is different from α and the polarity of

γ is the same as α. If α is of the form $\neg\beta$, then the polarity of β is different from α. Furthermore, A is any assignment of a justification variable to each subformula of ϕ of the form $\Box\alpha$ that is at the negative position. It is assumed that A is one-one. Relative to A, we define one mapping π_A as follows.

Definition 6.1 π_A assigns a set of $\mathsf{PJ_b}$ formulas to each subformula of ϕ:

(i) if p is an atomic subformula of ϕ, then $\pi_A(p) = \{p\}$;

(ii) if $\neg\alpha$ is a subformula of ϕ,
$\pi_A(\neg\alpha) = \{\neg\alpha' \mid \alpha' \in \pi_A(\alpha)\}$;

(iii) if $\alpha * \beta$ is a subformula of ϕ,
$\pi_A(\alpha * \beta) = \{\alpha' * \beta' \mid \alpha' \in \pi_A(\alpha) \text{ and } \beta' \in \pi_A(\beta)\}$, where $* \in \{\wedge, \rightarrow\}$;

(iv) if $\Box\alpha$ is a negative subformula of ϕ,
$\pi_A(\Box\alpha) = \{x : \alpha' \mid A(\Box\alpha) = x \text{ and } \alpha' \in \pi_A(\alpha)\}$;

(v) if $\Box\alpha$ is a positive subformula of ϕ,
$\pi_A(\Box\alpha)= \{t : (\alpha_1 \vee \cdots \vee \alpha_n) \mid \alpha_1, \ldots, \alpha_n \in \pi_A(\alpha)$ and t is any justification term$\}$.

Finally, for the rest of the section, we fix a constant specification $\mathcal{C}$. Assume $\mathcal{C}$ is strongly appropriate. Let $\mathcal{M} = \langle W, R, \mathcal{E}, v\rangle$ be the canonical model w.r.t $\mathcal{C}$ in $\mathsf{PJ_b}$. Define a $\mathsf{PE_b}$ model $\mathcal{N} = \langle W, R, v\rangle$. So, $\mathcal{N}$ is $\mathcal{M}$ dropping the evidence function $\mathcal{E}$. In addition, recall the following notation. Given a world u in $\mathcal{M}$ and a set Σ of formulas, $v^{\mathcal{M}}_u[\Sigma] = \{v^{\mathcal{M}}_u(\phi) \mid \phi \in \Sigma\}$.

Now, we are ready proving things. In [6], its Proposition 7.7 is the key to the quasi-realization theorem for $\langle \mathsf{S4}$, the *Logic of Proofs (without Plus)*$\rangle$. Here, the following lemma is the key to the quasi-realization theorem for $\langle \mathsf{PE_b}, \mathsf{PJ_b}\rangle$.

Lemma 6.2 *For all formulas ψ in $\mathsf{PE_b}$, for all $\langle\Gamma,\Gamma'\rangle \in W$, the following hold:*

(i) *if ψ is a positive subformula of ϕ and $v^{\mathcal{M}}_{\langle\Gamma,\Gamma'\rangle}[\pi_A(\psi)] \subseteq \{b, 0\}$, then $v^{\mathcal{N}}_{\langle\Gamma,\Gamma'\rangle}(\psi) \in \{b, 0\}$.*

(ii) *if ψ is a positive subformula of ϕ and $v^{\mathcal{M}}_{\langle\Gamma,\Gamma'\rangle}[\pi_A(\psi)] = \{0\}$, then $v^{\mathcal{N}}_{\langle\Gamma,\Gamma'\rangle}(\psi) = 0$.*

(iii) *if ψ is a negative subformula of ϕ and $v^{\mathcal{M}}_{\langle\Gamma,\Gamma'\rangle}[\pi_A(\psi)] \subseteq \{1, b\}$, then $v^{\mathcal{N}}_{\langle\Gamma,\Gamma'\rangle}(\psi) \in \{1, b\}$.*

(iv) *if ψ is a negative subformula of ϕ and $v^{\mathcal{M}}_{\langle\Gamma,\Gamma'\rangle}[\pi_A(\psi)] = \{1\}$, then $v^{\mathcal{N}}_{\langle\Gamma,\Gamma'\rangle}(\psi) = 1$.*

Proof. Please see Appendix B. $\Box$

The following three paragraphs might let readers see the rough shape of the proof for Lemma 6.2.

First of all, I would like to address one small difference between the statement of [6]'s Proposition 7.7 and the statement of Lemma 6.2 (of this paper). If looking at the statement of [6]'s Proposition 7.7, readers will find: under this classical two-valued context, positive subformulas correspond to and only correspond to the non-designated value 0, and negative subformulas correspond to and only correspond to the designated value 1. However, the correspondence between polarity and truth value is less perfect in the statement of Lemma 6.2.. More specifically, in the item (i) of Lemma 6.2, positive subformulas correspond to $\{b, 0\}$, so do not only correspond to the non-designated value. Besides, in the item (iv) of of Lemma 6.2, negative subformulas correspond to $\{1\}$, so do not correspond to all designated values.

The root that leads to the small difference just described is the non-classical negation used by both $\mathsf{PE_b}$ and $\mathsf{PJ_b}$. First, consider a simplified version of Lemma 6.2 that does not contain the items (i) and (iv). Actually, this simplified version of Lemma 6.2 is already sufficient for helping us prove the quasi-realization theorem for $\langle \mathsf{PE_b}, \mathsf{PJ_b} \rangle$. Second, the reason why we prove the more complicated statement, rather than the simplified version is that without the items (i) and (iv), the inductive proof will get stuck at the case of negation. Hence, the items (i) and (iv) of Lemma 6.2 are not redundant. In short, to handle the non-classical negation, we need Lemma 6.2's item (i) and item (iv), which breaks the perfect correspondence between polarity and truth value.

Now, we turn to the similarity between [6]'s proof for its Proposition 7.7 and the proof for Lemma 6.2. Both are proofs by induction on formulas. Except the case of negation, these two inductive proofs are very similar. The most complicated part of [6]'s proof for its Proposition 7.7 is the case of positive necessity, which relies on: the *Logic of Proofs* has Modus Ponens, deduction theorem and the validity of $s : (\phi \to \psi) \to (t : \phi \to (s \cdot t) : \psi)$. The way we interpret the conditional of $\mathsf{PJ_b}$ gives us these three. Therefore, we can also handle the case of positive necessity, when proving Lemma 6.2.

Applying Lemma 6.2 (and the truth lemma for $\mathsf{PJ_b}$), we can then prove the quasi-realization theorem.

Theorem 6.3 *Let $\mathcal{C}$ be a strongly appropriate constant specification. Let ϕ be a formula in $\mathsf{PE_b}$. If ϕ is valid in $\mathsf{PE_b}$, then there are $\phi_1, \dots, \phi_n \in \pi_A(\phi)$ such that $\phi_1 \vee \dots \vee \phi_n$ is $\mathcal{C}$-valid in $\mathsf{PJ_b}$.*

7 Conclusion

Based on $\mathsf{PE_b}$, we have constructed a paraconsistent justification logic system $\mathsf{PJ_b}$ by giving its Fitting models. $\mathsf{PJ_b}$ is paraconsistent in the sense that it is not the case that for all justification terms s_1, s_2, there exists a justification term t such that $(s_1 : \phi \wedge s_2 : \neg\phi) \to t : \psi$ is a valid scheme. Actually, $\mathsf{PJ_b}$ satisfies stronger paraconsistent properties (Fact 4.4's (iii), (iv)). Informally speaking, $\mathsf{PJ_b}$ reflects the idea: explicitly committing to a contradiction for two conflicting pieces of evidence does not imply that for all statement ψ, we have a corresponding evidence that forces us to commit to ψ. In addition, we motivate $\mathsf{PJ_b}$ from the angle of belief revision. We have argued that to better model belief revision, we might need to be able to model (1) inconsistent epistemic states and (2) the agent's justification structure. $\mathsf{PJ_b}$ is able to handle these two, so might help us better model belief revision. Furthermore, the main technical result of the paper is quasi-realization theorem for $\langle \mathsf{PE_b}, \mathsf{PJ_b} \rangle$, which establishes an embedding from $\mathsf{PE_b}$ into $\mathsf{PJ_b}$.

Possible future work: First of all, in standard justification logics, there is an algorithmic conversion from quasi-realization to realization [7], which is a simpler form of embedding from a modal logic into a justification logic. Hence, one natural next step is to check whether this conversion also works for $\langle \mathsf{PE_b}, \mathsf{PJ_b} \rangle$, so we can also have realization for $\langle \mathsf{PE_b}, \mathsf{PJ_b} \rangle$. Secondly, $\mathsf{PJ_b}$ performs nicely in the technical aspect because of the way we define the conditional. To avoid ad-hoc-ness, either a good story for motivating the conditional used by $\mathsf{PJ_b}$ should be further provided, or we should change to a conditional that is well-motivated. However, if we go for the second option, for the project to succeed, it is better that the resulting system still performs nicely in the technical aspect. Finally, we motivate paraconsistent justification logic from the angle of belief revision. Therefore, to finish the whole story, *dynamic* paraconsistent justification logic for belief revision should be developed.

Appendix

A Proofs for Fact 3.3 and Fact 4.4

Proof. [Fact 3.3] Proofs for (v)–(vii) are skipped. Example 3.2 provides a model for showing (i) and (ii). (iv) implies (iii), so for the rest, we focus on proving (iv).

Let ϕ, ψ be such that both ϕ, ψ contain no $\Box$; ϕ and ψ share no propositional variable; and ψ is not valid. Our goal is to construct a model showing $(\Box\phi \wedge \Box\neg\phi) \rightarrow \Box\psi$ is invalid.

Let $p_1, \ldots, p_n$ be all of the propositional variables occurring in ϕ; let $q_1, \ldots, p_m$ be those occurring in ψ. Let $(\mathcal{N}', u_3)$ be such that $\mathcal{N}' = \langle W', R', v' \rangle$ is a $\mathsf{PE_b}$ model, $u_3 \in W'$ and $v'^{\mathcal{N}'}_{u_3}(\psi) = 0$. Note that such a $(\mathcal{N}', u_3)$ exists, since ψ is assumed to be invalid. Define a $\mathsf{PE_b}$ model $\mathcal{N} = \langle W, R, v \rangle$, where $W = \{u_1, u_2\}$; $R = \{\langle u_1, u_2 \rangle\}$; $v^{\mathcal{N}}_{u_2}(p_i) = b$ for each p_i; $v^{\mathcal{N}}_{u_2}(q_i) = v'^{\mathcal{N}'}_{u_3}(q_i)$ for each q_i. How v assigns values to other propositional variables at u_2 and how v assigns values (to all propositional variables) at u_1 could be arbitrary.

Claim that $v^{\mathcal{N}}_{u_1}((\Box\phi \wedge \Box\neg\phi) \rightarrow \Box\psi) = 0$. Here is a proof for the claim. First, since ϕ contains no $\Box$, by the way $\mathsf{PE_b}$ interprets propositional connectives, it holds that $v^{\mathcal{N}}_{u_2}(\phi) = v^{\mathcal{N}}_{u_2}(\neg\phi) = b$. Secondly, because ψ contains no $\Box$, ψ's getting value 0 at u_3 is totally determined by how v' assign values to $q_1, \ldots, q_m$ at u_3. Since at u_2, v mimics how v' assigns values to $q_1, \ldots, q_m$ at u_3, it holds that $v^{\mathcal{N}}_{u_2}(\psi) = 0$. Third, by the previous two points, it follows that $v^{\mathcal{N}}_{u_1}(\Box\phi \wedge \Box\neg\phi) = 1$ and $v^{\mathcal{N}}_{u_1}(\Box\psi) = 0$. Therefore, the claim is proved. $\Box$

Proof. [Fact 4.4] Example 4.3 provides a model for showing (i) and (ii). (iv) implies (iii), so for the rest, we focus on proving (iv).

Let s_1, s_2, t be three justification terms. Assume that ϕ, ψ satisfy conditions listed in the antecedent of the conditional occurring in (iv). Our goal is to construct a model showing $\{s_1 : \phi, s_2 : \neg\phi\} \not\Vdash_{\mathcal{C}} t : \psi$.

Let $p_1, \ldots, p_n$ be all of the propositional variables occurring in ϕ; let

$q_1, \ldots, p_m$ be those occurring in ψ. Let $(\mathcal{N}', u_3)$ be such that $\mathcal{N}' = \langle W', R', \mathcal{E}', v' \rangle$ is a $\mathsf{PJ_b}$ model, $u_3 \in W'$ and $v'^{\mathcal{N}'}_{u_3}(\psi) = 0$. Note that such a $(\mathcal{N}', u_3)$ exists, since ψ is assumed to be not $\mathcal{C}$-valid. Define a $\mathsf{PJ_b}$ model $\mathcal{N} = \langle W, R, \mathcal{E}, v \rangle$, where $W = \{u_1, u_2\}$; $R = \{\langle u_1, u_2 \rangle\}$; $\mathcal{E}(u_1, j) = \mathcal{E}(u_2, j) = \{\tau \mid \tau \text{ is a } \mathsf{PJ_b} \text{ formula}\}$, for all justification terms j; $v^{\mathcal{N}}_{u_2}(p_i) = b$ for each p_i; $v^{\mathcal{N}}_{u_2}(q_i) = v'^{\mathcal{N}'}_{u_3}(q_i)$ for each q_i.

First, by the definition of $\mathcal{E}$ and a reasoning similar to the one we employ in proving (iv) of Fact 3.3, we can show that $v^{\mathcal{N}}_{u_1}(s_1 : \phi \wedge s_2 : \neg\phi) = 1$ and $v^{\mathcal{N}}_{u_1}(t : \psi) = 0$. Secondly, by the definition of $\mathcal{E}$, $\mathcal{M}$ meets any constant specification. By the previous two points, it follows that $\{s_1 : \phi, s_2 : \neg\phi\} \not\Vdash_{\mathcal{C}} t : \psi$. □

B Proof for Lemma 6.2

Proof. [Lemma 6.2] We prove the lemma by induction on formulas. The atomic case is trivial, so is skipped.

Let α and β be two formulas in $\mathsf{PE_b}$. Assume the following as the induction hypothesis (**IH**): for all $\langle \Gamma, \Gamma' \rangle \in W$,

- if α is a positive subformula of ϕ and $v^{\mathcal{M}}_{\langle \Gamma, \Gamma' \rangle}[\pi_A(\alpha)] \subseteq \{b, 0\}$, then $v^{\mathcal{N}}_{\langle \Gamma, \Gamma' \rangle}(\alpha) \in \{b, 0\}$.
- if α is a positive subformula of ϕ and $v^{\mathcal{M}}_{\langle \Gamma, \Gamma' \rangle}[\pi_A(\alpha)] = \{0\}$, then $v^{\mathcal{N}}_{\langle \Gamma, \Gamma' \rangle}(\alpha) = 0$.
- if α is a negative subformula of ϕ and $v^{\mathcal{M}}_{\langle \Gamma, \Gamma' \rangle}[\pi_A(\alpha)] \subseteq \{1, b\}$, then $v^{\mathcal{N}}_{\langle \Gamma, \Gamma' \rangle}(\alpha) \in \{1, b\}$.
- if α is a negative subformula of ϕ and $v^{\mathcal{M}}_{\langle \Gamma, \Gamma' \rangle}[\pi_A(\alpha)] = \{1\}$, then $v^{\mathcal{N}}_{\langle \Gamma, \Gamma' \rangle}(\alpha) = 1$.
- for ψ, we also assume similar four items of conditions as part of **IH**.
- case 1. $\neg\alpha$
  Let $\langle \Gamma, \Gamma' \rangle \in W$.
  · case 1(a). Positive Negation (I)
  The goal conditional we want to show is: if $\neg\alpha$ is a positive subformula of ϕ and $v^{\mathcal{M}}_{\langle \Gamma, \Gamma' \rangle}[\pi_A(\neg\alpha)] \subseteq \{b, 0\}$, then $v^{\mathcal{N}}_{\langle \Gamma, \Gamma' \rangle}(\neg\alpha) \in \{b, 0\}$. The third item of **IH** helps us prove this.

  First, the statement that $\neg\alpha$ is a positive subformula of ϕ implies the statement that α is a negative subformula of ϕ.

  Secondly, the statement that $v^{\mathcal{M}}_{\langle \Gamma, \Gamma' \rangle}[\pi_A(\neg\alpha)] \subseteq \{b, 0\}$ is equivalent to the statement that $v^{\mathcal{M}}_{\langle \Gamma, \Gamma' \rangle}[\pi_A(\alpha)] \subseteq \{1, b\}$.

  Assume that $\neg\alpha$ is a positive subformula of ϕ and $v^{\mathcal{M}}_{\langle \Gamma, \Gamma' \rangle}[\pi_A(\neg\alpha)] \subseteq \{b, 0\}$. By the first and the second point, the assumption implies that α is a negative subformula of ϕ and $v^{\mathcal{M}}_{\langle \Gamma, \Gamma' \rangle}[\pi_A(\alpha)] \subseteq \{1, b\}$. Then, by the third item of **IH**, we can derive that $v^{\mathcal{N}}_{\langle \Gamma, \Gamma' \rangle}(\alpha) \in \{1, b\}$. Therefore, $v^{\mathcal{N}}_{\langle \Gamma, \Gamma' \rangle}(\neg\alpha) \in \{b, 0\}$.
  · case 1(b). Positive Negation (II)

The goal conditional we want to show is: if $\neg\alpha$ is a positive subformula of ϕ and $v^{\mathcal{M}}_{\langle\Gamma,\Gamma'\rangle}[\pi_A(\neg\alpha)] = \{0\}$, then $v^{\mathcal{N}}_{\langle\Gamma,\Gamma'\rangle}(\neg\alpha) = 0$. By the fourth item of **IH** and reasoning similar to the previous case, this can be shown.

· case 1(c). Negative Negation (I)
The goal conditional we want to show is: if $\neg\alpha$ is a negative subformula of ϕ and $v^{\mathcal{M}}_{\langle\Gamma,\Gamma'\rangle}[\pi_A(\neg\alpha)] \subseteq \{1, b\}$, then $v^{\mathcal{N}}_{\langle\Gamma,\Gamma'\rangle}(\neg\alpha) \in \{1, b\}$. By the first item of **IH** and reasoning similar to case 1-(a), this can be shown.

· case 1(d). Negative Negation (II)
The goal conditional we want to show is: if $\neg\alpha$ is a negative subformula of ϕ and $v^{\mathcal{M}}_{\langle\Gamma,\Gamma'\rangle}[\pi_A(\neg\alpha)] = \{1\}$, then $v^{\mathcal{N}}_{\langle\Gamma,\Gamma'\rangle}(\neg\alpha) = 1$. By the second item of **IH** and reasoning similar to case 1-(a), this can be shown.

- case 2. $\Box\alpha$
Let $\langle\Gamma, \Gamma'\rangle \in W$.

· case 2(a). Positive Necessity (I)
Assume that $\Box\alpha$ is a positive subformula of ϕ and $v^{\mathcal{M}}_{\langle\Gamma,\Gamma'\rangle}[\pi_A(\Box\alpha)] \subseteq \{b, 0\}$. Note that here α is a positive subformula of ϕ. Our goal is to show that $v^{\mathcal{N}}_{\langle\Gamma,\Gamma'\rangle}(\Box\alpha) \in \{b, 0\}$. This case is similar the the next case, that is, the case 2-(b). Hence, we skip this case.

· case 2(b). Positive Necessity (II)
Assume that $\Box\alpha$ is a positive subformula of ϕ and $v^{\mathcal{M}}_{\langle\Gamma,\Gamma'\rangle}[\pi_A(\Box\alpha)] = \{0\}$. Our goal is to show that $v^{\mathcal{N}}_{\langle\Gamma,\Gamma'\rangle}(\Box\alpha) = 0$. Note that here α is a positive subformula of ϕ.

Claim that $\Gamma^\sharp \nvdash_{\mathcal{C}} \pi_A(\alpha)$ holds. Here is the proof for the claim. Suppose not. Then, there exist some finite subset $\Sigma \subseteq \Gamma^\sharp$ and some formulas $V_1, \ldots, V_m \in \pi_A(\alpha)$ such that $\Sigma \vdash_{\mathcal{C}} V_1 \vee \cdots \vee V_m$. Either $\Sigma \neq \emptyset$ or $\Sigma = \emptyset$.

In this paragraph, we consider the situation that $\Sigma \neq \emptyset$. Say $\Sigma = \{U_1, \ldots, U_n\}$. Then, it holds that $\{U_1, \ldots, U_n\} \vdash_{\mathcal{C}} V_1 \vee \cdots \vee V_m$. First, by the deduction theorem (Theorem 5.1), we have that $\vdash_{\mathcal{C}} U_1 \to (U_2 \to \cdots (U_n \to (V_1 \vee \cdots \vee V_m) \cdots)$. Let Λ stand for $U_1 \to (U_2 \to \cdots (U_n \to (V_1 \vee \cdots \vee V_m) \cdots)$. So, we have that $\vdash_{\mathcal{C}} \Lambda$. By internalization theorem (Theorem 5.2) and the assumption that $\mathcal{C}$ is strongly appropriate, it holds that $\vdash_{\mathcal{C}} t : \Lambda$, for some justification term t. Secondly, since $\{U_1, \ldots, U_n\} \subseteq \Gamma^\sharp$, we have that $\{s_1 : U_1, \ldots, s_n : U_n\} \subseteq \Gamma$, for some $s_1, \ldots, s_n$. Third, because our axiomatization of $\mathsf{PJ_b}$ has Modus Ponens (**R1**) and takes $j : (\tau_1 \to \tau_2) \to (j' : \tau_1 \to (j \cdot j') : \tau_2)$ as an axiom scheme (**A16**), it holds that $\{t : \Lambda, s_1 : U_1, \ldots, s_n : U_n\} \vdash_{\mathcal{C}} ((\ldots(t \cdot s_1) \cdot s_2 \ldots) \cdot s_n) : (V_1 \vee \cdots \vee V_m)$. Fourth, by the previous three points, we can conclude that $\Gamma \vdash_{\mathcal{C}} ((\ldots(t \cdot s_1) \cdot s_2 \ldots) \cdot s_n) : (V_1 \vee \cdots \vee V_m)$. Fifth, since $V_1, \ldots, V_m \in \pi_A(\alpha)$, it holds that $\{j : (V_1 \vee \cdots \vee V_m) \mid j \text{ is a term}\}$ is a subset of $\pi_A(\Box\alpha)$. Therefore, $(\ldots(t \cdot s_1) \cdot s_2 \ldots) \cdot s_n) : (V_1 \vee \cdots \vee V_m)$ is in $\pi_A(\Box\alpha)$. By the assumption that $v^{\mathcal{M}}_{\langle\Gamma,\Gamma'\rangle}[\pi_A(\Box\alpha)] = \{0\}$ and the truth lemma, we have that $((\ldots(t \cdot s_1) \cdot s_2 \ldots) \cdot s_n) : (V_1 \vee \cdots \vee V_m)$ is in Γ'. Sixth, by the fourth and fifth points, we can conclude that $\Gamma \vdash_{\mathcal{C}} \Gamma'$. This contradicts that Γ, Γ' is a $\mathcal{C}$-partition.

The situation that $\Sigma = \emptyset$ can be handled in a similar way, so is skipped.

By the claim we just prove and Lemma 5.3, there exists a $\mathcal{C}$-partition $\langle \Delta, \Delta' \rangle$ such that $\Gamma^\sharp \subseteq \Delta$ and $\pi_A(\alpha) \subseteq \Delta'$. Then, we can show that $v^{\mathcal{N}}_{\langle \Gamma, \Gamma' \rangle}(\Box\alpha) = 0$. Here is the proof. First, since $\Gamma^\sharp \subseteq \Delta$, we have that $\langle \Gamma, \Gamma' \rangle R \langle \Delta, \Delta' \rangle$. And, because $\mathcal{N}$ and $\mathcal{M}$ share the same R, in $\mathcal{N}$ it also holds that $\langle \Gamma, \Gamma' \rangle R \langle \Delta, \Delta' \rangle$. Second, since $\pi_A(\alpha) \subseteq \Delta'$, by the truth lemma it holds that $v^{\mathcal{M}}_{\langle \Delta, \Delta' \rangle}[\pi_A(\alpha)] = \{0\}$. Note that α is positive subformula of $\Box\alpha$. By **IH**, it follows that $v^{\mathcal{N}}_{\langle \Delta, \Delta' \rangle}(\alpha) = 0$. Third, by the previous two points, we can conclude that $v^{\mathcal{N}}_{\langle \Gamma, \Gamma' \rangle}(\Box\alpha) = 0$.

· case 2(c). Negative Necessity (I)
Assume that $\Box\alpha$ is a negative subformula of ϕ and $v^{\mathcal{M}}_{\langle \Gamma, \Gamma' \rangle}[\pi_A(\Box\alpha)] \subseteq \{1, b\}$. Our goal is to show that $v^{\mathcal{N}}_{\langle \Gamma, \Gamma' \rangle}(\Box\alpha) \in \{1, b\}$. Note that here α is a negative subformula of ϕ. This case is similar to the corresponding case in [6].

· case 2(d). Negative Necessity (II)
Assume that $\Box\alpha$ is a negative subformula of ϕ and $v^{\mathcal{M}}_{\langle \Gamma, \Gamma' \rangle}[\pi_A(\Box\alpha)] = \{1\}$. Note that here α is a negative subformula of ϕ. Our goal is to show that $v^{\mathcal{N}}_{\langle \Gamma, \Gamma' \rangle}(\Box\alpha) = 1$. This case is similar to the previous case, that is, the case 2-(c). Therefore, we skip it.

The case of $\to$ is similar to [6]; the case of $\wedge$ is similar to the case of $\to$. $\Box$

References

[1] Artemov, S., *Explicit provability and constructive semantics*, Bulletin of Symbolic logic **7** (2001), pp. 1–36.

[2] Avron, A., *On an implication connective of RM*, Notre Dame Journal of Formal Logic **27** (1986), pp. 201–209.

[3] Batens, D., *Paraconsistent extensional propositional logics*, Logique et Analyse Louvain **23** (1980), pp. 195–234.

[4] Beall, J., T. Forster and J. Seligman, *A note on freedom from detachment in the logic of paradox*, Notre Dame Journal of Formal Logic **54** (2013), pp. 15–20.

[5] Fagin, R. and J. Y. Halpern, *Belief, awareness, and limited reasoning*, Artificial intelligence **34** (1987), pp. 39–76.

[6] Fitting, M., *The logic of proofs, semantically*, Annals of Pure and Applied Logic **132** (2005), pp. 1–25.

[7] Fitting, M., *Realization implemented*, Technical Report TR-2013005, CUNY Ph.D Program in Computer Science (May 2013).

[8] Girard, P. and K. Tanaka, *Paraconsistent dynamics* (February 2014), manuscript, 15 pages.

[9] Hansson, S. O., *Ten philosophical problems in belief revision*, Journal of Logic and Computation **13** (2003), pp. 37–49.

[10] Jaspars, J. O., *Logical omniscience and inconsistent belief*, in: *Diamonds and Defaults*, Springer, 1993 pp. 129–146.

[11] Lurie, J., *Non-applicative justification logic* (June 2014), manuscript, 12 pages.

[12] Restall, G., *Ways things can't be*, Notre Dame Journal of Formal Logic **38** (1997), pp. 583–596.

[13] Van der Hoek, W. and J.-J. C. Meyer, *Possible logics for belief*, Logique et analyse **127** (1989), pp. 177–194.