Negative translations not intuitionistically equivalent to the usual ones*

Jaime Gaspar[†]

1 November 2010

Abstract

We refute the conjecture that all negative translations are intuitionistically equivalent by giving two counterexamples. Then we characterise the negative translations intuitionistically equivalent to the usual ones.

1 Introduction

Informally speaking, classical logic CL is the usual logic in mathematics, and intuitionistic logic IL is obtained from classical logic by omitting:

- reductio ad absurdum $\dfrac{\begin{array}{c}\neg A\\ \vdots\\ \bot\end{array}}{A}$;
- law of excluded middle $A \vee \neg A$;
- law of double negation $\neg\neg A \to A$.

In this sense, IL is a weakening of CL, that is IL proves less theorems than CL.

At first sight it seems that IL is just poorer than CL. However, there is a gain in moving from CL to IL: the theorems of IL have nicer properties. The main properties gained are

- disjunction property: if $\mathsf{IL} \vdash A \vee B$, then $\mathsf{IL} \vdash A$ or $\mathsf{IL} \vdash B$ (where A and B are sentences);
- existence property: if $\mathsf{IL} \vdash \exists x A(x)$, then $\mathsf{IL} \vdash A(t)$ for some term t (where $\exists x A$ is a sentence).

*Keywords: negative translation, classical logic, intuitionistic logic, minimal logic, negative fragment.
2000 Mathematics Subject Classification: 03F25.
[†]Arbeitsgruppe Logik, Fachbereich Mathematik, Technische Universität Darmstadt. Schlossgartenstrasse 7, 64289 Darmstadt, Germany. `mail@jaimegaspar.com`, `www.jaimegaspar.com`.
I'm grateful to Hajime Ishihara, Ulrich Kohlenbach and Benno van den Berg. This work was financially supported by the Portuguese Fundação para a Ciência e a Tecnologia, grant SFRH/BD/36358/2007.

Arguably, these two properties are the key criteria to say that a logic is constructive.

On the one hand IL is weaker than CL, on the other hand IL is constructive while CL is not. Given these differences, it is surprising that CL can be faithfully embedded in IL by the so-called negative translations into IL. Negative translations into IL are functions N, mapping a formula A to a formula A^N, that:

- embed CL into IL, that is $\mathsf{CL} \vdash A \;\Rightarrow\; \mathsf{IL} \vdash A^N$;

- are faithful, that is $\mathsf{CL} \vdash A^N \leftrightarrow A$.

The image of the usual negative translations is (essentially) the negative fragment NF, that is the set of all formulas without $\vee$ and $\exists$ and whose atomic formulas are all negated. So NF is a faithful copy of CL inside IL. This is pictured in figure 1.

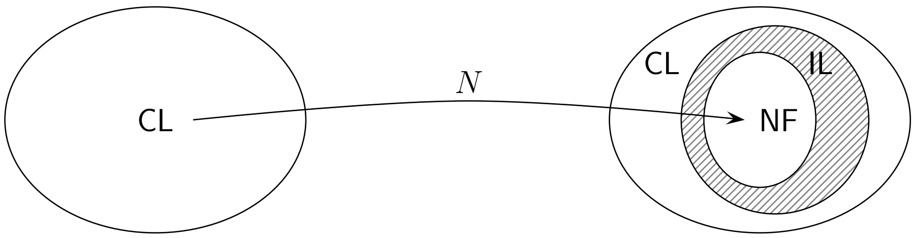

Figure 1: negative translation N into IL embedding CL in the fragment NF of IL.

There are four negative translations into IL usually found in the literature (and recently two new ones were presented [5]). They are introduced in table 1 and defined (by induction on the structure of formulas) in table 2. All these negative translations into IL are equivalent in IL: given any two of them, say M and N, we have $\mathsf{IL} \vdash A^M \leftrightarrow A^N$. This fact leads to the following conjecture that seems to be almost folklore:

> if we rigorously define the notion of a negative translation into IL, then we should be able to prove that all negative translations are equivalent in IL.

Curiously, this conjecture apparently has never been studied before. In this article we study it, reaching the following conclusions.

- The conjecture is false and we give two counterexamples.

- The usual negative translations into IL are characterised by the following two equivalent conditions:

  - to translate into NF in IL, that is A^N is equivalent in IL to a formula in NF;

  - to act as the identity on NF in IL, that is $\mathsf{IL} \vdash A^N \leftrightarrow A$ for all $A \in \mathsf{NF}$.

Year	Name	Symbol	Note
1925	Kolmogorov [14]	Ko	
1933	Gödel-Gentzen	G	One variant by Gödel [11] and another one independently by Gentzen [8]
1951	Kuroda [16]	Ku	
1998	Krivine [15]	Kr	Maybe better attributed to Streicher and Reus [17]

Table 1: the four usual negative translations.

$P^{Ko} :\equiv \neg\neg P \quad (P \not\equiv \bot \text{ atomic})$	$P^{G} :\equiv \neg\neg P \quad (P \not\equiv \bot \text{ atomic})$
$\bot^{Ko} :\equiv \bot$	$\bot^{G} :\equiv \bot$
$(A \wedge B)^{Ko} :\equiv \neg\neg(A^{Ko} \wedge B^{Ko})$	$(A \wedge B)^{G} :\equiv A^{G} \wedge B^{G}$
$(A \vee B)^{Ko} :\equiv \neg\neg(A^{Ko} \vee B^{Ko})$	$(A \vee B)^{G} :\equiv \neg(\neg A^{G} \wedge \neg B^{G})$
$(A \to B)^{Ko} :\equiv \neg\neg(A^{Ko} \to B^{Ko})$	$(A \to B)^{G} :\equiv A^{G} \to B^{G}$
$(\forall x A)^{Ko} :\equiv \neg\neg\forall x A^{Ko}$	$(\forall x A)^{G} :\equiv \forall x A^{G}$
$(\exists x A)^{Ko} :\equiv \neg\neg\exists x A^{Ko}$	$(\exists x A)^{G} :\equiv \neg\forall x \neg A^{G}$
$A^{Ku} :\equiv \neg\neg A_{Ku}$	$A^{Kr} :\equiv \neg A_{Kr}$
$P_{Ku} :\equiv P \quad (P \text{ atomic})$	$P_{Kr} :\equiv \neg P \quad (P \text{ atomic})$
$(A \wedge B)_{Ku} :\equiv A_{Ku} \wedge B_{Ku}$	$(A \wedge B)_{Kr} :\equiv A_{Kr} \vee B_{Kr}$
$(A \vee B)_{Ku} :\equiv A_{Ku} \vee B_{Ku}$	$(A \vee B)_{Kr} :\equiv A_{Kr} \wedge B_{Kr}$
$(A \to B)_{Ku} :\equiv A_{Ku} \to B_{Ku}$	$(A \to B)_{Kr} :\equiv \neg A_{Kr} \wedge B_{Kr}$
$(\forall x A)_{Ku} :\equiv \forall x \neg\neg A_{Ku}$	$(\forall x A)_{Kr} :\equiv \exists x A_{Kr}$
$(\exists x A)_{Ku} :\equiv \exists x A_{Ku}$	$(\exists x A)_{Kr} :\equiv \neg\exists x \neg A_{Kr}$

Table 2: definition of the four usual negative translations.

2 Notions

In the rest of this article, CL denotes the pure first order classical predicate logic based on $\perp$, $\wedge$, $\vee$, $\rightarrow$, $\forall$ and $\exists$ (where $\neg A :\equiv A \rightarrow \perp$, $A \leftrightarrow B :\equiv (A \rightarrow B) \wedge (B \rightarrow A)$ and $\equiv$ denotes syntactical equality) and IL and ML denote its intuitionistic and minimal counterparts, respectively. All formulas considered belong to the common language of CL, IL and ML. To save parentheses we adopt the convention that $\forall$ and $\exists$ bind stronger than $\wedge$ and $\vee$, which in turn bind stronger than $\rightarrow$.

Let us start by motivating our definition of negative translation.

The main feature of any negative translation N into IL is embedding CL into IL in the sense of $\mathsf{CL} \vdash A \;\Rightarrow\; \mathsf{IL} \vdash A^N$. We can be even more ambitious and ask for (1) $\mathsf{CL} + \Gamma \vdash A \;\Rightarrow\; \mathsf{IL} + \Gamma^N \vdash A^N$ where Γ is any set of formulas and $\Gamma^N := \{A^N : A \in \Gamma\}$.

But embedding CL into IL alone does not seem to capture our intuitive notion of a negative translation. For example, it includes the trivial example $A^N :\equiv \neg\perp$. The problem with this example is that the meaning of A^N is unrelated to the meaning of A. So require that a negative translation do not change the meaning of formulas, that is (2) $A^N \leftrightarrow A$. This equivalence must not be taken in IL or ML, otherwise from (1) and (2) we would get $\mathsf{CL} = \mathsf{IL}$. So we take the equivalence in CL, that is $\mathsf{CL} \vdash A \leftrightarrow A^N$.

Definition 1. Let N be a function mapping each formula A to a formula A^N.

- The following condition is called *soundness theorem into* IL (ML) *of* N: for all formulas A and for all sets Γ of possibly open formulas, we have the implication $\mathsf{CL} + \Gamma \vdash A \;\Rightarrow\; \mathsf{IL} + \Gamma^N \vdash A^N$ (respectively, $\mathsf{CL} + \Gamma \vdash A \;\Rightarrow\; \mathsf{ML} + \Gamma^N \vdash A^N$).

- The following condition is called *characterisation theorem of* N: for all formulas A we have $\mathsf{CL} \vdash A^N \leftrightarrow A$.

- We say that N is a *negative translation into* IL (ML) if and only if both the soundness theorem into IL (respectively, ML) of N and the characterisation theorem of N hold.

Remark 2. The soundness theorem into ML of N implies the soundness theorem into IL of N. So a negative translation into ML is in particular a negative translation into IL.

The conjecture that concerns us mentions equivalence in IL. For definiteness, we write down exactly what we mean by this.

Definition 3. We say that two negative translations M and N are *equivalent in* IL (ML) if and only if for all formulas A we have $\mathsf{IL} \vdash A^M \leftrightarrow A^N$ (respectively, $\mathsf{ML} \vdash A^M \leftrightarrow A^N$).

Later on we will see that what characterises the usual negative translations into IL are two properties related to NF. Again for definiteness we write down the definition of NF and of the two properties.

Definition 4. The *negative fragment* NF is the set of formulas inductively generated by:

- $\perp \in \mathsf{NF}$;
- if P is an atomic formula, then $\neg P \in \mathsf{NF}$;
- if $A, B \in \mathsf{NF}$, then $A \wedge B, A \to B, \forall x A \in \mathsf{NF}$.

Definition 5. Let N be a negative translation into IL.

- We say that N *translates into* NF *in* IL (ML) if and only if for all formulas A there exists a $B \in \mathsf{NF}$ such that $\mathsf{IL} \vdash A^N \leftrightarrow B$ (respectively, $\mathsf{ML} \vdash A^N \leftrightarrow B$).
- We say that N *acts as the identity on* NF *in* IL (ML) if and only if for all $A \in \mathsf{NF}$ we have $\mathsf{IL} \vdash A^N \leftrightarrow A$ (respectively, $\mathsf{ML} \vdash A^N \leftrightarrow A$).

3 Gödel-Gentzen negative translation

We will choose the Gödel-Gentzen negative translation G as a representative of the usual negative translations into IL, so let us take a closer look at it.

We start by motivating the definition of G. It is known from proof theory that CL is conservative over ML with respect to NF, that is (1) for all $A \in \mathsf{NF}$ we have the implication $\mathsf{CL} \vdash A \;\Rightarrow\; \mathsf{ML} \vdash A$. This suggests us that one way of constructing a negative translation into ML is to rewrite each formula A as a formula $A^N \in \mathsf{NF}$. By rewriting we mean that A^N still has the same meaning as A in the sense of (2) $\mathsf{CL} \vdash A^N \leftrightarrow A$. Then (1) would give us the soundness theorem into ML of N (almost, because there is no Γ) and (2) would give us the characterisation theorem of N. The natural way of rewriting a formula A as a classically equivalent formula $A^N \in \mathsf{NF}$ (that is having all atomic formulas $P \not\equiv \perp$ negated and using only $\perp$, $\wedge$, $\to$ and $\forall$) is:

- rewrite atomic formulas $P \not\equiv \perp$ as $\neg\neg P$;
- rewrite $A \vee B$ as $\neg(\neg A \wedge \neg B)$;
- rewrite $\exists x A$ as $\neg\forall x \neg A$;
- there's no need to rewrite $\perp$, $A \wedge B$, $A \to B$ and $\forall x A$.

If we formalise these rewritings as a definition of N by induction on the structure of formulas, then we get exactly G. As a "tagline" we can say: A^G is the natural rewriting of A into NF.

Incidentally, Gödel's and Gentzen's negative translations differ only in the way they translate $A \to B$: Gödel translates to $\neg(A^G \wedge \neg B^G)$ while Gentzen translates to $A^G \to B^G$. By the above discussion, we find Gentzen's variant more natural and so we adopt it.

Now we turn to the main properties of G. We can prove that G:

- is a negative translation into ML;
- translates into NF in ML;

- acts as the identity on NF in ML.

We can even prove strengthenings of the second and third properties above:

- for all formulas A we have $A^G \in \mathsf{NF}$;
- for all formulas $A \in \mathsf{NF}$ we have $A^G \equiv A$ (modulo identifying $\neg\neg\neg P$ with $\neg P$ for atomic formulas P).

These two strengthenings are specific of G: they do not hold for Ko, Ku and Kr.

To finish this section we discuss G as a representative of the usual negative translations into IL. We can prove that Ko, G, Ku and Kr are equivalent in IL. So any of them can be taken as a representative of the usual negative translations into IL. We choose to take G as a representative due to nice syntactical properties of G like $(A \leftrightarrow B)^G \equiv A^G \leftrightarrow B^G$ and the two strengthenings above. These properties allow us to work many times with syntactical equalities instead of equivalences, thus avoiding the question of where (CL, IL or ML) the equivalences are provable.

4 Two negative translations not intuitionistically equivalent to the usual ones

Before we present our two counterexamples to the conjecture, let us draw a scale to roughly measure how provable or refutable a formula F is. This scale will be useful to picture our main theorem about the counterexamples. We draw the scale following this set of instructions.

- We plot along an axis all possible pairs of combinations of

  $$\mathsf{CL} \vdash F, \qquad \mathsf{CL} \nvdash F \text{ and } \mathsf{CL} \nvdash \neg F, \qquad \mathsf{CL} \vdash \neg F$$

  with

  $$\mathsf{IL} \vdash F, \qquad \mathsf{IL} \nvdash F \text{ and } \mathsf{IL} \nvdash \neg F, \qquad \mathsf{IL} \vdash \neg F.$$

- Actually, we do not plot impossible pairs (for example, "$\mathsf{CL} \vdash F$ and $\mathsf{IL} \vdash \neg F$") and redundant entries in pairs (for example, the entry "$\mathsf{CL} \vdash F$" in the pair "$\mathsf{CL} \vdash F$ and $\mathsf{IL} \vdash F$").
- The plotting is ordered from provability of F on the left to refutability of F on the right (for example, "$\mathsf{IL} \vdash F$" stands on the left of "$\mathsf{CL} \vdash F$ and $\mathsf{IL} \nvdash F$" because "$\mathsf{IL} \vdash F$" states a stronger form of provability).

The resulting scale is pictured in figure 2.

Now we present our two counterexamples.

- The first counterexample N_1 is a weakening of G obtained by weakening A^G to $A^G \vee F$ (for suitable formulas F).
- The second counterexample N_2 is a variant of G obtained by making $\perp$ in A^G "less false" in the sense of replacing $\perp$ by F in A^G, that is $A^G[F/\perp]$ (again, for suitable F).

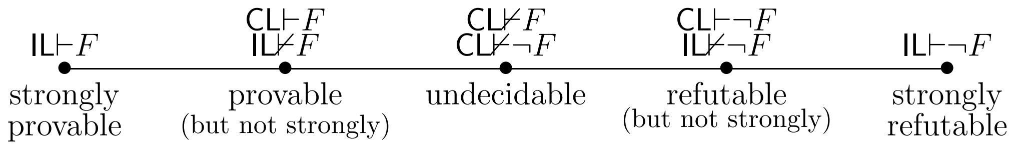

Figure 2: scale of provability-refutability.

Definition 6. Fix a formula F. We define two functions N_1 and N_2, mapping formulas to formulas, by

- $A^{N_1} :\equiv A^G \vee F$;
- $A^{N_2} :\equiv A^G[F/\bot]$.

Since N_1 and N_2 depend on the chosen F, in rigour we should write something like $A^{N_1(F)}$ and $A^{N_2(F)}$, but we avoid this cumbersome notation.

We found N_2 in an article by Ishihara [13] and in a book chapter by Coquand [2, section 2.3]. Maybe Ishihara drew inspiration from an article by Flagg and Friedman [6] where a similar translation appears. It is even possible that N_2 is folklore.

For our two counterexamples to work, we need the formula F to be classically refutable but intuitionistically acceptable. In the next lemma we prove that there are such formulas F.

Lemma 7.

1. *There exists a formula F such that* $\mathsf{CL} \vdash \neg F$ *but* $\mathsf{IL} \nvdash \neg F$.
2. *Any such formula F is not equivalent in* IL *to a formula in* NF.

Proof.

1. Let P be an unary predicate symbol. We are going to prove that $F \equiv \neg\forall x P(x) \wedge \forall x \neg\neg P(x)$ is such that $\mathsf{CL} \vdash \neg F$ but $\mathsf{IL} \nvdash \neg F$. Since $\mathsf{CL} \vdash \neg F$ is obvious, we move on to prove $\mathsf{IL} \nvdash \neg F$ by showing that the Kripke model $\mathcal{K}$ from figure 3 forces F.
   - $\mathcal{K}$ forces $\neg\forall x P(x)$ because no node forces $\forall x P(x)$.
   - $\mathcal{K}$ forces $\forall x \neg\neg P(x)$ because every node k forces $\neg\neg P(d)$ for all d in its domain $\{0, \ldots, k\}$ since the node $k+1$ forces $P(d)$.
2. If F were equivalent in IL to a formula in NF, then from $\mathsf{CL} \vdash \neg F$ and the fact that CL is conservative over IL with respect to NF we would get $\mathsf{IL} \vdash \neg F$, contradicting point 1. □

Now we prove our main theorem giving two counterexamples to the conjecture: N_1 and N_2 are negative translations into IL (even into ML) not equivalent in IL to the usual negative translations into IL (for suitable formulas F). The claims of this theorem are summarised in figure 4.

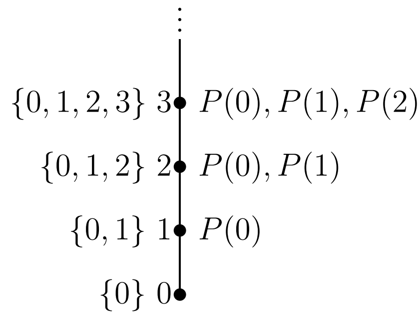

Figure 3: a Kripke model $\mathcal{K}$ forcing $\neg\forall x P(x) \wedge \forall x \neg\neg P(x)$.

Theorem 8. *The functions N_1 and N_2:*

1. *have a soundness theorem into* ML *for all formulas F;*
2. *have a characterisation theorem if and only if* $\mathsf{CL} \vdash \neg F$*;*
3. *are equivalent in* IL *to G if and only if* $\mathsf{IL} \vdash \neg F$.

So, if $\mathsf{CL} \vdash \neg F$ *but* $\mathsf{IL} \nvdash \neg F$, *then N_1 and N_2 are negative translations into* ML *not equivalent in* IL *to G.*

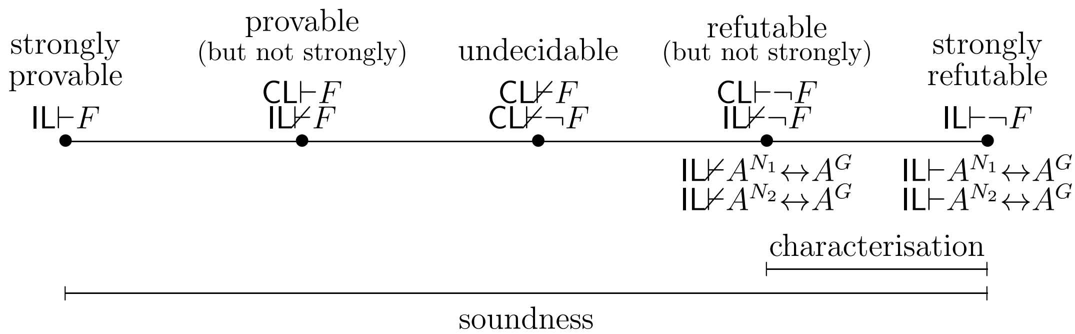

Figure 4: theorem 8 on the scale of provability-refutability.

Proof.

1. Consider an arbitrary formula F.

   First let us consider the case of N_1. By direct proof, consider an arbitrary set of formulas Γ and an arbitrary formula A, assume $\mathsf{CL} + \Gamma \vdash A$ and let us prove $\mathsf{ML} + \Gamma^{N_1} \vdash A^{N_1}$. Since a proof in CL of A uses only finitely many formulas $A_1, \ldots, A_n$ from Γ, then $\mathsf{CL} + A_1 + \cdots + A_n \vdash A$. By the soundness theorem into ML of G we get $\mathsf{ML} + A_1^G + \cdots + A_n^G \vdash A^G$ (where A_i^G abbreviates $(A_i)^G$), that is (1) $\mathsf{ML} \vdash A_1^G \wedge \cdots \wedge A_n^G \to A^G$ by the deduction theorem of ML.

   Let us show (2) $\mathsf{ML} + A_1^G \vee F + \cdots + A_n^G \vee F \vdash A^G \vee F$. We argue inside ML. Assume $A_1^G \vee F, \ldots, A_n^G \vee F$. Each $A_i^G \vee F$ gives us two cases: the case of A_i^G and the case of F.

- If for some $A_i^G \vee F$ we have the case F, then trivially $A^G \vee F$.
- Otherwise in all $A_i^G \vee F$ we have the case of A_i^G, so we have $A_1^G \wedge \cdots \wedge A_n^G$, thus A^G by (1), therefore trivially $A^G \vee F$.

So we have (2) as we wanted. This argument is illustrated for $n = 2$ in figure 5.

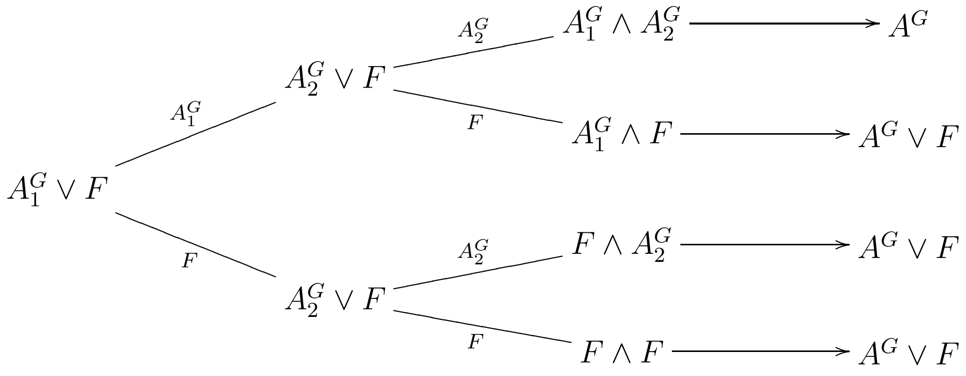

Figure 5: argument of $\mathsf{ML} + A_1^G \vee F + \cdots + A_n^G \vee F \vdash A^G \vee F$ for $n = 2$.

But (2) is $\mathsf{ML} + A^{N_1} + \cdots + A^{N_1} \vdash A^{N_1}$, so we get $\mathsf{ML} + \Gamma^{N_1} \vdash A^{N_1}$, as we wanted.

Now let us consider the case of N_2. By direct proof, consider an arbitrary set of formulas Γ and an arbitrary formula A, assume $\mathsf{CL} + \Gamma \vdash A$ and let us prove $\mathsf{ML} + \Gamma^{N_2} \vdash A^{N_2}$. By the soundness theorem into ML of G we get $\mathsf{ML} + A_1^G + \cdots + A_n^G \vdash A^G$. Since $\perp$ is treated as an arbitrary propositional letter in ML, we can replace $\perp$ by F getting $\mathsf{ML} + A_1^G[F/\perp] + \cdots + A_n^G[F/\perp] \vdash A^G[F/\perp]$, that is $\mathsf{ML} + A_1^{N_2} + \cdots + A_n^{N_2} \vdash A^{N_2}$, as we wanted.

2. First let us consider the case of N_1.

(⇒) By direct proof, assume that N_1 has a characterisation theorem and let us prove $\mathsf{CL} \vdash \neg F$. By the characterisation theorem of N_1 we have $\mathsf{CL} \vdash \perp^{N_1} \leftrightarrow \perp$ where $\perp^{N_1} \equiv \perp \vee F$, so $\mathsf{CL} \vdash \neg F$, as we wanted.

(⇐) By direct proof, assume $\mathsf{CL} \vdash \neg F$, consider an arbitrary formula A and let us prove $\mathsf{CL} \vdash A^{N_1} \leftrightarrow A$. By the characterisation theorem of G we have $\mathsf{CL} \vdash A^G \leftrightarrow A$. Since $\mathsf{CL} \vdash \neg F$ by assumption, it makes no difference in CL to replace A^G by $A^G \vee F$. So $\mathsf{CL} \vdash A^G \vee F \leftrightarrow A$, that is $\mathsf{CL} \vdash A^{N_1} \leftrightarrow A$, as we wanted.

Now let us consider the case of N_2.

(⇒) Analogous to the case of N_1.

(⇐) By direct proof, assume $\mathsf{CL} \vdash \neg F$, consider an arbitrary formula A and let us prove $\mathsf{CL} \vdash A^{N_2} \leftrightarrow A$. By the characterisation theorem of G we have $\mathsf{CL} \vdash A^G \leftrightarrow A$. Since $\mathsf{CL} \vdash \neg F$ by assumption, it makes no difference in CL to replace $\perp$ by F. So $\mathsf{CL} \vdash A^G[F/\perp] \leftrightarrow A$, that is $\mathsf{CL} \vdash A^{N_2} \leftrightarrow A$, as we wanted.

3. First let us consider the case of N_1.

   (⇒) By direct proof, assume that N_1 and G are equivalent in IL and let us prove $\mathsf{IL} \vdash \neg F$. By the assumption we have $\mathsf{IL} \vdash \bot^{N_1} \leftrightarrow \bot^G$ where $\bot^{N_1} \equiv \bot \vee F$ and $\bot^G \equiv \bot$. So $\mathsf{IL} \vdash \neg F$, as we wanted.

   (⇐) By direct proof, assume $\mathsf{IL} \vdash \neg F$, take an arbitrary formula A and let us prove $\mathsf{IL} \vdash A^{N_1} \leftrightarrow A^G$. By the assumption it makes no difference in IL to replace A^G by $A^G \vee F$. So $\mathsf{IL} \vdash A^G \vee F \leftrightarrow A^G$, that is $\mathsf{IL} \vdash A^{N_1} \leftrightarrow A^G$, as we wanted.

   Now let us consider the case of N_2.

   (⇒) Analogously to the case of N_1.

   (⇐) By direct proof, assume $\mathsf{IL} \vdash \neg F$, take an arbitrary formula A and let us prove $\mathsf{IL} \vdash A^{N_1} \leftrightarrow A^G$. By the assumption it makes no difference in IL to replace $\bot$ by F. So $\mathsf{IL} \vdash A^G[F/\bot] \leftrightarrow A^G$, that is $\mathsf{IL} \vdash A^{N_2} \leftrightarrow A^G$, as we wanted. □

We saw in theorem 8 that N_1 and N_2 are two counterexamples to the conjecture (for suitable F). Now in proposition 9 we clarify that these two counterexamples are different (for the same suitable F).

Proposition 9. *If* $\mathsf{CL} \vdash \neg F$ *but* $\mathsf{IL} \nvdash \neg F$*, then* N_1 *and* N_2 *are not equivalent in* IL.

Proof. By direct proof, assume that $\mathsf{CL} \vdash \neg F$ but $\mathsf{IL} \nvdash \neg F$ and let us prove $\mathsf{IL} \nvdash A^{N_1} \leftrightarrow A^{N_2}$. We start by making two observations about Kripke models.

1. There exists a Kripke model $\mathcal{K}$, with a bottom node, that forces $\neg F$.

   Let us prove this claim. Since $\mathsf{CL} \vdash \neg F$ by assumption, any classical model forces $\neg F$. Regarding a classical model as a Kripke model with only one node, we have a Kripke model, with a bottom node, forcing $\neg F$, as we wanted.

   For example, for the $F \equiv \neg\forall x P(x) \vee \forall x \neg\neg P(x)$ used in the proof of lemma 7, we can take $\mathcal{K}$ to be the Kripke model of figure 6.

$$\{0\} \bullet P(0)$$

Figure 6: a Kripke model $\mathcal{K}$ forcing $\neg F$ where $F \equiv \neg\forall x P(x) \vee \forall x \neg\neg P(x)$.

2. There exists a Kripke model $\mathcal{L}$, with a bottom node, that forces F.

   Let us prove this claim. Since $\mathsf{IL} \nvdash \neg F$ by assumption, there exists a Kripke model $\mathcal{L}'$ that does not force $\neg F$, that is some node n' of $\mathcal{L}'$ does not force $\neg F$. Then there exists a node n above or equal to n' that forces F. By restricting $\mathcal{L}'$ to all the nodes above or equal to n we get a Kripke model $\mathcal{L}$, with bottom node n, that forces F, as we wanted.

   For example, for the $F \equiv \neg\forall x P(x) \vee \forall x \neg\neg P(x)$ used in the proof of lemma 7, we can take $\mathcal{L}$ to be the Kripke model of figure 3.

Now let us return to our goal: $\mathsf{IL} \nvdash A^{N_1} \leftrightarrow A^{N_2}$. Consider a fresh nullary predicate $Q \not\equiv \bot$. Since Q is fresh and $Q \not\equiv \bot$,

- $\mathcal{L}$ forces $\neg Q$;
- we can force Q in $\mathcal{K}$;
- forcing Q in $\mathcal{K}$ will not collide with $\mathcal{K}$ forcing $\neg F$.

We will show $\mathsf{IL} \nvdash Q^{N_2} \to Q^{N_1}$, where $Q^{N_2} \equiv (Q \to F) \to F$ and $Q^{N_1} \equiv \neg\neg Q \vee F$, by presenting a Kripke model not forcing $(*)$ $((Q \to F) \to F) \to \neg\neg Q \vee F$.

The base nodes of $\mathcal{K}$ and $\mathcal{L}$ have (by definition of Kripke model) non empty domains. We can assume (renaming elements if necessary) that those domains share a common element d. Consider the Kripke model $\mathcal{M}$ from figure 7 obtained by:

- connecting a fresh bottom node 0, with domain $\{d\}$, to the bottom nodes of $\mathcal{K}$ and $\mathcal{L}$;
- for every node n of $\mathcal{M}$, forcing Q in n if and only if n forces $\neg F$; or equivalently, forcing Q in $\mathcal{K}$ but not in $\mathcal{L}$ and 0.

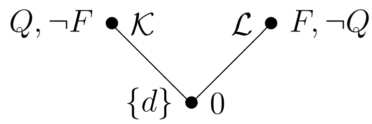

Figure 7: a Kripke model $\mathcal{M}$ not forcing $((Q \to F) \to F) \to \neg\neg Q \vee F$.

Note that $\mathcal{M}$ is well-defined because:

- the domains of $\mathcal{M}$ are monotone since $\{d\}$ is contained in the domains of $\mathcal{K}$ and $\mathcal{L}$;
- the forcing relation in $\mathcal{M}$ is monotone since Q is forced only in the entire $\mathcal{K}$.

Now we argue that $\mathcal{M}$ does not force $(*)$.

- The node 0 does not force $\neg\neg Q$ because $\mathcal{L}$ forces $\neg Q$.
- The node 0 does not force F because $\mathcal{K}$ forces $\neg F$.
- Let us show that the node 0 forces $(Q \to F) \to F$, that is any node n does not force $Q \to F$ or forces F. We consider the following three cases.
  - If n is in $\mathcal{K}$, then n does not force $Q \to F$ because $\mathcal{K}$ forces Q (by construction of $\mathcal{M}$) and $\neg F$.
  - If n is in $\mathcal{L}$, then n forces F because $\mathcal{L}$ forces F.
  - If n is 0, then n does not force $Q \to F$, otherwise $\mathcal{K}$ would force $Q \to F$ and we already saw that this is false.

We conclude that the node 0 does not force $(*)$, as we wanted. $\square$

As a curiosity, let us see that we have the factorisations $N_2 = FD \circ G$ and $N_2 = FD' \circ G$ of N_2 in terms of Friedman-Dragalin translation FD [4, 7] (better known as Friedman's A-translation), its refinement FD' [1] and G. The translation FD was used by Friedman and Dragalin to prove that certain intuitionistic theories IT are closed under Markov rule in the sense of $\mathsf{IT} \vdash \neg\neg\exists x P(x) \;\Rightarrow\; \mathsf{IT} \vdash \exists x P(x)$ where $P(x)$ is an atomic formula.

Definition 10. Fix a formula F.

- *Friedman-Dragalin translation* FD maps each formula A to the formula A^{FD} obtained from A by simultaneously replacing in A:
  - $\perp$ by F;
  - all atomic subformulas $P \not\equiv \perp$ by $P \vee F$.
- The *refined Friedman-Dragalin translation* FD' maps each formula A to the formula $A^{FD'} :\equiv A[F/\perp]$.

Naming FD' a refinement of FD is a little bit misleading, as we explain now. On the one hand, FD' simplifies FD by dropping the replacement of atomic subformulas $P \not\equiv \perp$ by $P \vee F$. On the other hand,

- FD is sound in the sense of $\mathsf{IL} \vdash A \;\Rightarrow\; \mathsf{ML} \vdash A^{FD}$;
- in general FD' is sound only in the weaker sense of $\mathsf{ML} \vdash A \;\Rightarrow\; \mathsf{ML} \vdash A^{FD'}$.

So we can say that FD' only really refines FD on ML, not on IL. This limitation of FD' is a problem if we want to apply a Friedman-Dragalin-like translation in IL. But it is not problem if we only want to apply a Friedman-Dragalin-like translation after a negative translation into ML (not just into IL).

Proposition 11 (factorisations $N_2 = FD \circ G$ and $N_2 = FD' \circ G$)**.**

1. *For all formulas A we have $\mathsf{ML} \vdash A^{N_2} \leftrightarrow (A^G)^{FD}$.*
2. *For all formulas A we have $A^{N_2} \equiv (A^G)^{FD'}$.*

Proof.

1. Let us abbreviate $(A^G)^{FD}$ by A^{GFD}. First we recall the definition of G writing all negations $\neg A$ in the form $A \to \perp$:

$$\begin{aligned}
P^G &:\equiv (P \to \perp) \to \perp \quad (P \not\equiv \perp \text{ atomic}),\\
\perp^G &:\equiv \perp,\\
(A \wedge B)^G &:\equiv A^G \wedge B^G,\\
(A \vee B)^G &:\equiv (A^G \to \perp) \wedge (B^G \to \perp) \to \perp,\\
(A \to B)^G &:\equiv A^G \to B^G,\\
(\forall x A)^G &:\equiv \forall x A^G,\\
(\exists x A)^G &:\equiv \exists x (A^G \to \perp) \to \perp.
\end{aligned}$$

Using this we unfold N_2 and GFD by induction on the structure of formulas:

$$
\begin{aligned}
P^{N_2} &:\equiv (P \to F) \to F \quad (P \not\equiv \perp \text{ atomic}),\\
\perp^{N_2} &:\equiv F,\\
(A \wedge B)^{N_2} &:\equiv A^{N_2} \wedge B^{N_2},\\
(A \vee B)^{N_2} &:\equiv (A^{N_2} \to F) \wedge (B^{N_2} \to F) \to F,\\
(A \to B)^{N_2} &:\equiv A^{N_2} \to B^{N_2},\\
(\forall x A)^{N_2} &:\equiv \forall x A^{N_2},\\
(\exists x A)^{N_2} &:\equiv \exists x (A^{N_2} \to F) \to F,
\end{aligned}
$$

$$
\begin{aligned}
P^{GFD} &:\equiv (P \vee F \to F) \to F \quad (P \not\equiv \perp \text{ atomic}),\\
\perp^{GFD} &:\equiv F,\\
(A \wedge B)^{GFD} &:\equiv A^{GFD} \wedge B^{GFD},\\
(A \vee B)^{GFD} &:\equiv (A^{GFD} \to F) \wedge (B^{GFD} \to F) \to F,\\
(A \to B)^{GFD} &:\equiv A^{GFD} \to B^{GFD},\\
(\forall x A)^{GFD} &:\equiv \forall x A^{GFD},\\
(\exists x A)^{GFD} &:\equiv \exists x (A^{GFD} \to F) \to F.
\end{aligned}
$$

Now we prove $\mathsf{ML} \vdash A^{N_2} \leftrightarrow A^{GFD}$ by induction on the structure of formulas. The only non-trivial case is the one of atomic formulas $P \not\equiv \perp$. In this case we argue $\mathsf{ML} \vdash P^{N_2} \leftrightarrow P^{GFD}$ using $\mathsf{ML} \vdash (P \to F) \leftrightarrow (P \vee F \to F)$.

2. Just note that A^{N_2} and $(A^G)^{FD'}$ are both syntactically equal to $A^G[F/\perp]$: we have we have $A^{N_2} \equiv A^G[F/\perp]$ by definition of N_2 and we have $(A^G)^{FD'} \equiv A^G[F/\perp]$ by definition of FD'. $\square$

5 Characterisation of the negative translations intuitionistically equivalent to the usual ones

There are two properties relative to NF that the usual negative translations share:

- to translate into NF in IL;
- to act as the identity on NF in IL.

We show that these two properties are not shared by N_1 and N_2.

Proposition 12. *If $\mathsf{CL} \vdash \neg F$ but $\mathsf{IL} \nvdash \neg F$, then N_1 and N_2:*

1. *do not translate into NF in IL;*
2. *do not act as the identity on NF in IL.*

Proof. We do the proof only for N_1 since the case of N_2 is analogous. By direct proof, assume $\mathsf{CL} \vdash \neg F$ but $\mathsf{IL} \nvdash \neg F$ and let us prove points 1 and 2.

1. If N_1 would translate into NF in IL, then $\bot^{N_1} \equiv \bot \vee F$, which is equivalent in IL to F, would be equivalent in IL to a formula in NF, contradicting point 2 of lemma 7.

2. If N_1 would act as the identity on NF in IL, then $\mathsf{IL} \vdash \bot^{N_1} \leftrightarrow \bot$ (since $\bot \in \mathsf{NF}$) where $\bot^{N_1} \equiv \bot \vee F$, so $\mathsf{IL} \vdash \neg F$, contradicting the assumption $\mathsf{IL} \nvdash \neg F$. $\square$

Proposition 12 suggests that the two properties relative to NF may tell the difference between the usual negative translations into IL and other negative translations into IL. Indeed, now we prove that they characterise the usual negative translations into IL.

Theorem 13. *Let N be a negative translation into IL (ML). The following properties are equivalent.*

1. *N is equivalent in IL (respectively, ML) to G.*
2. *N translates into NF in IL (respectively, ML).*
3. *N acts as the identity on NF in IL (respectively, ML).*

Proof. We do the proof only for negative translations into IL since the case of negative translations into ML is analogous.

($1 \Rightarrow 2$) By direct proof, if N is equivalent in IL to G, then N translates into NF in IL because G does so, as we wanted.

($2 \Rightarrow 3$) By direct proof, assume that N translates into NF in IL, consider an arbitrary formula $A \in \mathsf{NF}$ and let us prove $\mathsf{IL} \vdash A^N \leftrightarrow A$. By assumption the formula A^N is equivalent in IL to a formula in NF, and we have $A \in \mathsf{NF}$, so the formula $A^N \leftrightarrow A$ is equivalent in IL to a formula in NF. Since $\mathsf{CL} \vdash A^N \leftrightarrow A$ by the characterisation theorem of N, and since CL is conservative over IL with respect to NF, we have $\mathsf{IL} \vdash A^N \leftrightarrow A$, as we wanted.

($3 \Rightarrow 1$) By direct proof, assume that N acts as the identity on NF in IL, consider an arbitrary formula A and let us prove $\mathsf{IL} \vdash A^N \leftrightarrow A^G$. By the characterisation theorem of G we have $\mathsf{CL} + A \vdash A^G$ and $\mathsf{CL} + A^G \vdash A$. So by the soundness theorem into IL of N we get $\mathsf{IL} + A^N \vdash A^{GN}$ and $\mathsf{IL} + A^{GN} \vdash A^N$ (where A^{GN} abbreviates $(A^G)^N$). So by the deduction theorem of IL we have (1) $\mathsf{IL} \vdash A^N \leftrightarrow A^{GN}$. Since $A^G \in \mathsf{NF}$ by a property of G, by the assumption we have (2) $\mathsf{IL} \vdash A^{GN} \leftrightarrow A^G$. From (1) and (2) we get $\mathsf{IL} \vdash A^N \leftrightarrow A^G$, as we wanted. $\square$

Another property shared by the usual negative translations into IL is idempotence in IL, that is $N \circ N = N$ in the sense of: $\mathsf{IL} \vdash (A^N)^N \leftrightarrow A^N$ for all formulas A. Idempotence in IL is sometimes proved using the properties relative to NF. The proof roughly proceeds like this: if N is a negative translation into IL that (1) translates into NF in IL and (2) acts as the identity on NF in IL, then $A^N \in \mathsf{NF}$ by (1), so $\mathsf{IL} \vdash (A^N)^N \leftrightarrow A^N$ by (2). (This argument is not rigorous since from (1) we only get that A^N is equivalent in IL to a formula in NF, not that $A^N \in \mathsf{NF}$.) This relation of idempotence in IL with the properties relative to NF can make us suspect that idempotence in IL also characterises the usual negative translations into IL. But this is not so because, as we will show now, all negative translations into IL are idempotent in IL (but not equivalent in IL, as we already saw).

Definition 14. Let N be a negative translation into IL. We say that N is *idempotent in* IL (ML) if and only if for all formulas A we have $\mathsf{IL} \vdash (A^N)^N \leftrightarrow A^N$ (respectively, $\mathsf{ML} \vdash (A^N)^N \leftrightarrow A^N$).

Proposition 15. *All negative translations into* IL *(*ML*) are idempotent in* IL *(respectively,* ML*).*

Proof. We do the proof only for negative translations into IL since the case of negative translations into ML is analogous.

Consider an arbitrary negative translation N into IL, an arbitrary formula A and let us prove $\mathsf{IL} \vdash (A^N)^N \leftrightarrow A^N$. By the characterisation theorem of N we have $\mathsf{CL} + A^N \vdash A$ and $\mathsf{CL} + A \vdash A^N$. So by the soundness theorem into IL of N we get $\mathsf{IL} + (A^N)^N \vdash A^N$ and $\mathsf{IL} + A^N \vdash (A^N)^N$. Then by the deduction theorem of IL we have $\mathsf{IL} \vdash (A^N)^N \leftrightarrow A^N$, as we wanted. $\square$

6 Conclusion

The main three points of this article are the following.

Conjecture	The fact that the usual negative translations into IL are equivalent in IL leads to the conjecture: if we rigorously define the notion of a negative translation into IL, then we should be able to prove that all negative translations are equivalent in IL.
Refutation	We refuted the conjecture by presenting two counterexamples.
Characterisation	We characterised the usual negative translations into IL as being the ones that translate into NF in IL, or equivalently, that act as the identity on NF in IL.

References

[1] Ulrich Berger, Wilfried Buchholz, and Helmut Schwichtenberg. Refined program extraction from classical proofs. *Annals of Pure and Applied Logic*, 114(1):3–25, 2002.

[2] Thierry Coquand. Computational content of classical logic. In Andrew M. Pitts and Peter Dybjer, editors, *Semantics and logics of computation*, pages 33–78. Cambridge University Press, Cambridge, United Kingdom, 1997.

[3] Albert G. Dragalin. New forms of realizability and Markov's rule. *Soviet Mathematics Doklady*, 21:461–464, 1980.

[4] Albert G. Dragalin. New forms of realizability and Markov's rule (Russian). *Doklady Akademii Nauk SSSR*, 251:534–537, 1980. English translation: New forms of realizability and Markov's rule [3].

[5] Gilda Ferreira and Paulo Oliva. On various negative translations. In *Electronic Proceedings in Theoretical Computer Science*, to appear. Proceedings of the Third International Workshop on Classical Logic and Computation, Brno, Czech Republic 21–22 August 2010.

[6] Robert C. Flagg and Harvey Friedman. Epistemic and intuitionistic formal systems. *Annals of Pure and Applied Logic*, 32:53–60, 1986.

[7] Harvey Friedman. Classically and intuitionistically provably recursive functions. In Gert H. Müller and Dana S. Scott, editors, *Higher set theory*, Lecture notes in mathematics, pages 21–27, Berlin, 1978. Springer-Verlag. Proceedings of Higher Set Theory, Oberwolfach, Germany, 13–23 April 1977.

[8] Gerhard Gentzen. Über das Verhältnis zwischen intuitionistischer und klassischer Arithmetik, 1933. Galley proof from *Mathematische Annalen*. Appeared in *Archiv für mathematische Logik und Grundlagenforschung* [10]. English translation: On the relation between intuitionistic and classical arithmetic [9, pages 53–67].

[9] Gerhard Gentzen. *The collected papers of Gerhard Gentzen*. Studies in logic and the foundations of mathematics. North-Holland Publishing Company, Amsterdam, 1969.

[10] Gerhard Gentzen. Über das Verhältnis zwischen intuitionistischer und klassischer Arithmetik. *Archiv für mathematische Logik und Grundlagenforschung*, 16:119–132, 1974.

[11] Kurt Gödel. Zur intuitionistischen Arithmetik und Zahlentheorie. *Ergebnisse eines mathematischen Kolloquiums*, 4:34–38, 1933. English translation: On intuitionistic arithmetic and number theory [12, pages 286–295].

[12] Kurt Gödel. *Collected works*, volume 1. Oxford University Press, Oxford, 1986.

[13] Hajime Ishihara. A note on the Gödel-Gentzen translation. *Mathematical Logic Quarterly*, 46(1):135–137, 2000.

[14] Andrey N. Kolmogorov. On the principle of tertium non datur (Russian). *Matematicheskii Sbornik*, 32(4):646–667, 1925. English translation: On the principle of the excluded middle [18, pages 414–437].

[15] Jean-Louis Krivine. Opérateurs de mise en mémoire et traduction de Gödel. *Archive for Mathematical Logic*, 30(4):241–267, 1990.

[16] Sigekatu Kuroda. Intuitionistische Untersuchungen der formalistischen Logik. *Nagoya Mathematical Journal*, 2:35–47, 1951.

[17] Thomas Streicher and Bernhard Reus. Classical logic, continuation semantics and abstract machines. *Journal of functional programming*, 8(6):543–572, 1998.

[18] Jean van Heijenoort, editor. *From Frege to Gödel: a source book in mathematical logic, 1879–1931.* Harvard University Press, Cambridge, Massachusetts, 1967.