Institute of Philosophy
Russian Academy of Sciences

LOGICAL INVESTIGATIONS

Volume 27. Number 1

Special issue: Negation
(dedicated to the memory of Prof. J. Michael Dunn)

Edited by
H. Wansing, G. Olkhovikov, H. Omori

Moscow
2021

Федеральное государственное бюджетное учреждение науки
Институт философии Российской академии наук

ЛОГИЧЕСКИЕ ИССЛЕДОВАНИЯ

Том 27. Номер 1

Специальный выпуск: Отрицание
(посвящен памяти профессора Дж. М. Данна)

Приглашенные редакторы:
Г. Ванзинг, Г. Ольховиков, Х. Омори

Москва
2021

ISSN 2074-1472 (Print)
ISSN 2413-2713 (Online)

Logical Investigations

Scientific-Theoretical Journal

2021. Volume 27. Number 1

Publisher: Institute of Philosophy, Russian Academy of Sciences

Frequency: 2 times per year

First issue: 1993; the journal is a redesigned continuation of the annual *Logical Investigations* that has been published since 1993 till 2015

The journal is registered with the Federal Service for Supervision of Communications, Information Technology, and Mass Media (Roskomnadzor). The Mass Media Registration Certificate No. FS77-61228 on April 3, 2015

Abstracting and indexing: *Scopus*, *Zentralblatt MATH*, *Mathematical Reviews*, *Ulrich's Periodicals Directory*, *EBSCOhost (Philosopher's Index with Full Text)*

The journal is included in the list of peer-reviewed scientific editions acknowledged by the Higher Attestation Commission of the Ministry of Education and Science of the Russian Federation

Subscription index in the United Catalogue *The Russian Press* is 42046

All materials published in *Logical Investigations* undergo peer review process

Editorial address: 12/1 Goncharnaya St., Moscow 109240, Russian Federation
Tel.: +7 (495) 697-96-65; **e-mail:** logicalinvestigations@gmail.com
Website: https://logicalinvestigations.ru

ISSN 2074-1472 (Print)
ISSN 2413-2713 (Online)

Логические исследования

Научно-теоретический журнал

2021. Том 27. Номер 1

Учредитель и издатель: Федеральное государственное бюджетное учреждение науки Институт философии Российской академии наук

Периодичность: 2 раза в год

Выходит с 1993 г.; журнал является прямым продолжением ежегодника «Логические исследования», издававшегося с 1993 по 2015 г.

Журнал зарегистрирован Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации СМИ: ПИ № ФС77-61228 от 03 апреля 2015 г.

Журнал реферируется и индексируется: *Scopus*, *Mathematical Reviews*, *Zentralblatt MATH*, *Ulrich's Periodicals Directory*, *РИНЦ*, *EBSCOhost (Philosopher's Index with Full Text)*

Журнал включен в Перечень российских рецензируемых научных журналов, рекомендованных ВАК, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученых степеней доктора и кандидата наук (группа научных специальностей «09.00.00 – философские науки»)

Подписной индекс в Объединенном каталоге «Пресса России» — 42046

Публикуемые материалы прошли процедуру рецензирования и экспертного отбора

Адрес редакции: Российская Федерация, 109240, г. Москва, ул. Гончарная, д. 12, стр. 1, оф. 426

Тел.: +7 (495) 697-96-65; **e-mail:** `logicalinvestigations@gmail.com`

Сайт: `https://logicalinvestigations.ru`

TABLE OF CONTENTS

В НОМЕРЕ

Логические исследования
2021. Т. 27. № 1. С. 9–19
УДК 16+510.6

Logical Investigations
2021, Vol. 27, No. 1, pp. 9–19
DOI: 10.21146/2074-1472-2021-27-1-9-19

Heinrich Wansing, Grigory Olkhovikov, Hitoshi Omori

Questions to Michael Dunn

Heinrich Wansing
Ruhr University Bochum,
Universitätsstraße 150, 44801 Bochum, Germany.
E-mail: Heinrich.Wansing@rub.de

Grigory Olkhovikov
Ruhr University Bochum,
Universitätsstraße 150, 44801 Bochum, Germany.
E-mail: Grigory.Olkhovikov@rub.de

Hitoshi Omori
Ruhr University Bochum,
Universitätsstraße 150, 44801 Bochum, Germany.
E-mail: Hitoshi.Omori@rub.de

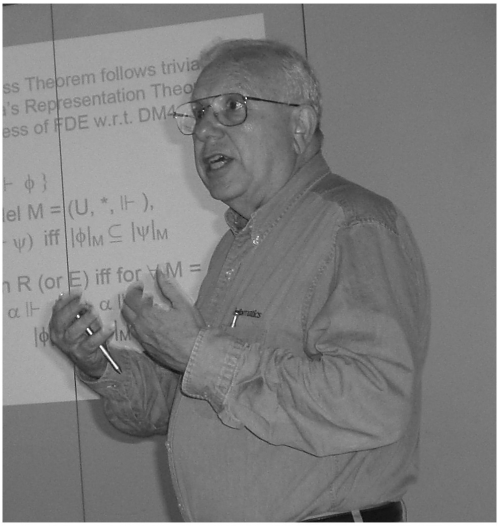
Abstract: We present nine questions related to the concept of negation and, in passing, we refer to connections with the essays in this special issue. The questions were submitted to one of the most eminent logicians who contributed to the theory of negation, Prof. (Jon) Michael Dunn, but, unfortunately, Prof. Dunn was no longer able to answer them. Michael Dunn passed away on 5 April 2021, and the present special issue of *Logical Investigations* is dedicated to his memory. The questions concern (i) negation-related topics that have particularly interested Michael Dunn or to which he has made important contributions, (ii) some controversial aspects of the logical analysis of the concept of negation, or (iii) simply properties of negation in which we are especially interested. Though sadly and regrettably unanswered by the distinguished scholar who intended to reply, the questions remain and might stimulate answers by other logicians and further research.

Keywords: Michael Dunn, Negation, Belnap-Dunn logic, American Plan, Australian Plan, Paraconsistency, Negation inconsistency, Rule of contraposition, Nelson's constructive logics with strong negation, Negation as cancellation, *Ex contradictione nihil sequitur*, demi-negation, Proof-theoretic bilateralism

For citation: Wansing H., Olkhovikov G., Omori H. "Questions to Michael Dunn", *Logicheskie Issledovaniya / Logical Investigations*, 2021, Vol. 27, No. 1, pp. 9–19. DOI: 10.21146/2074-1472-2021-27-1-9-19

1. Introduction

Instead of writing a detailed introduction to this special issue of *Logical Investigations* and thereby putting some timely research issues into perspective, the guest-editors decided to directly delve into the topic by posing a number of questions to one of the world-wide leading experts on the logical study of the concept of negation, Prof. J. Michael Dunn. The questions address focal and partly contentious issues related to negation. Unfortunately, Prof. Dunn was no longer able to answer the questions submitted to him. Michael Dunn passed away on 5 April 2021, and we decided to dedicate the present special issue of *Logical Investigations* to his memory and to present the unanswered questions. We hope they stimulate replies by other logicians interested in the notion of negation.



Prof. J. Michael Dunn (1941–2021)

2. Nine questions to Michael Dunn concerning negation

Michael Dunn has been known as one of the central and major contributors to Relevance Logic. Systems of relevance logic are paraconsistent logics since they reject the idea of *ex contradictione quodlibet* both as an inference, $\{A, \neg A\} \vdash B$, and as a theorem in formal languages with conjunction and implication, $(A \wedge \neg A) \rightarrow B$. Moreover, Dunn was one of the developers of a system that is frequently called *Belnap-Dunn logic*, or *first-degree entailment logic*, a fundamental and especially natural four-valued paraconsistent logic that has found many applications in various areas ranging form the philosophy of information to artificial intelligence, see, for example, [Omori and Wansing, 2017]. According to Graham Priest, there exists a slippery slope that leads from the endorsement of a paraconsistent logic, in order to enable non-trivial inconsistent theories, to a metaphysical position, *dialetheism*, according to which there exist true contradictions. Of course, advocates of a system of paraconsistent logic need not be dialetheists, yet it is interesting to see how scholars who investigate paraconsistent logics think about dialetheism. This prompted our first question to Michael Dunn, **Q1**. The topics of paraconsistency and relevance figure prominently in Arnon Avron's contribution to this special issue *Implication, equivalence, and negation.*

Q1 Can you briefly comment on your general attitude towards *paraconsistent logic* and *dialetheism*, and how would you locate the *Belnap-Dunn logic*, or the *Sanjaya-Belnap-Smiley-Dunn Four-valued Logic*, as you suggested in [Dunn, 2019], in your picture?

Some well motivated expansions of Belnap-Dunn logic turned out to be non-trivial *contradictory* logics, such as the logic of logical bilattices investigated by Ofer Arieli, Arnon Avron, Melvin Fitting, and other logicians. Negation inconsistency is a very remarkable property that makes these systems orthogonal to classical logic. According to Karl Popper, [Popper, 1962, p. 322], "The acceptance of contradictions must lead ... to the end of criticism, and thus to the collapse of science." If contradictions are provable in the logic on which a scientific theory is based, this is certainly a challenging feature. How are we to think about this thought-provoking property? Well, we wanted to know how Michael Dunn thought about it, **Q2**.

Q2 The *Belnap-Dunn logic* has been extremely fruitful in considering some expansions, such as the *logic of bilattices*, as well as the *connexive logic* **C**. We highlight these because of their shared feature of being *negation inconsistent*, namely for some formulas, both the formula and its negation are valid/derivable. Do you have any thoughts on this kind of *inconsistent logics*?

A semantics in terms of information states for a non-trivial contradictory logic has it that not only a state may both support the truth as well as the falsity of one and the same formula, but that there are formulas such that *every* state supports both their truth and their falsity. Thomas Studer in his paper *A conflict tolerant logic of explicit evidence* does not exactly take such a bold step. He presents a justification logic, CTJ, that accepts two different pieces of evidence such that one justifies a proposition whereas the other piece of evidence justifies the negation of that proposition. Yet, CTJ has no room for pieces of evidence that justify both a proposition and its negation. If one would think of justifications in terms of proofs in a logical system, however, two formulas A and $\neg A$ may both have different proofs.

Paraconsistency is a property a logical system has or fails to have chiefly in virtue of a negation connective occurring in its language. The famous paraconsistent Belnap-Dunn logic enjoys several semantical characterizations. In particular, negation can be captured by means of a four-valued semantics but also by means of a two-valued semantics. Negation on the so-called "Australian plan" models the negation connective as a point shift operator in a two-valued semantics, famously especially in the Routley star semantics. The four-valued approach of the so-called "American plan" can be combined with a state semantics, but negation essentially flip-flops (support of) truth and falsity at a given state, i.e., point of evaluation. Questions **Q3** and **Q4** address the two plans.

Q3 Here is another question continuing with the *Belnap-Dunn logic*. Given that there are two semantics for negation in the *Belnap-Dunn logic*, the *American plan* and the *Australian plan*, and that you have contributed immensely to the developments of both plans, we are interested in your basic picture concerning negation. Is the following quote from one of your earlier papers something you are willing to defend?

> Tim Smiley once good-naturedly accused me of being a kind of lawyer for various non-classical logics. He flattered me with his suggestion that I could make a case for anyone of them, and in particular provide it with a semantics, no matter what the merits of the case [. . .] But I must say that my own favourite is the 4-valued semantics. I am persuaded that '$\neg\phi$ is true iff ϕ is false', and that '$\neg\phi$ is false iff ϕ is true'. And now to paraphrase Pontius Pilate, we need to know more about 'What are truth and falsity?'. It is of course the common view that they divide up the states into two exclusive kingdoms. But there are lots of reasons, motivated by applications, for thinking that this is too simple-minded. [Dunn, 1999, p. 49]

Q4 A question related to the previous question concerns the recent discussions of the Australian plan of negation defended by Francesco Berto and Greg Restall in [Berto, 2015; Berto and Restall 2019] and criticized in [De and Omori, 2018]. What might be the lesson we should learn from the Australian plan, if one's preferred approach is the American plan?

Suppose we want to have a conditional that validates modus ponens and the deduction theorem. Then the two plans in many cases fall apart. Whereas the contraposition rule is valid on the Australian plan, on the American plan, in the presence of such an implication connective, the contraposition rule can be easily invalidated. The validity of contraposition is but one of several properties that have been discussed as characteristic features of negation. The double negation laws, the De Morgan laws, and the Law of Excluded Middle have also been discussed as criteria of negationhood. The latter principle is at center stage in the invited contribution by Jc Beall and Graham Priest, *A tale of excluding the middle*, in which the authors discuss an argument for the dialetheic nature of the liar sentence. Our next question, **Q5**, turns to properties of negation.

Q5 Continuing from the previous question, some defend the *rule of contraposition*, namely if $A \vdash B$ then $\neg B \vdash \neg A$ as the essential rule for negation, but as you have taught all of us, this rule requires some very delicate treatment in the context of the *Belnap-Dunn logic*. What is your opinion on the *rule of contraposition*? Do you think any of the properties of negation will stand out as essential? Do you also have thoughts on the minimal requirement for negation, i.e. for some atomic formulas p and q, $p \nvdash \neg p$ and $\neg q \nvdash q$, suggested by Wolfgang Lenzen [Lenzen, 1996] and João Marcos [Marcos, 2005] and later adopted by Ofer Arieli, Arnon Avron, and Anna Zamansky in [Arieli et al., 2011]?

Prof. J. Michael Dunn and Prof. Nuel D. Belnap
Pittsburgh, April 2018
(Photo: courtesy of Prof. Anil Gupta)

In §1 of *Entailment. Vol. I*, Alan Anderson and Nuel Belnap explain that they “take the heart of logic to lie in the notion ‘if ... then _’ ”. It may therefore seem especially significant to consider the relationship between implication and negation. If the negation connective is meant to express falsity, then the question about the relationship between implication and negation boils down to asking about the falsity conditions of implications. That’s what question **Q6** is doing by contrasting the negation in David Nelson’s constructive logics and the understanding of negated implications in certain (hyper)connexive logics.

Q6 Given the developments related to the relevant logic **R**, for example, there are some clear interactions between the treatment of negation and the treatment of implication. Do you have any thoughts how we should think about the interactions? Would there be any guiding principles for you to think about the interactions? In particular, in the Kripke semantics for David Nelson’s constructive logics with strong negation, usually denoted by “$\sim$”, we have the classical falsity condition for negated implications insofar as a state verifies $\sim(A \to B)$ iff it verifies A and falsifies B, so that $\sim(A \to B) \leftrightarrow (A \wedge \sim B)$ is valid. In certain connexive logics, however, the falsity condition of implications is such that $\sim(A \to B) \leftrightarrow (A \to \sim B)$ is valid, cf. [Omori and Wansing, 2019; Wansing, 2020]. Do you have an opinion on this matter?

The non-involutive weakening of strong negation in the three-valued paracomplete version of Nelson’s logic is investigated in the contribution by Thiago Nascimento da Silva and Umberto Rivieccio, *Negation and implication in quasi-Nelson logic*, where quasi-Nelson algebras are presented as a generalization of both Heyting algebras and Nelson algebras.

Question **Q6** gives rise to a plethora of further considerations. One of the semantical assumptions that underlie classical logic is bivalence, the view that there are exactly two truth values, *true* and *false*, and that in a given situation every meaningful statement takes exactly one of these values. In many-valued logic, this is a subtle issue as a distinction can be drawn between algebraic and inferential values, where the latter are usually given by a bi-partition of the set of algebraic values into a non-empty set of designated values and it’s nonempty complement, cf. [Malinowski, 1993; Malinowski, 1994; Malinowski, 2009; Blasio et al., 2017]. It is the latter sets of values that are used in definitions of semantical consequence. In a bivalent setting, a statement is false exactly when it is not true. In a many-valued setting, however, being untrue, respectively, non-designated, and being false, respectively anti-designated, fall apart. Gilberto Gomes in his contribution *Negation of conditionals in natural language and thought*, discusses the external negation of natural language conditionals expressed by means of “it is not the case that” and “it is not true that”, keeping in mind a distinction between implicative and concessive conditionals. Gomes concludes that in English conditionals *If A, then B* entail *It is not the case that if A, then not B*, but presents counterexamples to the converse. On may wonder whether native speakers of English are ready to draw a distinction between “it is not the case that A” and “it is definitely false that A”.

The non-contraposible negation as falsity in Nelson’s logics is but one type of negation, and several other ways of trying to capture semantic opposition by means of a unary sentential operator or formalizing negation phenomena in natural languages are known from the literature. Questions **Q7** and **Q8** address two such notions: negation as cancellation and demi-negation.

Q7 Quite a few different kinds of negation have been distinguished between in the literature. We would like to know your opinion on some of those listed in the *Stanford Encyclopedia of Philosophy* entry on negation [Horn and Wansing, 2020]. One such concept is the notion of negation as cancellation or erasure that has been discussed by Richard and Valerie Routley [Routley and Routley, 1985] and Graham Priest [Priest, 1999], and that has been very heavily criticized more recently in [Wansing and Skurt, 2018]. The Routleys [Routley and Routley, 1985, p. 205] characterize negation as cancellation as follows:

> $\sim A$ deletes, neutralizes, erases, cancels A (and similarly, since the relation is symmetrical, A erases $\sim A$), so that $\sim A$ together with A leaves nothing, no content. The conjunction of A and $\sim A$ says nothing, so nothing more specific follows. In particular, $A \wedge \sim A$ does not entail A and does not entail $\neg A$.

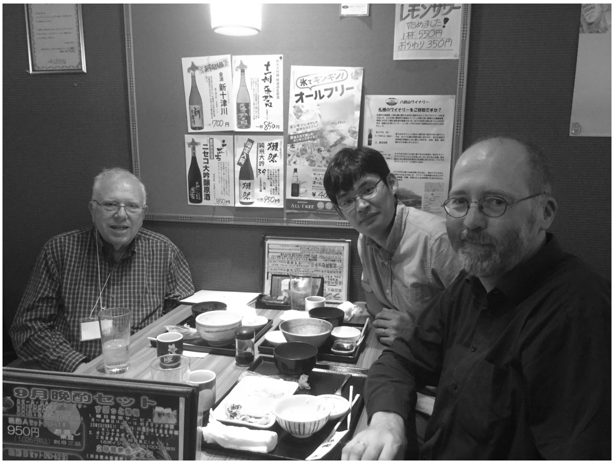
This idea is closely related to the slogan *ex contradictione nihil sequitur* (nothing follows from a contradiction), see [Wagner, 1991]. Do you think that this is a reasonable and viable conception of negation?

Q8 There is also the notion of negation by iteration, i.e., the idea to obtain a negation by a double application of a connective called “demi-negation” in [Humberstone, 1995], or “square root of negation”, $\sqrt{\texttt{not}}$, in quantum computational logic, e.g. in [Paoli, 2019]. In [Omori and Wansing, 2018] it is speculated that double demi-negation as negation could be used to analyze the phenomenon of negative concord in certain natural languages (or natural language dialects) such as in “She don’t eat no biscuit”, and it is discussed whether certain demi-negations are indeed negations. Have you ever thought about demi-negations?

The proper context of *ex contradictione nihil sequitur* is non-monotonic reasoning. If one wants to keep the reflexivity of the derivability relation, then, obviously, blocking the step from $A \vdash A$ to $A \wedge \neg A \vdash A$ will make inferences non-monotonic. Yet another type of negation is negation as failure to derive. It has been clear that logic programming needs in addition to negation as failure as a non-monotonic inference rule another kind of negation, explicit negation. Both types of negation are studied in Reinhard Kahle's paper *Default negation as explicit negation plus update.*

In the last question, **Q9**, the attention is directed to the treatment of negation in proof-theoretic semantics. The term "proof-theoretic semantics" was coined by Peter Schroeder-Heister, and not the least through his efforts, proof-theoretic semantics is now an established and very active research area. Like other areas within logic, it has seen a certain preoccupation with positive notions such as assertion and verification, but when negation comes into focus, their negative counterparts are in the spotlight. One logical system, in which there is a kind of duality between verification and refutation and between implication and a concept of co-implication is Heyting-Brouwer logic, also known as bi-intuitionistic logic, **BiInt**. Whereas in its Kripke semantics the co-implication of **BiInt** is a backwards-looking existential quantifier, the co-implication of the likewise bi-intuitionistic logic **2Int** is a forward-looking universal quantifier. Paolo Maffezioli and Luca Tranchini in their contribution *Equality and apartness in bi-intuitionistic logic* discuss the concepts of identity and apartness in the context of Heyting-Brouwer logic.

Q9 Let us finally come to the interplay between negation and consequence relations. In your paper "Partiality and its Dual" [Dunn, 2000] and elsewhere, you consider various semantically defined consequence relations that differ in whether valuations permit truth-value gaps, gluts, or both. The permission of gaps, gluts, or both is certainly a negation-related topic. One of the systems you deal with in the mentioned paper is a well-known expansion of *Belnap-Dunn logic*, viz. Almukdad and Nelson's constructive four-valued logic, nowadays known as **N4**. On the proof-theoretic side, **N4** has been arrived at via different roads, one being what is often called "bilateralism," the view that concepts like falsity, denial, and refutability should be considered in their own right and as being as important as their positive counterparts truth, assertion, and provability. As a result, this approach may lead one to making use of two separate derivability relations: provability and disprovability (or refutability). In that context, it has also been suggested, e.g. in [Wansing, 2013; Wansing, 2017; Drobyshevich, 2019], to consider in addition to implication as a connective that internalizes in the logical object language the preservation of (support of) truth, a co-implication connective that internalizes preservation of (support of) falsity. The meaning of a logical operation is then to be specified proof-theoretically by its inferential roles in *both* proofs and disproofs. What would you say about such a conception of meaning?



Prof. J. Michael Dunn meeting for lunch
with two of the guest editors during
Logic, Rationality, and Interaction.
6th International Workshop, LORI 2017,
Sapporo, Japan, September 11–14, 2017

Acknowledgements. Hitoshi Omori acknowledges support by a Sofja Kovalevskaja Award of the Alexander von Humboldt-Foundation, funded by the German Ministry for Education and Research. Grigory K. Olkhovikov acknowledges support by the German Research Foundation (DFG), grant OL-589/1-1. Heinrich Wansing, Hitoshi Omori, and Grigory K. Olkhovikov would like to thank Vladimir I. Shalak for giving them the opportunity to edit this special issue, and Natalya E. Tomova for her kind support.

References

Anderson and Belnap, 1975 – Anderson, A.R. and Belnap, N.D.Jr. *Entailment: The Logic of Relevance and Necessity. Volume I.* Princeton: Princeton University Press, 1975.

Arieli et al., 2011 – Arieli, O., Avron, A. and Zamansky, A., "Ideal paraconsistent logics", *Studia Logica*, 2011, Vol. 99, pp. 31–60.

Berto, 2015 – Berto, F. "A modality called 'negation'", *Mind*, 2015, Vol. 124, pp. 761–793.

Berto and Restall 2019 – Berto, F. and Restall, G., "Negation on the Australian plan", *Journal of Philosophical Logic*, 2019, Vol. 48, No. 6, pp. 1119–1144.

Blasio et al., 2017 – Blasio, C., Marcos, J. and Wansing, H., "An inferentially many-valued two-dimensional notion of entailment", *Bulletin of the Section of Logic*, 2017, Vol. 46, pp. 233–262.

De and Omori, 2018 – De, M. and Omori, H., "There is more to negation than modality", *Journal of Philosophical Logic*, 2018, Vol. 47, pp. 281–299.

Drobyshevich, 2019 – Drobyshevich, S., "A bilateral Hilbert-style investigation of 2-intuitionistic logic", *Journal of Logic and Computation*, 2019, Vol. 29, No. 5, pp. 665–692.

Dunn, 1999 – Dunn, J.M., "A comparative study of various model-theoretic treatments of negation: a history of formal negation", *What is Negation?*, ed. by D. Gabbay and H. Wansing, Dordrecht: Kluwer Academic Publishers, 1999, pp. 23–51.

Dunn, 2000 – Dunn, J.M., "Partiality and its dual", *Studia Logica*, 2000, Vol. 66, pp. 5–40.

Dunn, 2019 – Dunn, J.M., "Two, three, four, infinity: The path to the four-valued logic and beyond", *New Essays on Belnap-Dunn Logic*, ed. by H. Omori and H. Wansing, Cham: Springer, 2019, pp. 77–97.

Horn and Wansing, 2020 – Horn, L.R. and Wansing, H., "Negation", *The Stanford Encyclopedia of Philosophy*, [`https://plato.stanford.edu/archives/spr2020/entries/negation/`], Spring 2020 edition.

Humberstone, 1995 – Humberstone, L., "Negation by iteration", *Theoria*, 1995, Vol. 61, pp. 1–24.

Lenzen, 1996 – Lenzen, W., "Necessary conditions for negation operators", *Negation: A Notion in Focus*, ed. by H. Wansing. Berlin: Walter de Gruyter, 1996.

Malinowski, 1993 – Malinowski, G., *Many-valued logics*. Oxford: Clarendon Press, 1993.

Malinowski, 1994 – Malinowski, G., "Inferential many-valuedness", *Philosophical Logic in Poland*, ed. by J. Wolenski. Dordrecht: Kluwer, 1994, pp. 75–84.

Malinowski, 2009 – Malinowski, G., "Beyond three inferential values", *Studia Logica*, 2009, Vol. 92, pp. 203–213.

Marcos, 2005 – Marcos, J., "On negation: Pure local rules", *Journal of Applied Logic*, 2005, Vol. 3, pp. 185–219.

Omori and Wansing, 2017 – Omori, H. and Wansing, H., "40 years of FDE: An Introductory Overview", *Studia Logica*, 2017, Vol. 105, pp. 1021–1049.

Omori and Wansing, 2018 – Omori, H. and Wansing, H., "On contra-classical variants of Nelson logic **N4** and its classical extension", *Review of Symbolic Logic*, 2018, Vol. 11, No. 4, pp. 805–820.

Omori and Wansing, 2019 – Omori, H. and Wansing, H., "Connexive logics: An overview and current trends", *Logic and Logical Philosophy*, 2019, Vol. 28, pp. 371–387.

Paoli, 2019 – Paoli, F., "Bilattice logics and demi-negation", *New Essays on Belnap-Dunn Logic*, ed. by H. Omori and H. Wansing. Cham: Springer, 2019, pp. 223–243.

Popper, 1962 – Popper, K., *Conjectures and Refutations. The Growth of Scientific Knowledge*, London: Routledge, 1962.

Priest, 1999 – Priest, G., "Negation as Cancellation and Connexive Logic", *Topoi*, 1999, Vol. 18, No. 2, pp. 141–148.

Routley and Routley, 1985 – Routley, R. and and Routley, V., "Negation and contradiction", *Revista Colombiana de Matemáticas*, 1985, Vol. 19, No. 4, pp. 201–230.

Wagner, 1991 – Wagner, G., "Ex contradictione nihil sequitur", in: *Proceedings of the 12th International Joint Conference on Artificial Intelligence. Volume 1, IJCAI91.* Morgan Kaufmann Publishers Inc., San Francisco, 1991, pp. 538–543.

Wansing, 2013 – Wansing, H., "Falsification, natural deduction, and bi-intuitionistic logic", *Journal of Logic and Computation*, 2016, Vol. 26, No. 1, pp. 425–450, (first online 2013).

Wansing, 2017 – Wansing, H., "A more general general proof theory", *Journal of Applied Logic*, 2017, Vol. 25, pp. 23–46.

Wansing, 2020 – Wansing, H., "Connexive logic", *The Stanford Encyclopedia of Philosophy*, [http://plato.stanford.edu/archives/spr2020/entries/logic-connexive/], Spring 2020 edition, 2020.

Wansing and Skurt, 2018 – Wansing, H.and Skurt, D., "Negation as cancellation, connexive logic, and qLPm", *Australasian Journal of Logic*, 2018, Vol. 15, pp. 476–488.

Логические исследования
2021. Т. 27. № 1. С. 20–30
УДК 16+510.6

Logical Investigations
2021, Vol. 27, No. 1, pp. 20–30
DOI: 10.21146/2074-1472-2021-27-1-20-30

Jc Beall, Graham Priest

A Tale of Excluding the Middle

Jc Beall
University of Notre Dame,
Notre Dame, IN 46556, USA.
E-mail: jbeall@nd.edu

Graham Priest
CUNY Graduate Center,
365 5th Ave, New York, NY 10036, USA.
University of Melbourne,
Parkville, VIC 3010, Australia.
Ruhr University Bochum,
Universitätsstraße 150, 44801 Bochum, Germany.
E-mail: priest.graham@gmail.com

Abstract: The paper discusses a number of interconnected points concerning negation, truth, validity and the liar paradox. In particular, it discusses an argument for the dialetheic nature of the liar sentence which draws on Dummett's teleological account of truth. Though one way of formulating this fails, a different way succeeds. The paper then discusses the role of the Principle of Excluded Middle in the argument, and of the thought that truth in a model should be a model of truth.

Keywords: Negation, Principle of Excluded Middle, Liar Paradox, Teleological Account of Truth, Michael Dummett, Graham Priest, Truth, Validity, Truth in a Model

For citation: Beall Jc, Priest G. "A Tale of Excluding the Middle", *Logicheskie Issledovaniya / Logical Investigations*, 2021, Vol. 27, No. 1, pp. 20–30. DOI: 10.21146/2074-1472-2021-27-1-20-30

Prologue

Negation is one of the most important logical notions, and its properties are, and always have been, contentious. Many have held that satisfying the Principle of Excluded Middle is one of its properties. Many have denied this. Indeed, the Principle has been denied to provide an account of the open future, to allow for the identification of truth and provability in mathematics, to solve the liar paradox, and for a number of other reasons.

One could write a book (or two) on these matters. Our aim here is much less ambitious; it's to trace one strand of the story about the Principle. The liar paradox will be a central player in this, but so will truth and validity. The upshot of our story will be some important lessons about all these things.

Act I: In Which the Liar Paradox is Introduced

The Liar paradox is a familiar creature. For our purposes, we set it up as follows.

Angle brackets are a name-forming functor, and T is a truth predicate, satisfying the T-Schema rules of Capture and Release, respectively:

$$\frac{\alpha}{T\langle\alpha\rangle} \qquad \frac{T\langle\alpha\rangle}{\alpha}$$

A falsity predicate $F\langle\alpha\rangle$ may be defined as $T\langle\neg A\rangle$, so that the familiar dual F-Schema rules are enforced:

$$\frac{\neg\alpha}{F\langle\alpha\rangle} \qquad \frac{F\langle\alpha\rangle}{\neg\alpha}$$

The Principle of Excluded Middle (PEM), that every (declarative) sentence is either true or false, can be expressed by the schema $T\langle\alpha\rangle \vee F\langle\alpha\rangle$. Given Capture and Release, and standard rules concerning disjunction, the PEM can equivalently be expressed by the schema $\alpha \vee \neg\alpha$.[1]

The Liar sentence is a sentence λ equivalent to $F\langle\lambda\rangle$. The Liar paradox is now given by the following natural-deduction argument:

$$\dfrac{T\langle\lambda\rangle \vee F\langle\lambda\rangle \qquad \dfrac{\dfrac{\dfrac{\dfrac{\overline{T\langle\lambda\rangle}}{\lambda}}{F\langle\lambda\rangle}}{\neg\lambda} \quad \dfrac{\overline{T\langle\lambda\rangle}}{\lambda}}{\lambda\wedge\neg\lambda} \qquad \dfrac{\dfrac{\overline{F\langle\lambda\rangle}}{\neg\lambda} \quad \dfrac{\overline{F\langle\lambda\rangle}}{\lambda}}{\lambda\wedge\neg\lambda}}{\lambda\wedge\neg\lambda}$$

Clearly, one can break the argument if one rejects the PEM, as several logicians have argued should be done.[2] Those, like your narrators (JCB and GP), who think that the Liar serves as a strong witness to 'true contradictions' (gluts, dialethias) – true sentences of the form $\alpha\wedge\neg\alpha$ (dual of $\neg\alpha\vee\alpha$) – and who think that the Liar paradox (as *per* the above derivation) serves as a sound argument

[1] The two ways of expressing the PEM can, of course, come apart if the standard rules for disjunction fail, as they do, for example, on the familiar supervaluation account of van Fraassen [Fraassen, 1966], which validates the schema $\alpha\vee\neg\alpha$, though α may be neither true nor false in an interpretation.

[2] E.g., [Kripke, 1975; Field, 2008].

for just such a result, are naturally motivated to advance an argument for the PEM. And so the history unfolds.[3]

Act II: In Which the Teleological Account of Truth is Explained

In Chapter 4 of the first edition of *In Contradiction*,[4] GP, partly with a defense of the soundness of the Liar derivation in mind, advanced an argument for the PEM that rests on the teleological account of truth. This Act reviews the account; the next Act reviews the target argument for PEM.

The teleological account of truth is essentially Dummett's. Dummett notes that some notions are fully understandable only via an understanding of their point or *telos*. Thus, consider the notion of winning, as of a game. One could know what constitutes a winning position in a game of chess (or bridge, or cricket) and what constitutes a losing position, without understanding what it is to win. To understand this, one must know that the aim or point of playing a game, as such, is to achieve a winning position, not a losing one.

Similarly, says Dummett, one might know what constitutes a true statement and what constitutes a false statement without understanding what truth is. The T-Scheme for α states the condition under which α is true, and the F-Schema for α states the condition under which α is false. But if you don't understand what the point of calling something 'true' is, you don't understand truth.

What is the point (telos) of truth? The answer, according to Dummett, is that truth is the institutional (as opposed to personal) aim of assertion – the aim of asserting something.[5] As he puts it:[6]

> it is part of the concept of truth that we aim at making true assertions.

Or again:[7]

> the class of true sentences is the class the utterance a member of which a speaker of language is aiming at when he employs what is recognizably the assertoric use.

[3]One of us (viz., JCB) rejects the soundness of the Liar argument above without independent reason to accept the PEM, but JCB thinks that the methodological quest for completeness – sorting all sentences of the language of our theories into 'The True' or 'The False' – that drives systematic, truth-seeking theorists sufficiently pushes towards a glutty account of λ. We note this only to set it aside. For some discussion, see [Beall, 2017; Beall, 2018]. Both of us remain committed glut theorists, and equally interested in arguments for PEM.

[4][Priest, 1987]. In what follows, page references are to the *second* (viz., 2006) edition.

[5]And, one might add, other cognitive acts, such as believing.

[6][Dummett, 1959, p. 143].

[7][Dummett, 1973, p. 320].

Thus, as *In Contradiction* (§4.5) points out, if for whatever reason people played games in such a way as to realize what we currently take to be a losing position then winning would become losing, and vice versa. Similarly, if the institution of assertion morphed in such a way that it became the aim to assert the sentences we now take to be false ('the moon is a cube', '3+2 is not 5', etc.), people would understand these as expressing the opposite of what they are currently taken to express: the sense of a sentence would flip to what is currently expressed by its negation.

Now, one might think that the notion of truth (as applied to sentences, beliefs, propositions, etc.) is univocal, but one might not. Either way, at least with respect to the meaning of 'truth' that Dummett is talking about, his point is well taken. Without an understanding of the *point* of this notion, truth and falsity would be formally symmetrical notions, with nothing to break the symmetry. It is the way that they are applied in use that distinguishes them.

In what follows, it is truth in the given teleological sense with which we are chiefly concerned. Unless otherwise explicitly noted, when we speak of truth or falsity in what follows, we mean truth or falsity in this sense (if there is more than one).

Intermission: For the Sake of Transparency

Before our story continues, some points of clarification are useful. Call the following two-way rule the *Contraposed T-Schema Rule*:

$$\frac{\neg A}{\overline{\neg T\langle A\rangle}}$$

The T-Schema rules do not by themselves give the contraposed form. Nor is there anything in the teleological account of truth that delivers them; for the account says nothing about negation. On the other hand, just because of this, it is open to someone who endorses the teleological account of truth to endorse an account of negation that delivers the contraposed T-schema rule.

For the rest of this essay, we will call an account of truth *transparent* if for any A, A and $T\langle A\rangle$ are intersubstitutable *salva veritate* in all extensional contexts. If the only other contexts in play are those delivered by conjunction and disjunction (and those that can be defined from them and negation), and given that these behave in a standard fashion, transparency is equivalent to the T-schema rules plus the contraposed T-schema rules. Hence it is quite possible for someone to endorse an account of teleological truth that is transparent (the intersubstitutability holds), just as it's possible for someone to endorse an account of teleological truth that fails to be transparent (intersubstitutability fails).

JCB, in a variety of works, uses the word 'transparency' for an account of truth according to which there is neither more nor less to truth than is captured by (as he says) 'the transparency rules' – the given intersubstitutability rules in the above sense. Clearly, the transparency account of truth, so understood, is not compatible with the teleological account of truth. JCB has also argued that transparency in this sense – and so the relevant T-schema and F-schema rules – arise from the thought that (the target transparency notions of) truth and falsity are predicates that reflect (indeed, are just defined as 'abstractions from') the behavior of the logical connectives $\neg$ and † (the last of these being the monadic truth function whose output is identical with its input).[8]

Whether such an account of truth and the teleological account of truth are rivals or simply characterise different notions, is an issue on which we express no view here. As observed, both views deliver the T-Schema rules (viz., Capture and Release). For the teleological account of truth – our principal concern – it is exactly these rules that tell us which sentences *are* true. For JCB's account, it is the fact that truth 'supervenes' (in a sense that needn't detain us here) on †. Moreover, arguably Capture and Release govern any notion of truth worth the name. (As Tarski noted, such is a condition of adequacy on any account of truth.) And, given this, all notions of truth are extensionally equivalent. For if T_1 and T_2 are any such notions, $T_1 \langle\alpha\rangle \dashv\vdash \alpha \dashv\vdash T_2 \langle\alpha\rangle$. Given the F-Schema rules, a similar equivalence holds for falsity.

Act III: In Which an Argument for the PEM is Seen to Crash

Now to the argument for the PEM mentioned at the beginning of Act II.

Deploying the teleological account of truth, Dummett notes (with a qualification meant to take certain complications off of the agenda) that if one utters something, one either achieves the point of doing so or one fails. Anything less than success is failure. As he puts it – in distinctly Dummettian terms:[9]

> A sentence, so long as it is not ambiguous or vague, divides all possible states of affairs into just *two* classes. For a given state of affairs, either the statement is used in such a way that a man who asserted it but envisaged the possibility of that state of affairs as a possibility, would be held to have spoken misleadingly, or the assertion of the statement would not be taken as expressing the speaker's exclusion of that possibility. If a state of affairs of the

[8] Further on these matters, see [Beall, 2009, Chapter 1], and especially [Beall, 2021].

[9] [Dummett, 1959, p. 149] f. Italics original.

> the first kind obtains, the statement is false. If all actual states of affairs are of the second kind, it is true. It is therefore *prima facie* senseless to say of any statement that in such-and-such a state of affairs it would be neither true nor false.

The game analogy is illuminating here. In many games, there are three possible outcomes: winning, losing, and drawing. But in some games, there are only two: winning and losing. In such games, not to win is, *ipso facto*, to lose. In asserting, there is nothing that corresponds to drawing. Anything less than success is, *ipso facto*, failure. *In Contradiction* (§4.7) used this point to argue for the PEM. The last sentence of the last quote by Dummett clearly sounds like a version of this.

In 'True Contradictions'[10] Terry Parsons criticised the argument from the teleological account of truth to the PEM. For Dummett's point, 'false', he said, has to be interpreted as *untrue*, not as *having a true negation,* as required by the PEM. And that seems right: nothing in Dummett's reflections says anything about how sentences containing negation work. And to appeal to the inference from $\neg T\langle\alpha\rangle$ to $F\langle\alpha\rangle$ would clearly beg the question.

The argument from teleology for the PEM thus fails to achieve its goal. But the play goes on: there is importantly more to the matter than this.

Act IV: In Which the Liar Paradox Reappears

In the second edition of *In Contradiction* (§19.6) GP accepted Parson's criticism, and whilst pointing out that the teleological account of truth still puts the onus of proof on one who wishes to draw a distinction within the category of untruths, he agreed with Parsons that Dummett's point shows only that (p. 267):

> anything that fails to live up to the aim of assertion is *ipso facto* not true.

But then there is a footnote (fn 13):

> Note that this conclusion, on its own, is not without its sting. It establishes, even if one denies the Law of Excluded Middle in general, that particular instances of the form $T\langle\alpha\rangle \vee \neg T\langle\alpha\rangle$ still hold. These are precisely the ones that give the "strengthened liar" its punch, as I noted in discussing ch. 1.

[10] [Parsons, 1990].

Matters are left there; but the import of the point can be brought out much more simply and directly, as follows.[11]

Instead of considering the Liar paradox in the form ‘this sentence is false’, consider it in the form ‘this sentence is untrue’. Thus, let λ^* be a sentence of the form $\neg T\langle\lambda^*\rangle$. Given the preceding discussion, we have, as noted, $T\langle\lambda^*\rangle \vee \neg T\langle\lambda^*\rangle$, and a contradiction quickly follows. Merely consider the following argument in natural-deduction form:

$$\dfrac{T\langle\lambda^*\rangle \vee \neg T\langle\lambda^*\rangle \qquad \dfrac{\dfrac{\dfrac{\overline{T\langle\lambda^*\rangle}}{\lambda^*}}{\overline{\neg T\langle\lambda^*\rangle}} \quad \overline{T\langle\lambda^*\rangle}}{T\langle\lambda^*\rangle \wedge \neg T\langle\lambda^*\rangle} \qquad \dfrac{\dfrac{\dfrac{\overline{\neg T\langle\lambda^*\rangle}}{\lambda^*}}{\overline{T\langle\lambda^*\rangle}} \quad \overline{\neg T\langle\lambda^*\rangle}}{T\langle\lambda^*\rangle \wedge \neg T\langle\lambda^*\rangle}}{T\langle\lambda^*\rangle \wedge \neg T\langle\lambda^*\rangle}$$

Hence, *without relying on PEM the teleological notion of truth delivers a direct argument for true contradictions*, an argument for the truth and – assuming double negation – falsity of $\neg T\langle\lambda^*\rangle$, that is, λ^*, and so for the truth of the contradiction $\lambda^* \wedge \neg\lambda^*$.

We end this Act by noting a corollary concerning transparency. The scheme $T\langle\alpha\rangle \vee \neg T\langle\alpha\rangle$ is obviously a restricted case of PEM; let us call it *Restricted Excluded Middle* (REM). If truth is transparent[12] then $T\langle\neg\alpha\rangle$ is equivalent to $\neg\alpha$, which is equivalent to $\neg T\langle\alpha\rangle$. That is, truth commutes with negation. In this case, REM clearly delivers PEM. Hence the argument in Act I for the contradictory nature of λ goes through as well as that for λ^*.

Act V: In Which Validity Comes Onstage

What we have discussed so far concerns the truth of the REM and PEM, not their logical validity, which is a quite different matter. Indeed, the preceding discussion does *not* require these to be logically true (i.e., logically valid schemata, each instance of which is logically valid), simply true; and it is perfectly coherent to take them to be true without being logically so.

To show this, we need some machinery to deliver an account of validity; we shall take this to be model-theoretic machinery. To handle the inconsistency of Liars, a paraconsistent validity relation is required. Some paraconsistent logics, such as *LP*, defended by GP in *In Contradiction*, validate PEM, and so cannot be used to show this. But, *FDE*, preferred by JCB,[13] does not validate PEM.

[11]Perhaps GP overlooked the point; perhaps he forgot it. It was noted by JCB in a talk given at a conference at the National Autonomous University of Mexico in 2019, where GP was in the audience.

[12]As [Beall, 2009] and [Field, 2008] hold, though [Priest, 1987, §4.9] does not.

[13][Beall, 2017; Beall, 2018].

So we can use this. The semantics of FDE can be set up in many equivalent ways. We use the four-valued semantics, where the values are t (true only), f (false only), b (both), and n (neither). t and b are the designated values. LP has the same semantics, except that it does not use the value n.[14]

Let the language be a first-order one that contains a monadic truth predicate, T, and for every expression e in the language, a suitable name, $\langle e \rangle$. The extension of any monadic predicate P (including T) is the set of objects in the domain which the interpretation of P maps to t or b; the anti-extension is the set of objects it maps to f or b. If in every interpretation every sentence is in either the extension or anti-extension of T then REM is validated (that is, it is true on every interpretation over which the given validity relation is defined) but, crucially, PEM is not thereby validated, since the validity of REM is compatible with some formulas (not involving T) having value n.

Such interpretations may not validate the T-Schema rules, but they can be made to do so if, in addition, every sentence, α, has the value t or b iff $T\langle \alpha \rangle$ does. But PEM is still not validated. Again, suppose that P is some (other monadic) predicate and 'a' is the name of some object which is in neither the extension nor anti-extension of P. Then Pa has the value n, and neither Pa nor $\neg Pa$ is in the extension of T, and so (by assumption) is in its anti-extension.

Matters are more complicated if the truth predicate is transparent in every interpretation. For then, α and $T\langle \alpha \rangle$ have the same value in every interpretation. So if the REM holds in an interpretation then, for any α, $T\langle \alpha \rangle$ or $\neg T\langle \alpha \rangle$ holds. But then $T\langle \alpha \rangle$ has the value t, b, or f, and so, given transparency, α does too. Hence $\alpha \vee \neg\alpha$ is true in all interpretations.[15]

However, with a small change, one can accommodate the situation in which T behaves (let us say) 'transparently at a distinguished interpretation' (but not at all interpretations), REM is true at the given interpretation, but PEM is not valid (true at all interpretations). Fix on some interpretation, @. One may think of @ as an interpretation which verifies some true theory – maybe the theory of T itself – or is, in some other way, such that everything that holds in it is actually true (*simpliciter*) – whatever one takes that to mean. The way that T behaves in this interpretation may well not be the way in which it

[14]See [Priest, 2008, chs. 8 and 22]. For natural-deduction systems for for FDE, LP, and related systems, see [Priest, 2019]. For a wider discussion of FDE see [Omori and Wansing, 2017; Omori and Wansing, 2019].

[15]This result is implicit in the discussion of an FDE-based 'compromise' in [Beall, 2009, p. 104], which shows that if a PEM-demanding negation-like device ('exhaustive device' in said work) is added to a language with its own transparent truth predicate then the only models of the resulting theory are LP models: one kicks out any sentence that can be gappy, that is, has value n. In a language with its own transparent truth predicate, imposing REM is, then, tantamount to imposing a PEM-satisfying negation.

behaves at other interpretations. After all, interpretations represent situations. These may be actual, they may be possible, or they may be impossible. What validity gives is a way of preserving truth-in-an-interpretation in all of these.

In particular, then, @ may 'make true' REM and the transparency of truth (at @), and so likewise the schema $\alpha \vee \neg\alpha$. But there is no reason why these should hold in other interpretations too. In other words, they can hold (at @), but not be valid – not hold at all interpretations. And let us stress, again, that the argument that λ^* is a true contradiction requires only that the REM (or PEM) be *true*, not that it be valid.

Act VI: In Which an Old Adage is Dissected

Of course, whether the PEM *is* logically valid is another matter. FDE is *prima facie* a more attractive account of validity than LP, simply because of its symmetry. It also accommodates naturally the apparent duality between the liar and the truth-teller, 'this sentence is true'. The liar looks like a case of overdetermination (*both*); the latter looks like a case of underdetermination (*neither*).[16] So it is fair to say that if someone prefers LP to FDE as an account of logical validity, the onus is on them to make the case.

This is not the place to go into this matter, since it raises the whole question of the methodology determining the rational choice of logical theory.[17] But since truth has been very much a major concern of what has gone before, let us conclude this story by discussing just one aspect of the matter, which concerns truth. It involves the natural thought expressed in the old adage:

- truth-in-a-model ought to be a model of truth

that is, truth-in-an-interpretation should behave in the same way as does truth *simpliciter* (whatever, exactly, one takes this to be). Exactly how to understand this thought might be debated but one may naturally take it to have the following consequence. If PEM is true (i.e., all instances are true), so that for any α, either $T\langle\alpha\rangle$ or $T\langle\neg\alpha\rangle$ holds, then either α or $\neg\alpha$ should be true in an interpretation. But if either α or $\neg\alpha$ holds in every interpretation, then nothing has the value n in any interpretation, and we have the semantics of LP (not FDE). So on this understanding of the adage, the truth of the PEM entails its validity.

The adage may well be resisted, however. As observed, when we reason, we reason about all sorts of situations, actual, possible, and maybe impossible. And there is no reason why truth at such situations must behave like truth *simpliciter*. Thus, suppose the actual situation to be as classical as one likes.

[16]But on this, see *In Contradiction*, p. 66, and [Mortensen and Priest, 1981].

[17]For some discussion, see [Priest, 2014], and similarly [Beall, 2019].

We may yet want to reason about situations that are gappy or glutty. One may think of these as merely possible or impossible situations. And, one needs a canon of inference which preserves what holds at *every one* of these interpretations. Truth in some interpretations will not, then, mirror truth *simpliciter*.

Certain understandings of the adage may, in fact, be problematic even if one takes LP to deliver the correct notion of logical validity. For if both $T\langle\alpha\rangle$ and $\neg T\langle\alpha\rangle$ may hold then, it might be thought to follow, there should be interpretations where α can both hold and not hold; that is, where α may both have the value t or b, and not. Obviously this cannot be the case in a consistent semantics. The possibility may be accommodated by moving to a semantics itself formulated in a paraconsistent logic. However, that raises many issues of its own, and this is not the theatre for that tale.[18]

Epilogue

This brings our story to an end. Of course, we know that there are likely to be many other players in the wings who will want to jump on stage. And there may well be other acts to be written. Nonetheless, we think our tale an illuminating one, and trust that the reader has enjoyed it.

References

Beall, 2009 – Beall, J. *Spandrels of Truth*, Oxford: Oxford University Press, 2009.

Beall, 2017 – Beall, J. "There is no Logical Negation: True, False, Both and Neither", *Australasian Journal of Logic*, 2017, Vol. 14, `https://ojs.victoria.ac.nz/ajl/article/view/4025/3576`.

Beall, 2018 – Beall, J. "The Simple Argument for Subclassical Logic", *Philosophical Issues*, 2018, Vol. 28, pp. 30–54; reprinted in *The Philosopher's Annual*, 2018, `http://www.pgrim.org/philosophersannual/past.html`.

Beall, 2019 – Beall, J. "On Williamson's New Quinean Argument Against Nonclassical Logic", *Australasian Journal of Logic*, 2019, Vol. 16, `https://ojs.victoria.ac.nz/ajl/article/view/5915`.

Beall, 2021 – Beall, J. "Transparent Truth as a Logical Property", to appear in *The Nature of Truth: Classical and Contemporary Perspectives*, ed. by M. Lynch, J. Wyatt, J. Kim, and N. Kellen, Cambridge, MA: MIT Press, 2021.

Dummett, 1959 – Dummett, M. "Truth", *Proceedings of the Aristotelian Society*, 1959, Vol. 59, pp. 141–162; reprinted as ch. 1 of *Truth and Other Enigmas*, London: Duckworth, 1978.

Dummett, 1973 – Dummett, M. *Frege: Philosophy of Language*, London: Duckworth, 1973.

Field, 2008 – Field, H. *Saving Truth from Paradox*, Oxford: Oxford University Press, 2008.

[18] A discussion can be found in [Priest, 2020].

Fraassen, 1966 – Fraassen, van, B.C. "Singular Terms, Truth-Value Gaps, and Free Logic", *Journal of Philosophy*, 1966, Vol. 63, pp. 481–495.

Kripke, 1975 – Kripke, S. "Outline of a Theory of Truth", *Journal of Philosophy*, 1975, Vol. 72, pp. 690–716.

Mortensen and Priest, 1981 – Mortensen, C., and Priest, G. "The Truth Teller Paradox", *Logique et Analyse*, 1981, Vol. 95–96, pp. 381–388.

Omori and Wansing, 2017 – Omori, H., and Wansing, H. "40 Years of *FDE*: An Introductory Overview", *Studia Logica*, 2017, Vol. 105, pp. 1021–1049.

Omori and Wansing, 2019 – *New Essays on Belnap-Dunn Logic*, ed. by H. Omori and H. Wansing, New York: Springer, 2019.

Parsons, 1990 – Parsons, C. "True Contradictions", *Canadian Journal of Philosophy*, 1990, Vol. 20, pp. 35–53.

Priest, 1987 – Priest, G. *In Contradiction*, Dordrecht: Martinus Nijhoff, 1987; 2nd edn, Oxford: Oxford University Press, 2006.

Priest, 2008 – Priest, G. *Introduction to Non-Classical Logic: from If to Is*, Cambridge: Cambridge University Press, 2008.

Priest, 2014 – Priest, G. "Revising Logic", *The Metaphysics of Logic*, ed. by P. Rush, Cambridge: Cambridge University Press 2014, ch. 12.

Priest, 2019 – Priest, G. "Natural Deduction for Systems in the *FDE* Family", pp. 279–292 of Omori and Wansing (2019).

Priest, 2020 – Priest, G. "Metatheory and Dialetheism", *Logical Investigations*, 2020, Vol. 26, pp. 48–59.

Логические исследования
2021. Т. 27. № 1. С. 31–45
УДК 16+510.6

Logical Investigations
2021, Vol. 27, No. 1, pp. 31–45
DOI: 10.21146/2074-1472-2021-27-1-31-45

Arnon Avron

Implication, Equivalence, and Negation

Arnon Avron
School of Computer Science,
Tel Aviv University.
Tel Aviv 6997801, Israel.
E-mail: aa@cs.tau.ac.il

Abstract: A system $HCL_{\overset{\neg}{\leftrightarrow}}$ in the language of $\{\neg, \leftrightarrow\}$ is obtained by adding a single negation-less axiom schema to $HLL_{\overset{\neg}{\to}}$ (the standard Hilbert-type system for multiplicative linear logic without propositional constants), and changing $\to$ to $\leftrightarrow$. $HCL_{\overset{\neg}{\leftrightarrow}}$ is weakly, but not strongly, sound and complete for $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$ (the $\{\neg, \leftrightarrow\}$-fragment of classical logic). By adding the Ex Falso rule to $HCL_{\overset{\neg}{\leftrightarrow}}$ we get a system with is strongly sound and complete for $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$. It is shown that the use of a new rule cannot be replaced by the addition of axiom schemas. A simple semantics for which $HCL_{\overset{\neg}{\leftrightarrow}}$ itself is strongly sound and complete is given. It is also shown that $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$, the logic induced by $HCL_{\overset{\neg}{\leftrightarrow}}$, has a single non-trivial proper axiomatic extension, that this extension and $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$ are the only proper extensions in the language of $\{\neg, \leftrightarrow\}$ of $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$, and that $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$ and its single axiomatic extension are the only logics in $\{\neg, \leftrightarrow\}$ which have a connective with the relevant deduction property, but are not equivalent to an axiomatic extension of $\mathbf{R}_{\overset{\neg}{\to}}$ (the intensional fragment of the relevant logic $\mathbf{R}$). Finally, we discuss the question whether $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$ can be taken as a paraconsistent logic.

Keywords: Implication, Semi-implication, Negation, Equivalence, Biconditional, Classical propositional logic, Deduction theorems, Paraconsistent Logics

For citation: Avron A. "Implication, Equivalence, and Negation", *Logicheskie Issledovaniya / Logical Investigations*, 2021, Vol. 27, No. 1, pp. 31–45. DOI: 10.21146/2074-1472-2021-27-1-31-45

1. Introduction

The relevant deduction property (RDP) for a binary connective $\to$ is a weak form of the classical-intuitionistic deduction theorem which has (somewhat implicitly) motivated the design of the intensional fragments ($\mathbf{R}_{\to}$ and $\mathbf{R}_{\overset{\neg}{\to}}$) of the relevance logic $\mathbf{R}$ ([Anderson, Belnap, 1975; Dunn, Restall, 2002]).[1] In [Avron, 2015] we showed that with one exception, in the pure language of $\{\to\}$ $\to$ has in

[1] The RDP is also the key condition that should be satisfied by what is called in [Avron, 2015] 'semi-implication'. The other condition, included to avoid degenerate cases, is that not every case in which an implication holds is a case in which its converse also holds.

a finitary logic **L** the RDP iff **L** has a strongly sound and complete Hilbert-type system which is an axiomatic extension (i.e. an extension by axiom schemas) of $HR_{\rightarrow}$ (the standard axiomatization of $\mathbf{R}_{\rightarrow}$). The only exception is a logic which, like $\mathbf{R}_{\rightarrow}$, is an axiomatic extension of $\mathbf{LL}_{\rightarrow}$ (the purely implicational fragment of Linear Logic): $\mathbf{CL}_{\leftrightarrow}$, the pure equivalential fragment of classical logic (where the biconditional is denoted by $\rightarrow$).

The fact that $\leftrightarrow$ has in $\mathbf{CL}_{\leftrightarrow}$ the RDP raises the question to what extent it can actually be used as an implication connective. A crucial criterion here is the richness of the languages in which it might serve as such, that is: what useful connectives can be added to it. It can easily be seen that it is impossible to add to $\mathbf{CL}_{\leftrightarrow}$ a 'conjunction' $\wedge$ such that both $\varphi \wedge \psi \leftrightarrow \varphi$ and $\varphi \wedge \psi \leftrightarrow \psi$ would be valid, since this would immediately trivialize the logic. Similarly, one cannot add to $\mathbf{CL}_{\leftrightarrow}$ a 'disjunction' $\vee$ such that $\varphi \leftrightarrow \varphi \vee \psi$ and $\psi \leftrightarrow \varphi \vee \psi$ would be valid. It follows that among the connectives used in linear logic and in relevance logics, one may add to $\mathbf{CL}_{\leftrightarrow}$ only what are called in linear logic 'multiplicative' connectives, and in relevance logics 'intensional' connectives. The most basic such connective is negation (which together with the implication connective of linear and relevant logics suffices for defining the rest of them). Accordingly, the main goal of this paper is to investigate the equivalence-negation fragment of classical logic, and corresponding proof systems.

2. Preliminaries

In the sequel $\mathcal{L}$ is a propositional language, φ, ψ, θ vary over its formulas, p, q over its atomic formulas, and $\mathcal{T}, \mathcal{S}$ over its theories (i.e. sets of formulas).

Definition 1. A (Tarskian) *consequence relation* for a language $\mathcal{L}$ is a binary relation between theories in $\mathcal{L}$ and formulas in $\mathcal{L}$ satisfying the following three conditions:

[R]	*Reflexivity*:	$\psi \vdash \psi$ (i.e. $\{\psi\} \vdash \psi$).
[M]	*Monotonicity*:	if $\mathcal{T} \vdash \psi$ and $\mathcal{T} \subseteq \mathcal{T}'$, then $\mathcal{T}' \vdash \psi$.
[C]	*Cut (Transitivity)*:	if $\mathcal{T} \vdash \psi$ and $\mathcal{T}', \psi \vdash \varphi$ then $\mathcal{T} \cup \mathcal{T}' \vdash \varphi$.

Definition 2. Let $\vdash$ be a Tarskian consequence relation for $\mathcal{L}$.

- $\vdash$ is *structural*, if for every $\mathcal{L}$-substitution θ and every $\mathcal{T}$ and ψ, if $\mathcal{T} \vdash \psi$ then $\theta(\mathcal{T}) \vdash \theta(\psi)$.
- $\vdash$ is *non-trivial* if $p \nvdash q$ for distinct atomic formulas p, q.
- $\vdash$ is *finitary*, if for every theory $\mathcal{T}$ and every formula ψ such that $\mathcal{T} \vdash \psi$ there is a *finite* theory $\Gamma \subseteq \mathcal{T}$ such that $\Gamma \vdash \psi$.

Definition 3. A (propositional) *logic* is a pair $\mathbf{L} = \langle \mathcal{L}, \vdash_{\mathbf{L}} \rangle$, where $\mathcal{L}$ is a propositional language, and $\vdash_{\mathbf{L}}$ is a structural and non-trivial Tarskian consequence relation for $\mathcal{L}$.[2] A logic $\mathbf{L} = \langle \mathcal{L}, \vdash_{\mathbf{L}} \rangle$ is *finitary* if $\vdash_{\mathbf{L}}$ is finitary.

Definition 4. Let $\mathbf{L} = \langle \mathcal{L}, \vdash_{\mathbf{L}} \rangle$ be a propositional logic, and let $\supset$ be a (primitive or defined) connective of $\mathcal{L}$. $\mathbf{L}$ has the *relevant deduction property* (RDP) for $\supset$ if it satisfies the following condition:

$$\mathcal{T}, \varphi \vdash_{\mathbf{L}} \psi \text{ iff either } \mathcal{T} \vdash_{\mathbf{L}} \psi \text{ or } \mathcal{T} \vdash_{\mathbf{L}} \varphi \supset \psi.$$

Remark 1. If a finitary logic $\mathbf{L}$ has a connective $\supset$ with the RDP, then the following holds for every theory $\mathcal{T}$ and formula φ: $\mathcal{T} \vdash_{\mathbf{L}} \varphi$ iff there exist $\psi_1, \ldots, \psi_n \in \mathcal{T}$ $(n \geq 0)$ such that $\vdash_{\mathbf{L}} \psi_1 \supset (\psi_2 \supset (\cdots (\psi_n \supset \varphi) \cdots))$.

Definition 5. Let $\mathcal{L}_{\to} = \{\to\}$, $\mathcal{L}_{\leftrightarrow} = \{\leftrightarrow\}$.

1. $HLL_{\to}$ is the system in $\mathcal{L}_{\to}$ presented in Figure 1.

Axioms:

[Id]	$\varphi \to \varphi$	(Identity)
[Tr]	$(\varphi \to \psi) \to \big((\psi \to \theta) \to (\varphi \to \theta)\big)$	(Transitivity)
[Pe]	$\big(\varphi \to (\psi \to \theta)\big) \to \big(\psi \to (\varphi \to \theta)\big)$	(Permutation)

Rule of inference:

[MP] $\dfrac{\varphi \qquad \varphi \to \psi}{\psi}$

Fig. 1. The proof system $HLL_{\to}$

2. $HR_{\to}$ is the extension of $HLL_{\to}$ by the following axiom:

$$\text{[Ct]} \qquad (\varphi \to (\varphi \to \psi)) \to (\varphi \to \psi) \qquad \text{(Contraction)}$$

3. $HLL_{\leftrightarrow}$ is the system in $\mathcal{L}_{\leftrightarrow}$ which is obtained from $HLL_{\to}$ by using '$\leftrightarrow$' instead of '$\to$'. $HCL_{\leftrightarrow}$ is the extension of $HLL_{\leftrightarrow}$ by the following axiom:

$$\text{[Eq]} \qquad (\varphi \leftrightarrow (\varphi \leftrightarrow \psi)) \leftrightarrow \psi \qquad \text{(Equivalence)}$$

[2]This is the notion of propositional logic which has been used in [Avron, 2015], as well as in [Avron et al., 2018].

The following three theorems have been proved in [Avron, 2015].

Theorem 1. *A logic* $\mathbf{L}$ *is finitary and has a connective* $\to$ *which has in* $\mathbf{L}$ *the* RDP *iff* $\mathbf{L}$ *has a strongly sound and complete Hilbert-type system which is equivalent to an extension by axiom schemas of either* $HR_{\to}$ *or* $HCL_{\leftrightarrow}$.[3]

Theorem 2. $HCL_{\leftrightarrow}$ *is strongly sound and complete for the equivalence fragment of classical logic (i.e.* $\mathcal{T} \vdash_{HCL_{\leftrightarrow}} \varphi$ *iff by interpreting* $\to$ *as the classical biconditional* $\leftrightarrow$*, we get that every assignment that satisfies* $\mathcal{T}$ *also satisfies* φ*).*

Theorem 3. $\mathbf{CL}_{\leftrightarrow}$ *has no proper extension in its language.*[4]

3. The Logic $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$ and the System $HCL_{\overset{\neg}{\leftrightarrow}}$

Definition 6. $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$ is the equivalence-negation fragment of classical logic. $\mathcal{M}_{CL_{\overset{\neg}{\leftrightarrow}}}$ is the two-valued matrix which induces $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$.

Proposition 1. $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$ does not have the RDP.

Proof. Although $\neg p, p \vdash_{\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}} q$, neither $\neg p \nvdash_{\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}} q$, nor $\neg p \nvdash_{\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}} p \leftrightarrow q$. ■

Definition 7. Let $\mathcal{L} = \{\leftrightarrow, \neg\}$. The Hilbert-type proof system $HCL_{\overset{\neg}{\leftrightarrow}}$ is presented in Figure 2. $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$ is the logic induced by $HCL_{\overset{\neg}{\leftrightarrow}}$.

Remark 2. The axioms given in Figure 2 are actually not independent, since [N2] can be dropped. This can be seen by substituting $\neg\psi$ for φ in both [Id] and [N1]. By applying [MP] to the resulting formulas we get $\psi \leftrightarrow \neg\neg\psi$. From this we can get [N2] by using $HCL_{\leftrightarrow}$ and Theorem 2.[5]

Remark 3. In Chapter 11 of [Avron et al., 2018] a notion was introduced of a *negation associated with a given binary connective that has the RDP*. It was shown there that a logic possesses a binary connective $\to$ that has the RDP together with a negation $\neg$ associated with it iff it is induced by some axiomatic extension of the system $HLL_{\overset{\neg}{\to}}$. The latter is the standard Hilbert-type system (given in [Avron, 1988]) for the multiplicative fragment (without the multiplicative constants) of linear logic ([Girard, 1987]). It is obtained from $HCL_{\overset{\neg}{\leftrightarrow}}$ by deleting the axiom [Eq], and then changing $\leftrightarrow$ in all axioms and rules to $\to$. By adding to $HLL_{\overset{\neg}{\to}}$ the contraction axiom, we get the standard Hilbert-type system for $\mathbf{R}_{\overset{\neg}{\to}}$, the intensional fragment of the relevance logic $\mathbf{R}$.

---

[3]It seems that the RDP was first raised as being of interest to relevance logic in [Diaz, 1980].

[4] A weaker result, that $\mathbf{CL}_{\leftrightarrow}$ is Post-complete in the sense that one cannot (consistently) add any new axiom to it in its language, had already been shown by Prior in [Prior, 1962].

[5]I am indebted to an anonymous referee for this observation.

Axioms:

[Id]	$\varphi \leftrightarrow \varphi$	(Identity)
[Tr]	$(\varphi \leftrightarrow \psi) \leftrightarrow \big((\psi \leftrightarrow \theta) \leftrightarrow (\varphi \leftrightarrow \theta)\big)$	(Transitivity)
[Pe]	$\big(\varphi \leftrightarrow (\psi \leftrightarrow \theta)\big) \leftrightarrow \big(\psi \leftrightarrow (\varphi \leftrightarrow \theta)\big)$	(Permutation)
[Eq]	$(\varphi \leftrightarrow (\varphi \leftrightarrow \psi)) \leftrightarrow \psi$	(Equivalence)
[N1]	$(\varphi \leftrightarrow \neg\psi) \leftrightarrow (\psi \leftrightarrow \neg\varphi)$	(Contraposition)
[N2]	$\neg\neg\varphi \leftrightarrow \varphi$	(Double neg.)

Rule of inference:

$[\mathrm{MP}]_{\leftrightarrow}$ $$\frac{\varphi \qquad \varphi \leftrightarrow \psi}{\psi}$$

Fig. 2. The proof system $HCL_{\overset{\neg}{\leftrightarrow}}$

Theorem 4. *Let the logic* $\mathbf{L}$ *be induced by some axiomatic extension (i.e. extension by axiom schemas) of* $HCL_{\overset{\neg}{\leftrightarrow}}$. *Then* $\mathbf{L}$ *has the* RDP *for* $\leftrightarrow$.

Proof. Immediate from Theorem 1. ■

Theorem 5. $HCL_{\overset{\neg}{\leftrightarrow}}$ *is strongly sound for* $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$. *(I.e., if* $\mathcal{T} \vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} \varphi$ *then by interpreting* $\leftrightarrow$ *as the classical biconditional* $\leftrightarrow$, *and* $\neg$ *as the classical negation, we get that every assignment that satisfies* $\mathcal{T}$ *also satisfies* φ.*)*

Proof. Obviously, $[\mathrm{MP}]_{\leftrightarrow}$ is a valid rule of inference for the classical biconditional $\leftrightarrow$. It is also easy to check that every axiom of $HCL_{\overset{\neg}{\leftrightarrow}}$ becomes a classical tautology if $\leftrightarrow$ is interpreted as the classical biconditional. Hence $HCL_{\overset{\neg}{\leftrightarrow}}$ is strongly sound for the equivalence-negation fragment of classical logic. ■

Corollary 1. Let φ be a formula in the language of $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$.

1. There is a formula ψ in the language of $\mathbf{CL}_{\leftrightarrow}$ such that:
   - If the number of negations in φ is even then $\vdash_{\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}} \varphi \leftrightarrow \psi$.
   - If the number of negations in φ is odd then $\vdash_{\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}} \varphi \leftrightarrow \neg\psi$.
2. φ is a classical tautology iff the number of negations in φ is even, and for each atomic p the number of occurrences of p in φ is even too.[6]

[6] According to [Church, 1956], the second item of this Corollary has independently been observed by McKinsey and Mihailescu. As far as I know, the first item was first proved in [Mihailescu, 1937].

Proof. Obviously, $HCL_{\overset{\neg}{\leftrightarrow}}$ has the replacement property. (This is true for every axiomatic extension of $HLL_{\overset{\neg}{\leftrightarrow}}$.) It is also easy to show that $\theta_1 \leftrightarrow \neg\theta_2$ and $\neg\theta_1 \leftrightarrow \theta_2$ are both equivalent in $HCL_{\overset{\neg}{\leftrightarrow}}$ to $\neg(\theta_1 \leftrightarrow \theta_2)$. Using these two equivalences and axiom [N2], we can constructively find for φ a formula ψ in the language of $\mathbf{CL}_{\leftrightarrow}$ such that $\vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} \varphi \leftrightarrow \psi$ or $\vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} \varphi \leftrightarrow \neg\psi$ according to the parity of φ's number of negations. Hence Theorem 5 implies item 1.

For the second item, note that if ψ is in the language of $\mathbf{CL}_{\leftrightarrow}$, then $\neg\psi$ is not a tautology, since we can refute it by assigning t to all atomic formulas. Therefore the second part follows from the first, using Leśniewski's famous criterion for being a tautology in $\leftrightarrow$. (See e.g. Corollary 7.31.7 in [Humberstone, 2011].) ■

Theorem 6. *No Hilbert-type system which has* [MP] *for* $\leftrightarrow$ *as its sole rule of inference can be strongly sound and complete for* $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$.

Proof. Suppose that such a system H exists. From Theorem 5 it follows that we may assume that H is an axiomatic extension of $HCL_{\rightarrow}$. Hence Theorem 1 implies that H has the RDP. Therefore it follows from the strong soundness and completeness of H that so does $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$, This contradicts Proposition 1. ■

Corollary 2. $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}} \neq \mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$, i.e. $HCL_{\overset{\neg}{\leftrightarrow}}$ is not strongly complete for $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$.[7]

Next we present semantics for $HCL_{\overset{\neg}{\leftrightarrow}}$ for which this system *is* strongly sound and complete.

Definition 8. Let $\mathcal{L} = \{\leftrightarrow, \neg\}$. $\mathbf{CL}_{\{\leftrightarrow,id\}}$ is the two-valued logic which is obtained by interpreting $\leftrightarrow$ as the classical biconditional $\leftrightarrow$, and $\neg$ as the identity connective. $\mathcal{M}_{\mathbf{CL}_{\{\leftrightarrow,id\}}}$ is the two-valued matrix which induces $\mathbf{CL}_{\{\leftrightarrow,id\}}$.

For the next proof, we need the following easy lemma, which is directly proved in [Avron, 2015] (but also easily follows via an appeal to Lesniewski's criterion, mentioned at the end of the proof of Corollary 1).

Lemma 1.

1. $\vdash_{HCL_{\leftrightarrow}} (\varphi \leftrightarrow \psi) \leftrightarrow (\psi \leftrightarrow \varphi)$

2. $\vdash_{HCL_{\leftrightarrow}} \varphi \leftrightarrow (\psi \leftrightarrow (\varphi \leftrightarrow \psi))$

Theorem 7. $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}} = \mathbf{CL}_{\overset{\neg}{\leftrightarrow}} \cap \mathbf{CL}_{\{\leftrightarrow,id\}}$. *In other words:* $\mathcal{T} \vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} \varphi$ *iff* $\mathcal{T} \vdash_{\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}} \varphi$ *and also* $\mathcal{T} \vdash_{\mathbf{CL}_{\{\leftrightarrow,id\}}} \varphi$.

[7]This fact was first noticed in [Avron, 2020] (Corollary 14).

Proof. The strong soundness of $HCL_{\overset{\neg}{\leftrightarrow}}$ for $\mathbf{CL}_{\{\leftrightarrow,id\}}$ follows from Theorem 2. This and Theorem 5 imply the strong soundness of $HCL_{\overset{\neg}{\leftrightarrow}}$ for $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}} \cap \mathbf{CL}_{\{\leftrightarrow,id\}}$. To prove strong completeness, assume that $\mathcal{T} \nvdash_{HCL_{\overset{\neg}{\leftrightarrow}}} \theta$. Extend $\mathcal{T}$ to a maximal theory $\mathcal{T}^*$ such that $\mathcal{T}^* \nvdash_{HCL_{\overset{\neg}{\leftrightarrow}}} \theta$. Obviously, $\varphi \in \mathcal{T}^*$ iff $\mathcal{T}^* \vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} \varphi$, and $\varphi \notin \mathcal{T}^*$ iff $\mathcal{T}^*, \varphi \vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} \theta$. Therefore the RDP implies that

$$(*) \quad \varphi \notin \mathcal{T}^* \quad \text{iff} \quad \varphi \leftrightarrow \theta \in \mathcal{T}^*$$

Now define a valuation v as follows:

$$v(\varphi) = \begin{cases} t & \text{if } \varphi \in \mathcal{T}^* \\ f & \text{if } \varphi \notin \mathcal{T}^* \end{cases}$$

Obviously we have:

(**) $v(\varphi) = t$ for every $\varphi \in \mathcal{T}^*$, while $v(\theta) = f$.

Next we show that v respects the truth table of the classical biconditional.

- Suppose $v(\varphi) = v(\psi) = t$. Then $\varphi \in \mathcal{T}^*$ and $\psi \in \mathcal{T}^*$. Therefore it follows from the second item of Lemma 1 that $\mathcal{T}^* \vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} \varphi \leftrightarrow \psi$. Hence $\varphi \leftrightarrow \psi \in \mathcal{T}^*$, and so $v(\varphi \leftrightarrow \psi) = t$.

- Suppose $v(\varphi) = t$ and $v(\psi) = f$. Then $\varphi \in \mathcal{T}^*$, while $\psi \notin \mathcal{T}^*$. Because of the presence of $[\mathrm{MP}]_{\leftrightarrow}$, these facts immediately imply that $\varphi \leftrightarrow \psi \notin \mathcal{T}^*$, and so $v(\varphi \leftrightarrow \psi) = f$ in this case.

- Suppose $v(\varphi) = f$ and $v(\psi) = t$. By the previous item this implies that $\psi \leftrightarrow \varphi \notin \mathcal{T}^*$. Therefore the first item of Lemma 1 implies that $\varphi \leftrightarrow \psi \notin \mathcal{T}^*$, and so $v(\varphi \leftrightarrow \psi) = f$ in this case too.

- Suppose $v(\varphi) = v(\psi) = f$. Then $\varphi \notin \mathcal{T}^*$ and $\psi \notin \mathcal{T}^*$. By (*) above, it follows that $\varphi \leftrightarrow \theta \in \mathcal{T}^*$ and $\psi \leftrightarrow \theta \in \mathcal{T}^*$. By the first item of Lemma 1, the second fact implies that $\theta \leftrightarrow \psi \in \mathcal{T}^*$. Using [Tr] this last fact and the fact that $\varphi \leftrightarrow \theta \in \mathcal{T}^*$ together imply that $\varphi \leftrightarrow \psi \in \mathcal{T}^*$, and so $v(\varphi \leftrightarrow \psi) = t$ in this case.

To determine the behavior of v with respect to $\neg$ we have two cases to consider.

$\neg\theta \in \mathcal{T}^*$

- Suppose $v(\varphi) = t$. Then $\varphi \in \mathcal{T}^*$. By the second item of Lemma 1, this implies in this case that $\varphi \leftrightarrow \neg\theta \in \mathcal{T}^*$. It follows by [N1] that $\theta \leftrightarrow \neg\varphi \in \mathcal{T}^*$, and so $\neg\varphi \leftrightarrow \theta \in \mathcal{T}^*$ by the first item of Lemma 1. Hence (*) entails that $\neg\varphi \notin \mathcal{T}^*$, implying that $v(\neg\varphi) = f$ in this case.

- Suppose $v(\varphi) = f$. Then $\varphi \notin \mathcal{T}^*$. Hence (*) entails that $\varphi \leftrightarrow \theta \in \mathcal{T}^*$. By [N1] and [N2], this implies that $\neg\theta \leftrightarrow \neg\varphi \in \mathcal{T}^*$, and so $\neg\varphi \in \mathcal{T}^*$ (since we are assuming that $\neg\theta \in \mathcal{T}^*$). Therefore $v(\neg\varphi) = t$ in this case.

It follows that v is a legal valuation of the classical equivalence-negation matrix $\mathcal{M}_{CL^{\neg}_{\leftrightarrow}}$. Hence $\mathcal{T} \nvdash_{\mathbf{CL}^{\neg}_{\leftrightarrow}} \theta$ in this case.

$\neg\theta \notin \mathcal{T}^*$

- Suppose $v(\varphi) = t$. Then $\varphi \in \mathcal{T}^*$. Suppose that $\neg\varphi \notin \mathcal{T}^*$. Then by (*), $\neg\varphi \leftrightarrow \theta \in \mathcal{T}^*$, and so the first item of Lemma 1 implies that $\theta \leftrightarrow \neg\varphi \in \mathcal{T}^*$. Hence [N1] entails that $\varphi \leftrightarrow \neg\theta \in \mathcal{T}^*$. Since $\varphi \in \mathcal{T}^*$, $\neg\theta \in \mathcal{T}^*$ too. A contradiction. It follows that $\neg\varphi \in \mathcal{T}^*$, and so $v(\neg\varphi) = t$.
- Suppose $v(\varphi) = f$. Then $\varphi \notin \mathcal{T}^*$, implying by (*) that $\varphi \leftrightarrow \theta \in \mathcal{T}^*$. Hence the first item of Lemma 1 implies that $\theta \leftrightarrow \varphi \in \mathcal{T}^*$. Using [N1], [N2], and Lemma 1, we get from this that $\neg\varphi \leftrightarrow \neg\theta \in \mathcal{T}^*$. Since $\neg\theta \notin \mathcal{T}^*$, this means that $\neg\varphi \notin \mathcal{T}^*$ too, and so $v(\neg\varphi) = f$.

We have shown that v is a legal valuation of the classical equivalence-identity matrix $\mathcal{M}_{\mathbf{CL}_{\{\leftrightarrow,id\}}}$. Hence $\mathcal{T}^* \nvdash_{\mathbf{CL}_{\{\leftrightarrow,id\}}} \theta$ in this case.

It follows that if $\mathcal{T} \nvdash_{HCL^{\neg}_{\leftrightarrow}} \theta$ then either $\mathcal{T} \nvdash_{\mathbf{CL}^{\neg}_{\leftrightarrow}} \theta$ or $\mathcal{T} \nvdash_{\mathbf{CL}_{\{\leftrightarrow,id\}}} \theta$. ■

4. Proof-theoretical Characterizations of $\mathbf{CL}^{\neg}_{\leftrightarrow}$

Theorem 8. *$\mathcal{T} \vdash_{\mathbf{CL}^{\neg}_{\leftrightarrow}} \varphi$ iff $\mathcal{T}, \neg\varphi \vdash_{HCL^{\neg}_{\leftrightarrow}} \varphi$.*

Proof. Suppose $\mathcal{T} \vdash_{\mathbf{CL}^{\neg}_{\leftrightarrow}} \varphi$. Then $\mathcal{T}, \neg\varphi \vdash_{\mathbf{CL}^{\neg}_{\leftrightarrow}} \varphi$. Obviously, $\mathcal{T}, \neg\varphi \vdash_{\mathbf{CL}_{\{\leftrightarrow,id\}}} \varphi$ as well. Hence Theorem 7 implies that $\mathcal{T}, \neg\varphi \vdash_{HCL^{\neg}_{\leftrightarrow}} \varphi$.

For the converse, let $\mathcal{T}, \neg\varphi \vdash_{HCL^{\neg}_{\leftrightarrow}} \varphi$. Suppose that $\mathcal{T} \nvdash_{\mathbf{CL}^{\neg}_{\leftrightarrow}} \varphi$. Then there is a a valuation v in $\mathcal{M}_{CL^{\neg}_{\leftrightarrow}}$ such that $v(\psi) = t$ for every $\psi \in \mathcal{T}$, while $v(\varphi) = f$. The latter fact implies that $v(\neg\varphi) = t$, and so v is a model in $\mathcal{M}_{CL^{\neg}_{\leftrightarrow}}$ of $\mathcal{T} \cup \{\neg\varphi\}$ which is not a model in $\mathcal{M}_{CL^{\neg}_{\leftrightarrow}}$ of φ. Since $\mathcal{T}, \neg\varphi \vdash_{HCL^{\neg}_{\leftrightarrow}} \varphi$, this contradicts Theorem 5. ■

Theorem 9. *$HCL^{\neg}_{\leftrightarrow}$ is* weakly *complete for $\mathbf{CL}^{\neg}_{\leftrightarrow}$: $\vdash_{HCL^{\neg}_{\leftrightarrow}} \varphi$ iff $\vdash_{\mathbf{CL}^{\neg}_{\leftrightarrow}} \varphi$.*

Proof. Theorem 5 implies the 'only if' part. For the converse, let $\vdash_{\mathbf{CL}^{\neg}_{\leftrightarrow}} \varphi$. By Theorem 8, it follows that $\neg\varphi \vdash_{HCL^{\neg}_{\leftrightarrow}} \varphi$. Hence Theorem 4 implies that either $\vdash_{HCL^{\neg}_{\leftrightarrow}} \neg\varphi \leftrightarrow \varphi$, or $\vdash_{HCL^{\neg}_{\leftrightarrow}} \varphi$. The first option is impossible by Theorem 5, since $v(\neg\varphi \leftrightarrow \varphi) = f$ for every valuation v in $\mathcal{M}_{CL^{\neg}_{\leftrightarrow}}$. Hence $\vdash_{HCL^{\neg}_{\leftrightarrow}} \varphi$. ■

Remark 4. Theorem 9 easily follows also from Corollary 1 and Theorem 2.[8]

Theorem 10. *Let* $HCL^{\star}_{\overset{\neg}{\leftrightarrow}}$ *be obtained from* $HCL_{\overset{\neg}{\leftrightarrow}}$ *by adding to it the Ex Falso rule* $\frac{\neg\varphi \quad \varphi}{\psi}$ *as a rule of inference. Then* $HCL^{\star}_{\overset{\neg}{\leftrightarrow}}$ *is strongly sound and complete for* $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$.

Proof. That $HCL^{\star}_{\overset{\neg}{\leftrightarrow}}$ is strongly sound for $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$ follows from Theorem 5, and the strong soundness in classical logic of the special rule of $HCL^{\star}_{\overset{\neg}{\leftrightarrow}}$.

To show strong completeness, let $\mathcal{T} \vdash_{\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}} \varphi$. Then $\mathcal{T}, \neg\varphi \vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} \varphi$ by Theorem 8. By Theorem 4, either $\mathcal{T} \vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} \neg\varphi \leftrightarrow \varphi$, or $\mathcal{T} \vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} \varphi$. In the second case we are done. Assume the first. Since $\vdash_{\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}} \neg(\neg\varphi \leftrightarrow \varphi)$, it follows by Theorem 9 that $\mathcal{T} \vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} \neg(\neg\varphi \leftrightarrow \varphi)$ as well. Hence an application of the Ex Falso rule of $HCL^{\star}_{\overset{\neg}{\leftrightarrow}}$ yields that $\mathcal{T}^{\star} \vdash_{\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}} \varphi$. ■

Corollary 3.

1. If $\mathcal{T} \vdash_{HCL^{\star}_{\overset{\neg}{\leftrightarrow}}} \varphi$, then there is a proof of φ from $\mathcal{T}$ in $HCL^{\star}_{\overset{\neg}{\leftrightarrow}}$ that includes at most one application, made at the end of that proof, of its extra rule.
2. $\mathcal{T} \vdash_{\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}} \varphi$ iff either $\mathcal{T} \vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} \varphi$ or $\mathcal{T}$ is inconsistent in $HCL_{\overset{\neg}{\leftrightarrow}}$.

Proof. The first part easily follows from the proof of Theorem 10. The second one follows from that theorem together with the first part. ■

Turning to a corresponding Gentzen-type System, we note that in [Avron, Lev, 2005] an algorithm has been given for finding a cut-free sound and complete Gentzen-type system for every logic which has a two-valued characteristic matrix (or even non-deterministic matrix). By applying that algorithm to $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$, we get the system $GCL_{\overset{\neg}{\leftrightarrow}}$ presented at Figure 3. In this presentation Γ and Δ vary over finite *sets* of formulas.

Theorem 11.

1. $GCL_{\overset{\neg}{\leftrightarrow}}$ *is strongly sound and complete for* $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$.
2. *The cut-elimination theorem obtains for* $GCL_{\overset{\neg}{\leftrightarrow}}$*: if* $\vdash_{GCL_{\overset{\neg}{\leftrightarrow}}} \Gamma \Rightarrow \Delta$*, then* $\Gamma \Rightarrow \Delta$ *has a cut-free proof in* $GCL_{\overset{\neg}{\leftrightarrow}}$.

Proof. This is a special case of Theorem 4.7 of [Avron, Lev, 2005] and its proof. ■

[8] Theorem 9 was essentially first proved in [Mihailescu, 1937]. (See also [Bennett, 1937].) The system used there is easily seen to be equivalent to $HCL_{\overset{\neg}{\leftrightarrow}}$.

Axioms: $\varphi \Rightarrow \varphi$

Rules: cut, weakening, and the following logical rules:

$$\frac{\Gamma, \varphi, \psi \Rightarrow \Delta \quad \Gamma \Rightarrow \Delta, \varphi, \psi}{\Gamma, \varphi \leftrightarrow \psi \Rightarrow \Delta} \qquad \frac{\Gamma, \psi \Rightarrow \Delta, \varphi \quad \Gamma, \varphi \Rightarrow \Delta, \psi}{\Gamma \Rightarrow \Delta, \varphi \leftrightarrow \psi}$$

$$\frac{\Gamma \Rightarrow \Delta, \varphi}{\Gamma, \neg\varphi \Rightarrow \Delta} \qquad \frac{\Gamma, \varphi \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg\varphi}$$

Fig. 3. $GCL_{\overset{\neg}{\leftrightarrow}}$

Remark 5. A model of a sequent $\Gamma \Rightarrow \Delta$ in a two-valued matrix $\mathcal{M}$ is usually taken to be a valuation v in $\mathcal{M}$ such that $v(\varphi) = f$ for some $\varphi \in \Gamma$, or $v(\varphi) = t$ for some $\varphi \in \Delta$. Let $S \cup \{s\}$ be a set of sequents. Define: $S \vdash_{\mathcal{M}} s$ if every model of S in $\mathcal{M}$ is also a model of s. The first item of Theorem 11 means that $S \vdash_{\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}} s$ iff $S \vdash_{GCL_{\overset{\neg}{\leftrightarrow}}} s$.

5. The Expressive Power of $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$

To make our treatment of $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$ complete, we include also a characterization of the set of two-valued connectives that are definable in the language of $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$. For this it would be convenient to use 1 and -1 as our truth-values.

Definition 9. Let $H : \{1, -1\}^n \to \{1, -1\}$, and let $x_1, \ldots, x_n$ be n variables.

- For $1 \leq i \leq n$, x_i is a *dummy* variable of $H(x_1, \ldots, x_n)$ if

  $$H(x_1, \ldots, x_{i-1}, -x_i, x_{i+1}, \ldots, x_n) = H(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)$$

  for every $x_1, x_2, \ldots, x_n \in \{-1, 1\}$.

- For $1 \leq i \leq n$, x_i is a *flipping* variable of $H(x_1, \ldots, x_n)$ if

  $$H(x_1, \ldots, x_{i-1}, -x_i, x_{i+1}, \ldots, x_n) = -H(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)$$

  for every $x_1, x_2, \ldots, x_n \in \{-1, 1\}$.

Theorem 12. *A function $H : \{1, -1\}^n \to \{1, -1\}$ is definable by a formula of $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$ iff for every $1 \leq i \leq n$, x_i is either a dummy variable or a flipping variable of $H(x_1, \ldots, x_n)$.*[9]

[9]This theorem can most probably be extracted from Post's discussion in [Post, 1941].

Proof. That any function which is definable by a formula φ of $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$ satisfies the condition is easily proved by induction on the structure of φ.

For the converse, suppose $H : \{1,-1\}^n \to \{1,-1\}$ satisfies the condition. Let $\psi = \psi_1 \leftrightarrow \cdots \leftrightarrow \psi_n$, where $\psi_i = p_i \leftrightarrow p_i$ if x_i is a dummy variable of $H(x_1,\ldots,x_n)$, $\psi_i = p_i$ if x_i is a flipping variable of $H(x_1,\ldots,x_n)$. It is not difficult to see that ψ defines H in case $H(1,1,\ldots,1) = 1$, while $\neg\psi$ defines H in case $H(1,1,\ldots,1) = -1$. ■

Corollary 4. The classical conjunction, disjunction, and (material) implication are not definable by a formula of $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$.[10]

6. Extensions of $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$

Theorem 13. *Let $HCL^{id}_{\overset{\neg}{\leftrightarrow}}$ be obtained from $HCL_{\overset{\neg}{\leftrightarrow}}$ by adding to it $\varphi \leftrightarrow \neg\varphi$ as an axiom. Then $HCL^{id}_{\overset{\neg}{\leftrightarrow}}$ is strongly sound and complete for $\mathbf{CL}_{\{\leftrightarrow,id\}}$.*

Proof. The proof is almost identical to that of Theorem 7. We only need to observe that in the presence of the additional axiom, the case in which $\neg\theta \in \mathcal{T}^*$ is impossible. Hence we remain with the case in which $\mathcal{T}^* \not\vdash_{\mathbf{CL}_{\{\leftrightarrow,id\}}} \theta$. ■

Theorem 14. $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$ *and* $\mathbf{CL}_{\{\leftrightarrow,id\}}$ *are the sole non-trivial proper extensions of* $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$ *in its language.*

Proof. Let $\mathbf{L}$ be a logic in the language of $HCL_{\overset{\neg}{\leftrightarrow}}$ which is a proper extension of $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$. Then there is a theory $\mathcal{T}$ and a formula θ such that $\mathcal{T} \vdash_{\mathbf{L}} \theta$, but $\mathcal{T} \not\vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} \theta$. By Theorem 7 there is an assignment v in either $\mathcal{M}_{CL_{\overset{\neg}{\leftrightarrow}}}$ or $\mathcal{M}_{\mathbf{CL}_{\{\leftrightarrow,id\}}}$ such that $v(\varphi) = t$ for every $\varphi \in \mathcal{T}$, while $v(\theta) = f$. Pick some atomic formula p, and define a substitution S by:

$$S(q) = \begin{cases} p \leftrightarrow p & \text{if } v(q) = t \\ p & \text{if } v(q) = f \end{cases}$$

Let $\mathcal{T}^\star = \{S(\varphi) \mid \varphi \in \mathcal{T}\}$, $\theta^\star = S(\theta)$. Since $\mathbf{L}$ is a logic, $\mathcal{T}^\star \vdash_{\mathbf{L}} \theta^\star$. On the other hand, the obvious fact that $v(S(\varphi)) = v(\varphi)$ for every formula φ implies that $v \models \mathcal{T}^\star$, while $v \not\models \theta^\star$. It follows by Theorem 7 that $\mathcal{T}^\star \not\vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} \theta^\star$.

Now from Theorem 9 it follows (and it can also easily be shown directly) that every formula which has p as its sole atomic subformula is equivalent in $HCL_{\overset{\neg}{\leftrightarrow}}$ to one of the formulas in $\{p, \neg p, p \leftrightarrow p, \neg(p \leftrightarrow p)\}$. Hence $\mathcal{T}^\star \cup \{\theta^\star\}$ is a subset of this set. Moreover, the fact that $\vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} p \leftrightarrow p$ implies that $\theta^\star \neq p \leftrightarrow p$, and

[10] This result has been shown directly in [Massey, 1977]. It is also proved there that there is no single truth-function that generates precisely the functions definable by formulas of $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$.

that we may assume that $p \leftrightarrow p \notin \mathcal{T}^\star$. Hence $\mathcal{T}^\star \cup \{\theta^\star\} \subseteq \{p, \neg p, \neg(p \leftrightarrow p)\}$. However, since $\vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} p \leftrightarrow \neg p \leftrightarrow \neg(p \leftrightarrow p)$ by Theorem 9, any element of $\{p, \neg p, \neg(p \leftrightarrow p)\}$ follows in $HCL_{\overset{\neg}{\leftrightarrow}}$ from the other two. Therefore we remain with the following six cases.

$p \vdash_{\mathbf{L}} \neg(p \leftrightarrow p)$: Substituting $p \leftrightarrow p$ for p, we get that $\vdash_{\mathbf{L}} \neg((p \leftrightarrow p) \leftrightarrow (p \leftrightarrow p)$. But by Theorem 9, $\neg((p \leftrightarrow p) \leftrightarrow (p \leftrightarrow p)$ is equivalent in $HCL_{\overset{\neg}{\leftrightarrow}}$ to $\neg p \leftrightarrow p$, and so $\vdash_{\mathbf{L}} \neg p \leftrightarrow p$. Hence $\mathbf{L}$ extends $\mathbf{CL}_{\{\leftrightarrow, id\}}$.

$p \vdash_{\mathbf{L}} \neg p$: Substituting $p \leftrightarrow p$ for p, we again get that $\mathbf{CL}_{\{\leftrightarrow, id\}} \subseteq \mathbf{L}$.

$\neg p \vdash_{\mathbf{L}} \neg(p \leftrightarrow p)$: Substituting $\neg(p \leftrightarrow p)$ for p, we again get that $\mathbf{CL}_{\{\leftrightarrow, id\}} \subseteq \mathbf{L}$.

$\neg p \vdash_{\mathbf{L}} p$: Substituting $\neg(p \leftrightarrow p)$ for p, we again get that $\mathbf{CL}_{\{\leftrightarrow, id\}} \subseteq \mathbf{L}$.

$\neg(p \leftrightarrow p) \vdash_{\mathbf{L}} p$: Since $\vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} q \leftrightarrow \neg q \leftrightarrow (\neg(p \leftrightarrow p))$ by Theorem 9, we get that $q, \neg q \vdash_{\mathbf{L}} p$. Hence $\mathbf{L}$ extends $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$ in this case (by Theorem 10).

$\neg(p \leftrightarrow p) \vdash_{\mathbf{L}} \neg p$: Here we get by a similar argument that $q, \neg q \vdash_{\mathbf{L}} \neg p$. Substituting $\neg p$ for p, it follows (using [N2]) that $q, \neg q \vdash_{\mathbf{L}} p$. Hence $\mathbf{L}$ extends $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$ in this case too.

We have shown that $\mathbf{L}$ extends either $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$ or $\mathbf{CL}_{\{\leftrightarrow, id\}}$. However, since these are two-valued logics, neither of them has a proper non-trivial extension, by a general theorem in [Rautenberg, 1981]. (These two cases can also be shown directly using an analysis which is very similar to — though shorter and easier than — that given above for $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$.) Hence $\mathbf{L}$ is necessarily one of them. ■

Corollary 5. $HCL^{id}_{\overset{\neg}{\leftrightarrow}}$ is the sole non-trivial proper axiomatic extension of $HCL_{\overset{\neg}{\leftrightarrow}}$.[11]

Remark 6. Recall that theorems 2 and 3 imply Prior's result that $HCL_{\leftrightarrow}$ is Post-complete (footnote 4). Corollary 5 means that in contrast, $HCL_{\overset{\neg}{\leftrightarrow}}$ is not Post-complete.[12] However, the difference from $HCL_{\leftrightarrow}$ is small: $HCL_{\overset{\neg}{\leftrightarrow}}$ has just one proper axiomatic extension.

7. Is $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$ a Paraconsistent Logic?

In Chapter 2 of [Avron et al., 2018] a propositional $\mathbf{L}$ for a language with a unary connective $\neg$ is defined to be $\neg$-paraconsistent if $p, \neg p \nvdash_{\mathbf{L}} q$ whenever p and q are distinct variables, and $\neg$ *is a negation of* $\mathbf{L}$. $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$ certainly satisfies

[11] This fact was first proved in [Avron, 2020] (Theorem 16).

[12] That $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$ has no Post-complete axiomatization has already been noted in [Prior, 1962].

the first condition.[13] Hence the question whether it is $\neg$-paraconsistent depends on whether its connective $\neg$ can be viewed as a negation. This, in turn, depends of course on the definition of negation that one adopts.

In the literature one can find many different definitions of "negation" in $\mathbf{L}$. Some make very weak demands. The minimal ones might be that if p is atomic then $p \nvdash_{\mathbf{L}} \neg p$ and $\neg p \nvdash_{\mathbf{L}} p$. A more extensive set of negative conditions of this sort (divided into two groups, called 'verificatio' and 'falsificatio') is given in [Marcos, 2005]. It is also possible to add some positive conditions, like that $p \vdash_{\mathbf{L}} \neg\neg p$ and $\neg\neg p \vdash_{\mathbf{L}} p$. All these conditions are satisfied in $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$. So according to weak definitions of this sort, $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$ is indeed a $\neg$-paraconsistent logic.

In [Avron et al., 2018], a more restrictive definition of negation has been used. $\neg$ is called there a negation for $\mathbf{L}$ if it is possible to define in the language of $\mathbf{L}$ a binary connective $\diamond$ which is either a disjunction for $\mathbf{L}$, or a conjunction for $\mathbf{L}$, or a semi-implication for $\mathbf{L}$, such that the $\{\neg, \diamond\}$-fragment of $\mathbf{L}$ is contained in the corresponding fragment of classical logic. In order to see that no such connective $\diamond$ is available in $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$, we do not need to repeat the definitions given in Chapter 1 of [Avron et al., 2018] of these notions. It suffices to recall the following facts about them. (They all easily follow from the definitions.)

- If a connective $\wedge$ is a conjunction for a logic $\mathbf{L}$, then for every φ and ψ: $\varphi \wedge \psi \vdash_{\mathbf{L}} \varphi$, $\varphi \wedge \psi \vdash_{\mathbf{L}} \psi$, and $\varphi \vdash_{\mathbf{L}} \varphi \wedge \varphi$. From Theorem 7 it easily follows that such a connective should have in $\mathcal{M}_{CL_{\overset{\neg}{\leftrightarrow}}}$ the truth-table of the classical conjunction. Hence Corollary 4 implies that no such connective is available in $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$.
- If a connective $\vee$ is a disjunction for a logic $\mathbf{L}$, then for every φ and ψ: $\varphi \vdash_{\mathbf{L}} \varphi \vee \psi$, $\psi \vdash_{\mathbf{L}} \varphi \vee \psi$, and $\varphi \vee \varphi \vdash_{\mathbf{L}} \varphi$. From Theorem 7 it easily follows that such a connective should have in $\mathcal{M}_{CL_{\overset{\neg}{\leftrightarrow}}}$ the truth-table of the classical disjunction. Hence Corollary 4 implies that no such connective is available in $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$.
- If a connective $\supset$ is a semi-implication for a logic $\mathbf{L}$, then it has in $\mathbf{L}$ the RDP. Suppose now that $\supset$ is such a connective in $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$. Then $\vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} p \supset p$ for every atomic p. Hence $\vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} (p \supset q) \supset (p \supset q)$ (by structurality). Therefore two applications of the RDP for $\supset$ yield $p, p \supset q \vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} q$. By two other applications of the RDP, this time for $\leftrightarrow$,

[13] In fact, $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$ satisfies the stronger condition $p, \neg p \nvdash_{\mathbf{L}} \neg q$. So if we accept its connective $\neg$ as a negation, then according to [Avron et al., 2018] it would even be *strongly* $\neg$-paraconsistent.

we get that $\vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} (p \supset q) \leftrightarrow (p \leftrightarrow q)$. It follows that $\varphi \supset \psi$ and $\varphi \leftrightarrow \psi$ are equivalent in $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$. Therefore $\vdash_{HCL_{\overset{\neg}{\leftrightarrow}}} (p \supset (p \supset q)) \supset q$. Hence the $\{\neg, \supset\}$-fragment of $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$ is not contained in the corresponding fragment of classical logic.

It follows from the above considerations that $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$ is not a paraconsistent logic according to the definition used in [Avron et al., 2018].

Acknowledgements. I am very grateful to Lloyd Humberstone for many most helpful comments, suggestions, and pointers to the literature. This research was supported by The Israel Science Foundation (grant no. 817-15).

References

Anderson, Belnap, 1975 – Anderson, A.R., Belnap, N.D. *Entailment: The Logic of Relevance and Necessity.* Vol. I. Princeton University Press, 1975.

Avron, 1988 – Avron, A. "The Semantics and Proof Theory of Linear Logic", *Theoretical Computer Science*, 1988, Vol. 57, pp. 161–184.

Avron, 2015 – Avron, A. "Semi-implication: A chapter in universal logic", *The Road to Universal Logic, Volume I.*, ed. by A. Koslow and A. Buchsbaum. Studies in Universal Logic, Birkhäuser, 2015, pp. 59–72.

Avron, 2020 – Avron, A. "A Note on Semi-implication with Negation", *Abstract Consequence and Logics: Essays in Honor of Edelcio G. de Souza*, ed. by A. Costa-Leite. Tributes 42, College Publications, 2020, pp. 221–226.

Avron et al., 2018 – Avron, A., Arieli, O., Zamansky, A. *Theory of Effective Propositional Paraconsistent Logics.* volume 75 of Studies in Logic, (Mathematical Logic and Foundations), College Publications, 2018.

Avron, Lev, 2005 – Avron, A., Lev, I. "Non-Deterministic Multiple-valued Structures", *Journal of Logic and Computation*, 2005, Vol. 15, pp. 241–261.

Bennett, 1937 – Bennett, A.A. Review of [Mihailescu, 1937], *Journal of Symbolic Logic*, 1937, Vol. 2, No. 4, p. 173.

Church, 1956 – Church, A. *Introduction to Mathematical Logic.* Princeton University Press, 1956.

Diaz, 1980 – Diaz, M.R. "Deductive Completeness and Conditionalization in Systems of Weak Implication", *Notre Dame Journal of Formal Logic*, 1980, Vol. 21, pp. 119–130.

Dunn, Restall, 2002 – Dunn, J.M., Restall, G. "Relevance logic", *Handbook of Philosophical Logic*, Second edition, Vol. 6, ed. by D. Gabbay and F. Guenthner. Kluwer, 2002 pp. 1–136.

Girard, 1987 – Girard, J.Y. "Linear logic", *Theoretical Computer Science*, 1987, Vol. 50, pp. 1–102.

Humberstone, 2011 – Humberstone, L. *The Connectives.* The MIT Press, 2011.

Marcos, 2005 – Marcos, J. "On negation: Pure local rules", *Journal of Applied Logic*, 2005, Vol. 3, pp. 185–219.

Massey, 1977 – Massey, G.J. "Negation, Material Equivalence, and Conditioned Nonconjunction: Completeness and Duality", *Notre Dame Journal of Formal Logic*, 1977, Vol. 18, pp. 140–144.

Mihailescu, 1937 – Mihailescu, E.Gh. "Recherches sur la negation el l'équivalence dans le calcul des proposition", *Annales scientifiques de l'Université de Jassy*, 1937, Vol. 23, pp. 369–408.

Post, 1941 – Post, E. *The Two-Valued Iterative Systems of Mathematical Logic*. Annals of Mathematics Series, Vol. 5, Princeton University Press, Princeton, NJ, 1941.

Prior, 1962 – Prior, A.N. *Formal Logic*, second edition. Oxford, 1962.

Rautenberg, 1981 – Rautenberg, W. "2-Element Matrices", *Studia Logica*, 1981, Vol. 40, pp. 315–353.

Логические исследования
2021. Т. 27. № 1. С. 46–63
УДК 16+510.6

Logical Investigations
2021, Vol. 27, No. 1, pp. 46–63
DOI: 10.21146/2074-1472-2021-27-1-46-63

Gilberto Gomes

Negation of Conditionals in Natural Language and Thought

Gilberto Gomes
Northern Rio de Janeiro State University,
CCH, PGCL 28013-602, Campos RJ, Brazil.
E-mail: ggomes@uenf.br

Abstract: External negation of conditionals occurs in sentences beginning with 'It is not true that if' or similar phrases, and it is not rare in natural language. A conditional may also be denied by another with the same antecedent and opposite consequent. Most often, when the denied conditional is implicative, the denying one is concessive, and vice versa. Here I argue that, in natural language pragmatics, 'If A, $\sim B$' entails '$\sim$(if A, B)', but '$\sim$(if A, B)' does not entail 'If A, $\sim B$'. 'If A, B' and 'If A, $\sim B$' deny each other, but are contraries, not contradictories. Truth conditions that are relevant in human reasoning and discourse often depend not only on semantic but also on pragmatic factors. Examples are provided showing that sentences having the forms '$\sim$(if A, B)' and 'If A, $\sim B$' may have different pragmatic truth conditions. The principle of Conditional Excluded Middle, therefore, does not apply to natural language use of conditionals. Three squares of opposition provide a representation the aforementioned relations.

Keywords: External negation, Implicative conditionals, Concessive conditionals, Pragmatic truth conditions, Boethius' theses, Conditional Excluded Middle, Contraries, Contradictories, Squares of opposition

For citation: Gomes G. "Negation of Conditionals in Natural Language and Thought", *Logicheskie Issledovaniya / Logical Investigations*, 2021, Vol. 27, No. 1, pp. 46–63. DOI: 10.21146/2074-1472-2021-27-1-46-63

1. Introduction

Negation in natural language involves complexities that are not at first sight evident. Horn [Horn, 2001, xiii] notes that, by contrast with 'the simplicity of the one-place connective of propositional logic (...), the form and function of negative statements in ordinary language are far from simple'. Quantification is one reason for the need to distinguish contrary from contradictory opposition. *All A is B* and *No A is B* are contraries, since they cannot be both true, but not contradictories, since they can be both false. The contradictory of *All A is B* is *Some A is not B*, which is entailed by *No A is B*, while the contradictory of the latter is *Some A is B*, which is entailed by *All A is B*. This set of relations is summarized in the well-known *square of opposition.* Another complication involves the Aristotelian distinction between *term negation* (*A is not-B*) and *predicate denial* (*A is not B*), and the distinction between these and the external negation of the Stoics (*Not: A is B*), to which the modern Fregean negation can be traced [Horn, 2001, pp. 14–30].

Horn & Wansing [Horn and Wansing, 2020] note that we do not usually find negation in natural languages in the place propositional logic would lead us to look, that is, in "sentence- or clause- peripheral position, as an external one-place connective (...)". However, external negation can and does occur in natural languages by means of the phrases *it's not the case that*, *it isn't true that*, *it's false that* or similar ones. Horn [Horn, 2001, pp. 364–365] views sentences using such phrases as cases of metalinguistic negation, and notes that they do not guarantee a presupposition-free reading of the negated sentence. However, this does not exclude that they are a form of external negation, nor does it preclude their use in our discussion of the negation of conditionals.

Thus, we can identify a form of *external negation* which consists in a complex sentence in which the main clause denies the truth of the proposition expressed in the subordinate clause. For example:

(1) *It's not true that Phil was a good student.*

Sentences of this type are usually employed to deny something that was said by someone else, or something that might be thought by someone else. In this sense, they are metalinguistic. Most frequently, negation of a sentence is displayed as *internal negation*:

(2) *Phil wasn't a good student.*

Although (1) and (2) would be used in different situations, to express different speaker intentions, as far as truth/falsity conditions are concerned, it can be accepted that they usually express propositions having the same truth/falsity conditions: if Phil was not a good student, the proposition they both usually express is true; if Phil was a good student, this proposition is false.

In the case of complex sentences with a subordinate clause, negation is usually obtained by adding the negative particle to the verb of the main clause. For example:

(3) A: *When he lived with his father, Phil was a good student.*
B: *No. When he lived with his father, Phil wasn't a good student.*

As far as truth/falsity conditions are concerned, (3)B is usually equivalent to:

(4) *No. It isn't true that, when he lived with his father, Phil was a good student.*

The subordinate clause in examples (3) and (4) is an adverbial clause (of time). The antecedent of a conditional sentence is also an adverbial clause. However, conditional clauses are a peculiar type of adverbial clause. The proposition expressed by the main clause of a sentence that has an adverbial clause of time, for example, is usually independently asserted by the utterance of the sentence. If we suppress the adverbial in (3) A, for example, we are left with *Phil was a good student.* We no longer know when A asserts this to have been the case, but we still have the information that Phil was a good student, at some time in the past, according to A. Not so with conditional sentences. Consider for example:

(5) *If you study, you'll pass.*

The person who utters this sentence is not thereby asserting that you will pass. If we suppress the adverbial in (5), we are left with the sentence *You'll pass*, and the proposition usually expressed by this sentence was not independently asserted by the utterance of the original sentence. This may be the reason why external and internal negations of a conditional may have different meanings:

(6) *It's not true that if you study, you'll pass.*

(7) *If you study, you won't pass.*

One way to explain this difference is to accept that external negation involves a denial of the conditional relation between the two clauses, which is not equivalent to a conditional denial of the proposition expressed by the main clause, as given by internal negation. What is being denied in (6) is a dependable relation between studying and passing. (7), by contrast, either suggests a paradoxical causal relation between studying and not passing, or is intended to express a concessive conditional (*Even if you study, you still won't pass*). With regard to truth/falsity conditions, if in reality you study and pass, (7) will have been shown to be false, but not (6). According to (6), you might have studied and not passed, but the possibility of studying and passing was not excluded.

Some philosophers have argued that the natural way to deny a conditional is to deny its consequent. According to this view, external negation of a conditional is exceptional, and when it occurs, it has the same meaning as the conditional presenting the negation of the original consequent. Thus, the negation of a conditional is in fact a conditional negation of its consequent. I argue in this paper that this is wrong. In section 6, I give a long list of attested examples of external negation in three languages, and in many of these there is a clear difference in meaning, including truth/falsity conditions, between the external negation of a conditional and the conditional with a negated consequent.

It is important to make clear that this is not a paper about the semantics of conditionals, but rather about their pragmatics, and its interest for logic. There is no consensus among the authors on how the distinction between semantics and pragmatics should be drawn. Here I will treat semantics as the study of the coded meaning of words and sentences, that is, of that part of their meaning that pertains to the language spoken by a certain population at a certain extended period of time and is relatively independent of the context of utterance and of the intentions of the speaker. By contrast, pragmatics here refers to the use of language by a speaker, or the understanding of language by a listener, in a certain context. It studies the meaning intended by the speaker and understood by the listener, which depends on the particular contexts in which words and sentences are used.

Many philosophers of language believe that the truth conditions of sentences are entirely provided by semantics. Contrary to this prevalent conception, I accept that pragmatics is sometimes necessary to establish the truth conditions of a sentence (for a defense of truth-conditional pragmatics, see [Recanati, 2010]). In particular, the point of view adopted in this article is that semantics must be supplemented by pragmatics in order to be able to explain how conditionals are used in human reasoning. Accordingly, in this paper I repeatedly refer to the meaning intended by the speaker when s/he uses a conditional, rather than to the meaning that a conditional sentence may have in itself.

In the following, 'negation of a conditional' is understood as the external negation described above applied to a conditional sentence. By contrast, a conditional 'with opposite consequent' is one presenting internal negation of an originally affirmative consequent, or an affirmative consequent instead of an originally negative one.

2. A conditional entails the negation of a conditional with the same antecedent and opposite consequent

A distinction between two types of conditionals will be essential in the following discussion. This is the distinction between concessive and implicative conditionals.[1] In a recent paper, I have argued that this distinction is best viewed as a pragmatic distinction, not as a semantic one [Gomes, 2020]. There are semantic elements that contribute to a conditional receiving either a concessive or an implicative interpretation; for example, the presence of *then* for the latter, and of *even if* for the former. These semantic elements do not guarantee the respective interpretations, however, since there are pragmatically concessive conditionals without *even if* (some speakers may even admit in them the presence of *then*), and pragmatically implicative conditionals with *even if* [Gomes, 2020].

From a pragmatic point of view, implicative conditionals are those in which the truth of the antecedent appears as a sufficient condition for the truth of the consequent, in a given context [Gomes, 2009]. They are the conditionals that are most interesting from the point of view of logic, since they are the ones that are used in making inferences. By contrast, concessive conditionals, from the pragmatical point of view, are those in which the consequent is asserted as true (counterexamples in the literature are pragmatically implicative, though they have *even if*). In them, the antecedent usually expresses a condition that is somehow opposed to what the consequent expresses. While an implicative conditional conveys that the truth of its antecedent is a sufficient condition for the truth of the consequent, a concessive conditional conveys that the truth of its antecedent is an insufficient condition for the falsity of its consequent.

In so-called classic logic, the negation of a conditional entails the truth of its antecedent and the falsity of its consequent, a result that does not agree with the use of conditionals in natural language. This is a reason for those who think that the material conditional expresses the truth conditions of natural language conditionals to say that the negation of a conditional in natural language really means a conditional with opposite consequent. This is empirically false, however, or so I argue. Another interesting fact about classic logic is that it does not support Boethius' theses:

$$(A \to B) \to \sim(A \to \sim B) \text{ and } (A \to \sim B) \to \sim(A \to B)$$

Intuitively, however, these entailments are very appealing and seem to be supported in inferential reasoning and in natural language use of conditionals. Let us begin our examination of the relation between the a conditional and the negation of a conditional with the same antecedent and opposite consequent by the case in which the former conditional is concessive and the latter implicative. For example:

[1] Implicative conditionals have been called *standard* or *prototypical* by [Ducrot, 1972], *genuine* by J. Barker [Barker, 1973], *strong* by [Davis, 1983], *robust* by [Lycan, 2001] and *implicative* by [Declerck and Reed, 2001].

(8) *Even if she goes, I won't go.*

(9) *It's not the case that if she goes, then I'll go.*

In human reasoning and discourse, it seems natural to agree that (8) entails (9). The fact that a concessive conditional entails the negation of an implicative conditional with opposite consequent was noted by [Chisholm, 1946, p. 301] and [Goodman, 1947, p. 114]. Later developments in the theory of conditionals, however, tended to obscure the distinction between implicative and concessive conditionals, since it is difficult to attest by a purely semantic approach, and this observation by Chisholm and Goodman tended to be forgotten.

However, it is also to be noted that, conversely, implicative conditionals entail the negation of a concessive one with opposite consequent, as shown in the following examples.

(10) *If he comes, then, as a consequence, I'll leave. Therefore: It is not the case that even if he comes, I won't leave.*

(11) *If butter is heated to 150° F, then it melts. Therefore: It is not the case that even if butter is heated to 150° F, it still doesn't melt.*

(12) *If she had been invited, then she'd be here. Therefore: It is not the case that even if she had been invited, she wouldn't be here.*

This means that a concessive conditional may be denied by an implicative one with opposite consequent, and conversely, an implicative conditional may be denied by a concessive one with opposite consequent.

3. The contradictory of a conditional

The fact noted in the previous section raises the question of what the contradictory of a conditional is. A sure way to form the contradictory of a conditional, of course, is to use 'It is not the case that' before it:

(13) *If it rains, the match will be cancelled.*

(14) *It is not the case that if it rains, the match will be cancelled.*

According to Stalnaker's formal system, "the denial of a conditional is equivalent to a conditional with the same antecedent and opposite consequent (provided that the antecedent is not impossible)" [Stalnaker, 1968, pp. 48–49].

This equivalence involves the principle of *conditional excluded middle*, according to which either *If* A, B or *If* $A, \sim B$ must be true. So (14) would be equivalent to:

(15) *If it rains, the match will not be cancelled.*

According to this equivalence, (13) and (15) are contradictories. This example seems to confirm the proposed equivalence. It might be objected, however, that someone who asserts (14) may not have (15) in mind, but rather the following:

(16) *If it rains, the match may be cancelled (if it rains heavily), but it may also not be cancelled (if it rains lightly).*

In this case, (13) and (15) would be neither true nor false, because the expression *it rains*, in both of them, is not sufficiently precise for the case being considered. [Stalnaker, 1981] argues that conditional excluded middle is a principle of the abstract semantics of conditionals that may fail to apply to natural language sentences due to semantic indeterminacy. Some sentences should thus be considered as neither true nor false. However, there are examples for which Stalnaker's equivalence fails and do not seem to involve any indeterminacy.

(17) A:*If Bill goes to the party, Mary will go.*
B:*No. If Bill goes to the party, Mary won't go.*

From a pragmatical point of view, B may have three different reasons for asserting the conditional present in his answer: (i) it may be intended as a concessive conditional, equivalent to: *Even if Bill goes to the party, Mary still won't go*; (ii) it may be asserted because B thinks that *if Bill goes to the party, it is because Mary will not go* (since he will only go if she does not); and (iii) it may be intended as equivalent to: *If Bill goes to the party, then as a consequence Mary won't go* (e.g. because she does not want to meet him).[2]

Now, suppose that in fact Mary will go to the party if she finishes her work in time, regardless of Bill's going or not, and that Bill will go to the party if he gets better from his cold, regardless of Mary's going or not. In these circumstances, A's conditional is false since it is possible that Bill goes to the party and Mary does not. Consequently, its external negation is true:

(18) *It is not the case that if Bill goes to the party, Mary will go.*

[2]This means that (17) B will have different speaker meanings according to these three possibilities. The same sentence, with arguably the same semantic meaning, would be used in three different contexts with different speaker meanings. This difference is not relevant, however, to the argument made here, and is mentioned only to be sure that all possibilities of interpretation have been covered.

However, in the same circumstances, B's conditional is also false, whether it is asserted for any of the three reasons mentioned above, because Bill and Mary may very well both go to the party.

Now, is there any indeterminacy in either conditional that might justify their being considered as neither true nor false? In its natural interpretation, A's conditional establishes a relation between Bill's going to the party and Mary's going to the party. B's conditional may be interpreted as establishing the lack of such a relation (if interpreted as a concessive conditional) or as establishing a relation between Bill's going to the party and Mary's not going to the party. Saying that someone will go or not to a party does not involve any vagueness (at least not any that is relevant to the evaluation of the conditional). It will definitely prove true or false when the party occurs. As to the relation in question, different theories of conditionals will explain it differently. However, it is difficult to see what indeterminacy might be present in it that would justify these conditionals being considered as neither true nor false.

Of course, assuming Stalnaker's theory, one can always say that A's and B's conditionals in (17) do not establish in a determinate way the world in which Bill goes to the party that is most similar to ours, and are therefore neither true nor false. However, the argument is question-begging. It presupposes the theory that we want to put to test. Moreover, it contradicts the intuition that each of these sentences must be either true or false.

In order to make this point clearer, let's compare (13)–(15) with (17)–(18). There is a relevant vagueness in *If it rains* which accounts for the indeterminacy of (13) and (15) in truth value. If we precisify this antecedent in a way suitable for the context considered, e.g. with *heavily* or *lightly*, the indeterminacy vanishes. According to what was supposed in (16), if it rains heavily, the match will be cancelled, while if it rains lightly, it will not.

Regarding (17) (*If Bill goes to the party, Mary will/won't go*), by contrast, there is no available precisification of its antecedent that would render either conditional true in the context considered. If in reality Mary will go to the party if she finishes her work in time, regardless of Bill's going or not, there is no relevant vagueness in the conditional clause *If Bill goes to the party.*

We conclude that, pragmatically, *the contradictory of a conditional is not a conditional with same antecedent and opposite consequent.* This means that with regard to their use in reasoning and discourse, two conditionals with the same antecedent and opposite consequents are contraries, because they cannot be both true, but may be both false. A more complex form using *It's not the case that* or *It's not true that* is necessary for making the contradictory of a conditional.

4. A concessive conditional denies an implicative conditional with the same antecedent and opposite consequent and vice versa

Going back to the rain/match example, it is certain that (15) (*If it rains, the match will not be cancelled*) can also be used to deny (13) (*If it rains, the match will be cancelled*), because – in accordance with Boethius' thesis – it entails (14) (*It is not the case that if it rains, the match will be cancelled*). It was probably on the basis of such uses that Stalnaker proposed the previously mentioned equivalence, failing to acknowledge the difference between contrary and contradictory denials involved here.

However, it must be noted that, in their most natural pragmatic interpretation, (13) suggests a context in which it is used as an implicative conditional, while (15), in the same context, would be used as a concessive conditional, meaning the same as:

(19) *Even if it rains, the match will not be cancelled.*

However, both could have a different pragmatic meaning, in a different context. Suppose that the president of the club has asked the players to paint the exterior walls of the club and the best time to do this is the time the match is scheduled to be played. So the match will probably be cancelled. Then one of the players, who is eager to play, utters (15) (*If it rains, the match will not be cancelled*). Since painting the exterior walls would not be possible under the rain, he reasons, the match will not be cancelled if it rains. In this context, (15) is not pragmatically concessive, but implicative. This is shown by the fact that it would accept a paraphrase with *then, as a consequence*:

(20) *If it rains [oh, how nice!], then, as a consequence, the match will not be cancelled [since we won't be able to paint the exterior walls of the club].*

What conditional could be used to negate this sentence, in this context? Suppose a second player replies to the one who uttered (13) or (20):

(21) *No, you're wrong. If it rains, the match will be cancelled. [Because the interior walls of the club also need painting, and the rain will not prevent painting them.]*

We see that here it is (13) (*If it rains, the match will be cancelled*) that is used to deny (15) (*If it rains, the match will not be cancelled*), and not vice versa. The painting vs. playing context, however, imposes an interpretation to (13) which is different from the usual one. Here, the conditional does not accept a paraphrase with *then, as a consequence*. By contrast, it means the same as:

(22) *Even if it rains, the match will still be cancelled.*

We conclude that, when a conditional is used to deny another with the same antecedent and opposite consequent, if the latter is implicative, the former will usually be concessive. Moreover, when it is a concessive conditional that is denied by another with the same antecedent and opposite consequent, the latter will usually be implicative, as shown in the following example:

(23) A: *If it rains, the match won't be cancelled.*
B: *No. If it rains, the match will be cancelled.*

In their most probable interpretation, (23)A is pragmatically a concessive conditional (equivalent to a paraphrase with *Even if*) and (23)B is in this case an implicative conditional (equivalent to a paraphrase with *then as a consequence*).

There are also cases, however, in which an implicative conditional *If* A, B may be denied by a conditional *If* $A, {\sim}B$ that is also interpreted as implicative, but one needs additional reason for the latter interpretation. For example:

(24) A: *If Bob comes, Linda will come—because she loves him.*
B: *No, if Bob comes, then Linda WON'T come—because he has rejected her.*

Here, speaker B does not mean that Bob's coming will be an insufficient reason for Linda to come and [therefore] she will not come, as B would if s/he had used a concessive conditional. B is suggesting instead that Bob's coming, if it occurs, will be a sufficient reason for Linda NOT to come, that is, to refrain from coming (for the reason mentioned, that he has rejected her), even if she might have a different reason to do so. Although usual, it is not necessary, therefore, that the two conditionals with the same antecedent and opposite consequents that deny each other are one implicative and other concessive, from a pragmatic point of view. They may, in special cases, be both implicative, or both concessive.

5. The negation of a conditional does not entail a conditional with the same antecedent and opposite consequent

Some logical systems also accept the converses of Boethius' theses:

${\sim}(A \to B) \to (A \to {\sim}B)$ and ${\sim}(A \to {\sim}B) \to (A \to B)$

These are valid in classical logic, in Stalnaker's conditional logic (as discussed above), in Wansing's **C** system and the first of them is valid in intuitionist logic [Wansing, 2005]. However, I argue here that the use of conditionals in natural discourse and reasoning does not support the converses of Boethius's theses. Statements with the phrase *It's not true that* preceding an implicative conditional do not entail the same conditional with an opposite consequent.

It must be noted, however, that Wansing's connexive logic **C** uses a strong type of negation that does not express untruth, but rather definite falsity. It is the De Morgan negation from the four-valued first-degree entailment logic (**FDE**), in a language with implication. Four truth values are admitted: true, false, neither true nor false, and both true and false. Therefore, the falsity conditions in the semantics of **C** are not conditions of when a formula is not true, but conditions that specify when a formula is definitely false. Thus, negation in **C** is neither suitable nor has it been intended to formalize an external natural language negation expressed by *It's not true that* or *It's not the case that.* It must also be noted that, although it validates the converses of Boethius's theses, **C** does not validate conditional excluded middle.[3]

It must be recognized that there are many cases in which *If A, not B* seems equivalent to the external negation of *If* A, B. The following are two attested examples of such cases:

(25) *It's not true that if your employer isn't paying you, you're therefore unemployed.* (Paraphrasable as: *If your employer isn't paying you, you're not therefore unemployed.*)

(26) *If God does not exist, then it's not the case that if I pray, my prayers will be answered.* (Paraphrasable as: *If God does not exist, then if I pray, my prayers will not be answered.*)

It should also be noted, however, that the conditional with the same antecedent and opposite consequent that seems to be entailed by the negation of an implicative conditional is usually interpreted as a concessive conditional. In (25) and (26), for example, the paraphrases given are in turn paraphrasable as:

(27) *Even if your employer isn't paying you, you're not therefore unemployed.*)

(28) *If God does not exist, then even if I pray, my prayers will not be answered.*)

However, (6), (9), (18) are examples in which the negation of an conditional cannot be paraphrased by a conditional with opposite consequent. Here are other examples:

(29) *It is not the case that if Beth and Nick both have blood type AB, their child will not have blood type A.*

(30) *If Beth and Nick both have blood type AB, their child will have blood type A.*

[3] I am grateful to Heinrich Wansing for these observations.

(31) *Even if Beth and Nick both have blood type AB, their child will have blood type A.*

Example (29) does not entail either (30) or (31), because (29) is true (since it is possible for the child of such a couple to have blood type A), but (30) and (31) are certainly false (since it is also possible for such a child to have blood types B or AB).

(32) *It's false that if you use an underarm deodorant you'll get breast cancer.*

Agreeing with this sentence certainly does not entail believing that underarm deodorant use is an effective means to avoid the development of breast cancer:

(33) *If you use an underarm deodorant you won't get breast cancer.*

Nor does it entail believing the corresponding concessive conditional to be true:

(34) *Even if you use an underarm deodorant, you won't get breast cancer.*

The latter conveys that you will not get breast cancer, while (32) is a more cautious statement that does not implicate this conclusion.

6. Negation of a conditional in natural language

Although David Lewis (unlike Stalnaker) recognizes that conditionals with opposite consequents may be both false (see section 7), he also failed to properly evaluate the use of negation of conditionals in natural language, as shown in the following quotation, where he discusses a settlement of the problem of ties regarding the similarity of worlds (one in which both $A \Box\!\!\rightarrow F$ and $A \Box\!\!\rightarrow {\sim}F$ are considered false):

> This reasonable settlement, however, does not sound so good in words. $A \Box\!\!\rightarrow F$ and $A \Box\!\!\rightarrow {\sim}F$ are both false, so we want to assert their negations. But negate their English readings in any straightforward and natural way, and we do not get ${\sim}(A \Box\!\!\rightarrow F)$ and ${\sim}(A \Box\!\!\rightarrow {\sim}F)$ as desired. Rather negation moves in and attaches only to the consequent, and we get sentences that seem to mean $A \Box\!\!\rightarrow {\sim}F$ and $A \Box\!\!\rightarrow {\sim\sim}F$(...) ([Lewis, 1973, p. 61])

It is true that in many contexts *If* $A, {\sim}B$ seems a more straightforward and natural way of denying *If* A, B. This is not to say, however, that the more elaborate form ${\sim}($*if* $A, B)$ is not available and widely used in natural language (English or other). A search in the Internet provides one with lots of examples. Here is a sample of attested examples in three languages:

It is not true that if some is good then more is better.
Hence it is not true that if K is such a composant of M every point of M is a limit point of K.
It is not true that if you dream about someone they are thinking of you. It's not true that if I owned all of those things, that would make me happy.
It is not true that if a person talks about suicide, they will not kill themselves.
It is not true that if you touch a baby bird the parents will reject it.
It's not true that if you go to "any tiny village in China", they have an internet café.
It's not true that, if I'm rich, I'm happy.
It's not true that if a technology has benefits, it will automatically get accepted by the public.
It's not true that if you're unhappy, so are your children.
It isn't true that if we can't think of something right away then we don't know it.
It isn't true that if you liked the Don Williams album, then you'll buy the Vern Gosdin album.
I definitely know it isn't true that if you are having periods your estrogen is "fine".
It isn't true that if someone looking at you thinks you're beautiful, then you are.
However, it's not the case that if your camera does not have one or other of these modes then it cannot do that sort of photography.
It's not the case that if one candidate wins all three then he or she should be the nominee.
It's not the case that if we just feel bad enough about ourselves we will be motivated to change.
It's not the case that if you make it through the first 10 years, your marriage is divorce-proof.
It's not the case that if you're eligible you'll get the money.
It is not the case that if we don't immediately solve the Futenma issue we can't deepen relations.
It is not the case that if one has faith then one's reason is canceled out.
It is not the case that if the US isn't powerful enough, then China, Russia, or some caliphate will be.
It's not true that wisdom teeth crowd your teeth if they grow in impacted. It's not true that wasps don't sting you if you keep still.
In addition, it's not true that all intruders will flee if confronted.
It's not true that you will suffer if you come out with your ideas in public. It's not true that you gain weight if you're on birth control.
It's false that if others change, we will also change.
It's false that if your gutters are falling off, you must replace them.
It's false that if we keep running large deficits then we get increases to our debt-to-GDP ratio (...).
Ce n'est pas vrai que si dans l'année un homme marié n'a pas régularisé il ne fera jamais.

Ce n'est pas vrai que si le sexe va bien, tout va bien entre 2 personnes.
Ce n'est pas vrai que si les ordinateurs fonctionnent c'est grâce à notre maÎtrise de la physique quantique !
Ce n'est pas vrai que si le citoyen qui se fait arrêter n'est pas armé le policier ne devrait pas se servir de son arme.
Ce n'est pas vrai que si un enfant n'a pas suivit [sic] un programme éducatif avant l'école il aura des difficultés d'apprentissage.
C'est faux que si l'on fait un vœu, il se réalise.
[C]'est faux que si on ne paye pas la pension[,] il y a abandon.
(. . .) il est faux que si un sujet croit que a est F, alors il existe quelque chose a, à propos duquel le sujet croit qu'il est F.
No es verdad que, si se promueve el desarrollo económico, después se obtendrá la dictadura.
No es verdad que si un triángulo tiene tres ángulos entonces un cuadrado tiene cuatro ángulos.
No es verdad que si usas aceite sintético una vez tienes que seguir poniéndole para siempre.
No es verdad que si no eres gaditano no te respeten en el carnaval.
Por otro lado no es verdad que si dedicaran más dinero a mejorar sus productos no tendrían que anunciarlos.
No es verdad que "si estudias aprobas" [sic].
No es verdad que si el niño balbucea o emite sonidos, no tiene problemas en el oído.
[E]s falso que si cocemos una seta con una cuchara de plata y ésta ennegrece, se trata de una seta tóxica, y si no, es comestible.
Es falso que si te arrancas una cana, saldrán más (...).
Es falso que si uno fuma marihuana[,] termina consumiendo cocaína

It is thus clear that there are many, many cases in which negation does not – and sometimes indeed *cannot* – move in and attach only to the consequent, as Lewis states.

Consider one more example:

(35) *It's not true that if Jones loves his children, then he's not a criminal.*

It is quite obvious that the truth of (35) does not entail (36):

(36) *If Jones loves his children, then he's a criminal.*

And interpreting (36) as a concessive conditional would not help to make it viable as a paraphrase for (35), as shown by the following unambiguously concessive version of it:

(37) *Even if Jones loves his children, he's a criminal.*

Someone who asserts (37) is convinced that Jones is a criminal, while someone who asserts (35) is just admitting that this is possible, even if he loves his children.

7. Conditional excluded middle

Our conclusion entails that the principle of conditional excluded middle is not valid for natural language conditionals. According to this principle, either $A \to B$ or $A \to {\sim}B$ must be true. However, examples such as (5)–(7), (17)–(18), (29)–(30), (32)–(33) and (35)–(36) show that it is possible in natural language to consider two conditionals with same antecedent and opposite consequent as both false.

Lewis rejected the principle of conditional excluded middle, but for a different reason. He noted that there may be ties in comparative similarity among possible worlds, so that a world in which Bizet and Verdi are both French may be as similar to ours as one in which they are both Italian [Lewis, 1973, p. 61]. Thus, he does not believe that either *If Bizet and Verdi were compatriots, they would be French* or *If Bizet and Verdi were compatriots, they would not be French* must be true. As regards the external negation of conditionals in natural language, he seems rather (as shown in the previous section) to subscribe to the mistaken notion of an equivalence between the linguistic forms corresponding to ${\sim}(\textit{if } A, B)$ and *If* $A, {\sim}B$, respectively [Lewis, 1973, p. 61].

A problem for the rejection of conditional excluded middle seems to be the distributivity of the conditional operator over disjunction (see [Bennett, 2003, p. 185]):

$A \to (B \lor C) \therefore (A \to B) \lor (A \to C)$

Replacing C by ${\sim}B$, we get:

$A \to (B \lor {\sim}B) \therefore (A \to B) \lor (A \to {\sim}B)$

Since $A \to (B \lor {\sim}B)$ is a logical truth, the principle of conditional excluded middle seems to be proved. However, distributivity of the conditional operator over disjunction is valid in classic propositional logic, in which this operator is the material conditional operator, but need not be valid for other varieties of the conditional connective.

Concerning the implicative use of natural language conditionals in reasoning and discourse, distributivity over disjunction is certainly not valid, as shown by the following example:

(38) *If Anna has blood type A, then her genotype is either AA or AO.*

This is true, but does not entail that either (39) or (40) is true, since they are both pragmatically false:

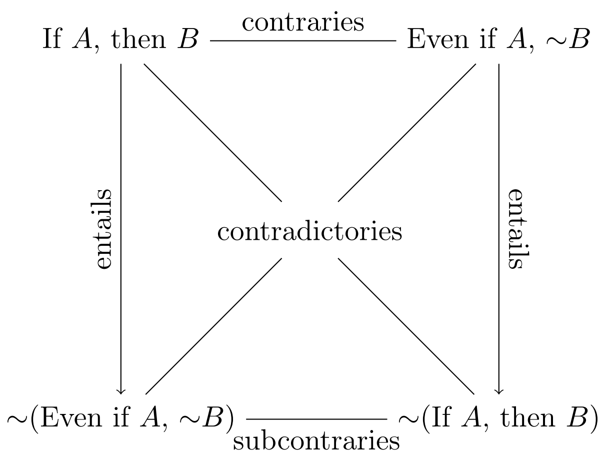
(39) *If Anna has blood type A, then her genotype is AA.*

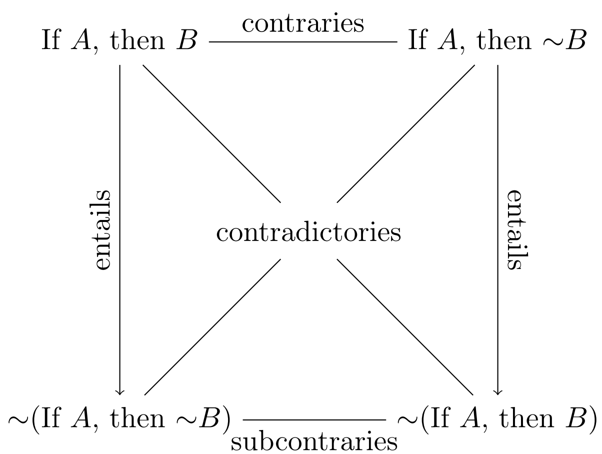
(40) *If Anna has blood type A, then her genotype is AO.*

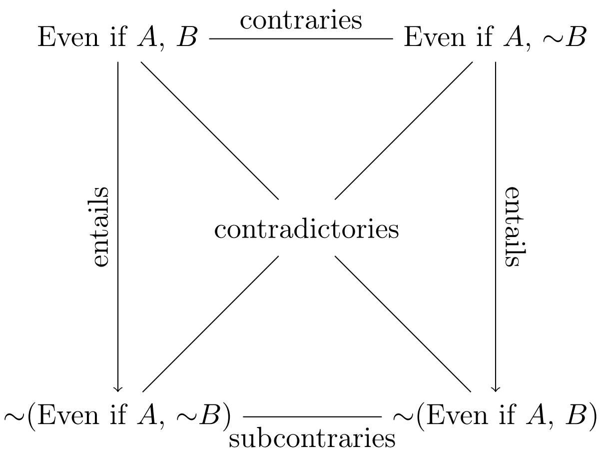
The proposition conveyed by the uttering of (39) would be pragmatically true if any person having blood type A (and not just Anna, if she does have it) had genotype AA (which is not the case), and similarly for (40) in relation to genotype AO. We conclude that distributivity over disjunction is not a reason for accepting conditional excluded middle for natural language conditionals, and that conditional excluded middle must in fact be rejected, as shown by the counterexamples given above.

8. Conditional squares of opposition

The relations among negations of conditionals and conditionals with opposite consequents may thus be summarized in three squares of opposition, as follows (contradictories indicated by the diagonals).







9. Conclusions

(i) *Not (if A, then B)* cannot always be paraphrased as *If A, then not B*, so the former does not entail the latter in natural language.

(ii) Similarly, *Not (if A, then B)* does not entail *Even if A, not B.*

(iii) Moreover, *Not (even if A,B)* does not entail *If A, then not B.*

(iv) Similarly, *Not (even if A,B)* does not entail *Even if A, not B.*

(v) On the other hand, *If A, then B* entails *Not (even if A, not B).*

(vi) Similarly, *Even if A, B* entails *Not (if A, then not B).*

(vii) Moreover, *If A, then B* also entails *Not (if A, then not B).*

(viii) And similarly, *Even if A, B* also entails *Not (even if A, not B).*

(ix) Therefore, *If A, then B* and *Even if A, not B* are contraries, not contradictories. Similarly, the pair *If A, then B* and *If A then not B* and the pair *Even if A, B* and *Even if A, not B* are also pairs of contraries, not of contradictories.

(x) Therefore, the principle of conditional excluded middle does not apply to the use of conditionals in natural language.

References

Barker, 1973 – Barker, J.A. "Hypotheticals: Conditionals and theticals", *Philosophical Quarterly*, 1973, Vol. 23, No. 93, pp. 335–345.

Bennett, 2003 – Bennett, J. *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press, 2003.

Chisholm, 1946 – Chisholm, R. "The contrary-to-fact conditional", *Mind*, 1946, Vol. 55, pp. 289–307.

Davis, 1983 – Davis, W. "Weak and strong conditionals", *Pacific Philosophical Quarterly*, 1983, Vol. 64, pp. 57–71.

Declerck and Reed, 2001 – Declerck, R. and Reed, S. "Some truths and nontruths about *even if*", *Linguistics*, 2001, Vol. 39, No. 2, pp. 203–255.

Ducrot, 1972 – Ducrot, O. *Dire et ne pas dire*. Paris: Hermann, 1972.

Gomes, 2009 – Gomes, G. "Are necessary and sufficient conditions converse relations?", *Australasian Journal of Philosophy*, 2009, Vol. 87, No. 3, pp. 375–387.

Gomes, 2020 – Gomes, G. "Concessive conditionals without *even if* and nonconcessive conditionals with *even if*", *Acta Analytica*, 2020, Vol. 35, No. 1, pp. 1–21.

Goodman, 1947 – Goodman, N. "The problem of counterfactual conditionals", *Journal of Philosophy*, 1947, Vol. XLIV, No. 5, pp. 113–128.

Horn, 2001 – Horn, L.R. *A Natural History of Negation*. Stanford (CA): CSLI Publications, 2001.

Horn and Wansing, 2020 – Horn, L.R. and Wansing, H. "Negation", *The Stanford Encyclopedia of Philosophy*, [https://plato.stanford.edu/archives/spr2020/entries/negation/], Spring 2020 edition.

Lewis, 1973 – Lewis, D. "Counterfactuals and comparative possibility", *Journal of Philosophical Logic*, 1973, Vol. 2, reprinted in *Ifs*, ed. by W.L. Harper, R. Stalnaker, and G. Pearce. Dordrecht: Reidel, 1981, pp. 57–85.

Lycan, 2001 – Lycan, W.G. *Real conditionals*. Oxford: Oxford University Press, 2001.

Recanati, 2010 – Recanati, F. *Truth-Conditional Pragmatics*. Oxford: Oxford University Press, 2010.

Stalnaker, 1968 – Stalnaker, R.C. "A theory of conditionals", in: *Studies in Logical Theory*, ed. by N. Rescher, Oxford: Blackwell, APQ Monograph No 2., reprinted in *Ifs*, ed. by W.L. Harper, R. Stalnaker, and G. Pearce, Dordrecht: Reidel, 1981, pp. 41–55.

Stalnaker, 1981 – Stalnaker, R.C. "A defense of conditional excluded middle", in: *Ifs*, ed. by W.L. Harper, R. Stalnaker, G. Pearce. Dordrecht: Reidel, 1981, pp. 87–104.

Wansing, 2005 – Wansing, H. "Connexive Modal Logic", in: *Advances in Modal Logic. Volume 5*, ed. by R. Schmidt et al. London: King's College Publications, 2005, pp. 367–383.

Логические исследования
2021. Т. 27. № 1. С. 64–81
УДК 16+510.6

Logical Investigations
2021, Vol. 27, No. 1, pp. 64–81
DOI: 10.21146/2074-1472-2021-27-1-64-81

REINHARD KAHLE

Default Negation as Explicit Negation plus Update

Reinhard Kahle
Universität Tübingen,
Theorie und Geschichte der Wissenschaften, Keplerstr.
2, D-72074 Tübingen, Germany.
Centro de Matemática e Aplicações,
FCT, Universidade Nova de Lisboa, P-2829-516 Caparica, Portugal.
E-mail: kahle@mat.uc.pt

Abstract: We argue that under the stable model semantics default negation can be read as explicit negation with update. We show that dynamic logic programming which is based on default negation, even in the heads, can be interpreted in a variant of updates with explicit negation only. As corollaries, we get an easy description of default negation in generalized and normal logic programming where initially negated literals are updated. These results are discussed with respect to the understanding of negation in logic programming.

Keywords: Default Negation, Explicit Negation, Logic Programming Update

For citation: Kahle R. "Default Negation as Explicit Negation plus Update", *Logicheskie Issledovaniya / Logical Investigations*, 2021, Vol. 27, No. 1, pp. 64–81. DOI: 10.21146/2074-1472-2021-27-1-64-81

1. Introduction

Negation is still one of the controversial concepts underlying logic programming. While the *negation-as-failure* interpretation is operationally well-understood, the logical interpretation of negation gives rise to discussions. Starting from the two most prominent approaches – Reiter's *closed world assumption* [Reiter, 1978] and Clark's *completion* [Clark, 1978] – there "are very successful attempts to discover the underlying logic of negation as failure. Their disadvantage is that the logics involved are more complicated and less familiar than classical logic so that they are not likely to help the naive programmer express his problem by means of a logic program, or to check the correctness of a program" [Shepherdson, 1998, p. 364]. In this paper we argue that *default* negation in logic programming can be understand as *explicit* negation in an *update* framework, as long as we consider stable model semantics. This reading may also give a "naive programmer" some help to deal with default negation in logic programming.

The question of treating updates in a logic programming framework is a research project in itself. There are several attempts to deal with updates and different semantics have been proposed, based on stable models or answer sets, cf. e.g. [Alferes et al., 2000; Buccafurri et al., 1999; Eiter et al., 2002; Leite and Pereira, 1998a; Leite and Pereira, 1998b; Leite, 2003; Sakama and Inoue, 1999; Zhang and Foo, 1998]. In all these frameworks updates are represented by a sequence of logic programs. Here, we focus on *dynamic logic programming*, introduced by Alferes, Leite, Pereira, Przymusinska, and Przymusinski [Alferes et al., 2000]. It is based on *generalized logic programs* which allow *default negation* in the head of rules. Its semantics is based on causal rejection, i.e., a rule can be rejected if there is a more recent one that conflicts with it.

We show that *default negation* as used in dynamic logic programming, generalized logic programming, and normal logic programming can be treated as *explicit negation* in an analogous update framework. For the technical result we can build on work of Leite in his dissertation [Leite, 2003]. However, with respect to the understanding of default negation, the result allows a different perspective to it. In particular, it questions the status of default negation as a special form of negation, different from the classical one. In contrast, default negation can be seen as involving an update aspect (which could also be considered as a temporal aspect), which can be expressed in an update framework with explicit negation.

In the following section we introduce the technical preliminaries for logic programs with default and with explicit negation. In the third section we introduce an update framework for explicit negation. With it we can state our main result in section 4, which also provides a short illustrating example. In section 5 we review dynamic logic programming, as it is defined in [Alferes et al., 2000] and [Leite, 2003]. It is used in the sixth section to prove a general translation of dynamic logic programs into the explicit update framework. From it, the main result follows as an immediate corollary. The final section is devoted to a discussion of the given result with respect to the closed world assumption, normal logic programming and the combination of default and explicit negation. We shortly address the question of well-founded semantics and the relation to the transformational semantics for dynamic logic programming and include a reference to abductive frameworks.

2. Preliminaries

In logic programming, for both, default and explicit negation, it is convenient to work syntactically with pure Horn theories. In this case, negated literals are introduced as new atoms (disjoint from all other atoms) which are formally positive. The difference of positive and negative literals is build in on the semantical level only. Thus, for default negation we will use atoms `not_a`, for explicit negation atoms `neg_a`.

Here, we do not consider programs which combine default and explicit negation. Let us first define the technicalities for the case of default negation.

2.1. Default negation

Given an arbitrary set $\mathcal{K}$ of propositional variables (which do not begin with "`not`"), the propositional language $\mathcal{L}^{\mathcal{K}}_{\mathtt{not}}$ is defined as the set $\{\mathtt{a} : \mathtt{a} \in \mathcal{K}\} \cup \{\mathtt{not\_a} : \mathtt{a} \in \mathcal{K}\}$. Elements of $\mathcal{L}^{\mathcal{K}}_{\mathtt{not}}$ are called *literals*, *positive literals* if they belong to $\mathcal{K}$, *negative literals* otherwise. While *normal logic programs* allow default negation only in the body of a clause, we consider *generalized logic programs* where default negation is also allowed in the heads of rules. Generalized logic programs, introduced in [Alferes et al., 2000], are a simplified version of the programs introduced by Lifschitz and Woo [Lifschitz and Woo, 1992]. Formally, a generalized logic program consists of a (finite or infinite) set of rules of the form L `<-` L_1`,`$\ldots$`,`L_n, where $L, L_1, \ldots, L_n$ are literals from $\mathcal{L}^{\mathcal{K}}_{\mathtt{not}}$.

We use the usual notational conventions in logic programming. If `a` is a positive literal, *not* `a` is `not_a`, and for a negative literal `not_b`, *not* `not_b` is `b`. If r is the rule L `<-` L_1`,`$\ldots$`,`L_n we write $\mathsf{H}(r)$ for the head L and $\mathsf{B}(r)$ for the body L_1`,`$\ldots$`,`L_n.

A *2-valued interpretation* M of $\mathcal{L}^{\mathcal{K}}_{\mathtt{not}}$ is a subset of $\mathcal{L}^{\mathcal{K}}_{\mathtt{not}}$ such that for all $\mathtt{a} \in \mathcal{K}$ precisely one of the literals `a` or `not_a` belong to M. An interpretation M of $\mathcal{L}^{\mathcal{K}}_{\mathtt{not}}$ satisfies a rule r if $\mathsf{H}(r)$ belong to M, or some literal in $\mathsf{B}(r)$ does not belong to M. A *model* M of a generalized logic program P is an interpretation which satisfies all its rules. Let M^+ be the set of positive literals in M, and M^- the set of negative literals in M. A model N is *p-smaller* than M, if N^+ is a proper subset of M^+, $N^+ \subset M^+$. A model of P is called *p-least* if it is the p-smallest model of P.[1] Now, the definition of stable models reads as follows, cf. [Alferes et al., 2000, Def. 2.1].

Definition 1. An interpretation M of $\mathcal{L}^{\mathcal{K}}_{\mathtt{not}}$ is a *stable model* of a generalized logic program P if M is the p-least model of the Horn theory $P \cup M^-$, i.e.,

$$M = \textit{p-least}(P \cup M^-).$$

For normal programs this definition is equivalent to the original definition of stable models by Gelfond and Lifschitz [Gelfond and Lifschitz, 1988]. In [Alferes et al., 2000] it is shown that this definition coincides with the answer set semantics given by Lifschitz and Woo, [Lifschitz and Woo, 1992], if it is restricted to generalized logic programs.

[1] We say "p-least", since we consider the positive literals, only. Later on, for explicit negation, we will define *leastness* with respect to all literals.

2.2. Explicit negation

Extended logic programs are well-known as the extension of normal logic programs with explicit negation. The *answer-set semantics* given for this class of programs allows the use of explicit negation in both, the body and the head of a clause, [Gelfond and Lifschitz, 1991].

As for the default literals, explicit negated literals are introduced as new atoms, and *consistency* (as well as *coherency* in the presence of default negation) is treated on the semantical level.

Here, we even dispense with default negation, and consider programs with explicit negation only. For convenience we tactically call them *explicit logic programs.* This class of programs, in itself, is without particular interest. But, in connection with update it allows us to give an easy interpretation of default negation.

The formal preliminaries are completely analogous to the case of dynamic logic programming. The only difference is that we replace the "prefix" `not` for default negation by `neg` for explicit negation.

Given a set of propositional variables $\mathcal{K}$, we consider now the language $\mathcal{L}^{\mathcal{K}}_{\texttt{neg}} = \{\texttt{a} : \texttt{a} \in \mathcal{K}\} \cup \{\texttt{neg\_a} : \texttt{a} \in \mathcal{K}\}$. As before, we will speak of positive and negative literals, and only if needed we speak about *default negated literals* or *explicit negated literals.* Instead of *not* `a`, which we used for default negation, we write *neg* `a` for the complementary literal with respect to explicit negation.

The notion of model differs from the default case. Since we cannot assume negative literals "by default" we have to ask for support of them as for positive literals. In the default case, we consider least models with respect to the positive literals only. Now, we have to consider least models with respect to both, positive and negative literals. Therefore, we start with consistent interpretations, which do not have to be 2-valued. Thus, an interpretation M is an subset of $\mathcal{L}^{\mathcal{K}}_{\texttt{neg}}$ such that for no literal $\texttt{a} \in \mathcal{K}$, `a` and `neg_a` belong to M. With the usual notion of satisfiability we define a *pre-model* M of an explicit logic program which is an interpretation satisfying all rules of the program. It is called "pre-model" since later we are interested in 2-valued models only. We order the pre-models with respect to both, positive and negative literals. A model N is *smaller* than M, if it is a proper subset of M. A pre-model of P is called *least* if it is a smallest pre-model of P. A *2-model* M of an explicit logic program is a least pre-model which is 2-valued with respect to explicit negation, i.e., M contains for all $\texttt{a} \in \mathcal{K}$ precisely one of the literals `a` or `neg_a`.

Definition 2. An interpretation M of $\mathcal{L}^{\mathcal{K}}_{\texttt{neg}}$ is the *2-model* of an explicit logic program P if M is 2-valued and the least pre-model of P, i.e., $M = \mathit{least}(P)$.

With the given definition the very most of explicit logic programs will not have a 2-model. As a very easy example, let us consider the generalized logic program P: `not_a <- not_a` and the explicit logic program E: `neg_a <- neg_a`. P has {`not_a`} as single stable model, while E does not have a 2-model. In general, for explicit logic programs, it would be more natural to consider a three valued semantics. However, for our specific aim to interprete default negation in an update framework with explicit negation, it turns out that the existence of 2-models is always guaranteed.

3. Explicit dynamic logic programs

In this section we will give an adaptation of *dynamic logic programming* as introduced in [Alferes et al., 2000] to explicit logic programs.[2]

An *explicit dynamic logic program* $\mathcal{E}$ is a finite sequence of *explicit logic programs* $E_1, \ldots, E_n$, written as

$$E_1 \otimes E_2 \otimes \cdots \otimes E_n.$$

Its semantics is – in analogy to dynamic logic programming – based on causal rejection: A rule can be rejected by a more recent one if the latter rule has as head the negated literal of the former one, and the body of the latter one is true in the considered model.

Formally we introduce the following notion of conflicting rules:

Definition 3. Two rules r and r' are called *conflicting with respect to explicit negation*, denoted by $r \stackrel{\text{neg}}{\bowtie} r'$ iff $\mathsf{H}(r) = neg\ \mathsf{H}(r')$.

Let $\bigcup \mathcal{E}$ be the union of all rules of all explicit logic programs of an explicit dynamic logic program. Then, we define 2-model of $\mathcal{E}$ as follows:

Definition 4. Let $\mathcal{E} = \{E_i : 0 \leq i \leq n\}$ be an explicit dynamic logic program. An interpretation M is a *2-model of* $\mathcal{E}$, if

1. M is 2-valued and
2. $M = least((\bigcup \mathcal{E}) \setminus Reject(\mathcal{E}, M))$,
   where $Reject(\mathcal{E}, M)$ is the set

$$\{r \in E_i : \exists r' \in E_j, i < j \leq n \,\&\, r \stackrel{\text{neg}}{\bowtie} r' \,\&\, M \models \mathsf{B}(r')\}.$$

[2] We will review dynamic logic programming only later in Section 5. It is not needed to state the main result of this paper (Proposition 1), but it will be instrumental for the proof of it.

4. Default Negation as Explicit Negation plus Update

We now state the main proposition of our paper. A generalized logic program P can be translated into an explicit dynamic logic program $\mathcal{E}$ such that the stable models of P coincide with the 2-models of $\mathcal{E}$. That means that the sets of positive literals in both sets are the same, and the default negated literals of a model of $\mathcal{P}$ coincide with the explicit negated literals in the corresponding model of $\mathcal{E}$. In fact, $\mathcal{E}$ is an explicit dynamic logic program with one update only.[3]

Definition 5. Let $\mathcal{K}$ be a set of propositional variables.

1. E_0 is defined as the set of all explicit negated literals in the language $\mathcal{K}$:
$$E_0 := \{\texttt{neg\_a} : \texttt{a} \in \mathcal{K}\}.$$
2. Given a generalized logic program P in $\mathcal{L}^{\mathcal{K}}_{\texttt{not}}$, we define the explicit logic program $E = \overline{P}$ in $\mathcal{L}^{\mathcal{K}}_{\texttt{neg}}$ as the program where every occurrence of a default negation in P is replaced by an explicit negation.
3. For a set S of literals of $\mathcal{L}^{\mathcal{K}}_{\texttt{not}}$, we define $\overline{S}$ as the set of literals of $\mathcal{L}^{\mathcal{K}}_{\texttt{neg}}$ where all default negated literals are replaced by their corresponding explicit negated ones.

Of course, E_0 has the trivial 2-model where all negative literals are true. Therefore, when we start an explicit dynamic logic program with the initial program E_0, we guarantee the 2-valuedness of its model.

Proposition 1. Let P be a generalized logic program in the language $\mathcal{K}$. Let $\mathcal{E}$ be the explicit dynamic logic program $E_0 \otimes \overline{P}$. Then,

$$M \text{ is a stable model of } P \text{ if and only if } \overline{M} \text{ is a 2-model of } \mathcal{E}.$$

Analogously we have for normal logic programs the following proposition:

Proposition 2. Let P be a normal logic program in the language $\mathcal{K}$. Let $\mathcal{E}$ be the explicit dynamic logic program $E_0 \otimes \overline{P}$. Then,

$$M \text{ is a stable model of } P \text{ if and only if } \overline{M} \text{ is a 2-model of } \mathcal{E},$$

and $\overline{P}$ does not use negation in the heads.

These propositions are corollaries of Theorem 1 which will be stated and proven below.

[3] To some extend, the study of a single update (instead of sequences of updates) has its own interest. For instance, in the original definition of dynamic logic programming, [Alferes et al., 2000], the authors even start with the definition of one update, and define dynamic logic programs as a generalization of it.

Example

To illustrate our result, let us consider the following normal logic program:

$$P = \{\texttt{a <- not\_b}, \texttt{b <- not\_a}\}.$$

It has the two stable models $M_1 = \{\texttt{a}, \texttt{not\_b}\}$ and $M_2 = \{\texttt{b}, \texttt{not\_a}\}$. The translation of P into an explicit dynamic logic programs yields

$$\mathcal{E} = \{\texttt{neg\_a}, \texttt{neg\_b}\} \otimes \{\texttt{a <- neg\_b}, \texttt{b <- neg\_a}\}.$$

Now, we have to check whether $\overline{M_1} = \{\texttt{a}, \texttt{neg\_b}\}$ and $\overline{M_2} = \{\texttt{b}, \texttt{neg\_a}\}$ are the only 2-models of $\mathcal{E}$ according to Definition 4. Of course, both are 2-valued.

For $\overline{M_1}$, we have that $\mathit{Reject}(\mathcal{E}, \overline{M_1}) = \{\texttt{neg\_a}\}$, since it is rejected by the rule `a <- neg_b` whose body is true in $\overline{M_1}$. Therefore, $\overline{M_1}$ has to be the least pre-model of $\{\texttt{neg\_b}, \texttt{a <- neg\_b}, \texttt{b <- neg\_a}\}$ which is the case.

Analogously, for $\overline{M_2}$, `neg_b` is rejected, and we get that it is the least pre-model of $\{\texttt{neg\_a}, \texttt{a <- neg\_b}, \texttt{b <- neg\_a}\}$.

There are only two other possibilities for 2-models of $\mathcal{E}$, $N_1 = \{\texttt{a}, \texttt{b}\}$ and $N_2 = \{\texttt{neg\_a}, \texttt{neg\_b}\}$. For N_1, $\mathit{Reject}(\mathcal{E}, N_1)$ is empty, so `neg_a` and `neg_b` are facts, and N_1 cannot be a 2-model. For N_2 the situation is different, since $\mathit{Reject}(\mathcal{E}, N_2) = \{\texttt{neg\_a}, \texttt{neg\_b}\}$. But the least pre-model of $\{\texttt{a <- neg\_b}, \texttt{b <- neg\_a}\}$ is the empty set, i.e., this program does not have a 2-model, in particular not N_2. Thus, $\overline{M_1}$ and $\overline{M_2}$ are the only 2-models of $\mathcal{E}$.

5. Dynamic logic programming

For the proof of the Propositions 1 and 2 we will use a more general result, translating dynamic logic programs into explicit dynamic logic programs. Therefore, we review briefly the definition of dynamic logic programs.

A *dynamic logic program* $\mathcal{P}$ consists of a finite sequence of generalized logic programs $P_1, \ldots, P_n$, written as

$$P_1 \oplus P_2 \oplus \cdots \oplus P_n.$$

As for the 2-models of extended dynamic logic programming, the *stable model semantics* of a dynamic logic program is based on causal rejection. The idea of default negation is build in by assuming all negated literals for which there is no rule with the positive literal as head and a true body.

For the formal definition, we need again the notion of conflicting rules, cf. [Leite, 2003, Def. 27, p. 35]:

Definition 6. Two rules r and r' are called *conflicting*, denoted by $r \bowtie r'$, iff $\mathsf{H}(r) = \mathit{not}\ \mathsf{H}(r')$.

Using $\bigcup \mathcal{P}$ for the union of all rules of all generalized logic programs of a dynamic logic program $\mathcal{P}$, we can define stable models of $\mathcal{P}$ as follows, cf. [Leite, 2003, Def. 37, p. 48].

Definition 7. Let $\mathcal{P} = \{P_i : 1 \leq i \leq n\}$ be a dynamic logic program. An interpretation M is a *stable model of* $\mathcal{P}$, if

$$M = p\text{-}least([(\bigcup \mathcal{P}) \setminus Reject(\mathcal{P}, M)] \cup Default(\mathcal{P}, M)),$$

where $Reject(\mathcal{P}, M)$ is the set

$$\{r \in P_i : \exists r' \in P_j, i < j \leq n \,\&\, r \bowtie r' \,\&\, M \models \mathsf{B}(r')\},$$

and $Default(\mathcal{P}, M)$ is the set

$$\{\texttt{not\_a} : \neg\exists r \in \bigcup \mathcal{P}.(\mathsf{H}(r) = \mathsf{a}) \,\&\, M \models \mathsf{B}(r)\}.$$

6. Embedding of dynamic logic programming in explicit dynamic logic programming

We now extend the translation of generalized logic programs in extended dynamic logic programs to dynamic logic programs. With the notation of Definition 5 we set:

Definition 8. For a dynamic logic program $\mathcal{P} = P_1 \oplus \ldots \oplus P_n$, $\overline{\mathcal{P}}$ is defined as $\overline{P_1} \otimes \cdots \otimes \overline{P_n}$.

The general theorem can now be stated as follows:

Theorem 1. *Let $\mathcal{K}$ be a set of propositional variables. Let $\mathcal{P} = P_1 \oplus \ldots \oplus P_n$ be a dynamic logic program in $\mathcal{L}^{\mathcal{K}}_{\texttt{not}}$. Let $\mathcal{E}$ be the explicit dynamic logic program $E_0 \otimes \overline{P_1} \otimes \cdots \otimes \overline{P_n}$. Then,*

$$M \text{ is a stable model of } \mathcal{P} \text{ if and only if } \overline{M} \text{ is a 2-model of } \mathcal{E}.$$

The theorem is a consequence of the following proposition, proven by Leite in his dissertation, [Leite, 2003, Prop. 28 and Cor. 29, p. 54].

Proposition 3. Let $\mathcal{P}$ be a dynamic logic program. Let $\mathcal{P}'$ be the dynamic logic program such that $\mathcal{P}' = P^{\mathcal{K}} \oplus P^{not\ \mathcal{K}} \oplus \mathcal{P}$. An interpretation M is a stable model of $\mathcal{P}$ iff

$$M = p\text{-}least((\bigcup \mathcal{P}') \setminus Reject(\mathcal{P}', M)),$$

where $P^{\mathcal{K}}$ is the set of all positive literals in the language $\mathcal{K}$ and $P^{not\ \mathcal{K}}$ the set of all default negated literals in the language $\mathcal{K}$.

Proof. (Theorem 1) We will use Proposition 3. However, the use of $P^{\mathcal{K}}$ is redundant, since $P^{not\ \mathcal{K}}$ is immediately rejecting all literals of $P^{\mathcal{K}}$. Therefore, we can choose for $\mathcal{P}'$ also the dynamic logic program $P^{not\ \mathcal{K}} \oplus \mathcal{P}$. So, $\mathcal{E} = \overline{\mathcal{P}'}$.

The assertion of the theorem seems to be a notational variant where default negation is replaced by explicit negation. However, since dynamic logic programming was formulated by use of p-least models we have to check that the least pre-models used for the semantics in explicit dynamic logic programming are coincide with them, modulo the substitution of default and explicit negation.

For the direction from the left to the right, let M be a stable model of $\mathcal{P}$. With the proposition we get that M is the p-least model of $P_M := (\bigcup \mathcal{P}') \setminus \mathit{Reject}(\mathcal{P}', M)$. Translating this program into an explicit logic program, we get that $\overline{M}$ is a 2-valued pre-model of $E_M := (\bigcup \mathcal{E}) \setminus \mathit{Reject}(\mathcal{P}', \overline{M})$. We have to show that it is the least pre-model. Assume there is another pre-model N of E_M, with $N \subset \overline{M}$. If N^+ is a proper subset of $\overline{M}^+$, then there is a model K of P_M with $N = \overline{K}$. So K^+ is a subset of M^+ which contradicts the assumption that M was the p-least model of P_M. Now, let N^- be a proper subset of $\overline{M}^-$. So there is a literal `neg_b` in $\overline{M}$ which is not in $\overline{N}$. Since E_0 contains the fact `neg_b`, there has to be a rule r in $\mathit{Reject}(\mathcal{P}', \overline{M})$ with $\mathsf{H}(r) = \mathtt{b}$ and the body of r is true in $\overline{M}$. However, with this condition $\overline{M}$ could not be a 2-model of E_M since it has to contain both `neg_b` and `b`. Thus, we have a contradiction, and $\overline{M}$ is the least pre-model of E_M, i.e., $\overline{M}$ is a 2-model of $\mathcal{E}$.

For the direction from the right to the left we start with a 2-model N of $\mathcal{E}$. So N is the least pre-model of $E_N := (\bigcup \mathcal{E}) \setminus \mathit{Reject}(\mathcal{E}, N)$. Of course, there is an interpretation M of $\mathcal{L}^{\mathcal{K}}_{\mathrm{not}}$ such that $N = \overline{M}$. We have to show that M is a stable model of $\mathcal{P}$. Using the proposition above, we have to show that M is a stable model of $P_M := (\bigcup \mathcal{P}') \setminus \mathit{Reject}(\mathcal{P}', M)$. Since $E_N = \overline{P_M}$, and N is a pre-model of E_N, M is a model of P_M. It remains to show that M is p-least. Assume that there is another model K of P_M with $K^+ \subset M^+$. Without loss of generality, we can assume that $K = M \setminus \{\mathtt{b}\}$ for some positive literal `b`. Since K is a model of P_M, there is no rule r in P_M with $\mathsf{H}(r) = \mathtt{b}$ and the body of r is true in K. But that means, there is no rule r' in E_N such that $\mathsf{H}(r') = \mathtt{b}$ and the body of r' is true in $\overline{K} = N \setminus \{\mathtt{b}\}$. Therefore, $N \setminus \{\mathtt{b}\}$ is a pre-model of E_N which contradicts the assumption that N was a 2-model of $\mathcal{E}$. Thus, M is the p-least model of P_M, i.e., M is a stable model of $\mathcal{P}'$ and $\mathcal{P}$. ■

7. Discussion

In this section we discuss our reading of default negation as explicit negation plus update in different respects.

7.1. The closed world assumption

Our explanation of default negation has a certain relation to the closed world assumption. In the closed world assumption "only" those negative literals are assumed for which the positive literal is not derivable from a program. In contrast, we assume all negative literals and update later on those for which the positive literal holds in a model. By this update step we avoid the inconsistencies which result sometimes from the closed world assumption in the case of indefinite information about ground literals, cf. [Shepherdson, 1998, p. 357].

7.2. Default negation in normal logic programming

Proposition 2 is more than a by-product of Theorem 1. It gives the default negation as it is used in normal logic programming a reading in terms of explicit negation with updates. Normal logic programming is the basis of logic programming, uncontroversial and well-understood. However, explicit negation seems to be closer to classical negation as used in standard logic, and therefore more favorable in discussions outside the logic programming community.[4] In fact, it is interesting to study the exact *logical* meanings of negation in logic programming, as it is done in the work of Pearce and others, cf. [Pearce, 1997; Pearce, 1999; Lifschitz et al., 2001]. Here, we have implemented explicitly the idea of defaults as a sceptical view of truth: Every literal, for which one cannot find a rule with a true body, is considered as false. In our reading just the perspective is changed: Every literal is initially considered as false (by use of the initial program E_0) and then we think of the original program as an update program which updates the initial scepticism.

There is a well-known correspondence, cf. [Bidoit and Froidevaux, 1991], between the stable model semantics of normal logic programs and *default logic*, [Reiter, 1980; Poole, 1994]. Now, we can even ask whether the reading of default as explicit plus updates allows for an interpretation of default reasoning as classical reasoning with *updates*. However, this question is outside the scope of this paper.

7.3. The combination of default and explicit negation

A naive extension of our translation does not work when we allow default and explicit negation together.

If one would translate both negations into explicit negation in explicit dynamic logic programming and keep the initial program E_0 in the way that it contains all negative literals, it would trivialize the difference between default and explicit negation. In fact, E_0 would give the original explicit negation the same meaning as default negation.

[4] But see the remarks about explicit negation in the following subsection.

Of course, one could think of restricting E_0 to the negative literals coming from default negation only. However, then, we will have problems to find 2-models for the resulting explicit dynamic logic program, if we have no support for neither `a` nor `neg_a`.

In fact, the understanding of *explicit negation* in logic programming is contrary to the idea of a 2-valued semantics: Only if we explicit information about `a` or `neg_a` we should accept either of it; if not, `a` should be considered as neither true or false. Thus, although explicit negation is in its operational behavior (as it should be guaranteed by the semantics) closer to "classical" negation,[5] it does not support the "classical" principle of bivalence. This principle is better supported by default negation, which guarantees the existence of 2-valued models.

In our translation of default negation in terms of explicit negation, we combine the two "classical" aspects: Since we use explicit negation, its operational behavior is classical; since we start with initial program E_0 which contains all negative literals, we have the 2-valuedness guaranteed. And, the non-monotonic aspect of default negation is resolved in the update step.

7.4. Default and explicit negation in dynamic logic programming

The original formulation of dynamic logic programming in [Alferes et al., 2000] is based on default negation only. But it is argued that the addition of explicit negation to dynamic logic programming is easy, cf. [Ibid., Sec. 5.2], [Leite, 2003, p. 68]. The rejection of rules is still carried out by default negation only. Therefore, *strong negation rules* `not_a <- neg_a` and `not_neg_a <- a` are added which propagate the explicit negation to default negation.

We do not treat this combinations for the reasons given in the preceding subsection.[6] But we like to note, that the use of default negation together with explicit negation gives dynamic logic programming interesting expressive power. In fact, when Leite claims that "logic program updates constitute the *killer application* for generalized logic programs" [Leite, 2003, p. 20], i.e., for the use of default negation in the heads, he uses an example which makes essential use of both, default and explicit negation.[7]

Here, we will not discuss the question of the meaning of default negation in the heads, as used in dynamic logic programming. However, obviously, the

[5]It is a separate discussion "how classical" explicit negation is at the end. What we mean when we say that its operational behavior is closer to classical negation (than the one of default negation) is that we require *only* the consistency for explicit negation and nothing else; in particular it does not involve a non-monotonic aspect.

[6]See [Slota et al., 2014] for a discussion of the combination; this paper also contains additional references to the literature.

[7]The example is also given in [Slota et al., 2014, Example 1].

proposed reading relates default negation in the heads to explicit negation, just with the extra aspect of updates. Maybe, this could be used as an additional justification for generalized logic programs. As general references, aside from [Alferes et al., 2000] and [Lifschitz and Woo, 1992] which were already mentioned above, we can give [Inoue and Sakama, 1998] and [Damásio and Pereira, 1996]. However, the last reference deals mainly with the well-founded semantics instead of the stable model semantics used here.

We should also mention that the treatment of default negation in the presence of explicit negation allows for variations. They relate to the question how the set of rejected literals is defined, cf. e.g., [Leite, 1997; Leite, 2003], and, for a comparison, [Leite, 2004].

As related work in this direction we like to mention the alternative approach to dynamic updates proposed by Eiter et al. [Eiter et al., 2002]. It is based on extended logic programming, i.e., it allows both negations, but only explicit negation in the heads. The authors give a detailed discussion of the relation to dynamic logic programming in the sense of [Alferes et al., 2000], cf. [Eiter et al., 2002, Sec. 7.3].

For the question of combining different forms of negation, the work of Jonker [Jonker, 1994] could be also of interest. She introduces a new form of negation, called *imex* negation which combines aspects of *im*plicit (default) and *ex*plicit negation. It is open whether updates based on this negation would yield different results.

7.5. Stable models versus well-founded semantics

The given reading of default negation in terms of explicit negation plus update is based on the stable model semantics for the default case. It suggests itself to ask how the situation is in the case of *well-founded semantics*, [Gelder, et al., 1991]. This question is open. A well-founded semantics for dynamic logic programs was proposed by Banti, Alferes and Brogi, [Banti et al., 2004], and it could serve as a starting point.

7.6. The transformational semantics

In contrast to [Leite, 2003], in [Alferes et al., 2000] dynamic logic programming is introduced via a *transformational semantics*. In this case (which is equivalent to the declarative semantics given above, cf. [Leite, 2003, Th. 40, p. 66]) a dynamic logic program is translated in a generalized logic program which is formulated in an extended language. It provides for every atom $\mathtt{a}$ a new one $\mathtt{a}^-$ for its explicit negation, and two pairs of them indexed by every separate generalized program and indexed by the stage of the program. Here, we do not give the (longish, but not complicated) definition of the transformational semantics, but just point out that it starts with an *initial state* 0 in which all positive literals are declared to be false in terms of the new negative atom. It is given by *default rules*: $\mathtt{a}_0^-$ for all positive literals of the language. This initial state can serve as a motivation for the definition of E_0 in our translation above.

7.7. Negation as failure as abduction

There is at least a conceptional relation between our reading and the treatment of logic programs as *abductive frameworks*, cf. [Kakas et al., 1998, Ch. 4; in particular 4.1]. An abductive framework $\langle P, A, I\rangle$ consisting of a logic program P, a set of *abducibles* A, and integrity constraints I. Given a query `q`, one tries to find one (or more) subset(s) Δ of A such that $P \cup \Delta \models \mathtt{q}$ and $P \cup \Delta$ satisfies I. A logic program with default negation can be translated into an abductive framework where A contains the negated literals; technically, one does not work with the negated literals themselves, but replaces them by new symbols, such that the related system is entirely positive; I contains constraints such that the new literals are correct and complete with respect to negation, cf. [Kakas et al., 1998, p. 255f]. It is a result by Eshghi and Kowalski [Eshghi and Kowalski, 1989] that there is a one to one correspondence between stable models of a logic program and the abductive extensions of its abductive framework. *A fortiori*, our reading can also be related to the abductive framework. Somehow, we just assume all possible abducibles in E_0, and the update step rejects those which can not be used in a particular stable model. But, of course, the remaining set of negative literals could be bigger than the ones needed in a Δ. Thus, abduction allows for a finer analysis of the negative information needed to derive something.

7.8. Logical properties

The interpretation of default negation as explicit negation with update carries over the logical properties of default negation to the use of explicit negation in an update of E_0. Contraposition, for example, does not hold for default negation: `{a <- not_b}` has the only stable model `{a, not_b}` while `{b <- not_a}` has only `{b, not_a}` as stable model. Equally, $\{\mathtt{neg\_a}, \mathtt{neg\_b}\} \otimes \{$`a <- neg_b`$\}$ has only `{a, neg_b}` and $\{\mathtt{neg\_a}, \mathtt{neg\_b}\} \otimes \{$`b <- neg_a`$\}$ has only `{b, neg_a}` as 2-models. This can be checked directly along the lines of the example in Section 4.

Remark 1. The issue of contrapositive in logic programming is widely discussed, see, for instance, [Pearce, 1997, § 7.2]. We give here an example of *pedestrian lights* which should illustrate how it 'works' in our case.

Pedestrian lights are characterized by the two atoms `red` and `green`. Whether one should cross the street is determined by two rules:

```
go :- green
neg_go :- red
```

Now, the idea of default negation implies that, as long as I don't see the lights, neither `green` nor `red` should be assumed but rather the contrary. Thus, in our initial programme $\mathcal{E}_0$ we will have the two atoms: `neg_red` and `neg_green`. In this way, one cannot conclude whether one could cross the street or not. In this context, one can safely (in a literal sense!) assume the following rule:

```
red :- neg_green
```

It expresses that, as long as I don't have *positive* information that the lights are green, I 'better' assume that it is red (and, with the rules above, I conclude not to go). In contrast, the contrapositive of the rule (i.e., `green :- neg_red`) should not be assume (just imagine the lights are broken).

We added this example, as it illustrates two aspects of default negation, as they becomes visible in our reading as explicit negation plus update: First, the initial 'agnostic' state gives (total) preference to negative information; this alone shoud not be used to conclude positive actions (here: to cross the street); if there should be consequences concluded, they have to be given explicitely (here: to consider the lights being red as long as one doesn't have positive information about green) — but these rules may be *intensional* as they should not imply all of their consequences in classical logic. Secondly, updates allows to overwrite default assumptions, for instance, when one is seeing the green light.

One may ask how our approach works for *double negation*. Our syntax does not allow to iterate `neg`, and one would, first, have to change the language, introducing negation as an operator (rather than a prefix). The concept of 2-models would profit from an annulation of double negation. If this is not the case, a new fact `neg neg a` would have to rule out `neg a` in a model (that's the minimum we would expect from a 'negation'), but would not give support for `a`, thus 'distroying' the 2-valuedness of a model. In consequence, our set-up for explicit dynamic logic programs would fail, as updates could result in the 'destruction' of all models. The question how to deal with double negation in our framework, thus, is subject to further investigation.

7.9. Disjunctive logic programming

The question whether we can extend our framework to *disjunctive logic programming* [Minker, 1994] is even more challenging. *Extended* disjunctive logic programming is insofar out of reach, as we do not combine explicit and default negation in the same framework (see § 7.3.). For normal disjunctive logic programs, we would expect that our translation should work conceptionally; however, in this case, where 'disjunctions' (given, for example, as lists of atoms) can occur in the heads of rules, one would have, first, to redefine the update operation, as the notion of *conflicting* rules (Definition 3 and 6) is not any longer straightforward and as, in addition, the *Reject* operation (Definition 4) needs to be significantly refined. This applies, of course, also to *generalized disjunctive logic programs* [Lifschitz, 1996] where default negation may occur in the disjunction of a head of a rule and *general disjunctive programs* [Shen and Eiter, 2019], which take even into consideration arbitrary first-order formulas. It is to expect that one can go along the increasing complexity of the heads in disjunctive logic programming to introduce corresponding update frameworks, but to develop such frameworks is future work.

8. Coda

In the present paper, we gave a reading of default negation as explicit negation with update, which is a form to formalize the '*commonsense law of inertia*, which is the principle that things do not change unless they are made to' [Przymusinski and Turner, 1997, p. 126].

Form the perspective of Computer Science, our approach may give rise to more investigations of update phenomena in *answer set programming* [Slota and Leite, 2010; Slota and Leite, 2014] and can be linked to *action languages* (stemming from [Gelfond and Lifschitz, 1998]). Also, a more profound comparison with other semantical approaches to Answer Set Programming, as *equilibrium logic* and 'Here-and-There'-models [Pearce, 1997; Odintsov and Pearce, 2005; Pearce, 2006] and the related *Strong-Equivalence*-models [Turner, 2003]. This is or particular interest in view of the skeptical evaluation of these accounts for updates in [Slota and Leite, 2014]. This will be investigated in the future.

The purpose of this paper could be characterized as more philosophical: the reading of default negation as explicit negation plus update illustrates how the non-monotonic nature of default negation can be located in the update step. Methodologically, it allows to *modularize* semantic questions, concerning the default assumptions and updates. In this way, we hope to contribute to the analysis of non-monotonicity, not only in logic programming, but in philosophical logic in general.

Acknowledgements. This work is partially supported by the Udo Keller Foundation and by the Portuguese Science Foundation, FCT, through the project

UID/MAT/00297/2020 (Centro de Matemática e Aplicações). The author is grateful to an anonymous referee for helpful comments.

References

Alferes et al., 2000 – Alferes, J.J., Leite, J.A., Pereira, L.M., Przymusinska, H. and Przymusinski, T.C. "Dynamic Updates of Non-Monotonic Knowledge Bases", *The Journal of Logic Programming*, 2000, Vol. 45, No. 1–2, pp. 43–70.

Banti et al., 2004 – Banti, F., Alferes, J.J. and Brogi, A. "Well Founded Semantics for Logic Program Updates", in: *Advances in Artificial Intelligence – IBERAMIA 2004*, Volume 3315 of *Lecture Notes in Computer Science*, ed. by Chr. LemaÎtre, C.A. Reyes, and J.A. González. Springer, 2004, pp. 397–407.

Bidoit and Froidevaux, 1991 – Bidoit, N., and Froidevaux, C. "General logic databases and programs: Default logic semantics and stratification", *Information and Computation*, 1991, Vol. 91, No. 1, pp. 15–54.

Buccafurri et al., 1999 – Buccafurri, F., Faber, W. and Leone, N. "Disjunctive Logic Programs with Inheritance", in: *Proceedings of the 1999 International Conference on Logic Programming (ICLP-99)*, ed. by D. De Schreye. Cambridge: MIT Press, 1999, pp. 79–93.

Clark, 1978 – Clark, K.L. "Negation as failure", in: *Logic and Data Bases*, ed. by H. Gallaire and J. Minker. Plenum, 1978, pp. 293–322.

Damásio and Pereira, 1996 – Damásio, C.V., and Pereira, L.M. "Default Negation in the heads: Why not?", in: *Extensions of Logic Programming, ELP'96*, ed. by R. Dyckhoff, H. Herre, and P. Schroeder-Heister, Volume 1050 of *Lecture Notes in Artificial Intelligence*. Springer, 1996, pp. 103–117.

Eiter et al., 2002 – Eiter, Th., Fink, M., Sabbatini, G. and Tompits, H. "On properties of update sequences based on causal rejection", *Theory and Practice of Logic Programming*, 2002, Vol. 2, No. 6, pp. 711–767.

Eshghi and Kowalski, 1989 – Eshghi, K., and Kowalski, R.A. "Abduction compared with negation as failure", in: *Proc. 6th International Conference on Logic Programming*, ed. by G. Levi and M. Martelli. MIT Press, 1989, pp. 234–255.

Gelder, et al., 1991 – Gelder, A. Van, Ross, K.A. and Schlipf, J.S. "The well-founded semantics for general logic programs", *Journal of the ACM*, 1991, Vol. 38, No. 3, pp. 620–650.

Gelfond and Lifschitz, 1988 – Gelfond, M., and Lifschitz, V. "The Stable Model Semantics for Logic Programming", in: *5th International Conference on Logic Programming*, ed. by R. Kowalski and K. A. Bowen. MIT Press, 1988, pp. 1070–1080.

Gelfond and Lifschitz, 1998 – Gelfond, M., and Lifschitz, V. "Action Languages", *Electronic Transactions on Artificial Intelligence*, 1998, Vol. 2, pp. 193–210.

Gelfond and Lifschitz, 1991 – Gelfond, M., and Lifschitz, V. "Classical negation in logic programs and disjunctive databases", *New Generation Computing*, 1991, Vol. 9, No. 3–4, pp. 365–385.

Inoue and Sakama, 1998 – Inoue, K., and Sakama, C. "Negation as failure in the head", *Journal of Logic Programming*, 1998, Vol. 35, pp. 39–78.

Jonker, 1994 – Jonker, C. “Constraints and Negations in Logic Programming”, Ph.D. diss., Department of Philosophy, Utrecht University, 1994.

Kakas et al., 1998 – Kakas, A.C., Kowalski, R.A. and Toni, F. “The Role of Abduction”, in: *Handbook of Logic in Artificial Intelligence and Logic Programming*, ed. by D.M. Gabbay, C.J. Hogger, and J.A. Robinson, Volume 5: Logic Programming. Oxford, 1998, pp. 235–324.

Leite, 1997 – Leite, J.A. “Logic Program Updates”, Master’s thesis, Dept. de Informática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 1997.

Leite, 2003 – Leite, J.A. *Evolving Knowledge Bases.* Volume 81 of *Frontiers in Artificial Intelligence and Applications.* IOS Press, 2003.

Leite, 2004 – Leite, J.A. “On Some Differences Between Semantics of Logic Program Updates”, *Advances in Artificial Intelligence – IBERAMIA 2004*, Volume 3315 of *Lecture Notes in Computer Science*, ed. by Chr. LemaÎtre, C.A. Reyes, and J.A. González. Springer, 2004, pp. 375–385.

Leite and Pereira, 1998a – Leite, J.A., and Pereira, L.M. “Generalizing updates: from models to programs”, *LPKR’97: ILPS’97 workshop on Logic Programming and Knowledge Representation.* Springer, 1998. pp. 224–246.

Leite and Pereira, 1998b – Leite, J.A., and Pereira, L.M. “Iterated Logic Program Updates”, in: *Proceedings of the 1998 Joint International Conference and Symposium on Logic Programming (JICSLP-98)*, ed. by J. Jaffar. Cambridge: MIT Press, 1998, pp. 265–278.

Lifschitz, 1996 – Lifschitz, V. “Foundations of logic programming”, in: *Principles of Knowledge Representation*, ed. by G. Brewka. Stanford, California: CSLI, 1996, pp. 69–128.

Lifschitz et al., 2001 – Lifschitz, V., Pearce, D. and Valverde, A. “Strongly equivalent logic programs”, *ACM Transactions on Computational Logic*, 2001, Vol. 2, pp. 526–541.

Lifschitz and Woo, 1992 – Lifschitz, V., and Woo, T. “Answer sets in general nonmonotonic reasoning (Preliminary report)”, in: *Principles of Knowledge Representation and Reasoning (KR92)*, ed. by B. Nebel, C. Rich, and W. Swartout. Morgan-Kaufmann, 1992, pp. 603–614.

Minker, 1994 – Minker, J. “Overview of disjunctive logic progamming”, *Annals of Mathematics and Artificial Intelligence*, 1994, Vol. 12, pp. 1–24.

Odintsov and Pearce, 2005 – Odintsov, S., and Pearce, D. “Routley Semantics for Answer Sets”, in: *Logic Programming and Nonmonotonic Reasoning*, ed. by Ch. Baral, G. Greco, N. Leone, and G. Terracina, Volume 3662 of *Lecture Notes in Artificial Intelligence.* Springer, 2005, pp. 343–355.

Pearce, 1997 – Pearce, D. “A new logical characterization of stable models and answer sets”, in: *Non-Monotonic Extensions of Logic Programming*, ed. by J. Dix, L.M. Pereira, and T. Przymusinski, Volume 1216 of *Lecture Notes in Computer Science.* Springer, 1997, pp. 57–70.

Pearce, 1999 – Pearce, D. "From here to there: Stable negation in logic programming", in: *What is Negation?*, ed. by D. Gabbay and H. Wansing. Kluwer, 1999, pp. 161–181.

Pearce, 2006 – Pearce, D. "Equilibrium logic", *Annals of Mathematics and Artificial Intelligence*, 2006, Vol. 47, pp. 3–41.

Poole, 1994 – Poole, D. "Default Logic", in: *Handbook of Logic in Artificial Intelligence and Logic Programming*, ed. by Dov Gabbay, C.J. Hogger, and J.A. Robinson, Vol. 3. Oxford, 1994, pp. 189–215.

Przymusinski and Turner, 1997 – Przymusinski, T.C., and Turner, H. "Update by means of inference rules", *The Journal of Logic Programming*, 1997, Vol. 20, pp. 125–143.

Reiter, 1978 – Reiter, R. "On closed world data bases", in: *Logic and Data Bases*, ed. by H. Gallaire and J. Minker. Plenum, 1978, pp. 55–76.

Reiter, 1980 – Reiter, R. "A Logic for Default-Reasoning", *Artificial Intelligence*, 1980, Vol. 13, pp. 81–132.

Sakama and Inoue, 1999 – Sakama, C., and Inoue, K. "Updating Extended Logic Programs through Abduction", *Proceedings of the 5th International Conference on Logic Programming and Nonmonotonic Reasoning (LPNMR-99)*, Volume 1730 of *LNAI*, ed. by M. Gelfond, N. Leone, and G. Pfeifer. Berlin: Springer, 1999, pp. 147–161.

Shen and Eiter, 2019 – Shen, Y.-D., and Eiter, T. "Determining inference semantics for disjunctive logic programs", *Artificial Intelligence*, 2019, Vol. 277, pp. 103–165.

Shepherdson, 1998 – Shepherdson, J.C. "Negation as Failure, Completion and Stratification", in: *Handbook of Logic in Artificial Intelligence and Logic Programming*, ed. by D.M. Gabbay, C.J. Hogger, and J.A. Robinson, Vol. 5: Logic Programming. Oxford, 1998, pp. 355–419.

Slota et al., 2014 – Slota, M., Baláž, M., and Leite, J. "Supporting Strong and Default Negation in Answer-Set Program Updates", *Advances in Artificial Intelligence – IBERAMIA 2014*, Volume 8864 of *Lecture Notes in Computer Science*, ed. by A. Bazzan and K. Pichara. Springer, 2014, pp. 41–53.

Slota and Leite, 2010 – Slota, M., and Leite, J. "On Semantic Update Operators for Answer-Set Programs", in: *Proceedings of the 2010 conference on ECAI 2010: 19th European Conference on Artificial Intelligence*. IOS Press, 2010, pp. 957–962.

Slota and Leite, 2014 – Slota, M., and Leite, J. "The Rise and Fall of Semantic Rule Updates Based on SE-Models", *Theory and Practice of Logic Programming*, 2014, Vol. 14, pp. 869–907.

Turner, 2003 – Turner, H. "Strong equivalence made easy: nested expressions and weight constraints", *Theory and Practice of Logic Programming*, 2003, Vol. 3, pp. 609–622.

Zhang and Foo, 1998 – Zhang, Y., and Foo, N.Y. "Updating Logic Programs", in: *Proceedings of the 13th European Conference on Artificial Intelligence (ECAI-98)*, ed. by Henri Prade. Chichester: John Wiley & Sons, 1998, pp. 403–407.

Логические исследования
2021. Т. 27. № 1. С. 82–106
УДК 16+510.6

Logical Investigations
2021, Vol. 27, No. 1, pp. 82–106
DOI: 10.21146/2074-1472-2021-27-1-82-106

Paolo Maffezioli, Luca Tranchini

Equality and Apartness in Bi-intuitinistic Logic

Paolo Maffezioli
Department of Computer Science
Università di Verona,
Strada le Grazie 15, 37134, Verona.
E-mail: paolo.maffezioli@univr.it

Luca Tranchini
Department of Computer Science
Universität Tübingen,
Sand 13 D-72076 Tübingen.
E-mail: luca.tranchini@uni-tuebingen.de

Abstract: In the present paper we argue that a symmetric picture of the relationships between equality and apartness can be attained by considering these notions on the background of bi-intuitionistic logic instead of the usual intuitionistic logic. In particular we show that, as the intuitionistic negation of a relation of apartness is an equality, the dual-intuitionistic co-negation of an equality is a relation of apartness. At the same time, as the intuitionistic negation of equality is not an apartness, the co-intuitionistic negation of an apartness is not an equality.

Keywords: Bi-intuitionistic logic, Equality, Apartness

For citation: Maffezioli P., Tranchini L. "Equality and Apartness in Bi-intuitinistic Logic", *Logicheskie Issledovaniya / Logical Investigations*, 2021, Vol. 27, No. 1, pp. 82–106. DOI: 10.21146/2074-1472-2021-27-1-82-106

1. Introduction

Bi-intuitionistic logic, introduced by [Rauszer, 1974a; Rauszer, 1974b] as "H-B logic" and also called "subtractive logic" in [Crolard, 2001], is a conservative extension of propositional intuitionistic logic obtained by the addition of a new connective: co-implication $\not\supset$ (also referred to as "pseudo-difference", e.g., in Rauszer's original work, or as "subtraction", e.g., by Crolard).

While implication relates to conjunction as follows:

$$A \wedge B \vdash C \;\text{ iff } A \vdash B \supset C$$

co-implication relates to disjunction as follows:

$$A \vdash B \vee C \;\text{ iff }\; A \not\supset B \vdash C$$

Intuitively, a formula $A \not\supset B$ can be read as "A but not B" or "A excludes B".

In the $\{\wedge, \vee, \sim\}$-fragment of the language of classical logic both implication and co-implication can be defined, respectively as $\sim A \vee B$ and $A \wedge \sim B$. On the other hand, in intuitionistic logic implication is independent from the other connectives and a well-known result is that also co-implication is undefinable in terms of the intuitionistic connectives, see e.g., Theorem 5.2 in [Urbas, 1996].

As implication is the distinctive connective of intuitionistic logic, the natural habitat of co-implication is dual-intuitionistic logic. In sequent calculi, a system for intuitionistic logic can be obtained by restricting all sequents in a calculus for classical logic to at most one formula in the succedent. Similarly, a sequent calculus for dual-intuitionistic logic can be obtained by imposing the dual restriction to classical sequents: at most one formula in the antecedent.

In classical logic, logical consequence can be equivalently characterized as "forward" truth-preservation or as "backwards" falsity preservation in all interpretations: a sequent $\Gamma \Rightarrow \Delta$ is classically valid iff for every interpretation, if all formulas in Γ are true, at least one formula in Δ is true; or equivalently iff for every interpretation, if all formulas in Δ are false at least one formula in Γ is false.

Informally, intuitionistic logic can be thought of as the result of replacing the classical notion of truth with the constructive notion of proof, so that $\Gamma \Rightarrow A$ is intuitionistically valid iff a A is provable whenever all formulas in Γ are provable. As argued by the second author [Tranchini, 2012], dual-intuitionistic logic can be thought as the result of replacing the classical notion of falsity with a constructive notion of *refutation* or *disproof*. Accordingly $A \Rightarrow \Gamma$ is dual-intuitionistically valid iff A is refutable whenever all formulas in Γ are (here refutation is understood as a primitive notion, not to be defined in terms of some object-language negation operator).

The constructive nature of the notions of proof and refutation induce the rejection of certain classically valid principles in each of the two logics. The rejection of these principle is rewarded by stronger meta-theoretical properties, such as the disjunction property of intuitionistic logic (if $A \vee B$ is intuitionistically provable either A or B is) and its dual-intuitionistic counterpart (if $A \wedge B$ is dual-intuitionistically refutable either A or B is), which represent the cornerstone of their constructive reading.

One may expect that bi-intuitionistic logic could be given an informal interpretation in terms of both proofs and refutations. This is, however, no obvious task, due to the fact that the disjunction property and its dual do not hold in bi-intuitionistic logic.

A further difficulty concerns the proper formulation of the duality underlying bi-intuitionistic logic. The duality between intuitionistic and dual-intuitionistic logic can be made precise by defining a mapping $()^*$ from the language of bi-intuitionistic logic to itself so that $A^* = A$ for atomic proposition, and

$$\begin{aligned}(\top)^* &= \bot\\ (\bot)^* &= \top\\ (A \wedge B)^* &= A^* \vee B^*\\ (A \vee B)^* &= A^* \wedge B^*\\ (A \supset B)^* &= B^* \not\supset A^*\\ (A \not\supset B)^* &= B^* \supset A^*\end{aligned}$$

The duality amounts to the fact that $\Gamma \Rightarrow A$ is intuitionistically valid iff $A^* \Rightarrow \Gamma^*$ is dual-intuitionistically valid and $A \Rightarrow \Gamma$ is dual-intuitionistically valid iff $\Gamma^* \Rightarrow A^*$ is intuitionistically valid. This duality extends to bi-intuitionistic logic itself, so that $\Gamma \Rightarrow \Delta$ is bi-intuitionistically valid iff $\Delta^* \Rightarrow \Gamma^*$ is.

The clauses for the connectives may suggest the following informal reading of the duality: a proof of A is a refutation of A^* and viceversa. This is however, incompatible with the base clause: whereas $()^*$ relates pairs of (distinct) connectives, the atomic propositions are the dual of themselves and this blocks the possibility of informally reading the duality as exchanging the role of proofs and refutations (for similar reasons [Crolard, 2001] refers to $()^*$ as a "pseudo-duality", rather than a genuine duality).

A natural, although so far unexplored, way of solving this second difficulty is that of considering the duality between theories rather than purely logical systems and by introducing the dual of each primitive notion used in the formalization of a given theory.

The present paper aims to be a preliminary investigation in this direction. In particular, we consider one of the most elementary theories, that of equality, and address the question as to whether the notion of apartness—well-investigated in constructive mathematics—can be taken to play the role of the dual of equality.

In constructive mathematics, the relation of apartness $a \neq b$ is a "positive" counterpart of the negative notion of inequality $\neg a = b$. In intuitionistic logic, $\neg A$ is short for $A \supset \bot$ and hence $\neg a = b$ means that the assumption $a = b$

leads to a contradiction. By contrast, two real numbers a and b are apart, $a \neq b$, when there is a third one c measuring the distance of b from a on the real line. On the constructive reading of the existential quantifier, $a \neq b$ is a stronger claim than just $\neg a = b$, since the latter does not imply that the distance between a and b on the real line can be effectively computed.

In fact, given intuitionistic logic, one *cannot define apartness* by negating equality, since the relation so defined fails to satisfy some characteristic principles of apartness. At the same time, the intuitionistic negation of apartness is a relation satisfying reflexivity, symmetry and transitivity, that is one *can define a notion of equality* as negated apartness. The defined notion, however, is not just an equivalence relation, but an equivalence relation which is also stable, i.e., it satisfies the law of double negation elimination.

Thus whereas on the background of classical logic the notions of equality and apartness are perfectly symmetric (in classical mathematics, two numbers are apart iff they are not equal and they are equal iff they are not apart), this is definitely not the case on the background of intuitionistic logic.

This asymmerty may suggest that apartness is not the best candidate to act as the dual of equality. However, once the base clause of $()^*$ is replaced with the clauses

$$(x = y)^* = x \neq y \qquad\qquad (x \neq y)^* = x = y$$

the relation between the axioms of the theories of equality and apartness is the same as the one observed above (i.e. if $\Gamma \Rightarrow \Delta$ is an axiom of equality, $\Delta^* \Rightarrow \Gamma^*$ is an axiom of apartness and vice versa).

Moreover, the bi-intuitionistic setting offers a natural way to remedy the asymmetry between identity and apartness by considering two further notions besides equality, apartness and their intuitionistic negations, namely their dual-intuitionistic co-negations. In dual- and bi-intuitionistic logic one can define a unary connective called co-negation $\llcorner$ using $\not\supset$ and $\top$, by taking $\llcorner A$ as short for $\top \not\supset A$. Co-negation is the dual of intuitionistic negation, i.e. $\llcorner A = (\neg A)^*$ and $\neg A = (\llcorner A)^*$.

In the present paper we will show that, as the intuitionistic negation of a relation of apartness is an equality, the co-negation of an equality is a relation of apartness. At the same time, as the intuitionistic negation of equality is not an apartness, the co-intuitionistic negation of an apartness is not an equality.

Although the results presented do not exhaust the possibilities of investigating equality and apartness in the context of bi-intuitionistic logic, we believe that suggest so far unexplored, lines of research. In particular, they demonstrate that bi-intuitionistic logic is not only interesting as a logical system, but that it can be fruitfully applied to the study of mathematical theories.

The paper is structured as follows. In Section 2. we present the semantics and proof theory of bi-intuitionistic logic and in Section 3. we introduce the theories of equality and apartness that will be discussed in the present paper. In Section 4. we summarize the results concerning the intuitionistic theory of equality and apratness as well as a strengthening of the former (the theory of stable equality) and a weakening of the latter (here called the theory of weak apartness). Although most results of this section are not new, they are here embedded in a systematic picture and clearly formulated as (in some cases faithful) interpretations of the above mentioned theories into each other. In Section 5. we consider the four theories above on the background of bi-intuitionistic logic and we show how co-negation allows to establish further relationships between them. The resulting picture, however, does not seem to suggest a real symmetry between $=$ and $\neq$ in bi-intuitionistic logic. In Section 6., we dispel this impression by considering two further theories (a weakening of the theory of equality, that we call the theory of weak equality; and a strengthening of that of apartness, the theory of co-stable apartness) and we show how intuitionistic negation and dual-intuitionistic co-negation allow to interpret each theory into any other. In Section 7. we discuss the significance of the results presented and indicate several directions along which the present work can be further developed.

2. Preliminaries

Assumed countably many individual variables, to be indicated with $x, y, z \ldots$, let $\mathcal{L}$ be the language defined by the following grammar (we indicate formulas of $\mathcal{L}$ with $A, B, C \ldots$):

$$\mathcal{L} ::= \ x = y \mid x \neq y \mid (A \wedge B) \mid (A \vee B) \mid (A \supset B) \mid A \not\supset B \mid \bot \mid \top$$

The negation $\neg A$ of a formula A is defined as $A \supset \bot$ and its co-negation $\urcorner A$ as $\top \not\supset A$. We indicate with $\mathcal{L}^{=}$ and $\mathcal{L}^{\neq}$ (respectively) the $\neq$-free and $=$-free fragments of $\mathcal{L}$, and with $\mathcal{L}^{i=}$ and $\mathcal{L}^{i\neq}$ the $\not\supset$-free fragments of the latter languages.

A *sequent over* $\mathcal{L}$ is an expression of the form $(\Gamma \Rightarrow \Delta)$, where Γ and Δ are multisets of formulas of $\mathcal{L}$, called the antecedent and succedent of the sequent, respectively. Outermost parenthesis will be mostly omitted. We indicate with $\mathcal{S}(\mathcal{L})$ the set of sequents over $\mathcal{L}$. We use similar notions and notation for the fragments of $\mathcal{L}$ introduced above.

The semantics of $\mathcal{L}$ and its fragments is based on Kripke models. A bi-intuitionistic Kripke frame consists of a non-empty set K of worlds $\alpha, \beta, \gamma, \ldots$

pre-ordered by $\leq$ and a set D of objects $a, b, c, \ldots$ called domain. A bi-intuitionistic Kripke model M for $\mathcal{L}$ is a bi-intuitionistic Kripke frame equipped with an interpretation function I assigning in each world α two relations $=_\alpha$ and $\neq_\alpha$ on D to the symbols $=$ and $\neq$, respectively. It is assumed that if $\alpha \leq \beta$ and $a =_\alpha b$, then $a =_\beta b$ and if $\alpha \leq \beta$ and $a \neq_\alpha b$, then $a \neq_\beta b$. Sometimes the subscript in $=_\alpha$ and $\neq_\alpha$ will be omitted.

An assignment φ maps variables to elements in D. We define inductively what it means for a formula A of $\mathcal{L}$ to hold at a possible world α with respect to an assignment φ, in symbols $\alpha \Vdash^\varphi A$.

$$
\begin{array}{lll}
\alpha \nVdash^\varphi \bot & & \\
\alpha \Vdash^\varphi \top & & \\
\alpha \Vdash^\varphi x = y & \text{iff} & \varphi(x) =_\alpha \varphi(y) \\
\alpha \Vdash^\varphi x \neq y & \text{iff} & \varphi(x) \neq_\alpha \varphi(y) \\
\alpha \Vdash^\varphi A \wedge B & \text{iff} & \alpha \Vdash^\varphi A \text{ and } \alpha \Vdash^\varphi B \\
\alpha \Vdash^\varphi A \vee B & \text{iff} & \alpha \Vdash^\varphi A \text{ or } \alpha \Vdash^\varphi B \\
\alpha \Vdash^\varphi A \supset B & \text{iff} & \beta \Vdash^\varphi B \text{ for all } \beta \geq \alpha \text{ s.t. } \beta \Vdash^\varphi A \\
\alpha \Vdash^\varphi A \not\supset B & \text{iff} & \beta \nVdash^\varphi B \text{ for some } \beta \leq \alpha \text{ s.t. } \beta \Vdash^\varphi A
\end{array}
$$

It is easily verified that if $\alpha \leq \beta$ and $\alpha \Vdash^\varphi A$ then $\beta \Vdash^\varphi A$. Moreover, observe that $\alpha \Vdash^\varphi \neg A$ iff $\beta \nVdash^\varphi A$, for all $\beta \geq \alpha$; and that $\alpha \Vdash^\varphi \urcorner A$ iff $\beta \nVdash^\varphi A$ for some $\beta \leq \alpha$.

We can now define the notion of validity in a class of models. A sequent holds at α with respect to φ, written $\alpha \Vdash^\varphi (\Gamma \Rightarrow \Delta)$, when for all $A \in \Gamma$, if $\alpha \Vdash^\varphi A$, then $\alpha \Vdash^\varphi B$, for some $B \in \Delta$. A sequent $(\Gamma \Rightarrow \Delta)$ is valid in a model M, written $M \Vdash (\Gamma \Rightarrow \Delta)$, when $\alpha \Vdash^\varphi (\Gamma \Rightarrow \Delta)$ for all $\alpha \in K$ and φ. Notice that the notion of validity in a model given here is *local*, see, for a discussion[Goré et al., 2020]. A sequent is valid in a class of models $\mathscr{C}$, indicated as $\mathscr{C} \Vdash (\Gamma \Rightarrow \Delta)$, when $M \Vdash (\Gamma \Rightarrow \Delta)$ for all $M \in \mathscr{C}$. Opportune restrictions of these notions to the various fragments of $\mathcal{L}$ will be used throughout.

By a theory over $\mathcal{L}$ we understand a subset of $\mathcal{S}(\mathcal{L})$ and similarly for its fragments. We shall conveniently use Gentzen calculi to describe theories, so that a theory will be identified with the set obtained by closing under the rules of inference the set of initial sequents of the calculus. In the present paper we will be almost exclusively concerned with three theories over $\mathcal{L}^{i=}$ and three theories over $\mathcal{L}^{i\neq}$. We describe them as extensions of a basic sequent calculus for bi-intuitionistic logic with non-logical initial sequents corresponding to the properties of equality and apartness. The underlying logical calculus G consists of the initial sequents and logical rules given in Table 1 (this is in fact the system $\mathbf{LBJ}_1$ of [Kowalski et al. 2017]).

$$\Rightarrow \top \qquad\qquad A \Rightarrow A \qquad\qquad \bot \Rightarrow$$

$$\frac{A, \Gamma \Rightarrow \Delta}{A \wedge B, \Gamma \Rightarrow \Delta}\,L\wedge 1 \qquad \frac{B, \Gamma \Rightarrow \Delta}{A \wedge B, \Gamma \Rightarrow \Delta}\,L\wedge 2 \qquad \frac{\Gamma \Rightarrow \Delta, A \quad \Gamma \Rightarrow \Delta, B}{\Gamma \Rightarrow \Delta, A \wedge B}\,R\wedge$$

$$\frac{A, \Gamma \Rightarrow \Delta \quad B, \Gamma \Rightarrow \Delta}{A \vee B, \Gamma \Rightarrow \Delta}\,L\vee \qquad \frac{\Gamma \Rightarrow \Delta, A}{\Gamma \Rightarrow \Delta, A \vee B}\,R\vee 1 \qquad \frac{\Gamma \Rightarrow \Delta, B}{\Gamma \Rightarrow \Delta, A \vee B}\,R\vee 2$$

$$\frac{\Gamma \Rightarrow \Delta, A \quad B, \Sigma \Rightarrow \Pi}{A \supset B, \Gamma, \Sigma \Rightarrow \Delta, \Pi}\,L\supset \qquad \frac{A, \Gamma \Rightarrow B}{\Gamma \Rightarrow A \supset B}\,R\supset$$

$$\frac{A \Rightarrow \Delta, B}{A \not\supset B \Rightarrow \Delta}\,L\not\supset \qquad \frac{\Gamma \Rightarrow \Delta, A \quad B, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, A \not\supset B}\,R\not\supset$$

$$\frac{\Gamma \Rightarrow \Delta}{A, \Gamma \Rightarrow \Delta}\,LW \qquad \frac{\Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, A}\,RW \qquad \frac{A, A, \Gamma \Rightarrow \Delta}{A, \Gamma \Rightarrow \Delta}\,LC \qquad \frac{\Gamma \Rightarrow \Delta, A, A}{\Gamma \Rightarrow \Delta, A}\,RC \qquad \frac{\Gamma \Rightarrow \Delta, A \quad A, \Sigma \Rightarrow \Pi}{\Gamma, \Sigma \Rightarrow \Delta, \Pi}\,cut$$

Table 1. The sequent calculus G

weq := $\left\{\begin{array}{ll} \Rightarrow x = x & \text{(Reflexivity)} \\ ⌐x = y \Rightarrow ⌐x = z, ⌐z = y & \text{(Co-negative Co-transitivity)} \\ ⌐x = y \Rightarrow ⌐y = x & \text{(Co-negative symmetry)} \end{array}\right.$	wap := $\left\{\begin{array}{ll} x \neq x \Rightarrow & \text{(Irreflexivity)} \\ \neg x \neq z, \neg z \neq y \Rightarrow \neg x \neq y & \text{(Negative transitivity)} \\ \neg x \neq y \Rightarrow \neg y \neq x & \text{(Negative symmetry)} \end{array}\right.$
eq := $\left\{\begin{array}{ll} \Rightarrow x = x & \text{(Reflexivity)} \\ x = z, z = y \Rightarrow x = y & \text{(Transitivity)} \\ x = y \Rightarrow y = x & \text{(Symmetry)} \end{array}\right.$	ap := $\left\{\begin{array}{ll} x \neq x \Rightarrow & \text{(Irreflexivity)} \\ x \neq y \Rightarrow x \neq z, z \neq y & \text{(Co-transitivity)} \\ x \neq y \Rightarrow y \neq x & \text{(Symmetry)} \end{array}\right.$
seq := $\left\{\begin{array}{ll} \mathsf{eq} \cup & \\ \neg\neg x = y \Rightarrow x = y & \text{(Stability)} \end{array}\right.$	sap := $\left\{\begin{array}{ll} \mathsf{ap} \cup & \\ x \neq y \Rightarrow ⌐⌐x \neq y & \text{(Co-stability)} \end{array}\right.$

Table 2. The sets of initial sequents yielding the different theories to be considered

We observe that by the definition of (dual-)intuitionistic (co-)negation, the following rules are derivable:

$$\frac{\Gamma \Rightarrow \Delta, A}{\neg A, \Gamma \Rightarrow \Delta}\ L\neg \qquad \frac{A, \Gamma \Rightarrow}{\Gamma \Rightarrow \neg A}\ R\neg$$

$$\frac{\Rightarrow \Delta, A}{\llcorner A \Rightarrow \Delta}\ L\llcorner \qquad \frac{A, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \llcorner A}\ R\llcorner$$

We indicate with superscripts the restriction of G to the corresponding fragments of $\mathcal{L}$ (so that e.g. $\mathsf{G}^{i=}$ is the restriction of G to $\mathcal{L}^{i=}$).

3. Six theories and their duality

We will consider theories obtained by extending (fragments of) G with the sets of initial sequents schematically depicted in Table 2. For each such theory T, we write $\mathsf{T} \vdash (\Gamma \Rightarrow \Delta)$ iff the sequent belongs to the theory, i.e., it is derivable in the sequent calculus describing T.

In particular, we call the result of adding to $\mathsf{G}^{=}$ the set of initial sequents eq (resp. $\mathsf{weq}/\mathsf{seq}$) the *bi-intuitionistic theory of (resp. weak/stable) equality*, to be indicated with EQ (resp. $\mathsf{WEQ}/\mathsf{SEQ}$).

Similarly, we call the result of adding to $\mathsf{G}^{\neq}$ the set of initial sequents ap (resp. $\mathsf{wap}/\mathsf{sap}$) the *bi-intuitionistic theory of (resp. weak/stable) equality*, to be indicated with AP (resp. $\mathsf{WAP}/\mathsf{SAP}$).

Note that, given the rules for dual-intuitionistic co-negation, transitivity and symmetry imply co-negative co-transitivity and co-negative symmetry (respectively); and that, given the rules of intuitionistic negation, co-transitivity and symmetry imply negative transitivity and negative symmetry (respectively). Hence $\mathsf{WEQ} \subset \mathsf{EQ} \subset \mathsf{SEQ}$ and $\mathsf{WAP} \subset \mathsf{AP} \subset \mathsf{SAP}$.

We will refer to the restrictions of EQ, (respectively SEQ) and AP (respectively WAP) to the languages $\mathcal{L}^{i=}$ and $\mathcal{L}^{i\neq}$ as the intuitionistic theories of (resp. stable) equality and (resp. weak) apartness, to be indicated with EQ^i, (resp. SEQ^i) and AP^i (resp. WAP^i). These four theories have been extensively investigated in the literature, and the intuitionistic theory of weak apartness is sometimes called "negative equality" in [Negri et al., 2001] or "defined equality" in [Negri, 1999].[1]

From the perspective of bi-intuitionistic logic, the set of axioms eq and ap suggests that the relationship between $=$ and $\neq$ should be the same as the one between $\wedge$ and $\vee$ and $\supset$ and $\not\supset$ that we described in the introduction as a duality. To spell it out properly we consider a further theory on the full

[1] Our terminological choice has the only purpose of making easier for the reader to remember which theories are based on the language $\mathcal{L}^{i\neq}$ and which are based on $\mathcal{L}^{i=}$.

language $\mathcal{L}$, obtained by extending G with both eq and ap, to be referred to as EA. Let $()^* : \mathcal{L} \mapsto \mathcal{L}$ be defined as described in the introduction, namely:

$$
\begin{array}{cc}
(x = y)^* := x \neq y & (x \neq y)^* := x = y \\
(\top)^* := \bot & (\bot)^* := \top \\
(A \wedge B)^* := A^* \vee B^* & (A \vee B)^* := A^* \wedge B^* \\
(A \supset B)^* := B^* \not\supset A^* & (A \not\supset B)^* := B^* \supset A^*
\end{array}
$$

and let $(\Gamma \Rightarrow \Delta)^* := (\Delta^* \Rightarrow \Gamma^*)$, where Σ^* is the multiset of all A^* such that $A \in \Sigma$. The following holds:

Theorem 1.

$$\mathsf{EA} \vdash (\Gamma \Rightarrow \Delta) \quad \textit{iff} \quad \mathsf{EA} \vdash (\Gamma \Rightarrow \Delta)^*$$

Proof. The proposition is established by induction on the derivation of $\Gamma \Rightarrow \Delta$, by constructing a derivation of $(\Gamma \Rightarrow \Delta)^*$ that we call *the dual of the given derivation* of $\Gamma \Rightarrow \Delta$. If $\Gamma \Rightarrow \Delta$ is an initial sequent it is easily verified that $(\Gamma \Rightarrow \Delta)^*$ is an initial sequent as well. If the derivation of $\Gamma \Rightarrow \Delta$ ends with an application of, e.g., $R\supset$ it suffice to apply the induction hypothesis to the immediate sub-derivation and the dual of the original derivation can be obtained by opportunely applying $L\not\supset$. The other cases are treated analogously, by exchanging an application of a left/right operational rules by an application of the right/left rule for the dual connective (in the case of structural rules it is enough to exchange left/right). ■

As suggested in the introduction, in bi-intuitionistic logic the duality can be informally understood as the possibility of obtaining meaning explanation for the dual of a give formula by replacing the notions of proof and refutation in the meaning explanations of the original formula.[2] For example, the proof-conditions of $A \supset B$ can be expressed by saying that a proof of $A \supset B$ is a method to transform any proof of A into a proof of B; and the refutation-conditions of $A^* \not\supset B^*$ can be expressed by saying that a refutation of $A^* \not\supset B^*$ is a method to transform any refutation of B^* into a refutation of A^*. In particular, a proof of $\neg A$ is a method to transform proofs of A into proofs of $\bot$ (which is by definition the proposition of which there is no proof); dually a refutation of $\llcorner A^*$ is a method to transform refutations of A^* into refutations of $\top$ (which is by definition the proposition of which there is no refutation).

[2]However, as observed in the introduction, due to the fact that in bi-intuitionistic logic both disjunction property and its dual fail, it is not wholly clear how should the notions of proof of a disjunction and of refutation of a conjunction be informally characterized so as to fit the bi-intuitionistic setting. Addressing this additional difficulty goes beyond the scope of the present paper.

Proposition 1 suggests to extend this informal interpretation to the case of $=$ and $\neq$, so that as we can read e.g., transitivity as saying that given proofs of $x = z$ and $z = y$ we can construct a proof of $x = z$, we can read co-transitivity as saying that given refutations of $x \neq z$ and $z \neq y$ then we can construct a refutation of $x \neq y$.

As the results of Section 4. show, the set of initial sequents seq and wap naturally aroused by studying the relationship between $=$ and $\neq$ using intutuionistic negation. The two sets of initial sequents weq and sap are motivated by considerations of duality: as the reader can easily verify, Proposition 1 holds if one replaces the theory EA with either of the two theories that one obtains by extending G with either weq and wap, or with seq and sap. The notion of dual of a derivation, as defined in the proof of Proposition 1 extends to these further two theories as well. We observe that by construction, if the derivation of $\Gamma \Rightarrow \Delta$ is a derivation in EQ (resp. AP), its dual is a derivation in AP (resp. EQ), and similarly for WEQ and WAP and for SEQ and SAP.

In this case as well, by duality we obtain an informal interpretation of e.g. co-stability as warranting that from a refutation of $\ulcorner\ulcorner A$ one can obtain a refutation of A (thus dualizing the informal reading of stability), and similarly for co-negative co-transitivity and co-negative symmetry.

To establish the results below we will rely on the soundness of these theories with respect to certain classes of bi-intuitionistic models. In particular, we indicate with $\mathscr{WEQ}$, $\mathscr{EQ}$, $\mathscr{SEQ}$, $\mathscr{WAP}$ $\mathscr{AP}$ and $\mathscr{SAP}$ the classes of models in which all sequents in weq, eq, seq, wap, ap and sap (respectively) are valid.

It is easy to see that each of the four theories considered is sound in the corresponding class of models, that is:

Theorem 2 (soundeness)**.** *The following hold:*

1. *if* $\mathsf{WEQ} \vdash (\Gamma \Rightarrow \Delta)$*, then* $\mathscr{WEQ} \Vdash (\Gamma \Rightarrow \Delta)$*;*
2. *if* $\mathsf{EQ} \vdash (\Gamma \Rightarrow \Delta)$*, then* $\mathscr{EQ} \Vdash (\Gamma \Rightarrow \Delta)$*;*
3. *if* $\mathsf{SEQ} \vdash (\Gamma \Rightarrow \Delta)$*, then* $\mathscr{SEQ} \Vdash (\Gamma \Rightarrow \Delta)$*;*
4. *if* $\mathsf{WAP} \vdash (\Gamma \Rightarrow \Delta)$*, then* $\mathscr{WAP} \Vdash (\Gamma \Rightarrow \Delta)$*;*
5. *if* $\mathsf{AP} \vdash (\Gamma \Rightarrow \Delta)$*, then* $\mathscr{AP} \Vdash (\Gamma \Rightarrow \Delta)$*.*
6. *if* $\mathsf{SAP} \vdash (\Gamma \Rightarrow \Delta)$*, then* $\mathscr{SAP} \Vdash (\Gamma \Rightarrow \Delta)$*.*

Proof. For each theory T the proof is by induction on the length of the derivation of $\Gamma \Rightarrow \Delta$. If $\Gamma \Rightarrow \Delta$ is an initial sequent that is not of the form

$A \Rightarrow A$ the proposition is an immediate consequence of the way in which the classes of models have been defined. The remaining cases are standard. ■

If $\sigma : \mathcal{L}_1 \to \mathcal{L}_2$ is a translation from $\mathcal{L}_1$ to $\mathcal{L}_2$ and $(\Gamma \Rightarrow \Delta) = (A_1, \dots, A_n \Rightarrow B_1, \dots, B_m)$ we write $\sigma(\Gamma \Rightarrow \Delta)$ for $\sigma(A_1), \dots, \sigma(A_n) \Rightarrow \sigma(B_1), \dots, \sigma(B_m)$. Let X and Y be two theories over $\mathcal{L}_1$ and $\mathcal{L}_2$, respectively. We say that $\sigma : \mathcal{L}_1 \to \mathcal{L}_2$ is an *interpretation* of X in Y when for all $(\Gamma \Rightarrow \Delta) \in \mathcal{S}(\mathcal{L}_1)$, if $\mathsf{X} \vdash (\Gamma \Rightarrow \Delta)$, then $\mathsf{Y} \vdash \sigma(\Gamma \Rightarrow \Delta)$. Moreover, we say that σ *preserves* X in Y when the converse holds, namely for no sequent $(\Gamma \Rightarrow \Delta) \in \mathcal{S}(\mathcal{L}_\infty)$, $\mathsf{Y} \vdash \sigma(\Gamma \Rightarrow \Delta)$ and $\mathsf{X} \nvdash (\Gamma \Rightarrow \Delta)$ (we call such sequents, if they exists, *counterexamples* to preservation). A preserving interpretation σ is said to be *faithful*.

We shall assume throughout that translations are structural, i.e., $\sigma(A \circ B) := \sigma(A) \circ \sigma(B)$, where $\circ$ is a binary connective, $\sigma(\bot) := \bot$ and $\sigma(\top) := \top$.

We observe the following facts (proofs are obvious and left to the reader):

Fact 1. If X and Y are two theories over the same language $\mathcal{L}$ and $\mathsf{X} \subseteq \mathsf{Y}$, then $id_{\mathcal{L}}$, (the identity function on $\mathcal{L}$) is an interpretation of X in Y, but not necessarily of Y in X. Moreover, $id_{\mathcal{L}}$ preserves X in Y iff $X = Y$.

Fact 2. If σ and τ are interpretations of X in Y and of Y in Z, respectively, then the composition $\tau \circ \sigma$ is an interpretation of X in Z.

Fact 3. If σ is a faithful interpretation of X in Y and τ is a non-faithful interpretation of Y in Z, $\tau \circ \sigma$ is a faithful interpretation of X in Z iff no counterexample $(\Gamma \Rightarrow \Delta)$ to the faithfulness of τ is in the range of σ, i.e., there is no $(\Sigma \Rightarrow \Theta)$ such that $\sigma(\Sigma \Rightarrow \Theta) = (\Gamma \Rightarrow \Delta)$.

We conclude this section by observing that although the theories we consider could be formulated in a first-order setting as well, most of the results below hold only for the propositional versions of the theories (see in particular footnote 3 below). Moreover, bi-intuitionistic logic is non-conservative over first-order intutionistic logic (but only over the first-order logic of constant domains). For these reasons we decided to limit our attention to propositional theories.

4. Relating $=$ and $\neq$ with intuitionistic negation

In this section we will establish to which extent the two translations $\sigma^\neg : \mathcal{L}^{i=} \to \mathcal{L}^{i\neq}$ and $\tau^\neg : \mathcal{L}^{i\neq} \to \mathcal{L}^{i=}$

$$\sigma^\neg(x = y) := \neg x \neq y \qquad \qquad \tau^\neg(x \neq y) := \neg x = y$$

can be used to (faithfully) interpret the four theories EQ^i, SEQ^i, AP^i and WAP^i into each other.

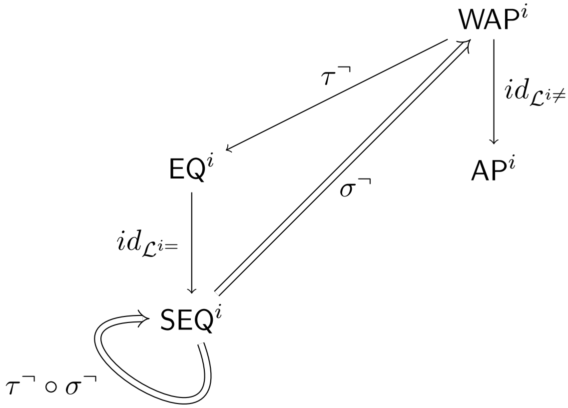


Fig. 1. Relating EQ^i, SEQ^i, AP^i and WAP^i with intuitionistic negation

Informally, the translation $\sigma^\neg$ can be seen as an attempt to define of equality in terms of apartness and negation by taking a proof of $x = y$ to be a method to transform a proof of $x \neq y$ into a proof of $\bot$; and the translation $\tau^\neg$ can be read as an attempt to define of apartness in terms of equality and negation by taking a proof of $x \neq y$ to be a method to transform a proof of $x = y$ into a proof of $\bot$.

To enhance readability, we summarize the results to be established in Figure 1, where single/double arrows indicate non-faithful/faithful interpretations and in which we also indicate the non-faithful interpretations $id_{\mathcal{L}^{i=}}$ and $id_{\mathcal{L}^{i\neq}}$ of EQ^i in SEQ^i and of WAP^i in AP^i respectively (see Fact 1 above).

Theorem 3. *$\tau^\neg$ is a non-faithful interpretation of WAP^i in EQ^i.*

Proof. To show that $\tau^\neg$ is an interpretation of WAP^i in EQ^i, it is enough to show that $\mathsf{EQ}^i \vdash \tau^\neg(x \neq x \Rightarrow)$, $\mathsf{EQ}^i \vdash \tau^\neg(\neg x \neq z, \neg z \neq y \Rightarrow \neg x \neq y)$ and $\mathsf{EQ}^i \vdash \tau^\neg(\neg x \neq y \Rightarrow \neg y \neq x)$:

$$\dfrac{\Rightarrow x = x}{\neg x = x \Rightarrow}\,L\neg$$

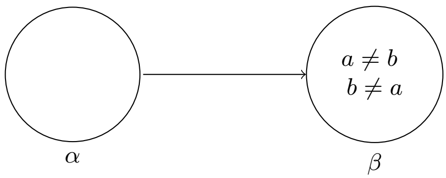
$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{x = z, z = y \Rightarrow x = y}{\neg x = y, x = z, z = y \Rightarrow}\,L\neg}{\neg x = y, x = z \Rightarrow \neg z = y}\,R\neg}{\neg x = y, x = z, \neg\neg z = y \Rightarrow}\,L\neg}{\neg x = y, \neg\neg z = y \Rightarrow \neg x = z}\,R\neg}{\neg x = y, \neg\neg x = z, \neg\neg z = y \Rightarrow}\,L\neg}{\neg\neg x = z, \neg\neg z = y \Rightarrow \neg\neg x = y}\,R\neg$$

$$\dfrac{\dfrac{\dfrac{\dfrac{x = y \Rightarrow y = x}{\neg y = x, x = y \Rightarrow}\,L\neg}{\neg y = x \Rightarrow \neg x = y}\,R\neg}{\neg\neg x = y, \neg y = x \Rightarrow}\,L\neg}{\neg\neg x = y \Rightarrow \neg\neg y = x}\,R\neg$$

To show that the interpretation is non-faithful, i.e., that $\tau^\neg$ does not preserve WAP^i in EQ^i, we consider the sequent $(\neg\neg x \neq y \Rightarrow x \neq y)$. Clearly, $\mathsf{EQ}^i \vdash \tau^\neg(\neg\neg x \neq y \Rightarrow x \neq y)$:

$$\dfrac{\dfrac{\dfrac{\dfrac{x = y \Rightarrow x = y}{x = y, \neg x = y \Rightarrow}\,L\neg}{x = y \Rightarrow \neg\neg x = y}\,R\neg}{x = y, \neg\neg\neg x = y \Rightarrow}\,L\neg}{\neg\neg\neg x = y \Rightarrow \neg x = y}\,R\neg$$

However, $\mathsf{WAP}^i \nvdash (\neg\neg x \neq y \Rightarrow x \neq y)$. Consider a Kripke model M_1^i with two worlds α and β such that $\alpha \leq \beta$ and two objects a and b in D such that $a \neq_\beta b$ and $b \neq_\beta a$, i.e.,



Let $\varphi(x)$ and $\varphi(y)$ be a and b, respectively. Thus, $\alpha \Vdash^\varphi \neg\neg x \neq y$ since for all worlds $\delta \geq \alpha$ there is a world $\epsilon \geq \delta$ such that $\epsilon \Vdash^\varphi x \neq y$. But clearly $\alpha \nVdash^\varphi x \neq y$. Therefore $M_1^i \nVdash \neg\neg x \neq y \Rightarrow x \neq y$. Moreover, it is easy to see that $M_1^i \in \mathscr{WAP}^i$. We only show that M_1^i satisfies the negative transitivity principle, namely $M_1^i \Vdash (\neg x \neq y, \neg y \neq z \Rightarrow \neg x \neq z)$. It suffices to show that $\alpha \Vdash^{\varphi_i} (\neg x \neq y, \neg y \neq z \Rightarrow \neg x \neq z)$, for all $i = 1, \dots, 8$ such that:

1. $\varphi_1(x)$, $\varphi_1(y)$ and $\varphi_1(z)$ are a
2. $\varphi_2(x)$, $\varphi_2(y)$ are a, $\varphi_2(z)$ is b
3. $\varphi_3(x)$, $\varphi_3(z)$ are a, $\varphi_3(y)$ is b
4. $\varphi_4(y)$, $\varphi_4(z)$ are b, $\varphi_4(x)$ is a
5. $\varphi_5(y)$, $\varphi_5(z)$ are a, $\varphi_5(x)$ is b
6. $\varphi_6(x)$, $\varphi_6(z)$ are b, $\varphi_6(y)$ is a
7. $\varphi_7(x)$, $\varphi_7(y)$ are b, $\varphi_7(z)$ is a
8. $\varphi_1(x)$, $\varphi_1(y)$, $\varphi_1(z)$ are b

Cases (3) and (6) hold since the succedent is valid; all the remaining cases are valid since they can be obtained from initial sequents using weakening. We leave to the reader to verify that also the negative symmetry principles is valid in M_2^i. Thus, we conclude that $\mathscr{WAP}^i \nVdash (\neg\neg x \neq y \Rightarrow x \neq y)$, hence by soundness $\mathsf{WAP}^i \nvdash (\neg\neg x \neq y \Rightarrow x \neq y)$. ■

To establish the next proposition we will need the following:

Lemma 1. *For all $A \in \mathcal{L}^{i=}$:*

1. $\mathsf{SEQ}^i \vdash (A \Rightarrow \tau^\neg \circ \sigma^\neg(A))$

2. $\mathsf{SEQ}^i \vdash (\tau^\neg \circ \sigma^\neg(A) \Rightarrow A)$

Proof. We establish the two claims by simultaneous induction on A:

- A is $x = y$. Clearly, $\mathsf{SEQ}^i \vdash (x = y \Rightarrow \neg\neg x = y)$:

$$\dfrac{\dfrac{x = y \Rightarrow x = y}{x = y, \neg x = y \Rightarrow}\ L\neg}{x = y \Rightarrow \neg\neg x = y}\ R\neg$$

  and obviously $\mathsf{SEQ}^i \vdash (\neg\neg x = y \Rightarrow x = y)$ given the stability initial sequents.

- A is P,$\top$ or $\bot$. Obvious.

- A is $B \supset C$. We have that $\tau^\neg \circ \sigma^\neg(A) = \tau^\neg(\sigma^\neg(B \supset C)) = \tau^\neg \circ \sigma^\neg(B) \supset \tau^\neg \circ \sigma^\neg(C)$. By induction hypothesis we have that

  $\mathsf{SEQ}^\mathsf{i} \vdash (B \Rightarrow \tau^\neg \circ \sigma^\neg(B))$ $\mathsf{SEQ}^\mathsf{i} \vdash (\tau^\neg \circ \sigma^\neg(B) \Rightarrow B)$

  $\mathsf{SEQ}^\mathsf{i} \vdash (C \Rightarrow \tau^\neg \circ \sigma^\neg(C))$ $\mathsf{SEQ}^\mathsf{i} \vdash (\tau^\neg \circ \sigma^\neg(C) \Rightarrow C)$ and hence:

$$\dfrac{\dfrac{B \Rightarrow \tau^\neg \circ \sigma^\neg(B) \quad \tau^\neg \circ \sigma^\neg(C) \Rightarrow C}{\tau^\neg \circ \sigma^\neg(B) \supset \tau^\neg \circ \sigma^\neg(C), B \Rightarrow C}\ L\supset}{\tau^\neg \circ \sigma^\neg(B) \supset \tau^\neg \circ \sigma^\neg(C) \Rightarrow B \supset C}\ R\supset \qquad \dfrac{\dfrac{\tau^\neg \circ \sigma^\neg(B) \Rightarrow B \quad C \Rightarrow \tau^\neg \circ \sigma^\neg(C)}{B \supset C, \tau^\neg \circ \sigma^\neg(B) \Rightarrow \tau^\neg \circ \sigma^\neg(C)}\ L\supset}{B \supset C \Rightarrow \tau^\neg \circ \sigma^\neg(B) \supset \tau^\neg \circ \sigma^\neg(C)}\ R\supset$$

- A is $B \wedge C$ or $B \vee C$. Similar to the previous case.

■

Corollary 1. $\tau^\neg \circ \sigma^\neg$ is a faithful interpretation of SEQ^i into itself.

Proof. From the previous lemma it is almost immediate that $\mathsf{SEQ}^i \vdash (\Gamma \Rightarrow \Delta)$ iff $\mathsf{SEQ}^i \vdash \tau^\neg \circ \sigma^\neg(\Gamma \Rightarrow \Delta)$ for any $(\Gamma \Rightarrow \Delta) \in \mathcal{L}^{i=}$. Let $(\Gamma \Rightarrow \Delta)$ be $(A_1, \ldots A_n, \Rightarrow B_1, \ldots, B_m)$. The corollary follows by $n + m$ applications of the *Cut* rule using derivations of $A_i \Rightarrow \tau^\neg \circ \sigma^\neg(A_i)$ and $\tau^\neg \circ \sigma^\neg(B_j) \Rightarrow B_j$ for the one direction, and derivations of $B_j \Rightarrow \tau^\neg \circ \sigma^\neg(B_j)$ and $\tau^\neg \circ \sigma^\neg(A_i) \Rightarrow A_i$ for the other direction. ■

Theorem 4. *$\sigma^\neg$ is a faithful interpretation of SEQ^i in WAP^i.*

Proof. To show that $\tau^\neg$ is an interpretation of SEQ^i in WAP^i, it is enough to prove that $\mathsf{WAP}^i \vdash \sigma^\neg(\Rightarrow x = x)$, $\mathsf{WAP}^i \vdash \sigma^\neg(x = z, z = y \Rightarrow x = y)$, $\mathsf{WAP}^i \vdash \sigma^\neg(x = y \Rightarrow y = x)$, as well as $\mathsf{WAP}^i \vdash \sigma^\neg(\neg\neg x = y \Rightarrow x = y)$. The first three claims are obvious since the translation of each such sequent is an initial sequent of WAP^i. The last claim is established by a derivation of

$(\neg\neg\neg x \neq y \Rightarrow \neg x \neq y)$ that can be obtained by replacing $=$ with $\neq$ in the last derivation used in the proof of Proposition 3.

To show faithfulness, we need to show that, for all $(\Gamma \Rightarrow \Delta) \in \mathcal{S}(\mathcal{L}^{i=})$, if $\mathsf{WAP}^i \vdash \sigma^\neg(\Gamma \Rightarrow \Delta)$, then $\mathsf{SEQ}^i \vdash (\Gamma \Rightarrow \Delta)$. We reason as follows. If $\mathsf{WAP}^i \vdash \sigma^\neg(\Gamma \Rightarrow \Delta)$, then $\mathsf{SEQ}^i \vdash \tau^\neg \circ \sigma^\neg(\Gamma \Rightarrow \Delta)$ by Proposition 3 and hence $\mathsf{SEQ}^i \vdash (\Gamma \Rightarrow \Delta)$ by (the faithfulness part of) Corollary 1. ■

By composing the interpretations depicted in Figure 1, we obtain further interpretations:

Corollary 2. The following hold:

1. $\tau^\neg$ is an interpretation of WAP^i in SEQ^i;
2. $\sigma^\neg$ is an interpretation of EQ^i in WAP^i;
3. $\sigma^\neg$ is an interpretation of EQ^i in AP^i;
4. $\sigma^\neg$ is an interpretation of SEQ^i in AP^i.
5. $\tau^\neg \circ \sigma^\neg$ is an interpretation of SEQ^i in EQ^i;

Proof. That the translations in (a)-(e) are interpretation of the appropriate theories follows by the above theorems and by Fact 2. ■

Theorem 5. *The interpretations in 1-3 of Corollary 2 are not faithful. The interpretations in 4 and 5 are faithful.*

Proof. We discuss the interpretations separately:

1. Faithfulness fails for the same reasons seen in the proof of Proposition 3.

2./3. As seen in the proof of Proposition 3, the sequent $\sigma^\neg(\neg\neg x = y \Rightarrow x = y)$ is derivable in WAP^i (and hence it is derivable in AP^i as well). Stability of equality is clearly underivable in EQ^i (to see this replace apartness with equality in M_1^i, the countermodel to the stability of apartness in the proof of Proposition 3), we conclude that $\sigma^\neg$ does not preserve EQ^i in either WAP^i or AP^i.

4. The faithfulness is a consequence of the main result of [Negri, 1999]. Negri calls a formula $A \in \mathcal{L}^{i\neq}$ *negatomic* iff all occurrences of $x \neq y$ in A are negated, and a sequent is called negatomic just in case it contains only negatomic formulas. Negri shows that if $\mathsf{AP} \vdash (\Gamma \Rightarrow \Delta)$ then $\mathsf{WAP} \vdash (\Gamma \Rightarrow \Delta)$ if $(\Gamma \Rightarrow \Delta)$ is negatomic, that is, that $id_{\mathcal{L}^{i\neq}}$ is a faithful interpretation

of WAP into AP if one restricts the attention to negatomic sequents. Since $\sigma^{\neg}(\Gamma \Rightarrow \Delta)$ is negatomic for any $(\Gamma \Rightarrow \Delta) \in \mathcal{S}(\mathcal{L}^{i=})$, by Fact 3 we have that $\sigma = id_{\mathcal{L}^{i\neq}} \circ \sigma$ is a faithful interpretation of SEQ into AP.[3]

5. Suppose $\mathsf{EQ}^i \vdash \tau^{\neg} \circ \sigma^{\neg}(\Gamma \Rightarrow \Delta)$. Then, $\mathsf{SEQ}^i \vdash \tau^{\neg} \circ \sigma^{\neg}(\Gamma \Rightarrow \Delta)$ as well. By Corollary 1 we have that also $\mathsf{SEQ}^i \vdash (\Gamma \Rightarrow \Delta)$.[4]

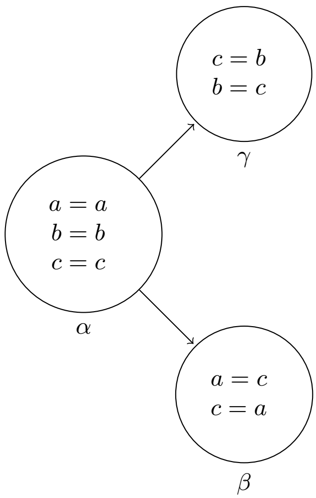
■

Thus, intuitionistic negation permits to interpret all theories we are considering in the theory of apartness AP^i. However, using $\sigma^{\neg}$ and $\tau^{\neg}$ the theory of apartenss AP^i cannot be interpreted into any other theory (except, of course, itself).

This follows from the following:

Theorem 6. *$\tau^{\neg}$ is not an interpretation of AP^i in SEQ^i.*

Proof. We need to show that there is a sequent $(\Gamma \Rightarrow \Delta) \in \mathcal{L}^{i\neq}$ such that $\mathsf{AP}^i \vdash (\Gamma \Rightarrow \Delta)$ and $\mathsf{SEQ}^{\mathsf{i}} \nvdash \tau^{\neg}(\Gamma \Rightarrow \Delta)$. Let $(\Gamma \Rightarrow \Delta)$ be an instance of co-transitivity $x \neq y \Rightarrow x \neq z, z \neq y$. Obviously, $\mathsf{AP}^i \vdash (\Gamma \Rightarrow \Delta)$. However, its $\tau^{\neg}$-translation, namely $(\neg x = y \Rightarrow \neg x = z, \neg z = y)$ is not derivable in SEQ^i. Let M_2^i be a model with three worlds α, β and γ such that $\alpha \leq \beta$ and $\alpha \leq \gamma$ and three objects a, b and c in D such that:



[3]We observe that Negri's result fails for the first-order version of the theory, with the negatomic sequent $\neg\forall z(\neg\neg z \neq x \vee \neg\neg z \neq y) \Rightarrow \neg x \neq y$ being a counterexample, see [van Dalen et al., 1979, p. 95].

[4]We thank one of the referees for suggesting the proof of this point.

Let $\varphi(x)$, $\varphi(y)$ and $\varphi(z)$ be a, b and c, respectively. Clearly, $\alpha \Vdash^{\varphi} \neg x = y$, for $\delta \nVdash^{\varphi} x = y$ for all $\delta \geq \alpha$. However, $\alpha \nVdash^{\varphi} \neg x = z$ and $\alpha \nVdash^{\varphi} \neg z = y$, for there exist two worlds $\epsilon, \epsilon' \geq \alpha$ such that $\epsilon \Vdash^{\varphi} x = z$ and $\epsilon' \Vdash^{\varphi} y = z$. Therefore, $M_2^i \nVdash (\neg x = y \Rightarrow \neg x = z, \neg z = y)$. Additionally, we need to show that M_2^i is in $\mathscr{SEQ}^i$. We leave to the reader to verify that $=$ is an equivalence relation, whereas to see that $M_2^i \Vdash (\neg\neg x = y \Rightarrow x = y)$, i.e., it satisfies also the stability principle, it is enough to show that $\alpha \Vdash^{\varphi_i} (\neg\neg x = y \Rightarrow x = y)$, for all $i = 1, \dots, 9$ such that:

1. $\varphi_1(x)$ and $\varphi_1(y)$ are a
2. $\varphi_2(x)$ is a and $\varphi_2(y)$ is b
3. $\varphi_3(x)$ is a and $\varphi_3(y)$ is c
4. $\varphi_4(x)$ is b and $\varphi_4(y)$ is a
5. $\varphi_5(x)$ and $\varphi_5(y)$ are b
6. $\varphi_6(x)$ is b and $\varphi_6(y)$ is c
7. $\varphi_7(x)$ is c and $\varphi_7(y)$ is a
8. $\varphi_8(x)$ is c and $\varphi_8(y)$ is b
9. $\varphi_9(x)$ and $\varphi_9(y)$ are c

With respect to assignments in (1), (5) and (9), the sequent holds at α since the formula in their succedent is true at all worlds $\geq \alpha$. All other cases the sequent has a formula in the antecedent which is false at both α and β, so the whole sequent holds at α. ■

We conclude this section establishing the following.

Theorem 7. *$\tau^{\neg}$ does not preserve AP^i in EQ^i.*

Proof. We need to show that there exists a sequent $(\Gamma \Rightarrow \Delta)$ in $\mathcal{L}^{i\neq}$ such that $\mathsf{EQ}^i \vdash \tau^{\neg}(\Gamma \Rightarrow \Delta)$ and $\mathsf{AP}^i \nvdash (\Gamma \Rightarrow \Delta)$. Let $(\Gamma \Rightarrow \Delta)$ be the stability of apartness, nameley $\neg\neg x \neq y \Rightarrow x \neq y$. We have that $\tau^{\neg}(\Gamma \Rightarrow \Delta) := \neg\neg\neg x = y \Rightarrow \neg x = y)$ which can be derived in EQ^i (see proof of Proposition 3 above) and that $(\Gamma \Rightarrow \Delta)$ is not derivable in AP^i (as shown by M_2^i, the countermodel to the stability of apartness in WAP^i given in the proof of Proposition 6, that is actually a model of AP^i). ■

Since EQ^i is stricly included in SEQ^i, we have that:

Corollary 3. The following holds:

1. $\tau^{\neg}$ is not an interpretation of AP^i in EQ^i.
2. $\tau^{\neg}$ does not preserve AP^i in SEQ^i.

5. Relating = and ≠ with co-negation

By extending $\sigma^\neg$ and $\tau^\neg$ to $\mathcal{L}^=$ and $\mathcal{L}^{\neq}$ the results established in the previous section also applies to the bi-intuitionistic versions of the four theories of equality and apartness considered so far. Moreover, using $\ulcorner$ in place of $\neg$ one can define the two further translations $\sigma^\ulcorner : \mathcal{L}^= \to \mathcal{L}^{\neq}$ as follows:

$$\sigma^\ulcorner(x = y) := \ulcorner x \neq y \qquad\qquad \tau^\ulcorner(x \neq y) := \ulcorner x = y$$

Informally, one can read the translation $\sigma^\ulcorner$ as an attempt to define equality in terms of aparntess and co-negation by taking a refutation of $x = y$ to be a method to transform a refutation of $x \neq y$ into a refutation of $\top$; and the translation $\tau^\ulcorner$ can be read as an attempt to define apartness in terms of equality and co-negation by taking a refutation of $x \neq y$ as a method to transform a refutation of $x = y$ into a refutation of $\top$.

One may expect that these two translations can help in getting a more symmetric picture of the relationship between theories of equality and theory of apartness. This is to some extent the case. As in the previous section, to improve readability we summarize the results to be established in this section in Figure 2.

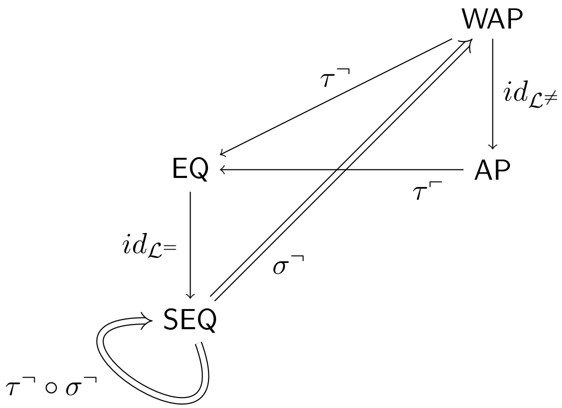


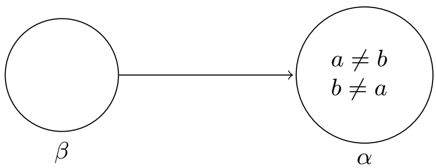
Fig. 2. Relating EQ, SEQ, AP and WAP with $\neg$ and $\ulcorner$

Theorem 8. *$\tau^\ulcorner$ is a non-faithful interpretation of* AP *in* EQ.

Proof. To show that $\tau^\ulcorner$ is an interpretation it is enough to prove: EQ $\vdash \tau^\ulcorner(x \neq x \Rightarrow)$, EQ $\vdash \tau^\ulcorner(x \neq y \Rightarrow x \neq z, z \neq y)$ and EQ $\vdash \tau^\ulcorner(x \neq y \Rightarrow y \neq x)$:

$$\dfrac{\Rightarrow x = x}{\ulcorner x = x \Rightarrow}\,L\ulcorner \qquad \dfrac{\dfrac{\dfrac{x = z, z = y \Rightarrow x = y}{x = z \Rightarrow x = y, \ulcorner z = y}\,R\ulcorner}{\Rightarrow x = y, \ulcorner x = z, \ulcorner z = y}\,R\ulcorner}{\ulcorner x = y \Rightarrow \ulcorner x = z, \ulcorner z = y}\,L\ulcorner \qquad \dfrac{\dfrac{y = x \Rightarrow x = y}{\Rightarrow \ulcorner y = x, x = y}\,R\ulcorner}{\ulcorner x = y \Rightarrow \ulcorner y = x}\,L\ulcorner$$

To show that $\tau^{\ulcorner}$ is not faithful it is enough to find a sequent $(\Gamma \Rightarrow \Delta) \in \mathcal{S}(\mathcal{L}^{\neq})$ such that $\mathsf{EQ} \vdash \tau^{\ulcorner}(\Gamma \Rightarrow \Delta)$ and $\mathsf{AP} \nvdash (\Gamma \Rightarrow \Delta)$. Take $(\Gamma \Rightarrow \Delta)$ to be an instance of co-stability $x \neq y \Rightarrow \ulcorner\ulcorner x \neq y$. Now, $\tau^{\ulcorner}(x \neq y \Rightarrow \ulcorner\ulcorner x \neq y) := (\ulcorner x = y \Rightarrow \ulcorner\ulcorner\ulcorner x = y)$. But $\mathsf{EQ} \vdash (\ulcorner x = y \Rightarrow \ulcorner\ulcorner\ulcorner x = y)$ (to see this take the dual of the derivation of $\neg\neg\neg x = y \Rightarrow \neg x = y$ in AP, see the proof of Proposition 3). However, apartness is not co-stable in AP. To see this, consider a Kripke model M_3 with two worlds α and β such that $\beta \leq \alpha$ and two objects a and b in D such that:



Let $\varphi(x)$ and $\varphi(y)$ be a and b, respectively. Clearly $\alpha \nVdash^{\varphi} \ulcorner\ulcorner x \neq y$, since for all worlds $\delta \leq \alpha$, namely β and α itself, there is a world $\epsilon \leq \delta$ such that $\epsilon \nVdash^{\varphi} x \neq y$. However, $\alpha \nVdash^{\varphi} x \neq y$. We leave to the reader to check that $M_3 \in \mathscr{A}\mathscr{P}$. ■

Thus using both intuitionistic negation and dual-intuitionistic co-negation we can interpret not only WAP (as we already could using $\tau^{\neg}$) but also AP into EQ, and actually, by composing the different embeddings depicted in Figure 2, we have that *all* four theories can be interpreted into one another:

Corollary 4. The following hold:

1. $\sigma^{\neg} \circ \tau^{\ulcorner}$ is an interpretation of AP in WAP;

2. $\tau^{\ulcorner}$ is an interpretation of AP in SEQ.

3. $\tau^{\ulcorner}$ is an interpretation of WAP in EQ.

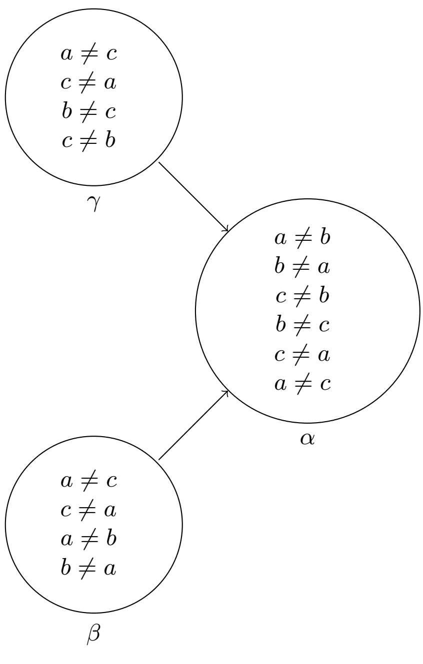
4. $\tau^{\ulcorner}$ is an interpretation of WAP in SEQ.

The picture is however far from symmetric. Although we can embed the theories of apartness into those of equality, none of the former ones can be faithfully interpreted into any of the latter ones (in contrast to the fact that

SEQ can be faithfully interpreted into WAP and AP). Moreover, whereas both $\tau^{\neg}$ and $\sigma^{\neg}$ interpret at least some theory based on one language into one based on the other language, we have that $\sigma^{\ulcorner}$ does not interpret EQ into AP, and hence a fortiori neither EQ or SEQ into either WAP or AP:

Theorem 9. *$\sigma^{\ulcorner}$ is not an interpretation of* EQ *in* AP*.*

Proof. We need to find a seqent $(\Gamma \Rightarrow \Delta) \in \mathcal{S}(\mathcal{L}^{=})$ such that $\mathsf{SEQ} \vdash (\Gamma \Rightarrow \Delta)$ and $\mathsf{AP} \nvdash \sigma^{\ulcorner}(\Gamma \Rightarrow \Delta)$. Let $(\Gamma \Rightarrow \Delta)$ be $x = z, z = y \Rightarrow x = y$. Clearly, $\mathsf{EQ} \vdash (\Gamma \Rightarrow \Delta)$. We need to show that $\mathsf{AP} \nvdash \sigma^{\ulcorner}(x = z, z = y \Rightarrow x = y)$. Since $\sigma^{\ulcorner}(x = z, z = y \Rightarrow x = y) := (\ulcorner x \neq z, \ulcorner z \neq y \Rightarrow \ulcorner x \neq y)$, we only need to show that there is a Kripke model satisfying the axioms of apartness in which $\ulcorner x \neq z, \ulcorner z \neq y \Rightarrow \ulcorner x \neq y$ fails. Let M_4 be a Kripke model with three worlds α, β and γ such that $\beta \leq \alpha$ and $\gamma \leq \alpha$ and three objects a, b and c in D_γ and D such that:



Again let $\varphi(x)$, $\varphi(y)$ and $\varphi(z)$ be a, b and c, respectively. Clearly, $\alpha \Vdash^{\varphi} \ulcorner x \neq y$, for there exists a world $\delta \leq \alpha$, namely γ, such that $\delta \nVdash^{\varphi} x \neq y$. Similarly, $\alpha \Vdash^{\varphi} \ulcorner x \neq y$. However, $\alpha \nVdash^{\varphi} \ulcorner x \neq z$, since for all worlds $\delta \leq \alpha$, namely γ, β and α itself, $\delta \Vdash^{\varphi} x \neq z$. We leave to the reader to verify that $M_4 \in \mathscr{A}\mathscr{P}$. ■

6. Enriching the picture

As intuitionistic negation does not allow to interpret AP, but only its weakening WAP into some theory of equality, one can expect that co-negation can be used to interpret only a *weakening* of the theory of equality into some theory of apartness.

Similarly, the non-faithfulness of the interpretation $\tau^{\urcorner}$ of AP into EQ witnessed by the co-stability of apartness clearly mirrors the non-faithfulness of the interpretation $\sigma^{\neg}$ of EQ into WAP and AP resulting by the fact that the translation of the stability of equality holds in both theories. This suggests that in order for $\tau^{\urcorner}$ to faithfully interpret some theory of apartness into a theory of equality, the former must at least validate co-stability.

Given these considerations, we take into account two further theories based on the languages $\mathcal{L}^{=}$ and $\mathcal{L}^{\neq}$. One, to be called SAP is the strengthening of AP obtained by adding to it initial sequents expressing the co-stability of apartness

$$x \neq y \Rightarrow \ulcorner\ulcorner x \neq y$$

The other one, to be called WEQ is the weakening of EQ obtained by replacing the initial sequents expressing transitivity and symmetry with initial sequents expressing what we may call "co-negative co-transitivity" and "co-negative symmetry":

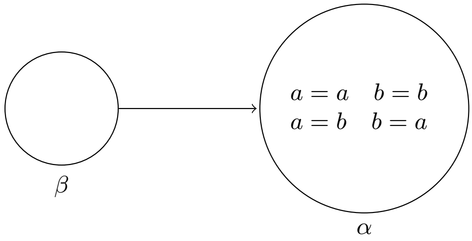
$$\ulcorner x = y \Rightarrow \ulcorner x = z, \ulcorner z = y$$

$$\ulcorner x = y \Rightarrow \ulcorner y = x$$

Theorem 10. $\sigma^{\ulcorner}$ *is a non-faithful interpretation of* WEQ *into* AP.

Proof. We have to show that AP $\vdash \sigma^{\ulcorner}(\Rightarrow x = x)$, AP $\vdash \sigma^{\ulcorner}(\ulcorner x = y \Rightarrow \ulcorner x = z, \ulcorner z = y)$ and AP $\vdash \sigma^{\ulcorner}(\ulcorner x = y \Rightarrow \ulcorner y = x)$. This sequents can be derived using the dual of the derivations used in the proof of Proposition 3.

To establish the non-failtfulness we reason as in Propostion 8. We need to find a sequent $(\Gamma \Rightarrow \Delta) \in \mathcal{S}(\mathcal{L}^{=})$ such that AP $\vdash \sigma^{\ulcorner}(\Gamma \Rightarrow \Delta)$ and WEQ $\nvdash (\Gamma \Rightarrow \Delta)$. Take $(\Gamma \Rightarrow \Delta)$ to be $x = y \Rightarrow \ulcorner\ulcorner x = y$. Now, $\sigma^{\ulcorner}(x = y \Rightarrow \ulcorner\ulcorner x = y) := (\ulcorner x \neq y \Rightarrow \ulcorner\ulcorner\ulcorner x \neq y)$ and AP $\vdash (\ulcorner x \neq y \Rightarrow \ulcorner\ulcorner\ulcorner x \neq y)$ (to see this take replace $=$ with $\neq$ in the dual of the derivation of $\neg\neg\neg x = y \Rightarrow \neg x = y$ in AP, see the proof of Proposition 3). However, equality is not "co-stable" in WEQ. To see this, consider a Kripke model M_4 with two worlds α and β such that $\beta \leq \alpha$ and two objects a and b in D such that:



Let $\varphi(x)$ and $\varphi(y)$ be a and b, respectively. Clearly $\alpha \nVdash^{\varphi} \ulcorner\ulcorner x = y$, since for all worlds $\delta \leq \alpha$, namely β and α itself, there is a world $\epsilon \leq \delta$ such that $\epsilon \nVdash^{\varphi} x = y$. However, $\alpha \nVdash^{\varphi} x = y$. We leave to the reader to check that $M_4 \in \mathscr{W}\mathscr{E}\mathscr{Q}$ (in fact $M_4 \in \mathscr{E}\mathscr{Q}$). ■

To establish the next proposition we will need the following:

Lemma 2. *For all* $A \in \mathcal{L}^{\neq}$*:*

1. $\mathsf{SAP} \vdash (A \Rightarrow \sigma^{\ulcorner} \circ \tau^{\ulcorner}(A))$
2. $\mathsf{SAP} \vdash (\sigma^{\ulcorner} \circ \tau^{\ulcorner}(A) \Rightarrow A)$

Proof. As in the proof of Lemma 1, we establish the two claims by simultaneous induction on A:

- A is $x \neq y$. Clearly, $\mathsf{SAP} \vdash (\ulcorner\ulcorner x = y \Rightarrow x = y)$:

$$\dfrac{\dfrac{x \neq y \Rightarrow x \neq y}{\Rightarrow x \neq y, \ulcorner x \neq y}\, R\ulcorner}{\ulcorner\ulcorner x \neq y \Rightarrow x \neq y}\, L\ulcorner$$

  and obviously $\mathsf{SAP} \vdash (x = y \Rightarrow \ulcorner\ulcorner x = y)$ given the co-stability initial sequents.

- A is P,$\top$ or $\bot$. Obvious.
- A is $B \supset C$. The case is proved as the corresponding case of Lemma 1, it suffice to replace $\tau^{\neg} \circ \sigma^{\neg}$ with $\sigma^{\ulcorner} \circ \tau^{\ulcorner}$.
- A is $B \wedge C$ or $B \vee C$. Similar to the previous case.

■

Corollary 5. $\sigma^{\ulcorner} \circ \tau^{\ulcorner}$ is a faithful interpretation of SAP into itself.

Proof. The proof follows the same pattern of that of Corollary 1. ■

We can now show that:

Theorem 11. *$\tau^{\ulcorner}$ is a faithful interpretation of* SAP *into* WEQ.

Proof. To show that $\tau^{\ulcorner}$ is an interpretation, we have to show that $\mathsf{WEQ} \vdash \tau^{\ulcorner}(x \neq x \Rightarrow)$, $\mathsf{WEQ} \vdash \tau^{\ulcorner}(x \neq y \Rightarrow x \neq z, z \neq y)$, $\mathsf{WEQ} \vdash \tau^{\ulcorner}(x \neq y \Rightarrow y \neq x)$ and $\mathsf{WEQ} \vdash \tau^{\ulcorner}(x \neq y \Rightarrow \ulcorner\ulcorner x \neq y)$.

The sequent $\tau^{\ulcorner}(x \neq x \Rightarrow)$ can be derived as in EQ (see proof of Proposition 8), while $\tau^{\ulcorner}(x \neq y \Rightarrow x \neq z, z \neq y)$ and $\tau^{\ulcorner}(x \neq y \Rightarrow y \neq x)$ are obviously derivable in WEQ, being initial sequents. Finally, one can see that $\mathsf{WEQ} \vdash \ulcorner x \neq y \Rightarrow \ulcorner\ulcorner\ulcorner x = y$ by taking the dual of the derivation of $\neg\neg\neg x \neq y \Rightarrow \neg x \neq y$ in WAP in the proof of Proposition 4.

To show faithfulness, we need to show that, for all $(\Gamma \Rightarrow \Delta) \in \mathcal{S}(\mathcal{L}^{\neq})$, if $\mathsf{WEQ} \vdash \tau^{\ulcorner}(\Gamma \Rightarrow \Delta)$, then $\mathsf{SAP} \vdash (\Gamma \Rightarrow \Delta)$. We reason as in the proof of Proposition 4. If $\mathsf{WEQ} \vdash \tau^{\ulcorner}(\Gamma \Rightarrow \Delta)$, then by Proposition 10, $\mathsf{SAP} \vdash \sigma^{\ulcorner}(\tau^{\ulcorner}(\Gamma \Rightarrow \Delta))$ and by Corollary 5 $\mathsf{SAP} \vdash (\Gamma \Rightarrow \Delta)$. ■

We summarize our results in Figure 3.

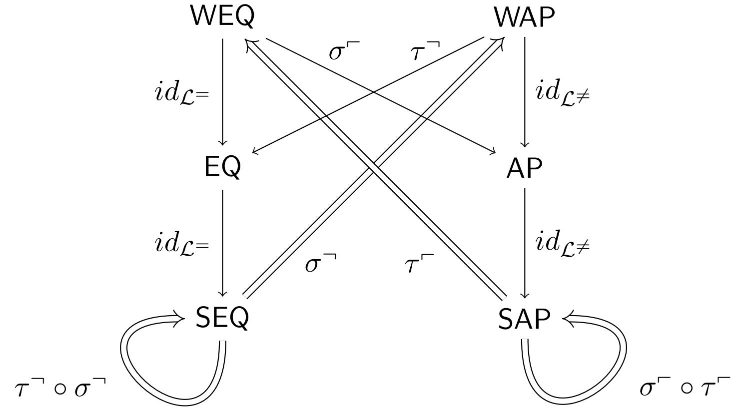


Fig. 3. Relating WEQ, EQ, SEQ, WAP, AP and SAP with $\neg$ and $\ulcorner$

7. Concluding remarks

By considering theories of equality and apartness on the background of bi-intuitionistic logic we attained a fully symmetric picture of the relationship between the two notions. In particular we could show that not only the theories of equality can be embedded into those of apartness, but the other way around as well. However, we still lack a faithful interpretation of EQ into any theory of apartness, and of AP in any theory of equality. We leave this to future work.

The investigation undertaken in the present work can be further pursued into different directions.

First, notice that *a priori* there are eight possibile translations, i.e., $\sigma^i \circ \tau^j$ for $i, j \in \{\neg, \neg\}$. However, only two have been explicitly considered here. Our choice was motivated primarily by the fact that these two translation are enough to provide a fully symmetric picture of the relationships among the various theories presented, but of course it would be certainly interesting to consider translations combining intuitionistic and co-intuitionistic negation. We expect that in this case, bi-intuitionistic "mixed" double-negation laws such as $\neg\neg A \Rightarrow A$ and $A \Rightarrow \neg\neg A$ will play a prominent role.[5]

Secondly, the partial faithfulness results of [Negri, 1999] for the negatomic fragment suggest the possibility of establishing similar results using co-negation.

Thirdly, it seems natural to consider further bi-intuitionistic theories beyond those considered in the present paper. On the one hand, one may consider the theory of paraconsistent apartness as based on dual-intuitionistic logic, i.e., in the implication-free fragment of $\mathcal{L}^{\neq}$, as is done in [Brunner, 2004]. Another possibility is to apply the idea underlying the present paper to the investigation of the theory of positive lattices [von Plato, 2001] by extending our theories with operators of join and meet.

Finally, one could investigate the notions of equality and apartness (as well as other notions) on the basis of other constructive systems based on a symmetry between positive and negative notions, such as Nelson's logic of constructible falsity [Nelson, 1949], or Wansing's 2-intuitionistic logic [Wansing, 2016].

References

Brunner, 2004 – Brunner, A. "First-order anti-intuitionistic logic with apartness", *Logic and Logical Philosophy*, 2004, Vol. 13, pp. 77–88.

Crolard, 2001 – Crolard, T. "Subtractive logic", *Theoretical Computer Science*, 2001, Vol. 254, pp. 151–185.

van Dalen et al., 1979 – Van Dalen, D. and Statman, R. "Equality in presence of apartness", in: *Essays in Mathematical and Philosophical Logic*, ed. by J. Hintikka, I. Niniluoto and E. Saarinen, Dordrecht: Reidel, 1979, pp. 95–116.

Goré et al., 2020 – Goré, R. and Shillito, I. "Bi-Intuitionistic logics: A new Instance of an old problem", in: *AiML 2020*, ed. by N. Olivetti, R. Verbrugge, S. Negri, G. Sandu, London: College Publications, 2020, pp. 269–288.

Kowalski et al. 2017 – Kowalski, T. and Ono, H. "Analytic cut and interpolation for bi-intuitionistic logic", *The Review of Symbolic Logic*, 2017, Vol. 10, pp. 259–283.

Negri, 1999 – Negri, S. "Sequent calculus proof theory of intuitionistic apartness and order relations", *Archive for Mathematical Logic*, 1999, Vol. 38, pp. 521–547.

[5] Thanks to a referee for bringing this to our attention.

Negri et al., 2001 – Negri, S. and von Plato, J. *Structural Proof Theory*, Cambridge: Cambridge University Press, 2001.

Nelson, 1949 – Nelson, D. "Constructible Falsity", *The Journal of Symbolic Logic*, 1949, Vol. 14, pp. 16–26.

von Plato, 2001 – von Plato, J. "Positive Lattices", in: *Reuniting the Antipodes: Constructive and Nonstandard Views of the Continuum*, ed. by P. Schuster, U. Berger and H. Osswald, Dordrecht: Springer, 2001, pp. 185–197.

Rauszer, 1980 – Rauszer, C. *An algebraic and Kripke-style approach to a certain extension of intuitionistic logic*, Warszawa: Instytut Matematyczny Polskiej Akademi Nauk, 1980.

Rauszer, 1974a – Rauszer, C. "A formalization of the propositional calculus of H-B logic", *Studia Logica*, 1974, Vol. 33, pp. 23–34.

Rauszer, 1974b – Rauszer, C. "Semi-boolean algebras and their applications to intuitionistic logic with dual operations", *Fundamaenta Mathematicae*, 1974, Vol. 83, pp. 219–249.

Tranchini, 2012 – Tranchini, L. "Natural Deduction for Dual-Intuitionistic Logic", *Studia Logica*, 2012, Vol. 100, pp. 631–648.

Urbas, 1996 – Urbas, I. "Dual-Intuitionistic Logic", *Notre Dame Journal of Formal Logic*, 1996, Vol. 37, pp. 440–451.

Wansing, 2010 – Wansing, H. "Proofs, Disproofs, and Their Duals", in: *AiML 2010*, ed. by M. Kracht, M. de Rijke, H. Wansing and M. Zakharyaschev, Stanford: CSLI Publications, 2010, pp. 483–505.

Wansing, 2016 – Wansing, H. "Falsification, natural deduction and bi-intuitionistic logic", *Journal of Logic and Computation*, 2016, Vol. 26, pp. 425–450.

Логические исследования
2021. Т. 27. № 1. С. 107–123
УДК 16+510.6

Logical Investigations
2021, Vol. 27, No. 1, pp. 107–123
DOI: 10.21146/2074-1472-2021-27-1-107-123

Thiago Nascimento, Umberto Rivieccio

Negation and Implication in Quasi-Nelson Logic

Thiago Nascimento
Universidade Federal do Rio Grande do Norte,
59078-970 Natal, Brazil.
E-mail: thiagnascsilva@gmail.com

Umberto Rivieccio
Universidade Federal do Rio Grande do Norte,
59078-970 Natal, Brazil.
E-mail: umberto.rivieccio@ufrn.br

Abstract: Quasi-Nelson logic is a recently-introduced generalization of Nelson's constructive logic with strong negation to a non-involutive setting. In the present paper we axiomatize the negation-implication fragment of quasi-Nelson logic (QNI-logic), which constitutes in a sense the algebraizable core of quasi-Nelson logic. We introduce a finite Hilbert-style calculus for QNI-logic, showing completeness and algebraizability with respect to the variety of QNI-algebras. Members of the latter class, also introduced and investigated in a recent paper, are precisely the negation-implication subreducts of quasi-Nelson algebras. Relying on our completeness result, we also show how the negation-implication fragments of intuitionistic logic and Nelson's constructive logic may both be obtained as schematic extensions of QNI-logic.

Keywords: Nelson's constructive logic with strong negation, quasi-Nelson algebras, implication-negation subreducts, QNI-algebra, quasi-Nelson logic, algebraizable logics

For citation: Nascimento T., Rivieccio U. " Negation and Implication in Quasi-Nelson Logic", *Logicheskie Issledovaniya / Logical Investigations*, 2021, Vol. 27, No. 1, pp. 107–123. DOI: 10.21146/2074-1472-2021-27-1-107-123

1. Introduction

Quasi-Nelson algebras are the subvariety of commutative integral bounded residuated lattices (CIBRLs, see [Galatos et al., 2007]) obtained by adding the *Nelson identity*:

$$(x \Rightarrow (x \Rightarrow y)) \wedge ({\sim} y \Rightarrow ({\sim} y \Rightarrow {\sim} x)) \approx x \Rightarrow y.$$

In an involutive context (i.e. if the double negation law ${\sim}{\sim} x \approx x$ is also satisfied), the Nelson identity characterizes the class of Nelson algebras, the equivalent algebraic semantics of Nelson's constructive logic with strong negation [Nelson, 1949]. Quasi-Nelson algebras, however, need not be involutive; indeed, it is easy to verify that every Heyting algebra (viewed as a CIBRL) satisfies the Nelson identity. Quasi-Nelson algebras are thus a common generalization of Heyting algebras and Nelson algebras.

The logic of quasi-Nelson algebras corresponds to the axiomatic extension of the *Full Lambek Calculus with exchange (e) and weakening (w)*, FL_{ew} ([Galatos et al., 2007]) obtained by adding the *Nelson Axiom*:

$$(\varphi \Rightarrow (\varphi \Rightarrow \psi)) \land ({\sim}\psi \Rightarrow ({\sim}\psi \Rightarrow {\sim}\varphi)) \Rightarrow (\varphi \Rightarrow \psi).$$

As such, quasi-Nelson logic is algebraizable, and has the class of quasi-Nelson algebras as its equivalent algebraic semantics (see [Liang, Nascimento, 2019]; for further information and motivation on quasi-Nelson algebras, see also [Rivieccio, Spinks, 2018; Rivieccio, Jansana, 2020; Rivieccio, Spinks, 2021]).

The language of (quasi-)Nelson logic includes two implication connectives, the *strong implication* ($\Rightarrow$) that satisfies the residuation property and the *weak implication* ($\to$) that enjoys the standard version of the Deduction-Detachment Theorem. Both quasi-Nelson and Nelson's logic can be axiomatized by taking any of the two implications as primitive, defining the other through the following terms: $p \to q := p \Rightarrow (p \Rightarrow q)$ and $p \Rightarrow q := (p \to q) \land ({\sim}q \to {\sim}p)$. From each of the above implications (and the falsity constant 0) a negation can be defined in the standard way. In the case of Nelson's logic, the definition ${\sim}p := p \Rightarrow 0$ yields the strong involutive negation, whereas $\neg p := p \to 0$ is sometimes referred to as the "intuitionistic negation". However, in the case of quasi-Nelson logics the above terminology is less meaningful because neither of the two negations is involutive.

Like Nelson algebras, also quasi-Nelson algebras can be represented as so-called *twist-structures*, though the twist construction needs to be generalized to account for the non-involutivity of the negation (see [Rivieccio, Spinks, 2018; Rivieccio, Spinks, 2021]). Further generalizations allow us to give twist representations for some subreducts of quasi-Nelson algebras: see the recent papers [Rivieccio, Jansana, 2020; Rivieccio, 2020a; Rivieccio, 2020b; Rivieccio, 2020c]. In the present paper we are in particular interested in the representation of QNI-algebras given in [Rivieccio, Jansana, 2020, Theorem 5], corresponding to the $\{\to, \sim\}$-subreducts of quasi-Nelson algebras. This twist construction proved to be quite useful for establishing results regarding congruences, subdirectly irreducible algebras and subvarieties of QNI-algebras; see [Rivieccio, 2020a] for more details.

As mentioned earlier, quasi-Nelson logic is algebraizable in the sense of [Blok, Pigozzi, 1989]. As a set of equivalence formulas one can take $\Delta(\varphi, \psi) = \{\varphi \to \psi, \psi \to \varphi, \sim\varphi \to \sim\psi, \sim\psi \to \sim\varphi\}$, and as defining equation $E(\varphi) = \{\varphi \approx \varphi \to \varphi\}$. These observations entail that the $\{\to, \sim\}$-fragment of quasi-Nelson logic must also be algebraizable, with the same translations, with respect to the corresponding subreducts of quasi-Nelson algebras (see e.g. [Font, 2016, Proposition 3.29]). In the present paper, we introduce a finite Hilbert-style calculus that characterizes the $\{\to, \sim\}$-fragment of quasi-Nelson logic. Indeed, we show that our calculus is algebraizable with respect to the variety of QNI-algebras introduced in [Rivieccio, Jansana, 2020].

We shall proceed in the following way. In Section 2. we introduce QNI-algebras and a lemma that will be useful in order to prove the equivalence between the variety of QNI-algebras and $\mathtt{Alg}^*(\mathbf{L}_{\mathrm{QNI}})$, the equivalent quasivariety semantics corresponding to our calculus. In Section 3. we introduce the QNI-calculus and state the Deduction Theorem. In Section 4. we prove that QNI-logic is algebraizable and give a quasi-equational presentation for $\mathtt{Alg}^*(\mathbf{L}_{\mathrm{QNI}})$. In Section 5. we prove that the class of QNI-algebras and $\mathtt{Alg}^*(\mathbf{L}_{\mathrm{QNI}})$ coincide. We obtain as a corollary that the class $\mathtt{Alg}^*(\mathbf{L}_{\mathrm{QNI}})$ has equationally definable principal congruences (EDPC). In Section 6. we give the equation that defines principal congruences on $\mathtt{Alg}^*(\mathbf{L}_{\mathrm{QNI}})$, and we consider a few axiomatic extensions of QNI-logic, including (the negation-implication fragments of) intutionistic logic and Nelson's constructive logic with strong negation.

2. Preliminaries

In this section we recall the definition and a few properties of the class of $\{\to, \sim\}$-subreducts of quasi-Nelson algebras introduced in [Rivieccio, 2020a].

Given an algebra $\mathbf{A} = \langle A; \to, \sim, 0, 1\rangle$ and elements $a, b \in A$, we will write $a \equiv b$ as a shorthand for $a \to b = b \to a = 1$. We will also employ the following abbreviations: $a \odot b := \sim(a \to \sim b)$ and

$$\beta(a, b, c) := (a \to b) \to ((b \to a) \to ((\sim a \to \sim b) \to ((\sim b \to \sim a) \to c))).$$

Definition 1 ([Rivieccio, 2020a], Definition 3.1)**.** An algebra $\mathbf{A} = \langle A; \to, \sim, 0, 1\rangle$ of type $\langle 2, 1, 0, 0\rangle$ is a *quasi-Nelson implication algebra* (QNI-algebra) if the following properties are satisfied, for all $a, b, c, d \in A$:

(1) $1 \to a = a$

(2) $a \to (b \to a) = a \to a = 0 \to a = 1$

(3) $a \to (b \to c) = b \to (a \to c) = (a \to b) \to (a \to c)$

(4) $\sim a \to (\sim b \to c) = (\sim a \odot \sim b) \to c$

(5) $\beta(a, b, a) = \beta(a, b, b)$

(6) if $\sim a \to \sim b = 1$, then $\sim a \to (\sim a \odot \sim b) = 1$

(7) $a \odot (b \odot c) \equiv (a \odot b) \odot c$

(8) $a \odot b \equiv b \odot a$

(9) if $a \equiv b$ and $c \equiv d$, then $a \to c \equiv b \to d$ and $a \odot c \equiv b \odot d$

(10) $\sim a = \sim\sim\sim a$

(11) $\sim 1 = 0$ and $\sim 0 = 1$

(12) $(a \to b) \to (\sim\sim a \to \sim\sim b) = 1$.

(13) $a \to \sim\sim a = 1$

(14) if $a \to b = 1$, then $a \odot c \to b \odot c = 1$ and $c \odot a \to c \odot b = 1$

(15) $a \odot (a \to b) \equiv a \odot b$

(16) $a \odot b \equiv \sim\sim a \odot \sim\sim b$

(17) $\sim(a \to b) \equiv \sim(\sim\sim a \to \sim\sim b)$.

We shall denote by QNI the class of QNI-algebras. By definition, QNI is a quasivariety; it was shown in [Rivieccio, 2020a, Corollary 3.15] that QNI is in fact a variety.

Lemma 1 ([Rivieccio, 2020a], Lemma 3.3)**.** *Let* $\mathbf{A} \in \mathsf{QNI}$ *and* $a, b, c \in A$. *Then:*

(1) $(\sim a \odot \sim b) \to \sim a = (\sim a \odot \sim b) \to \sim b = 1$.

(2) If $a \to b = 1$ *and* $b \to c = 1$*, then* $a \to c = 1$.

(3) The relation $\leq$ *defined by* $a \leq b$ *iff (*$a \to b = 1$ *and* $\sim b \to \sim a = 1$*) is a partial order on* A*, with minimum* 0 *and maximum* 1.

3. A Hilbert calculus for QNI logic

In this section we introduce a Hilbert-style calculus that determines a logic (in the sense of [Hamilton, 1978]) henceforth denoted by $\mathbf{L}_{\mathrm{QNI}}$. Our aim is to show that $\mathbf{L}_{\mathrm{QNI}}$ is regularly algebraizable, and that its equivalent algebraic semantics is precisely the variety QNI.

Fix a denumerable set Atprop of propositional variables. We use letters p, q, r etc. to refer to generic elements of Atprop. The propositional language $\mathcal{L}$ of $\mathbf{L}_{\mathrm{QNI}}$ over Atprop is defined recursively as follows:

$$\varphi ::= p \mid {\sim}\varphi \mid \varphi \to \varphi.$$

Consistently with the above-introduced notation for QNI-algebras, we abbreviate $\varphi \odot \psi := {\sim}(\varphi \to {\sim}\psi)$. The Hilbert-calculus for $\mathbf{L}_{\mathrm{QNI}}$ consists of the following axiom schemes:

AX1 $\varphi \to (\psi \to \varphi)$

AX2 $(\varphi \to (\psi \to \gamma)) \to ((\varphi \to \psi) \to (\varphi \to \gamma))$

AX3 ${\sim}{\sim}{\sim}\varphi \to {\sim}\varphi$

AX4 $(\varphi \to \psi) \to ({\sim}{\sim}\varphi \to {\sim}{\sim}\psi)$

AX5 $\varphi \to {\sim}{\sim}\varphi$

AX6 $(\varphi \odot (\varphi \to \psi)) \to (\varphi \odot \psi)$

AX7 ${\sim}{\sim}\varphi \to ({\sim}\psi \to {\sim}(\varphi \to \psi))$

AX8 ${\sim}(\varphi \to \psi) \to {\sim}\psi$

AX9 ${\sim}(\varphi \to \psi) \to {\sim}{\sim}\varphi$

AX10 ${\sim}(\varphi \to \varphi) \to \psi$.

The only rule is *modus ponens* (MP): from φ and $\varphi \to \psi$, derive ψ.

The proof of the following result is the standard one by induction on the length of derivations.

Theorem 1 (Deduction-Detachment Theorem). *If* $\Phi \cup \{\varphi\} \vdash_{\mathbf{L}_{\mathrm{QNI}}} \psi$, *then* $\Phi \vdash_{\mathbf{L}_{\mathrm{QNI}}} \varphi \to \psi$.

Lemma 2.

(1) $\emptyset \vdash_{\mathbf{L}_{\mathrm{QNI}}} \varphi \to \varphi$

(2) $\{\varphi \to \psi, \psi \to \chi\} \vdash_{\mathbf{L}_{QNI}} \varphi \to \chi$.

4. $\mathbf{L}_{\mathbf{QNI}}$ is (regularly) algebraizable

In this section we prove that $\mathbf{L}_{\mathrm{QNI}}$ is regularly algebraizable. Using this result, we axiomatize the equivalent algebraic semantics of $\mathbf{L}_{\mathrm{QNI}}$ via the algorithm of [Blok, Pigozzi, 1989, Theorem 2.17]. We then prove that the class of algebras thus obtained is equivalent to the class QNI given in Definition 1. Given the formula algebra $\mathbf{Fm}$, the associated set of equations, $Fm \times Fm$, will henceforth be denoted by Eq. We abbreviate an equation $\langle \varphi, \psi \rangle$ as $\varphi \approx \psi$.

Theorem 2. *A logic* $\mathbf{L}$ *is algebraizable if and only if there are a set of equations* $E(\varphi) \subseteq Eq$ *and a set of formulas* $\Delta(\varphi, \psi) \subseteq Fm$*, such that:*

(**Alg**) $\varphi \dashv\vdash_{\mathbf{L}} \Delta(E(\varphi))$

(**Ref**) $\emptyset \vdash_{\mathbf{L}} \Delta(\varphi, \varphi)$

(**MP**) $\varphi, \Delta(\varphi, \psi) \vdash_{\mathbf{L}} \psi$

(**Cong**) *for each n-ary operator* $\bullet$*,*

$$\bigcup_{i=1}^{n} \Delta(\varphi_i, \psi_i) \vdash_{\mathbf{L}} \Delta(\bullet(\varphi_1, \ldots, \varphi_n), \bullet(\psi_1, \ldots, \psi_n)).$$

We call any such $E(\varphi)$ *a set of defining equations and any such* $\Delta(\varphi, \psi)$ *a set of equivalence formulas of* $\mathbf{L}$*.*

Definition 2. A logic $\mathbf{L}$ is regularly algebraizable when it is algebraizable and satisfies:

(G) $\varphi, \psi \vdash_{\mathbf{L}} \Delta(\varphi, \psi)$

for any non-empty set $\Delta(\varphi, \psi)$ of equivalence formulas.

Proposition 1. $\mathbf{L}_{\mathrm{QNI}}$ is regularly algebraizable with $\Delta(\varphi, \psi) := \{\varphi \to \psi, \psi \to \varphi, {\sim}\varphi \to {\sim}\psi, {\sim}\psi \to {\sim}\varphi\}$ and $E(\varphi) := \{\varphi \approx \varphi \to \varphi\}$.

Proof. As to (**Alg**), it suffices to prove that $\varphi \dashv\vdash_{\mathbf{L}_{\mathrm{QNI}}} \{\varphi \to (\varphi \to \varphi), (\varphi \to \varphi) \to \varphi, {\sim}\varphi \to {\sim}(\varphi \to \varphi), {\sim}(\varphi \to \varphi) \to {\sim}\varphi\}$. From right to left, thanks to Lemma 2.1, we have that $\varphi \to \varphi$ is a theorem and from $\varphi \to \varphi$ and $(\varphi \to \varphi) \to \varphi$ we get φ by using modus ponens. From left to right, we will prove that (i) $\varphi \vdash_{\mathbf{L}_{\mathrm{QNI}}} {\sim}\varphi \to {\sim}(\varphi \to \varphi)$ and (ii) $\varphi \vdash_{\mathbf{L}_{\mathrm{QNI}}} {\sim}(\varphi \to \varphi) \to {\sim}\varphi$, the other two proofs easily follow from **AX1**. For (i),

1.	φ	Assumption
2.	$\varphi \to ({\sim}\varphi \to {\sim}(\varphi \to \varphi))$	**AX7**
3.	${\sim}\varphi \to {\sim}(\varphi \to \varphi)$	1, 2, MP

For (ii), notice that it is just **AX10**.

In order to prove **Ref**, it is necessary to show that $\emptyset \vdash_{\mathbf{L}_{\mathrm{QNI}}} \{\varphi \to \varphi, {\sim}\varphi \to {\sim}\varphi\}$, and it is Lemma 2.1. (**MP**) is a straightforward consequence of modus ponens.

As to (**Cong**), we need to prove for each connective $\bullet \in \{\to, \sim\}$. For ($\sim$), we need to prove that: (i) $\{\varphi \to \psi, \psi \to \varphi, {\sim}\varphi \to {\sim}\psi, {\sim}\psi \to {\sim}\varphi\} \vdash_{\mathbf{L}_{\mathrm{QNI}}} {\sim}\varphi \to {\sim}\psi$ and $\{\varphi \to \psi, \psi \to \varphi, {\sim}\varphi \to {\sim}\psi, {\sim}\psi \to {\sim}\varphi\} \vdash_{\mathbf{L}_{\mathrm{QNI}}} {\sim}\psi \to {\sim}\varphi$. The two deductions follow by our hypothesis. And we also need to prove (ii) $\{\varphi \to \psi, \psi \to \varphi, {\sim}\varphi \to {\sim}\psi, {\sim}\psi \to {\sim}\varphi\} \vdash_{\mathbf{L}_{\mathrm{QNI}}} {\sim}{\sim}\varphi \to {\sim}{\sim}\psi$ and $\{\varphi \to \psi, \psi \to \varphi, {\sim}\varphi \to {\sim}\psi, {\sim}\psi \to {\sim}\varphi\} \vdash_{\mathbf{L}_{\mathrm{QNI}}} {\sim}{\sim}\psi \to {\sim}{\sim}\varphi$. The two deductions follow from **AX4** together with our hypothesis. For ($\to$), we need to prove that: (i) $\{\varphi_1 \to \psi_1, \psi_1 \to \varphi_1, {\sim}\varphi_1 \to {\sim}\psi_1, {\sim}\psi_1 \to {\sim}\varphi_1\} \cup \{\varphi_2 \to \psi_2, \psi_2 \to \varphi_2, {\sim}\varphi_2 \to {\sim}\psi_2, {\sim}\psi_2 \to {\sim}\varphi_2\} \vdash_{\mathbf{L}_{\mathrm{QNI}}} (\varphi_1 \to \varphi_2) \to (\psi_1 \to \psi_2)$ and $\{\varphi_1 \to \psi_1, \psi_1 \to \varphi_1, {\sim}\varphi_1 \to {\sim}\psi_1, {\sim}\psi_1 \to {\sim}\varphi_1\} \cup \{\varphi_2 \to \psi_2, \psi_2 \to \varphi_2, {\sim}\varphi_2 \to {\sim}\psi_2, {\sim}\psi_2 \to {\sim}\varphi_2\} \vdash_{\mathbf{L}_{\mathrm{QNI}}} (\psi_1 \to \psi_2) \to (\varphi_1 \to \varphi_2)$, they can be shown by Lemma 2.2; (ii) $\{\varphi_1 \to \psi_1, \psi_1 \to \varphi_1, {\sim}\varphi_1 \to {\sim}\psi_1, {\sim}\psi_1 \to {\sim}\varphi_1\} \cup \{\varphi_2 \to \psi_2, \psi_2 \to \varphi_2, {\sim}\varphi_2 \to {\sim}\psi_2, {\sim}\psi_2 \to {\sim}\varphi_2\} \vdash_{\mathbf{L}_{\mathrm{QNI}}} {\sim}(\varphi_1 \to \varphi_2) \to {\sim}(\psi_1 \to \psi_2)$ and $\{\varphi_1 \to \psi_1, \psi_1 \to \varphi_1, {\sim}\varphi_1 \to {\sim}\psi_1, {\sim}\psi_1 \to {\sim}\varphi_1\} \cup \{\varphi_2 \to \psi_2, \psi_2 \to \varphi_2, {\sim}\varphi_2 \to {\sim}\psi_2, {\sim}\psi_2 \to {\sim}\varphi_2\} \vdash_{\mathbf{L}_{\mathrm{QNI}}} {\sim}(\psi_1 \to \psi_2) \to {\sim}(\varphi_1 \to \varphi_2)$, we only prove the first one, the other proof is similar and hence omitted. Thanks to Theorem 1, in order to prove that $\{\varphi_1 \to \psi_1, \psi_1 \to \varphi_1, {\sim}\varphi_1 \to {\sim}\psi_1, {\sim}\psi_1 \to {\sim}\varphi_1\} \cup \{\varphi_2 \to \psi_2, \psi_2 \to \varphi_2, {\sim}\varphi_2 \to {\sim}\psi_2, {\sim}\psi_2 \to {\sim}\varphi_2\} \vdash_{\mathbf{L}_{\mathrm{QNI}}} {\sim}(\varphi_1 \to \varphi_2) \to {\sim}(\psi_1 \to \psi_2)$ is sufficient to prove that $\{\varphi_1 \to \psi_1, \psi_1 \to \varphi_1, {\sim}\varphi_1 \to {\sim}\psi_1, {\sim}\psi_1 \to {\sim}\varphi_1\} \cup \{\varphi_2 \to \psi_2, \psi_2 \to \varphi_2, {\sim}\varphi_2 \to {\sim}\psi_2, {\sim}\psi_2 \to {\sim}\varphi_2\} \cup \{{\sim}(\varphi_1 \to \varphi_2)\} \vdash_{\mathbf{L}_{\mathrm{QNI}}} {\sim}(\psi_1 \to \psi_2)$

1.	$\varphi_1 \to \psi_1$	Assumption
2.	${\sim}{\sim}\varphi_1 \to {\sim}{\sim}\psi_1$	Lemma 3.6
3.	${\sim}(\varphi_1 \to \varphi_2)$	Assumption
4.	${\sim}\varphi_2$	Lemma 3.2
5.	${\sim}{\sim}\varphi_1$	Lemma 3.3
6.	${\sim}\varphi_2 \to {\sim}\psi_2$	Assumption
7.	${\sim}\psi_2$	4, 6, MP
8.	${\sim}{\sim}\psi_1$	2, 5, MP
9.	${\sim}(\psi_1 \to \psi_2)$	7, 8, Lemma 3.1

It remains to prove (**G**), that is, that $\varphi, \psi \vdash_{\mathbf{L}_{\mathrm{QNI}}} \{\varphi \to \psi, \psi \to \varphi, {\sim}\psi \to {\sim}\varphi, {\sim}\varphi \to {\sim}\psi\}$. The deductions $\varphi, \psi \vdash_{\mathbf{L}_{\mathrm{QNI}}} \varphi \to \psi$ and $\varphi, \psi \vdash_{\mathbf{L}_{\mathrm{QNI}}} \psi \to \varphi$ follow from the deduction theorem. Now, that $\varphi, \psi \vdash_{\mathbf{L}_{\mathrm{QNI}}} {\sim}\varphi \to {\sim}\psi$ and $\varphi, \psi \vdash_{\mathbf{L}_{\mathrm{QNI}}} {\sim}\psi \to {\sim}\varphi$ follow from deduction theorem together with **AX7** and **AX10**. ■

We now apply [Czelakowski, Pigozzi, 2004, Theorem 30] to obtain a presentation of the equivalent algebraic semantics of $\mathbf{L}_{\mathrm{QNI}}$.

Theorem 3. *Let* $\mathbf{L}$ *be a logic axiomatized by a set* $\boldsymbol{Ax}$ *of axioms and a set* $\boldsymbol{Ru}$ *of proper inference rules. Assume* $\mathbf{L}$ *is regularly algebraizable with a finite set of equivalence formulas* $\Delta(\varphi, \psi) = \{\varepsilon_0(\varphi, \psi), \cdots, \varepsilon_{n-1}(\varphi, \psi)\}$. *Let* $\top$ *be a fixed but arbitrary theorem of* $\mathbf{L}$. *Then the unique equivalent quasivariety semantics of* $\mathbf{L}$ *is defined by the following identities and quasi-identities:*

(1) $\varphi \approx \top$ *for each* $\varphi \in \boldsymbol{Ax}$.

(2) $(\psi_0 \approx \top, \cdots, \psi_p \approx \top)$ *implies* $\varphi \approx \top$, *for each inference rule* $\psi_0, \cdots, \psi_p \vdash \varphi$ *in* $\boldsymbol{Ru}$.

(3) $\Delta(\varphi, \psi) \approx \top$ *implies* $\varphi \approx \psi$.

In the above theorem, $\Delta(\varphi, \psi) \approx \top$ means that $\gamma \approx \top$ for each $\gamma \in \Delta(\varphi, \psi)$. Thanks to [Font, 2016, Proposition 3.47], we know that given a logic $\mathbf{L}$ that is regularly algebraizable with equivalent algebraic semantics the class K, if φ is any theorem of $\mathbf{L}$, then φ is an algebraic constant of the class K. Having this in mind, from now on we will let $1 := \varphi \to \varphi$ and $0 := \sim(\varphi \to \varphi)$. Applying Theorem 3 to our calculus $\mathbf{L}_{\mathrm{QNI}}$, we obtain the following axiomatization of $\mathtt{Alg}^*(\mathbf{L}_{\mathrm{QNI}})$.

Proposition 2. The class $\mathtt{Alg}^*(\mathbf{L}_{\mathrm{QNI}})$ is axiomatized in the following way:

(1) $\varphi \approx 1$ for each axiom φ of $\mathbf{L}_{\mathrm{QNI}}$.

(2) If $\varphi \approx 1$ and $\varphi \to \psi \approx 1$, then $\psi \approx 1$.

(3) If $\varphi \to \psi \approx \psi \to \varphi \approx \sim\varphi \to \sim\psi \approx \sim\psi \to \sim\varphi \approx 1$, then $\varphi \approx \psi$.

5. $\mathtt{Alg}^*(\mathbf{L}_{\mathbf{QNI}}) = \mathsf{QNI}$

In this section we prove that the class of QNI-algebras (Definition 1) coincides with the class $\mathtt{Alg}^*(\mathbf{L}_{\mathrm{QNI}})$ given in Proposition 2.

Proposition 3. $\mathtt{Alg}^*(\mathbf{L}_{\mathrm{QNI}}) \subseteq \mathsf{QNI}$.

Proof. Given $\mathbf{A} \in \mathtt{Alg}^*(\mathbf{L}_{\mathrm{QNI}})$, we will prove that $\mathbf{A}$ satisfies all equations and quasi-equations given in Definition 1. Recall that from Proposition 2.3, we have that if $\varphi \to \psi \approx \psi \to \varphi \approx \sim\varphi \to \sim\psi \approx \sim\psi \to \sim\varphi \approx 1$, then $\varphi \approx \psi$.

The proofs of (1), (3), (4) and (5) are very similar to one another; we show the proof of (3) by way of an example. We have to prove that $(\varphi \to (\psi \to \gamma)) \approx ((\varphi \to \psi) \to (\varphi \to \gamma))$. We are going to show that $(\varphi \to (\psi \to \gamma)) \to$

$((\varphi \to \psi) \to (\varphi \to \gamma)) \approx 1$, $((\varphi \to \psi) \to (\varphi \to \gamma)) \to (\varphi \to (\psi \to \gamma)) \approx 1$, ${\sim}(\varphi \to (\psi \to \gamma)) \to {\sim}((\varphi \to \psi) \to (\varphi \to \gamma)) \approx 1$ and ${\sim}((\varphi \to \psi) \to (\varphi \to \gamma)) \to {\sim}(\varphi \to (\psi \to \gamma)) \approx 1$. The desired result will then follow from Proposition 2.3. Thanks to Lemma 4.1 and Lemma 4.2, ${\sim}(\varphi \to (\psi \to \gamma)) \to {\sim}((\varphi \to \psi) \to (\varphi \to \gamma)), {\sim}((\varphi \to \psi) \to (\varphi \to \gamma)) \to {\sim}(\varphi \to (\psi \to \gamma))$ are theorems of $\mathbf{L}_{\mathrm{QNI}}$ and therefore by Proposition 2.1 we conclude that ${\sim}(\varphi \to (\psi \to \gamma)) \to {\sim}((\varphi \to \psi) \to (\varphi \to \gamma)) = 1, {\sim}((\varphi \to \psi) \to (\varphi \to \gamma)) \to {\sim}(\varphi \to (\psi \to \gamma)) \approx 1$, the same applies to $(\varphi \to (\psi \to \gamma)) \to ((\varphi \to \psi) \to (\varphi \to \gamma))$ and $((\varphi \to \psi) \to (\varphi \to \gamma)) \to (\varphi \to (\psi \to \gamma))$. As to (2), we have to prove that $\varphi \to (\psi \to \varphi) \approx 1$, $\varphi \to \varphi \approx 1$ and $0 \to \varphi \approx 1$, since $\varphi \to (\psi \to \varphi)$, $\varphi \to \varphi$ and $0 \to \varphi$ are theorems of $\mathbf{L}_{\mathrm{QNI}}$ (**AX1**, Lemma 2.1 and **AX10**, respectively), thanks to Proposition 2.1 we have the desired equalities. As to (6), we have to prove that if ${\sim}\varphi \to {\sim}\psi \approx 1$, then ${\sim}\varphi \to ({\sim}\varphi \odot {\sim}\psi) \approx 1$. Since $({\sim}\varphi \to {\sim}\psi) \to ({\sim}\varphi \to ({\sim}\varphi \odot {\sim}\psi))$ is a theorem of $\mathbf{L}_{\mathrm{QNI}}$, Lemma 4.6, we conclude from Proposition 2.2 the equality. As to (7), it is just application of Lemmas 4.10 and 4.11. And for (8), notice that is just application of Lemma 4.3. As to (9), supposing that $\varphi \to \psi \approx 1, \psi \to \varphi \approx 1, \gamma \to \delta, \delta \to \gamma \approx 1$, we want to prove that $(\varphi \to \gamma) \to (\psi \to \delta) \approx 1, (\psi \to \delta) \to (\varphi \to \gamma) \approx 1$ and that $(\varphi \to \gamma) \odot (\psi \to \delta) \approx 1, (\psi \to \delta) \odot (\varphi \to \gamma) \approx 1$, since $(\psi \to \varphi) \to ((\gamma \to \delta) \to ((\varphi \to \gamma) \to (\psi \to \delta))$ and $(\psi \to \varphi) \to ((\delta \to \gamma) \to ((\psi \to \delta) \to (\varphi \to \gamma)))$ are theorems of $\mathbf{L}_{\mathrm{QNI}}$, Proposition 2.2 give us the desired equalities. The same idea is applied to $(\varphi \to \gamma) \odot (\psi \to \delta) \approx 1, (\psi \to \delta) \odot (\varphi \to \gamma) \approx 1$. In order to prove (10) notice that Lemma 3.4 and 3.5 give us that ${\sim}\varphi \to {\sim}{\sim}{\sim}\varphi, {\sim}{\sim}{\sim}\varphi \to {\sim}\varphi, {\sim}{\sim}\varphi \to {\sim}{\sim}{\sim}{\sim}\varphi, {\sim}{\sim}{\sim}{\sim}\varphi \to {\sim}{\sim}\varphi$ are theorems of $\mathbf{L}_{\mathrm{QNI}}$ and therefore from Proposition 2.3, we conclude that ${\sim}\varphi \approx {\sim}{\sim}{\sim}\varphi$. In order to prove (11), notice that $0 := {\sim}(\varphi \to \varphi)$ and that ${\sim}{\sim}(\varphi \to \varphi)$ is a theorem of $\mathbf{L}_{\mathrm{QNI}}$, then thanks to Proposition 2.1, ${\sim}{\sim}(\varphi \to \varphi) \approx 1$, i.e, ${\sim}0 \approx 1$. As to (12), it is just application of **AX4**. From **AX5** and Proposition 2.1, we have that $\varphi \to {\sim}{\sim}\varphi \approx 1$, and it proves (13). As to (14), supposing that $\varphi \to \psi \approx 1$, we have to prove that $(\varphi \odot \gamma) \to (\psi \odot \gamma) \approx 1$ and that $(\gamma \odot \varphi) \to (\gamma \odot \psi) \approx 1$, and it is Lemma 4.3, Lemma 4.7 and Proposition 2.1. (15), (16) and (17) are **AX6**, Lemma 4.4 and Lemma 4.5 together with Proposition 2.2. ■

Proposition 4. $\mathsf{QNI} \subseteq \mathtt{Alg}^*(\mathbf{L}_{\mathrm{QNI}})$.

Proof. First for all, we will prove that $\mathbf{A}$ satisfies $\varphi \approx 1$ for each axiom φ of $\mathbf{L}_{\mathrm{QNI}}$. In the proof below, the notation $E(\mathbf{AX1})$ means $\varphi \approx 1$ being φ the first axiom of $\mathbf{L}_{\mathrm{QNI}}$ and so on.

As to $E(\mathbf{AX1})$ and $E(\mathbf{AX2})$, notice that thanks to (3), $\varphi \to (\psi \to \varphi) \approx 1$ and $(\varphi \to (\psi \to \gamma)) \to ((\varphi \to \psi) \to (\varphi \to \gamma)) \approx 1$. As to $E(\mathbf{AX3})$, notice that thanks to (10) we have that $\sim\varphi \to \sim\sim\sim\varphi \approx 1$. As to $E(\mathbf{AX4})$, $E(\mathbf{AX5})$ and $E(\mathbf{AX6})$ notice that thanks to (12), (13) and (15), respectively, we have the equality. As to $E(\mathbf{AX7})$, notice that thanks to (4), we have that $\sim\sim\varphi \to (\sim\psi \to (\sim\sim\varphi \odot \sim\psi)) \approx (\sim\sim\varphi \odot \sim\psi) \to (\sim\sim\varphi \odot \sim\psi)$ and since $\varphi \to \varphi \approx 1$ from (2), we conclude that $\sim\sim\varphi \to (\sim\psi \to (\sim\sim\varphi \odot \sim\psi)) \approx 1$. Now, notice that $(\sim\sim\varphi \odot \sim\psi) \approx \sim(\sim\sim\varphi \to \sim\sim\psi)$ and thanks to (17), $\sim(\sim\sim\varphi \to \sim\sim\psi) \equiv \sim(\varphi \to \gamma)$ and now from Lemma 1.2 we conclude that $\sim\sim\varphi \to (\sim\psi \to \sim(\varphi \to \psi)) \approx 1$. As to $E(\mathbf{AX8})$ and $E(\mathbf{AX9})$, notice that thanks to (16) and Lemma 1.1 we have that $\sim(\varphi \to \psi) \to \sim\sim\varphi \approx 1$ and $\sim(\varphi \to \psi) \to \sim\psi \approx 1$. As to $E(\mathbf{AX10})$, it is (2). In order to prove that If $\varphi \approx 1$ and $\varphi \to \psi \approx 1$, then $\psi \approx 1$, thanks to (1), since $\varphi \to \psi \approx \psi$, given that $\varphi \to \psi \approx 1$ we conclude that $\psi \approx \varphi \to \psi \approx 1$. It remains to prove that if $\varphi \to \psi \approx 1, \psi \to \varphi \approx 1, \sim\varphi \to \sim\psi \approx 1$ and $\sim\psi \to \sim\varphi \approx 1$, then $\varphi \approx \psi$. Thanks to Lemma 1.3 we conclude that from $\varphi \to \psi \approx 1$ and $\sim\psi \to \sim\varphi \approx 1$, $\varphi \leq \psi$ and from $\psi \to \varphi \approx 1$ and $\sim\varphi \to \sim\psi \approx 1$, $\psi \leq \varphi$. From the two inequalities we conclude that $\varphi \approx \psi$. ■

6. Extensions of $\mathbf{L}_{\mathbf{QNI}}$ and EDPC

In this section we look at some extensions of $\mathbf{L}_{\mathrm{QNI}}$.

Proposition 5. The $\{\to, \sim\}$-fragment of intuitionistic propositional logic is a strengthening of $\mathbf{L}_{\mathrm{QNI}}$ obtained by adding any of the axioms below:

(1) $(\varphi \to \psi) \to (\sim\psi \to \sim\varphi)$

(2) $(\varphi \to \sim\psi) \to (\psi \to \sim\varphi)$

(3) $(\varphi \to \sim(\varphi \to \varphi)) \to \sim\varphi$

(4) $\sim(\varphi \odot (\varphi \to \varphi)) \to \sim\varphi$

(5) $\sim(\varphi \odot \psi) \to \sim(\psi \odot \varphi)$

(6) $\sim(\sim\sim\varphi \odot (\varphi \to \varphi)) \to \sim((\varphi \to \varphi) \odot \sim\sim\varphi)$

(7) $(\varphi \to \sim\varphi) \to \sim\varphi$

(8) $\sim(\varphi \odot \varphi) \to \sim\varphi$.

Proof. Thanks to [Rivieccio, 2020a, Proposition 4.26 (iii)], we know that a QNI-algebra $\mathbf{A}$ is a bounded Hilbert algebra iff $\mathbf{A}$ satisfies any of the equations in [Rivieccio, 2020a, Lemma 4.25], the axioms (1), (2), (3), (4), (5) and (6) are equivalent to the equations (i), (ii), (iii), (v), (vi) and (vii), respectively.

In order to prove that (7) and (8) are equivalent, observe that $(\varphi \to \sim\varphi) \to \sim(\varphi \odot \varphi)$ and $\sim(\varphi \odot \varphi) \to (\varphi \to \sim\varphi)$ (See Lemma 4.9) are theorems of $\mathbf{L}_{\mathrm{QNI}}$. In order to finish the equivalence, observe that $(\varphi \to \sim\varphi) \to (\varphi \to \sim(\varphi \to \varphi))$ and $(\varphi \to \sim(\varphi \to \varphi)) \to (\varphi \to \sim\varphi)$ are theorems of $\mathbf{L}_{\mathrm{QNI}}$. ■

In order to axiomatize the $\{\to, \sim\}$-fragment of Nelson's constructive logic with strong negation, it is enough to add the involutive axiom $(\sim\sim\varphi \to \varphi)$ to $\mathbf{L}_{\mathrm{QNI}}$. In order to obtain classical logic, we can add any of the following axioms to $\mathbf{L}_{\mathrm{QNI}}$:

(1) $\{(\varphi \odot (\varphi \to \varphi)) \to \varphi, \sim(\varphi \odot (\varphi \to \varphi)) \to \sim\varphi\}$

(2) $(\sim\varphi \to \sim\psi) \to (\psi \to \varphi)$

(3) $\{(\varphi \odot \varphi) \to \varphi, \sim(\varphi \odot \varphi) \to \sim\varphi\}$.

Proof. (1) and (2) are the logical counterparts identities in [Rivieccio, 2020a, Proposition 4.26 (iv))]. In order to prove that (1) and (3) are equivalent, suppose that $(\varphi \odot \varphi) \to \varphi$. Thanks to **AX6**, it follows that $(\varphi \odot (\varphi \to \varphi)) \to (\varphi \odot \varphi)$ and by using Lemma 2.2 it follows that $(\varphi \odot (\varphi \to \varphi)) \to \varphi$. Now, suppose that $\sim(\varphi \odot \varphi) \to \sim\varphi$, as $\sim(\varphi \odot (\varphi \to \varphi)) \to \sim(\varphi \odot \varphi)$ is a theorem of $\mathbf{L}_{\mathrm{QNI}}$ (Lemma 4.8), it follows by Lemma 2.2 that $\sim(\varphi \odot (\varphi \to \varphi)) \to \varphi$. ■

Remark 1. Consider the algebra $\mathbf{A}$ with universe $\{0, \frac{1}{2}, 1\}$ and operations given by the following tables:

$\sim$	
0	1
$\frac{1}{2}$	$\frac{1}{2}$
1	0

$\to$	0	$\frac{1}{2}$	1
0	1	1	1
$\frac{1}{2}$	1	1	1
1	0	$\frac{1}{2}$	1

It is not difficult to show that $\mathbf{A}$ is a QNI-algebra that witnesses $\sim(\varphi \odot \psi) \nvdash_{\mathbf{L}_{\mathrm{QNI}}} \sim(\psi \odot \varphi)$. This entails that $\mathbf{L}_{\mathrm{QNI}}$ is not self-extensional, see [Font, 2016, Definition 5.24] for a definition of self-extensionality.

It is well known that an algebraizable logic $\mathbf{L}$ with equivalent algebraic semantics a variety K has the Deduction-Detachment Theorem if and only if K has EDPC. Since $\mathbf{L}_{\mathrm{QNI}}$ enjoys the DDT (Theorem 1), we know that $\mathtt{Alg}^*(\mathbf{L}_{\mathrm{QNI}})$ has EDPC. We give below the equations witnessing this.

Definition 3 ([Blok, Pigozzi, 1994])**.** An algebra $\mathbf{A}$ has definable principal congruences (DPC) if there is a first-order formula $\varphi(x, y, z, w)$ such that, for all $a, b, c, d \in \mathbf{A}$,

$$A \models \varphi[a, b, c, d] \iff c \equiv d \pmod{\Theta(a, b)}.$$

A has equationally definable principal congruences (EDPC) if φ can be taken to be a finite conjunction of identities $p_i(x, y, z, w) \approx q_i(x, y, z, w)$, for $i \in \omega$. A class K of algebras has EDPC if a single equation or conjunction of equations defines principal congruences on every member of K.

Proposition 6 ([Font, 2016] Corollary 3.81)**.** Let **L** be an algebraizable logic with equivalent algebraic semantics a variety K. Then **L** satisfies the DDT if and only if K has EDPC.

Theorem 4 ([Rivieccio, Jansana, 2020] Corollary 34)**.** *The term* $\beta(x, y, z)$ *satisfies:*

$$c \equiv d \ (\mathrm{mod}\ \Theta(a, b)) \Longleftrightarrow \beta(a, b, c) \approx \beta(a, b, d).$$

7. Conclusions and future work

As observed in [Rivieccio, 2020a], the translations that witness the algebraizability of (quasi-)Nelson logic can be defined using different choices of connectives. For the defining equation, one can let $E(\varphi) := \{\varphi \approx \varphi \to \varphi\}$, or $E(\varphi) := \{\varphi \approx 1\}$, or $E(\varphi) := \{\varphi \approx \varphi \to \varphi\}$, or $E(\varphi) := \{\varphi \approx \varphi \leftrightarrow \varphi\}$, or $E(\varphi) := \{\varphi \approx \varphi \Leftrightarrow \varphi\}$, where $\varphi \leftrightarrow \psi := (\varphi \to \psi) \wedge (\psi \to \varphi)$ and $\varphi \Leftrightarrow \psi := (\varphi \to \psi) * (\psi \to \varphi)$, etc. For the equivalence formulas, one has for instance the following options: $\Delta(\varphi, \psi) := \{\varphi \Leftrightarrow \psi\}$, $\Delta(\varphi, \psi) := \{\varphi \Rightarrow \psi, \psi \Rightarrow \varphi\}$, $\Delta(\varphi, \psi) := \{\varphi \to \psi, \psi \to \varphi, {\sim}\varphi \to {\sim}\psi, {\sim}\psi \to {\sim}\varphi\}$, etc. Indeed, one can show that every fragment of quasi-Nelson logic containing either $\{\Rightarrow\}$ or $\{\Leftrightarrow\}$ or $\{\to, \sim\}$ is algebraizable.

With regards to future research, the above considerations suggest that a number of further fragments may be worthwhile looking at from the point of view of algebraizability. We note that, on the one hand, a successful characterization of a given fragment of quasi-Nelson logic may be easily specialized to the involutive case, therefore also yielding a characterization of the corresponding fragment of Nelson's constructive logic with strong negation[1]. On the other hand, quasi-Nelson logic certainly has a greater number of non-equivalent fragments than Nelson's (and intuitionistic) logic, suggesting that the landscape may be quite complex.

While we know that the fragments containing either of the quasi-Nelson implications must be algebraizable, finding a complete axiomatization for them is a different question. Recent research experience on the topic suggests that the latter issue is related to the problem of giving a twist representation for the

[1] As far as we know, the only studies on fragments of Nelson's logic are those by A. Monteiro's school [Monteiro, 1963; Brignole, Monteiro, 1967; Cignoli, 1986] on the Kleene algebra subreducts of Nelson algebras and Sendlewski's [Sendlewski, 1991].

corresponding classes of subreducts of quasi-Nelson algebras (see e.g. [Rivieccio, Spinks, 2021; Rivieccio, Jansana, 2020; Rivieccio, 2020a]). This endeavour, in turn, may require non-trivial adaptations of the algebraic constructions employed so far in the study of (quasi-)Nelson algebras. We mention below a few fragments regarding which a successful outcome may be anticipated, as well as some that seem less tractable.

If we enrich the $\{\rightarrow, \sim\}$-fragment with the monoid or the lattice connectives of quasi-Nelson algebras, then we obtain logics/classes of algebras that appear to be well-behaved from the point of view that concerns us here. In particular, the $\{*, \rightarrow, \sim\}$-fragment of quasi-Nelson logic (which can be shown to be equivalent to the $\{*, \Rightarrow, \sim\}$-fragment) appears to be a particularly interesting one because of its connection with the theory of other residuated structures; this is currently the subject of ongoing research.

By contrast, the fragments that do not contain the negation ($\sim$) seem to lie beyond the applicability of the methods employed so far. The reason for this is too technical to be discussed in detail here, but it appears to be related to the observation that the property corresponding to the Nelson identity (not only as it appears in the present paper, but also in any of its alternative formulations: see [Rivieccio, Spinks, 2018; Rivieccio, Spinks, 2021]) can only be stated through the interaction of the negation with some other connective. In other words, what makes Nelson (and quasi-Nelson) algebras 'Nelson' seems to be precisely the interaction of the negation with some other connective (e.g. the interaction of the negation with the implications, of the negation with the lattice connectives, of the negation with the monoid conjunction, etc.).

Between the two extreme cases mentioned above, there also appear to be fragments that are not intractable in principle but (also for technical reasons) may prove to be hard to tackle. Of these we mention, as an example, the $\{\Rightarrow, \sim\}$-fragment: this may turn out to be a particularly interesting case study, for it is certainly algebraizable and it contains the $\{\rightarrow, \sim\}$-fragment, in the sense that the weak implication is definable. We leave this suggestion as a challenge for future investigations.

8. Appendix

The rules and valid formulas below are used in the proofs of Proposition 1 and Proposition 3.

Lemma 3.

(1) $\sim\sim\psi, \sim\sim\varphi \vdash_{\mathbf{L}_{\mathrm{QNI}}} \varphi \odot \psi$

(2) $\sim(\varphi \rightarrow \psi) \vdash_{\mathbf{L}_{\mathrm{QNI}}} \sim\psi$

(3) $\sim(\varphi \to \psi) \vdash_{\mathbf{L}_{\mathrm{QNI}}} \sim\sim\varphi$

(4) $\varphi \vdash_{\mathbf{L}_{\mathrm{QNI}}} \sim\sim\varphi$

(5) $\sim\sim\sim\varphi \vdash_{\mathbf{L}_{\mathrm{QNI}}} \sim\varphi$

(6) $(\varphi \to \psi) \vdash_{\mathbf{L}_{\mathrm{QNI}}} (\sim\sim\varphi \to \sim\sim\psi)$

Lemma 4.

(1) $\emptyset \vdash_{\mathbf{L}_{\mathrm{QNI}}} \sim(\varphi \to (\psi \to \gamma)) \to \sim((\varphi \to \psi) \to (\varphi \to \gamma))$

(2) $\emptyset \vdash_{\mathbf{L}_{\mathrm{QNI}}} \sim((\varphi \to \psi) \to (\varphi \to \gamma)) \to \sim(\varphi \to (\psi \to \gamma))$

(3) $\emptyset \vdash_{\mathbf{L}_{\mathrm{QNI}}} (\varphi \odot \psi) \to (\psi \odot \varphi)$

(4) $\emptyset \vdash_{\mathbf{L}_{\mathrm{QNI}}} (\varphi \odot \psi) \to (\varphi \odot (\varphi \to \psi))$

(5) $\emptyset \vdash_{\mathbf{L}_{\mathrm{QNI}}} (\sim\sim\varphi \odot \sim\psi) \to \sim(\varphi \to \psi)$

(6) $\emptyset \vdash_{\mathbf{L}_{\mathrm{QNI}}} (\sim\varphi \to \sim\psi) \to (\sim\varphi \to (\sim\varphi \odot \sim\psi))$

(7) $\emptyset \vdash_{\mathbf{L}_{\mathrm{QNI}}} (\varphi \to \psi) \to (\varphi \odot \gamma) \to (\psi \odot \gamma)$

(8) $\emptyset \vdash_{\mathbf{L}_{\mathrm{QNI}}} \sim(\varphi \odot (\varphi \to \varphi)) \to \sim(\varphi \odot \varphi)$

(9) $\emptyset \vdash_{\mathbf{L}_{\mathrm{QNI}}} \sim\sim(\varphi \to \sim\psi) \to (\varphi \to \sim\psi)$

(10) $\emptyset \vdash_{\mathbf{L}_{\mathrm{QNI}}} (\varphi \odot (\psi \odot \gamma)) \to ((\varphi \odot \psi) \odot \gamma)$

(11) $\emptyset \vdash_{\mathbf{L}_{\mathrm{QNI}}} ((\varphi \odot \psi) \odot \gamma) \to (\varphi \odot (\psi \odot \gamma))$

Proof.[3.1]

1.	$\sim(\varphi \to (\psi \to \gamma))$	Assumption
2.	$\sim\sim\varphi$	1, Lemma 3.3
3.	$\sim(\psi \to \gamma)$	1, Lemma 3.2
4.	$\sim\sim\psi$	3, Lemma 3.3
5.	$\sim\gamma$	3, Lemma 3.2
6.	$\psi \to (\varphi \to \psi)$	**AX1**
7.	$\sim\sim\psi \to \sim\sim(\varphi \to \psi)$	6, Lemma 3.6
8.	$\sim\sim(\varphi \to \psi)$	4, 7, MP
9.	$\sim(\varphi \to \gamma)$	2, 5, **AX7**
10.	$\sim((\varphi \to \psi) \to (\varphi \to \gamma))$	8, 9, **AX7**

■

Proof.[3.3]

1.	$\varphi \odot \psi$	Assumption
2.	$\sim\sim\varphi$	1, Lemma 3.3
3.	$\sim\sim\psi$	2, Lemma 3.2
4.	$\psi \odot \varphi$	2, 3, Lemma 3.1

■

Proof. [3.4]

1.	$\varphi \odot \psi$	Assumption
2.	$\sim\sim\varphi$	1, Lemma 3.3
3.	$\sim\sim\psi$	2, Lemma 3.2
4.	$\psi \to (\varphi \to \psi)$	**AX1**
5.	$\sim\sim\psi \to \sim\sim(\varphi \to \psi)$	**AX4**
6.	$\sim\sim(\varphi \to \psi)$	3, 5, MP
7.	$\varphi \odot (\varphi \to \psi)$	2, 6, Lemma 3.1

■

Proof. [3.5]

1.	$\sim\sim\varphi \odot \sim\psi$	Assumption
2.	$\sim\sim\sim\sim\varphi$	1, Lemma 3.3
3.	$\sim\sim\sim\psi$	2, Lemma 3.2
4.	$\sim\sim\varphi$	3, Lemma 3.4
5.	$\sim\psi$	4, Lemma 3.5
6.	$\sim\sim\varphi \to (\sim\psi \to \sim(\varphi \to \psi))$	**AX7**
7.	$\sim\psi \to \sim(\varphi \to \psi)$	4, 6, MP
8.	$\sim(\varphi \to \psi)$	5, 7, MP

■

Proof. [3.6]

1.	$\sim\varphi \to \sim\psi$	Assumption
2.	$\sim\varphi$	Assumption
3.	$\sim\psi$	1, 2, MP
4.	$\sim\sim\sim\varphi$	2, Lemma 3.5
5.	$\sim\sim\sim\psi$	3, Lemma 3.5
6.	$\sim\sim\sim\varphi \to (\sim\sim\sim\psi \to \sim(\sim\varphi \to \sim\sim\psi))$	**AX7**
7.	$\sim\sim\sim\psi \to \sim(\sim\varphi \to \sim\sim\psi)$	4, 6, MP
8.	$\sim(\sim\varphi \to \sim\sim\psi)$	5, 7, MP

■

Proof. [3.7]

1.	$\varphi \to \psi$	Assumption
2.	$\varphi \odot \gamma$	Assumption
3.	$\sim\sim\varphi$	2, Lemma 3.3
4.	$\sim\sim\gamma$	2, Lemma 3.2
5.	$(\varphi \to \psi) \to (\sim\sim\varphi \to \sim\sim\psi)$	**AX4**
6.	$\sim\sim\varphi \to \sim\sim\psi$	1, 5, MP
7.	$\sim\sim\psi$	3, 6, MP
8.	$\psi \odot \gamma$	4, 7, Lemma 3.1

■

Proof.[3.10]

1.	$(\varphi \odot (\psi \odot \gamma))$	Assumption
2.	$\varphi \odot \sim(\psi \to \sim\gamma)$	Definition
3.	$\sim(\varphi \to \sim\sim(\psi \to \sim\gamma))$	Definition
4.	$\sim\sim\varphi$	3, Lemma 3.3
5.	$\sim\sim\sim(\psi \to \sim\gamma)$	4, Lemma 3.2
6.	$\sim(\psi \to \sim\gamma)$	5, Lemma 3.5
7.	$\sim\sim\psi$	6, Lemma 3.3
8.	$\sim\sim\gamma$	7, Lemma 3.2
9.	$(\varphi \odot \psi)$	4, 7, Lemma 3.1
10.	$\sim\sim(\varphi \odot \psi)$	9, Lemma 3.4
11.	$((\varphi \odot \psi) \odot \gamma)$	8, 10, Lemma 3.1

■

Acknowledgements. T. Nascimento was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior – Brasil (CAPES) – Finance Code 001. U. Rivieccio acknowledges partial funding by the Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq, Brazil), under the grant 313643/2017-2 (Bolsas de Produtividade em Pesquisa – PQ). Special thanks are due to Ramon Jansana, for several useful comments on earlier versions of the paper.

References

Blok, Pigozzi, 1989 – Blok, W.J., Pigozzi, D. "Algebraizable Logics", *Memoirs of the American Mathematical Society*, 1989, Vol. 77.

Blok, Pigozzi, 1994 – Blok, W.J., Pigozzi, D. "On the structure of varieties with equationally definable principal congruences III", *Algebra Universalis*, 1994, Vol. 32, pp. 545–608.

Brignole, Monteiro, 1967 – Brignole, D., Monteiro, A. "Caracterisation des algèbres de Nelson par des egalités. I, II", *Proceedings of the Japan Academy*, 1964, Vol. 43, pp. 279–285.

Cignoli, 1986 – Cignoli, R. "The class of Kleene algebras satisfying an interpolation property and Nelson algebras", *Algebra Universalis*, 1986, Vol. 23, pp. 262–292.

Czelakowski, Pigozzi, 2004 – Czelakowski, J., Pigozzi, D. "Fregean logics", *Annals of Pure and Applied Logic*, 2004, Vol. 127, pp. 17–76.

Font, 2016 – Font, J.M. "Abstract Algebraic Logic. An Introductory Textbook", *College Publications*, 2016.

Galatos et al., 2007 – Galatos, N., Jipsen, P., Kowalski, T., and Ono, H. "Residuated Lattices: an algebraic glimpse at substructural logics", *Studies in Logic and the Foundations of Mathematics*, 2007.

Hamilton, 1978 – Hamilton, A.G. "Logic for Mathematicians", *Cambridge University Press*, 1978.

Liang, Nascimento, 2019 – Liang, F., Nascimento, T. "Algebraic Semantics for quasi-Nelson Logic", *Lecture Notes in Computer Science*, 2019, Vol. 11541, pp. 450–466.

Monteiro, 1963 – Monteiro, A. "Construction des algèbres de Nelson finies", *Bulletin of the Polish Academy of Sciences*, 1963, Vol. 11, pp. 359–362.

Nelson, 1949 – Nelson, D. "Constructible falsity", *The Journal of Symbolic Logic*, 1949, Vol. 14, pp. 16–26.

Rivieccio, 2020a – Rivieccio, U. "Fragments of quasi-Nelson: The Algebraizable core", *submitted.*

Rivieccio, 2020b – Rivieccio, U. "Representation of De Morgan and (Semi-)Kleene lattices", *Soft Computing*, 2020, Vol. 24 (12), pp. 8685–8716.

Rivieccio, 2020c – Rivieccio, U. "Fragments of quasi-Nelson: two negations", *Journal of Applied Logic*, 2020, Vol. 7, No. 4, pp. 499–559.

Rivieccio, Jansana, 2020 – Rivieccio, U., Jansana, R. "Quasi-Nelson algebras and fragments", *Mathematical Structures in Computer Science*, pp. 1–29. DOI: 10.1017/S0960129521000049.

Rivieccio, Spinks, 2018 – Rivieccio, U., Spinks, M. "Quasi-Nelson algebras", *Electronic Notes in Theoretical Computer Science*, 2019, Vol. 344, pp. 169–188.

Rivieccio, Spinks, 2021 – Rivieccio, U., Spinks, M. "Quasi-Nelson; or, non-involutive Nelson algebras.", *Algebraic Perspectives on Substructural Logics*, ed. by D. Fazio, A. Ledda, and F. Paoli, Cham: Springer 2021, pp. 133–168.

Sendlewski, 1991 – Sendlewski, A. "Topologicality of Kleene algebras with a weak pseudocomplementation over distributive p-algebras", *Reports on Mathematical Logic*, 1991, Vol. 25, pp. 13–56.

Логические исследования
2021. Т. 27. № 1. С. 124–144
УДК 16+510.6

Logical Investigations
2021, Vol. 27, No. 1, pp. 124–144
DOI: 10.21146/2074-1472-2021-27-1-124-144

Thomas Studer

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

For citation: Studer T. "A Conflict Tolerant Logic of Explicit Evidence", *Logicheskie Issledovaniya / Logical Investigations*, 2021, Vol. 27, No. 1, pp. 124–144. DOI: 10.21146/2074-1472-2021-27-1-124-144

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 CTJ_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 CTJ_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* $A, B_1, \ldots, B_n$*, arbitrary terms* $s_1, \ldots, s_n$*, and an arbitrary agent* a*, if*

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

then there is a term t *such that*

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

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* $\mathsf{CTJ}_{\mathsf{CS}}$

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

Proof. By internalization, there exists a term r with

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

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.

[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*.

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*

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

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

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

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

$$[k]_a (A \wedge \neg A \rightarrow P).$$

Since CS is schematic, we find by Lemma 3 that

$$[k]_a (A \wedge \neg A \rightarrow \neg P)$$

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

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

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

$$[s]_a A \vee [t]_a A \rightarrow [s+t]_a A.$$

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

$$([s]_a A \wedge \neg [t]_a \neg A) \rightarrow ([s+t]_a A \wedge [t+s]_a A).$$

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 $W_0 \neq \emptyset$.
- $V : W \times \mathsf{Fml} \rightarrow \{0, 1\}$ such that for all $\omega \in W_0$, $t \in \mathsf{Tm}$, $F, G \in \mathsf{Fml}$:
  - $V(\omega, \bot) = 0$;
  - $V(\omega, F \rightarrow G) = 1$ iff $V(\omega, F) = 0$ or $V(\omega, G) = 1$;
  - $V(\omega, [t]_a F) = 1$ iff $E_a(\omega, t) \subseteq \{\upsilon \in W \mid V(\upsilon, F) = 1\}$.
- $E : \mathsf{Ag} \rightarrow (W \times \mathsf{Tm} \rightarrow \mathcal{P}(W))$ 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 $s, t \in \mathsf{Tm}$:

  - $E_a(\omega, s \cdot t) \subseteq \{\upsilon \in W \mid \forall F \in \mathsf{APP}_{a,\omega}(s, t)(\upsilon \in [F])\}$ 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 $[c_n]_{a_n} \ldots [c_1]_{a_1} A \in \mathsf{CS}$:

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

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 $\nvdash_\emptyset ([s]_a A \wedge [t]_a \neg A) \to \bot$.

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 $[t]_a F$ 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 $[t]_a F \to F$.[2] The

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

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 $[t]_a F$ 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

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

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 $\mathsf{CTJ}_{\mathsf{CS}}$. 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. $[r]_bF$. Thus agent b can announce F to agent a. Then this announcement, represented by ann, is a's evidence for F, i.e. $[\mathsf{ann}]_aF$.

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:

If Ursula observes outcome ok, then Bob obtained outcome 1. (7)

If Bob observes outcome 1, then Alice obtained outcome 1. (8)

If Alice observes outcome 1, then Wigner will obtain outcome fail. (9)

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

if Ursula observes outcome ok, then Wigner will obtain outcome fail. (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 $\mathbf{j}$ 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 $\mathsf{CTJ_{CS}}$, 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 $\mathsf{CTJ_{CS}}$. 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 $\mathsf{CTJ_{CS}}$ 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 $\mathsf{CTJ_{CS}}$ 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

$$\mathbf{ju} \quad [s]_a[t]_b A \to [\downarrow s]_a A)$$

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

$$[t]_b A \to [\uparrow^a_b t]_a A.$$

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

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

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:

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

Applying the trust axiom several times leads to

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

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 $\neg(\Box A \land \Box\neg A)$, 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 – Arrow, K.J. "A difficulty in the concept of social welfare", *Journal of Political Economy*, 1950, Vol. 58, No. 4, pp. 328–346.

Artemov, 2001 – Artemov, S. "Explicit provability and constructive semantics", *Bulletin of Symbolic Logic*, 2001, Vol. 7, No. 1, pp. 1–36.

Artemov, 2006 – Artemov, S.N. "Justified common knowledge", *TCS*, 2006, Vol. 357, No. 1–3, pp. 4–22.

Artemov, 2008 – Artemov, S.N. "The logic of justification", *RSL*, 2008, Vol. 1, No. 4, pp. 477–513.

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

Brezhnev, Kuznets, 2006 – Brezhnev, V.N. and Kuznets, R. "Making knowledge explicit: How hard it is", *TCS*, 2006, Vol. 357, No. 1–3, pp. 23–34.

Bucheli et al., 2011 – Bucheli, S., Kuznets, R. and Studer, T. "Justifications for common knowledge", *Applied Non-Classical Logics*, 2011, Vol. 21, No. 1, pp. 35–60.

Bucheli et al., 2014 – Bucheli, S., Kuznets, R. and Studer, T. "Realizing public announcements by justifications", *Journal of Computer and System Sciences*, 2014, Vol. 80, No. 6, pp. 1046–1066.

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

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

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

Fitting, 2005 – Fitting, M. "The logic of proofs, semantically", *APAL*, 2005, Vol. 132, No. 1, pp. 1–25.

Fitting, 2016 – Fitting, M. "Modal logics, justification logics, and realization", *Annals of Pure and Applied Logic*, 2016, Vol. 167, No. 8, pp. 615–648.

Frauchiger, Renner, 2018 – Frauchiger, D. and Renner, R. "Quantum theory cannot consistently describe the use of itself", *Nature Communications* 9, 2018, article number 3711.

Kuznets, Studer, 2012 – Kuznets, R. and Studer, T. "Justifications, ontology, and conservativity", in: *Advances in Modal Logic, Volume 9*, T. Bolander, T. Braüner, S. Ghilardi, and L. Moss (eds.). College Publications, 2012, pp. 437–458.

Kuznets, Studer, 2016 – Kuznets, R. and Studer, T. "Weak arithmetical interpretations for the Logic of Proofs", *Logic Journal of IGPL*, 2016, Vol. 24, No. 3, pp. 4243–440.

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

Lazarovici, Hubert, 2019 – Lazarovici, D. and Hubert, M. "How quantum mechanics can consistently describe the use of itself", *Scientific Reports* 9, 2019, article number 470.

Lehmann, Studer, 2019 – Lehmann, E. and Studer, T. "Subset models for justification logic", in: R. Iemhoff, M. Moortgat, and R. de Queiroz (eds.), *Logic, Language, Information, and Computation – WoLLIC 2019*. Springer, 2019, pp. 433–449.

Lehmann, Studer, 2020 – Lehmann, E. and Studer, T. "Belief expansion in subset models", in: *Proceedings of Logical Foundations of Computer Science LFCS'20*. Springer, 2020, pp. 85–97.

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

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

Schroeter, 2019 – Schroeter, L. "Two-Dimensional Semantics", in: E. N. Zalta, (ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2019.

Studer, 2013 – Studer, T. "Decidability for some justification logics with negative introspection", *JSL*, 2013, Vol. 78, No. 2, pp. 388–402.

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

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

Yavorskaya, 2007 – Yavorskaya (Sidon),T. "Interacting explicit evidence systems", *Theory of Computing Systems*, 2008, Vol. 43, No. 2, pp. 272–293. Published online October 2007.

Приложение

Appendix

Вопросы Майклу Данну

Г. Ванзинг
E-mail: `Heinrich.Wansing@rub.de`

Г.К. Ольховиков
E-mail: `Grigory.Olkhovikov@rub.de`

Х. Омори
E-mail: `Hitoshi.Omori@rub.de`

Аннотация. Мы представляем девять вопросов, относящихся к понятию отрицания, и, в процессе, мы указываем на их связи со статьями в этом специальном выпуске. Эти вопросы были высланы одному из самых блестящих логиков, внесших вклад в теорию отрицания, профессору Джону Майклу Данну, однако, к несчастью, профессор Данн уже не мог ответить на них. Майкл Данн скончался 5 апреля 2021 г., и настоящий специальный выпуск «Логических исследований» посвящен его памяти. Поднимаемые вопросы затрагивают (i) связанные с отрицанием темы, которые особенно занимали Майкла Данна или в которые он внес важный вклад, (ii) некоторые противоречивые аспекты логического анализа понятия отрицания, или (iii) попросту свойства отрицания, в которых мы особенно заинтересованы. Хотя, к большому сожалению, эти вопросы остались без ответа от выдающегося ученого, которому были адресованы, они остаются актуальны и могут повлечь ответы со стороны других логиков, а также дальнейшие исследования.

***Ключевые слова*:** Майкл Данн, отрицание, логика Белнапа – Данна, американский план, австралийский план, паранепротиворечивость, противоречивость отрицания, правило контрапозиции, конструктивные логики Нельсона с сильным отрицанием, отрицание как аннулирование, *Ex contradictione nihil sequitur*, полуотрицание, теоретико-доказательный билатерализм.

Повесть об исключении третьего

Дж. С. Билл
E-mail: `jbeall@nd.edu`

Г. Прист
E-mail: `priest.graham@gmail.com`

Аннотация. В статье обсуждается несколько взаимосвязанных идей, касающихся отрицания, истины, общезначимости и парадокса лжеца. В частности, обсуждается аргумент в пользу диалетической природы «предложения лжеца», который опирается на телеологическую концепцию истины Даммета. Хотя один подход к его формулировке проваливается, другой оказывается успешным. Далее в статье обсуждается роль закона исключенного третьего в данном аргументе, а также той идеи, что истинность в модели должна быть моделью истины.

***Ключевые слова*:** отрицание, закон исключенного третьего, парадокс лжеца, телеологическая концепция истины, Майкл Даммет, Грэм Прист, истина, общезначимость, истинность в модели.

Импликация, эквивалентность и отрицание

А. Аврон
E-mail: `aa@cs.tau.ac.il`

Аннотация. Система $HCL_{\overset{\neg}{\leftrightarrow}}$ в языке $\{\neg, \leftrightarrow\}$ получена добавлением единственной схемы аксиом, не содержащей отрицания, к $HLL_{\overset{\neg}{\to}}$ (стандартной системе гильбертовского типа для мультипликативной линейной логики без пропозициональных констант), а также заменой $\to$ на $\leftrightarrow$. $HCL_{\overset{\neg}{\leftrightarrow}}$ слабо, но не сильно, непротиворечива и полна для $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$ ($\{\neg, \leftrightarrow\}$-фрагмента классической логики). Добавляя правило *Ex Falso* к $HCL_{\overset{\neg}{\leftrightarrow}}$, мы получаем систему, которая сильно непротиворечива и полна для $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$. Демонстрируется, что использование нового правила нельзя заменить схемами аксиом. Дана простая семантика, для которой сама $HCL_{\overset{\neg}{\leftrightarrow}}$ сильно непротиворечива и полна. Также показано, что $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$, логика, порожденная $HCL_{\overset{\neg}{\leftrightarrow}}$, имеет единственное не тривиальное собственное аксиоматическое расширение, что это расширение и $\mathbf{CL}_{\overset{\neg}{\leftrightarrow}}$ являются единственными собственными расширениями $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$ в языке $\{\neg, \leftrightarrow\}$, а также что только $\mathbf{L}_{HCL_{\overset{\neg}{\leftrightarrow}}}$ и ее единственное аксиоматическое расширение являются логиками в языке $\{\neg, \leftrightarrow\}$, которые содержат связку, обладающую свойством релевантной дедукции, но не эквивалентны какому-либо аксиоматическому расширению $\mathbf{R}_{\overset{\neg}{\to}}$ (интенсионального фрагмента релевантной логики $\mathbf{R}$). В заключение мы обсуждаем вопрос о том, может ли $\mathbf{L}_{HCL_{\vec{\leftrightarrow}}}$ рассматриваться как паранепротиворечивая логика.

***Ключевые слова*:** импликация, полуимпликация, отрицание, эквивалентность, бикондиционал, классическая пропозициональная логика, теоремы о дедукции, паранепротиворечивые логики.

Отрицание кондиционалов в естественном языке и мышлении

Г. Гомес
E-mail: ggomes@uenf.br

Аннотация. Внешнее отрицание кондиционалов возникает в предложениях, начинающихся с «Неверно, что если» или схожих фраз, и оно нередко в естественном языке. Один кондиционал может также отрицаться другим с таким же антецедентом и противоположным консеквентом. Чаще всего, когда отрицаемый кондиционал импликативен, отрицающий его концессивен и наоборот. Здесь я обосновываю, что в прагматике натурального языка «Если A, $\sim B$» влечет «$\sim$(если A, B)», но «$\sim$(если A, B)» не влечет «Если A, $\sim B$». «Если A, B» и «Если A, $\sim B$» отрицают друг друга, но контрарны, а не контрадикторны. Условия истинности, которые играют роль в человеческих рассуждениях, часто зависят не только от семантических, но также от прагматических факторов. Даются примеры, показывающие, что предложения, имеющие формы «$\sim$(если A, B)» и «Если A, $\sim B$», могут иметь разные прагматические условия истинности. Принцип кондиционального исключенного третьего, таким образом, не применим к использованию кондиционалов в натуральном языке. Три квадрата противоположностей дают представление упомянутых выше отношений.

***Ключевые слова*:** внешнее отрицание, импликативные кондиционалы, концессивные кондиционалы, прагматические условия истинности, тезис Боэция, кондициональное исключенное третье, контрарность, контрадикторность, квадраты противоречия.

Отрицание по умолчанию как явное отрицание плюс обновление

Р. Кале
E-mail: kahle@mat.uc.pt

Аннотация. Мы утверждаем, что в рамках семантики стабильной модели отрицание по умолчанию может читаться как явное отрицание с обновлением. Мы показываем, что динамическое логическое программирование, которое основано на отрицании по умолчанию, можно трактовать в варианте обновлений с только явным отрицанием. В качестве следствия мы получаем простое описание отрицания по умолчанию в обобщенном и нормальном логическом программировании, при котором первоначально отрицаемые литералы обновляются. Эти результаты обсуждаются в связи с пониманием отрицания в логическом программировании.

***Ключевые слова*:** отрицание по умолчанию, явное отрицание, логическое программирование, обновление.

Равенство и отделенность в биинтуиционистской логике

П. Маффециоли
E-mail: paolo.maffezioli@univr.it
Л. Транчини
E-mail: luca.tranchini@uni-tuebingen.de

Аннотация. В данной статье мы утверждаем, что симметричная картина отношений между равенством и отделенностью может быть достигнута путем рассмотрения этих понятий на фоне биинтуиционистской логики вместо обычной интуиционистской логики. В частности, мы показываем, что, поскольку интуиционистское отрицание отношения отделенности является равенством, дуально-интуиционистское коотрицание равенства является отношением отделенности. В то же время, поскольку интуиционистское отрицание равенства не является отделенностью, коинтуиционистское отрицание отделенности не является равенством.

***Ключевые слова*:** биинтуиционистская логика, равенство, отделенность.

Отрицание и импликация в квази-нельсоновой логике

Т. Нассименту
E-mail: thiagnascsilva@gmail.com

У. Ривеччио
E-mail: umberto.rivieccio@ufrn.br

Аннотация. Квази-нельсонова логика – это недавно введенное обобщение конструктивной логики Нельсона с сильным отрицанием для неинволютивных условий. В настоящей статье мы аксиоматизируем импликативно-негативный фрагмент квази-нельсоновой логики (QNI-логики), который представляет собой, в определенном смысле, алгебраизируемое ядро квази-нельсоновой логики. Мы представляем конечное исчисление гильбертовского типа для QNI-логики, демонстрируя полноту и алгебраизируемость относительно многообразия QNI-алгебр. Элементы этого класса, также введенного и исследованного в недавней статье, суть в точности негативно-импликативные подредукты квази-нельсоновых алгебр. Опираясь на наш результат о полноте, мы также демонстрируем, как негативно-импликативные фрагменты интуиционистской логики и конструктивной логики Нельсона могут быть получены в форме расширений QNI-логики схемами аксиом.

***Ключевые слова*:** конструктивная логика Нельсона с сильным отрицанием, квази-нельсоновы алгебры, импликативно-негативные подредукты, QNI-алгебра, квази-нельсонова логика, алгебраизируемые логики.

Конфликтоустойчивая логика явных свидетельств

Т. Штудер
E-mail: thomas.studer@inf.unibe.ch

Аннотация. Стандартная эпистемическая модальная логика неспособна адекватно справиться с парадоксом Фраушигер – Реннера в квантовой физике. Мы вводим новую логику обоснования CTJ, в которой этот парадокс можно формализовать, не приходя к противоречию. В то же время, CTJ достаточно сильна, чтобы моделировать традиционные эпистемические рассуждения. Наша логика устойчива к наличию двух свидетельств, таких что одно обосновывает высказывание, а другое обосновывает отрицание этого высказывания. Однако, наша логика запрещает одному свидетельству обосновывать одновременно высказывание и его отрицание. Мы представляем синтаксис и семантику для CTJ и описываем ее основные свойства. Затем мы приводим пример эпистемического рассуждения в CTJ, который иллюстрирует взаимодействие различных принципов CTJ между собой. В продолжение, мы приводим формализацию мысленного эксперимента Фраушигер–Реннера и обсуждаем его в детялях. Далее, мы добавляем аксиому доверия к CTJ и вновь рассматриваем эпистемические рассуждения и парадокс в этих расширенных условиях.

***Ключевые слова*:** противоречивые свидетельства, логика обоснования, эпистемическая логика, парадокс Фраушигер–Реннера, квантовая физика.

Информация для авторов

• Журнал «Логические исследования» принимает к публикации рукописи, содержащие изложение оригинальных результатов из различных областей современной логики, ранее не публиковавшиеся и не представленные к публикации в других изданиях. (Рубрики см. на сайте журнала `http://logicalinvestigations.ru`)

• Все статьи, планируемые к публикации в журнале «Логические исследования», проходят процедуру анонимного рецензирования.

• Решение о публикации текста принимается главным редактором с учетом мнения редколлегии и оценки рецензентов. Решение о публикации принимается в течение двух месяцев с момента предоставления рукописи.

• Плата за опубликование рукописей не взимается.

• Рукопись должна быть представлена в электронном виде и оформлена в формате LaTeX 2_ε (по согласованию с редколлегией – в MS Word с обязательным предоставлением pdf-файла).

• При подготовке рукописи в LaTeX 2_ε необходимо использовать стилевой класс `LIarticle.cls` и шаблон `LI_template.tex` (если рукопись на русском языке) или `LI_template_eng.tex` (если рукопись на английском языке), которые размещены в правилах предоставления рукописей на сайте `http://logicalinvestigations.ru`. Здесь же размещены подробные рекомендации по подготовке рукописи.

• Объем рукописи не должен превышать 25 стр. (60 тыс. знаков), включая ссылки, примечания, списоки литературы, аннотации (на русском и английском языках).

Статьи следует направлять в электронном виде в редакцию через сайт
`http://logicalinvestigations.ru`

Information for authors

• *Logical Investigations* accepts for submission papers containing original results in all areas of logic. The papers should not have been published or simultaneously submitted to another publication. (Sections of the journal: `http://logicalinvestigations.ru`)

• All materials published in *Logical Investigations* undergo peer review process.

• The Editor in Chief makes the decision which of the submitted articles should be published, with due account for opinions of the Editorial Board and the reviewers. The decision is made within two months since the date of submission of the manuscript.

• Authors are not charged for the publication.

• Papers should be submitted electronically in the $\LaTeX 2_\varepsilon$ format (special permission of the editorial board is needed for submissions to be made in the MS Word format).

• While typesetting a paper, the class file `LIarticle.cls` and the template `LI_template_eng.tex` should be used; both files can be accessed at `http://logicalinvestigations.ru`. Here you can also find detailed guidelines for preparing your paper.

• Papers should not exceed 25 pages in the above mentioned format (including the notes, the bibliography, the abstract).

Submissions should be sent in electronic form through the website:
`http://logicalinvestigations.ru`

Научно-теоретический журнал

Логические исследования / Logical Investigations
2021. Том 27. Номер 1

Учредитель и издатель: Федеральное государственное бюджетное учреждение науки Институт философии Российской академии наук

Свидетельство о регистрации СМИ: *ПИ № ФС77-61228 от 03.04.2015 г.*

Главный редактор: *В.И. Шалак*

Ответственный секретарь: *Н.Е. Томова*

Технический редактор: *Е.А. Морозова*

Художник: *Н.Н. Попов*, *С.Ю. Растегина*

Подписано в печать с оригинал-макета 27.05.2021.
Формат 70x100 1/16. Печать офсетная. Гарнитура Computer Modern.
Для набора греческого текста исользован пакет Teubner.
Усл. печ. л. 12,57. Уч.-изд. л. 7,7. Тираж 1 000 экз. Заказ № 10.

Оригинал-макет изготовлен в Институте философии РАН

Разработка LaTeX-класса стилевого оформления оригинал-макета: *А.Е. Коньков*

Компьютерная верстка: *Н.Е. Томова*

Отпечатано в ЦОП Института философии РАН
109240, г. Москва, ул. Гончарная, д. 12, стр. 1

Информацию о журнале «Логические исследования» см. на сайте:
`http://logicalinvestigations.ru`