Logical Form and Reflective Equilibrium

Jaroslav Peregrin Vladimír Svoboda*

Abstract

Though, at first sight, logical formalization of natural language sentences and arguments might look like an unproblematic enterprise, the criteria of its success are far from clear and, surprisingly, there have only been a few attempts at making them explicit. This paper provides a picture of the enterprise of logical formalization that does not conceive of it as a kind of translation from one language (a natural one) into another language (a logical one), but rather as a construction of a 'map' of (a piece of) the 'inferential landscape' of the natural language. The criteria that appear to govern the enterprise are labeled as those of reliability, ambitiousness, transparency and parsimony. These criteria, it is argued, do not provide for an excavation of a ready-made logical structure, but rather help us achieve a "reflective equilibrium" between the normative authority of logic and the answerability of logic to a natural language.

Keywords: logical analysis, logical form, reflective equilibrium, reasoning

1 Introduction

One of the most characteristic types of tasks that students of logic must deal with is usually articulated as follows: "Rewrite the following argument in logical notation and then decide whether it is valid or not". This seems quite natural—the ability to examine the correctness of argumentation is precisely what students of logic are supposed to learn. Fulfilling this kind of task consists of two parts,

*Work on this paper was supported by the research grant No. P401/10/1279 of the Czech Science Foundation.

each of which requires a somewhat different skill. First, students must rewrite the natural language sentence into a logical formalism (chosen by the teacher) and then they must employ a method (either already implied by the previous choice or chosen by the teacher) leading to the decision.

The first part of this enterprise is usually called *logical analysis* or *logical formalization*. Though teachers usually allow for some variations in fulfilling this task, they are generally supposed to be able to tell whether what the students provide as their solution is correct or not. In this sense, logical analysis might seem to merely be an unproblematic enterprise that might appear difficult to the students but not to the teacher, who knows the criteria of its success. Yet, if we were to ask a randomly picked teacher how she decides whether a given formalization is correct, what criteria she actually employs, she is likely to be surprised by the question. She would probably say something to the effect that anyone who masters a logical system acquires an insight that enables her to recognize the correct formalization, similarly as a good translator is able to recognize a correct translation (even without being able to explicitly articulate any general criteria).

But can we accept a response of this kind? Is formalization simply a translation from one language (a natural one) into another language (a formal one)? And is it enough to leave it on the level of an implicit know-how? We think that the answers to both these questions are negative. We do not believe that formalization is very similar to translation. And even if it were, we do not believe that it could, in general, be left at the level of practical know-how without an explicit reflection of its criteria—even translation from one natural language into another must be explicitly reflected upon once competing proposals appear.

In this respect, we find it surprising how little attention questions of this kind receive in the (meta)logical literature. In fact, the only book-length treatment explicitly devoted to the criteria of logical formalization that we know of is (Brun, 2003).[1] This is also peculiar in view of the fact that this issue is closely related to questions regarding the very nature of logic, especially the question to what extent logic is merely descriptive and to what extent is it prescriptive.

[1]Many not so systematic considerations of this kind are, of course, scattered throughout the literature, especially about 'logical form', ranging from the seminal paper of Russell (1905) to more contemporary treatises like Sainsbury (1991).

In this paper, we want to contribute to the self-understanding of logical analysis of natural language by examining the practice of formalizing natural language sentences and arguments more explicitly than is common among the practitioners of logical analysis. We will try to articulate the criteria of correctness implicit to this practice and we will also try to draw some conclusions regarding the nature of logic. Some of the conclusions that we reach may appear as sheer platitudes, others as quite controversial. We believe that even the platitudinous parts are something that must be stated explicitly, and we believe that the controversial parts can be justified. This is what we will attempt to show in the rest of the paper.

2 Translation or not?

The usual idea is that as formalization, *viz.* rewriting a given sentence or argument into a logical language, is a kind of translation, it can be assessed like any other translation—the most basic criterion of the success of the enterprise being the preservation of meaning. What we want to point out in this section is that this is misguided. Though formalization may involve a translation-like step, it should not, in general, be considered as translation. We must therefore look for different criteria of its success than those governing translation.[2]

A student who faces the task of formalizing a sentence of natural language usually starts from paraphrasing the sentence in such a way that the resulting sentence (still in the natural language) reflects the shape of a suitable logical formula, and then proceeds to an expression of a language that mixes expressions of natural language with logical symbols. We will call expressions mixing natural language phrases in their raw form with phrases of a logical language *hybrid expressions*.

[2]Let us note that the problem of adequate formalization of a sentence formulated in a natural language can be considered from two different perspectives: a narrower one and a wider one. The narrower, *internal* one consists in accepting a framework of a particular logical language; the wider, *external* one does not presuppose such a framework and considers the choice of the language as a part of the task to be solved. (And we should keep in mind that we need not even restrict ourselves to the spectrum of existing languages—we could even consider the possibility of inventing a new one.) In this article we will, most of the time, be adopting the internal perspective. We will operate within the framework of classical predicate logic (CPL).

The process in which we 'translate' a natural language sentence into such a formula can be called, following Quine, *regimentation*.

Let us, for example, consider the following simple sentence

(S1) *Dogs have legs.*

A student who is to formalize it in classical predicate logic (CPL) (and is already sufficiently indoctrinated) is likely to see it as 'in fact' a shortcut for

(S1′) *All dogs have legs*

and to see this sentence as saying what is more precisely expressed by:

(S1″) *For every individual it is the case that if it is a dog, then it has legs.*

So far, what has been going on is paraphrasing *within English*, but now it seems natural (to logicians) to take a further step and make the meaning of the last sentence even more precise and transparent by means of CPL. Thus, we get the following hybrid expression

(HF1) $\forall x(\boldsymbol{Dog}(x) \rightarrow \boldsymbol{Has\text{-}legs}(x))$

where $\forall$ and $\rightarrow$ are the constants whose meanings (or 'meanings') are exactly delimited, whereas ***Dog*** and ***Has-legs*** are terms about which we merely presuppose that they inherit the meaning of the English expressions *is a dog* and *has legs*.

Thus we achieve a logical regimentation.[3] Note that its outcome—a hybrid formula—is not an expression of a language with a coherent semantics, for it consists of two different kinds of constituents, the respective semantics of which are of very different natures. Due to the fact that it contains elements of natural language (with their natural meanings) it is not a formula of a logical calculus, while due to the

[3]Of course, that regimentation may involve significant shifts in meaning. While (HF1) is false once there is a single individual dog which is—perhaps by some strange coincidence—legless, hardly anybody would consider (S1) as false in such a situation. Moreover, regimenting (S1) as (HF1) involves another meaning shift: While normal speakers would probably not hesitate to infer that there are some dogs that have legs from (S1), it is not correct to infer $\exists x(\boldsymbol{Dog}(x) \wedge \boldsymbol{Legs}(x))$ from (HF1).

fact that it contains artificially introduced symbols it does not have (strictly speaking) a natural meaning. Nevertheless, people conversant with the corresponding logical system can understand them very well—or at least this is what they feel.

There are two obvious pathways leading from such a hybrid language to a logical language whose semantics is fully and explicitly delimited (i.e. its expressions are put together according to explicit formation rules and their functions or semantic values are explicitly given). The first one consists in also rectifying the 'extralogical' vocabulary of natural language, thus gaining, aside from *logical* constants, *extralogical* ones as well and, as a consequence, formulas that contain no elements of the vocabulary of natural language.[4] (In the case of extralogical vocabulary it is, however, much less clear how to capture its functioning.) In this way, we reach a language which is *formalized*, though not *formal*—in the sense of Tarski (1933).[5]

Taking this kind of step in our example yields the formula

(CF1) $\forall x(\boldsymbol{D}(x) \rightarrow \boldsymbol{L}(x))$

What are $\boldsymbol{D}$ and $\boldsymbol{L}$ here? As they must be expressions belonging to a language within the framework of CPL and hence with logical (mathematized) semantics, there seems to be only one option—they are unary predicate letters, and hence they denote subsets of the universe. Of course, if the bold letters are given this meaning, (CF1) will be true or false (unchangeably) depending on the relations of the particular subsets they represent.

The second pathway consists in *dismissing* the extralogical constants. This is a natural thing to do if what we are after is a fully-fledged *formalization* that leads to a language that is not just formalized, but *formal*, in the sense that its formulas do not correspond to natural language sentences but rather to sentence *forms*. The step from an expression like (HF1) to a formal language expression consists in dropping the terms borrowed from natural language and replacing them with utterly meaningless symbols, which we will call *parameters*. Let us call this step away from the hybrid language *abstraction* (as

[4] The boundary between the 'logical' and 'extralogical' vocabulary of natural language is, of course, blurry.

[5] A paradigmatically clear case of this is Peano arithmetic: its language consists, aside from the logical constants, of the extralogical constants $\mathbf{0}$, $\mathbf{S}$, $+$, and $\cdot$, whose functioning is exactly stipulated.

we abstract from meanings of certain expressions). In our case, we obtain the traditional formalization

(FF1) $\forall x(F(x) \rightarrow G(x))$

Let us stress, once again, that (FF1) is no longer a meaningful sentence, but rather a *pure* formula, i.e. an articulation of a mere sentence *form*, containing meaningless parameters (we will speak simply about a *formula*, where no confusion is likely). It is sometimes also referred to as the *logical form* of the sentence out of the regimentation of which it has been abstracted. The hybrid formula that served as the input of the abstraction may be called an *instance* of the form; other instances are all those hybrid formulas that result from the replacement of the parameters of the formula by natural language terms of suitable grammatical categories. For simplicity, we will call the natural language sentences that verbalize the instances of a formula its *natural language instances*.

Thus, for us, the outcome of the formalization of a sentence like (S1) is a (pure, i.e. 'uninterpreted') formula such as (FF1). However, logicians engaged in logical analysis of language sometimes do not see it in this way: what they consider as the result of formalization is rather a ('fully interpreted', possibly hybrid) formula of the kind of (CF1) or (HF1). This might be an innocent terminological clash solvable simply by acknowledging the ambiguity of the term *formalization*, choosing one of the senses and introducing a different word for the other. Curiously enough, however, some of the most prominent logicians engaged in the logical analysis of language try to avoid this choice. They propose something that appears to be a somewhat strange compromise between the two options. This approach, promoted, among others, by Sainsbury (1991); Brun (2003) or Baumgartner and Lampert (2008), would result in the articulation of the result of regimentation of (S1) in the following shape:

(FC1) $\forall x(F(x) \rightarrow G(x))$
F: ... is a dog; G: ... has legs

Here the first line, which is nothing other than (FF1), is complemented by the second one, which is called the *correspondence scheme*.[6] Thus,

[6]Sainsbury (1991) and Brun (2003) use the term "correspondence scheme", Baumgartner and Lampert (2008) call the same thing "realization".

the outcome of the formalization of a natural language sentence is (somewhat surprisingly) not a formula of a logical language but something more complex. What is the nature of such complexes? One possibility of how to read them is to take the correspondence scheme as simply an instruction for the interpretation of the parameters; hence, in our case, as an instruction to interpret F by a certain set (the extension of *is a dog* at our world at some time point) and G as another one (that of *has legs*). In this case, (FC1) would simply collapse into (CF1), and there would seem to be no reason for presenting the result in this complex form.

Another option is to read (FC1) as establishing a 'dynamic' connection between the parameters and their natural language counterparts—in the sense that the latter confer their extensions on the former not on a one-time basis, but continually, which results in a situation where the extensions of F and G repetitively change. In this way we can say that (FC1) has not only a constant truth *value*, but truth *conditions*—it has (similarly as (S1)) different truth values in different situations/possible worlds. It is, however, surprising that formalization of a sentence in CPL does not yield a formula with the standard semantics but one with a kind of 'intensional' semantics. We are afraid that this institutes a dangerous Janus-facedness of (FC1): on the one hand, it is seen as a formula of (CPL) (disregarding, in effect, the second line), while on the other it is seen as an 'intensional' formula with non-trivial truth-conditions.[7]

We think this is a mere trick: a trick that supplies a first-order formula with truth *conditions* when, in fact, it has merely an (unchangeable) truth *value*. Moreover, we think it is an *unnecessary* trick: the criteria of adequacy of a logical formalization need not be (and, in fact, are not) based on the comparison of truth conditions, but rather on the comparison of behavior within arguments. Recognizing the logical form of a sentence is, first and foremost, recognizing the correctness/incorrectness of the arguments in which the sentence features, i.e. identifying its inferential role. Doing logical formalization, we start from a natural language argument, move to its logical form in a logical system and then use the means of the logical system to decide whether the argument form is logically valid—where the move from natural language to the formal one is usually not di-

[7]See Peregrin and Svoboda (in press) for a more detailed discussion.

rect but leads via the intermediate level of a hybrid language. And while the first part of the move (from natural to the hybrid language) might perhaps be considered as a kind of translation, the second part (from the hybrid to the formal language) certainly does not have this character.

3 How can we assess adequacy of logical formalization?

Suppose that three students are given the task to formalize the sentence

(S2) *No red snakes are dangerous*

and they came up with the following respective proposals:

(FFS2a) $\neg\exists x((Fx \wedge Gx) \rightarrow Hx)$

(FFS2b) $\neg\exists x(Fx \wedge Gx \wedge Hx)$

(FFS2c) $\forall x((\neg Gx \vee \neg Fx) \rightarrow \neg Hx)$

(where the parameter F replaces the expression ***is red***, G replaces ***is a snake***, and H replaces ***is dangerous***). How could we find out which of the proposals is to be preferred?

Unlike those who propose the 'corresponding schemes' as part of the result of formalization, we cannot take recourse to the sameness of truth conditions—the above formulas, not being sentences of a fully-interpreted language, simply do not have any. But we have already indicated what we should focus on instead: the behavior in arguments, i.e., in effect, inferential roles. What we usually do, as a matter of fact, is a careful reflection on arguments of a certain kind. We can consider (implicitly or explicitly) a sample list of natural language 'reference arguments' that we intuitively hold for falling into the intended scope of the logical system we use (here CPL)[8] and that are *perspicuous* in

[8]We assume that each logical system has been conceived with the goal of accounting for the behavior of a certain part of the logical vocabulary of natural language and the arguments that hold in virtue of this very vocabulary. Classical propositional logic focuses on the behavior of the well known connectives, classical predicate logic adds the basic quantifiers to this and modal logic further adds a certain modal vocabulary, etc. The intended scope of the system is then constituted by the arguments that are correct solely in virtue of the specific kind of vocabulary that the logical system is supposed to capture.

the sense that each of them is clearly intuitively correct or incorrect and in which the sentence we are considering (here S2) features as a premise or as the conclusion. Let us call the arguments on such a list *reference arguments of the sentence*. In our case, for example, a list of reference arguments can contain the following (correct and incorrect) cases:

Kaa is red
Kaa is a snake
Kaa is not dangerous

No red snakes are dangerous

Every snake is a reptile
No reptile is dangerous

No red snakes are dangerous

No red snakes are dangerous
Kaa is not red

Kaa is not dangerous

No red snakes are dangerous
Kaa is not dangerous

Kaa is red

No red snakes are dangerous
Kaa is a red snake

Kaa is not dangerous

If we now, next to the arguments, put parallel lists consisting of argument forms composed of the corresponding formulas of CPL, in which the sentence *No red snakes are dangerous* is formalized in each of the three proposed ways respectively, we get the table printed on the next page. For a better orientation we write those sample arguments that are (intuitively) correct with bold font and similarly for the argument forms that are valid in CPL.[9]

How does this list help us decide which of the proposed formalizations of (S1) is the most adequate one? The general answer is obvious: Where we have an intuitively incorrect argument that is rendered as valid by its formalization, or where we have, conversely, an intuitively

[9] F, G, H are as before, I replaces ***is a reptile*** and k replaces the name Kaa.

Kaa is red	Fk	Fk	Fk
Kaa is a snake	Gk	Gk	Gk
Kaa is not dangerous	Hk	Hk	Hk
No red snakes are dangerous	$\neg\exists x((Fx \wedge Gx) \rightarrow Hx)$	$\neg\exists x(Fx \wedge Gx \wedge Hx)$	$\forall x((\neg Gx \vee \neg Fx) \rightarrow \neg Hx)$

Every snake is a reptile	$\boldsymbol{\forall x(Gx \rightarrow Ix)}$	$\boldsymbol{\forall x(Gx \rightarrow Ix)}$	$\forall x(Gx \rightarrow Ix)$
No reptile is dangerous	$\boldsymbol{\neg\exists x(Ix \wedge Hx)}$	$\boldsymbol{\neg\exists x(Ix \wedge Hx)}$	$\neg\exists x(Ix \wedge Hx)$
No red snakes are dangerous	$\boldsymbol{\neg\exists x((Fx \wedge Gx) \rightarrow Hx)}$	$\boldsymbol{\neg\exists x(Fx \wedge Gx \wedge Hx)}$	$\forall x((\neg Gx \vee \neg Fx) \rightarrow \neg Hx)$

No red snakes are dangerous	$\neg\exists x((Fx \wedge Gx) \rightarrow Hx)$	$\neg\exists x(Fx \wedge Gx \wedge Hx)$	$\boldsymbol{\forall x((\neg Gx \vee \neg Fx) \rightarrow \neg Hx)}$
Kaa is not red	$\neg Fk$	$\neg Fk$	$\boldsymbol{\neg Fk}$
Kaa is not dangerous	$\neg Hk$	$\neg Hk$	$\boldsymbol{\neg Hk}$

No red snakes are dangerous	$\boldsymbol{\neg\exists x((Fx \wedge Gx) \rightarrow Hx)}$	$\neg\exists x(Fx \wedge Gx \wedge Hx)$	$\forall x((\neg Gx \vee \neg Fx) \rightarrow \neg Hx)$
Kaa is not dangerous	$\boldsymbol{\neg Hk}$	$\neg Hk$	$\neg Hk$
Kaa is not red	$\boldsymbol{\neg Fk}$	$\neg Fk$	$\neg Fk$

No red snakes are dangerous	$\neg\exists x((Fx \wedge Gx) \rightarrow Hx)$	$\boldsymbol{\neg\exists x(Fx \wedge Gx \wedge Hx)}$	$\forall x((\neg Gx \vee \neg Fx) \rightarrow \neg Hx)$
Kaa is a red snake	$Fk \wedge Gk$	$\boldsymbol{Fk \wedge Gk}$	$Fk \wedge Gk$
Kaa is not dangerous	$\neg Hk$	$\boldsymbol{\neg Hk}$	$\neg Hk$

correct argument that is rendered as incorrect, the formalization becomes suspicious. Thus, the fourth and the fifth case suggest that we have a reason to reject the formalization (FFS2a), whereas the second and the third cases provide reasons for rejecting (FFS2c). Hence the victorious formalization that we (tentatively) embrace is (FFS2b), which was not 'disproved' by the reference arguments.

Let us note that this method of selecting the best formalization is, in fact, not so different from that employed by the adherents of 'correspondence schemes'. The point is that inspecting the correctness of arguments, such as that which we have just been engaged in, can be seen as inspecting truth conditions. For example, the claim that the second argument is correct can be read as the claim that (S2) is true in all situations where all snakes are reptiles and no reptile is dangerous. If we consider formulas (FFS2a), (FFS2b) and (FFS2c) furnished by the 'correspondence schemes', we can say that the first two of them are also true in all such situations (where the situations are described in terms of the corresponding language). The same thing, however, cannot be said about (FFS2c). (Inspecting other arguments may directly amount to inspecting the truth conditions of sentences other than (S2), with (S2) taking part in the characterization of the situations considered).

In general, what we actually do when we check for the truth value of a sentence in a certain situation is, in fact, hardly distinguishable from checking inferences. We must somehow characterize the situation which we are considering, and we can hardly do it otherwise than in terms of some sentences; hence, when we then ask whether a sentence is true in the situation, we can be seen as asking whether the latter sentence follows from the former ones. (It is true that the sentence in which we characterize the situation can be couched in a metalanguage rather than in the object language we are analyzing; but, if it is natural language that is our ultimate target, then we cannot count on a metalanguage different from it.)

4 Criteria

How to articulate criteria of adequacy of formalization based on the above insights? If we generalize the lesson from the sketch of the method presented in the previous section, we can say that the point

of the formalization is to make explicit the place of a natural language sentence A within the inferential structure of its natural language, by means of associating A with a formula of the logical system **S** the position of which within the inferential structure of **S** is explicit and definite. Hence, with the help of **S** we construct a 'map' of the 'inferential surroundings' of A, making it possible for us to gain an overview over this 'inferential landscape'. This allows us to spot the inferential interrelationships of A with other sentences, which would be not so easily discernible otherwise.

However, it is crucial to keep in mind that if we try to identify the inferential (sub)structures of a natural language we want to make explicit, we will necessarily uncover a slightly fuzzy and gappy network of relations among sets (or sequences) of sentences (premises) and individual sentences (conclusions). The inferential structure of **S** will be, on the other hand, definite, determinate and much simpler.

To be able to formulate the criteria of adequacy of logical formalization that has issued from the above considerations, we introduce some terminology. A $[\Phi/A]$-*formalization* of an argument containing A will be a formalization with the formula Φ in place of A; conversely, a $[\Phi/A]$-*instance* of an argument form containing Φ will be any natural language instance of the form with A in place of Φ. Thus, given that A is *All dogs have legs* and Φ is $\forall x(P(x) \rightarrow Q(x))$, the $[\Phi/A]$-*formalization* of the argument

$$\text{(A1)}\ \frac{\begin{array}{l}\textit{All dogs have legs}\\ \textit{Fido is a dog}\end{array}}{\textit{Fido has legs}}$$

will be (given that the formalizations of *Fido is a dog* and *Fido has legs* are fixed as $P(a)$ resp. $Q(a)$):

$$\text{(AF1)}\ \frac{\begin{array}{l}\forall x(P(x) \rightarrow Q(x))\\ P(a)\end{array}}{Q(a)}$$

Conversely, (A1) will be an $[\Phi/A]$-*instance* of (AF1).

Now an argument form containing Φ is $[\Phi/A]$-*defeated* if it has an intuitively incorrect $[\Phi/A]$-*instance* among the reference arguments representing the intended scope of the actual logic (otherwise it is

[Φ/A]-*undefeated*). Given this terminology, we can articulate the most fundamental criterion of adequacy of formalization, which we will call the *principle of reliability*, rather succinctly:

(REL) Φ is a *proto-adequate formalization* of A in **S** iff no argument form valid in **S** and containing Φ is [Φ/A]-defeated.

The other criterion implicit to our proceedings envisaged in the previous section can be termed the *principle of ambitiousness*:

(AMB) Among the proto-adequate formalizations of A, Φ is the more adequate formalization of A in **S** the more intuitively correct arguments belonging to the intended scope of **S** in which A features as a premise or a conclusion are rendered as valid argument forms of **S**.[10]

To complete a truly comprehensive set of criteria we should add some principles guiding the choice for the cases undecided by the previous criteria. They can be called *the principle of transparency* and *the principle of parsimony*. We can articulate the first principle, for example, in this way:

(PT) (Other things being equal,) Φ is the more preferable formalization of the sentence A in the logical system **S** the more the grammatical structure of Φ is similar to that of A.

The second principle can then be formulated as follows:

(PP) (Other things being equal,) Φ is the more preferable formalization of the sentence A in the logical system **S** the more it is parsimonious as concerns the number of (types as well as tokens) of logical symbols it employs.

The import of the principles should be seen as decreasing in the order in which they have been presented. The first of them is close to

[10]These two principles are similar to (COR) and (COM) of Brun (2003). We have chosen different labels as we do not want to suggest that the first of them must be inevitably fulfilled for a formalization to count as *correct* in the ordinary sense of the word (valid argument forms to which we have natural language counterexamples might be a price we are willing to pay for having a particularly simple and perspicuous logical system); and that the second one marks a *completion* which we must achieve to be successful—we rather think that it spells out an ideal which we usually want merely to more or less approximate.

a *sine qua non* matter (though keep in mind that this holds only in the realm of the intended scope of the logic in question). The second is essential as well, as it suggests that the logician should not search just for 'the safest' formalization but also for the inferentially most 'fruitful' one—the one that makes explicit more relevant valid inferences than competing ones. The last two principles are more-or-less auxiliary (though they can be given more weight within analyses made for certain specific purposes). Thus, especially in the case of the last three, there might be various trade-offs (we might, for example, want to have a regimentation that is not quite transparent if it is exceptionally parsimonious.).

5 Bootstrapping

Now, however, we must return to various simplifying assumptions that we made throughout the course of our way from the description of the praxis of logical analysis to our articulation of the criteria.

First, the *principle of correctness* states that we can consider Φ as a candidate for the formalization of A only if *no* argument form containing Φ is $[\Phi/A]$-defeated. In fact, this is not quite realistic. Sometimes we may encounter what look to be invalid instances of argument forms that we hold for valid without putting their validity into doubt. Thus, consider the following argument, which looks, at least *prima facie*, as an instance of (AF1):

$$\text{(A2)}\ \frac{\begin{array}{l}\textit{All dogs have common genes}\\ \textit{Fido is a dog}\end{array}}{\textit{Fido has a common gene}}$$

This is clearly not a valid argument. Yet, its existence is not likely to make us conclude that (AF1) is defeated by (A2)—we would rather conclude that (A2) is, despite appearances, *not* an instance of (AF1), in particular that the logical form of *All dogs have common genes* is *not* $\forall x(P(x) \rightarrow Q(x))$. Why? We will probably say something to the effect that the predicate *to have common genes* is not an 'individual-level' (but rather 'group-level') predicate and thus should be represented, on the level of logical form, in a way different from the individual-level ones. However, how do we tell such an individual-level predicate from a group-level one? We might well say that a predicate

is individual-level if its use in the place of Q in instances of (AF1) yields correct arguments. But we would then have a vicious circle: an argument form is valid *because* all its instances are correct, but to be an instance of a valid form appears to *involve* being correct. (Of course the circle need not be so straightforward—we need not take directly (AF1) as the hallmark of individual-levelness of the predicate involved. However, we think that in the end *some* kind of circle is inevitable, for the distinction between individual-level and group-level predicates is not syntactic in the sense that it would be discernible by studying the predicates aside of their inferential properties.)

Is this circle vicious? Not necessarily. We think that it only points out that what we see as valid forms is not something which we can directly read off natural language, but rather that it is something that must be bootstrapped into existence. It is okay to explain away *some* invalid *prima facie* instances of an allegedly valid schema provided they can be plausibly taken as something marginal; however, if there is no way of moving them into a marginal position, we must retract the validity of the form.

Similar kinds of bootstrapping, in our view, penetrate the whole enterprise of logical formalization. Thus, we have to return to another unrealistic assumption that we have tacitly made when we started to look for the criteria of adequacy of formalization, *viz.* the assumption that the formalizations of all other sentences, save the one whose formalization we are pondering, are fixed.[11] Taken literally, it would, of course, once again lead us into a vicious circle: if we had to base the regimentation of any sentence on already accomplished formalizations of other sentences, the whole enterprise would never really be able to get out of the ground.

And once again the solution is, of course, a bootstrapping: we start with mere tentative regimentations of some simple sentences, basing the regimentations of others on them. Hence, if we are considering Φ as a possible formalization of A and we find out that some argument form involving Φ as a counterpart of A is valid, whereas there is a natural language instance that provides a counterexample (defeats the argument form), we will not only consider dropping the hypothesis that Φ is an adequate formalization of A, but will also take into

[11] In our case the trick was not so obvious as we only employed, in our test examples, formalizations of simple sentences that seem quite straightforward, like *Kaa is (not) a snake*.

account the possibility of keeping the hypothesis at the cost of dispensing with formalizations of some of the other sentences involved in the counterexample. Again, the process of formalization of sentences and arguments is, in fact, a holistic, give-and-take enterprise.

The third simplifying assumption was implicit to assuming our internal perspective, i.e. assuming that the logical language we use for the formalization is fixed. A formal language used as the tool of formalization is always more or less Procrustean, and to a certain extent this may be seen as its *virtue*: it lets us get rid of those elements of natural language that are irrelevant from the viewpoint of argumentation or of semantics and lets us clearly see the relevant backbone. But it might well happen that it may come to be Procrustean to the extent that it becomes a *vice*: it makes us neglect or obscure some important feature of natural language. In such a case, we need to ascend to the external perspective and look for a more suitable language.[12]

Hence, even the language we use for the formalization must be bootstrapped into existence: to a certain extent the features of natural language that do not fit into the mould of such language, are tolerable if they can be explained away as irrelevant or marginal. Once this discrepancy becomes excessive, however, it may be wise to give up on the language and upgrade. (The fact is, the standard logical languages, like those of classical propositional and predicate logic, have come to be taken so much for granted that we often take their adequacy as self-evident and tend to ignore discrepancies between them and natural language.)

6 Reflective equilibrium

The considerations of the previous section indicate that logic, though in a sense dealing with inferential patterns extracted from natural language (and thus answerable to how the language, in fact, works), also has a normative role to play: once it acquires a definite shape, it assumes the role of a standard which can be used to adjudicate individual cases of argumentation not only within a hybrid language but also in the natural one. As long as logical rules are in force, they decide what a correct argument is. But once a logical system urges us

[12]The common logical languages are, of course, common exactly for the reason that they turned out to be *tolerably* Procrustean.

to correct intuitions of competent speakers too frequently or in such a way that that we perceive the corrections as too counterintuitive, we have a serious reason to amend some rules of the logical system or to abandon the system as a whole. Hence, we have here the most basic give-and-take. And this is where, we believe, we must see it as a matter of what Goodman (1955) aptly called the *reflective equilibrium*.

We can, in general, say that the laws articulated by logic are not merely a reflection of something that exists, in a wholly articulated shape, either within our thinking or somewhere under the surface of our language. There is no way of merely extracting already completed laws of logic directly from there—what we can get as the starting point of logic are certain patterns of valid inferences that are accepted across different domains of our discourse and reasoning but which are not quite definite (both in the sense of not being exceptionless, and in the sense of not having an utterly clear-cut semantics).

This implies that any kind of logical system may only partially be based on patterns which logicians simply *find* and *report*—it must *also* be based on *completions* and *streamlinings* that logicians perform. Hence the laws of logic, as articulated by logicians, though crucially reflecting pre-existing patterns of valid inference, go well beyond them. Thanks to this and also to the—modest but extant—feedback that the work of logicians receives, logic influences the language of science and consequently even—slightly—the colloquial idiom, and comes to be taken as a *norm*. It acts as a norm of what is to be seen as regular and what is to be seen as 'irregular', and what is a lawful usage and what is an exception. (In this way, it ties together a framework for adjudicating various disputes that would hardly be resolvable otherwise.)

We have tried to portray how this works in terms of the dialectics of correct inferences and valid forms. Some inferences (in natural language) are *prima facie correct*, which makes us see some forms of inferences (namely those which have correct instances) as *prima facie valid*. However, we take the quest for (getting a grasp on) validity as an instance of a quest for *e pluribus unum*, as a quest for finding a perspicuous order within the *prima facie* messy vastness of individual cases of more or less correct or incorrect inferences; this makes us impose more order on our language and our reasoning than we are able to *find* there, even at the cost of some Procrustean trimming and stretching. Hence, upon reflection, a form of inference comes to be

taken as valid not exactly in those cases when all its natural language instances are correct, but in cases when those which are not can be reasonably explained away.

More traditional approaches to logical formalization often create the illusion that, behind or beneath the surface form of our language, there is a definite deeper and more substantial logical form. However, we do not believe that anybody could get to such a form by a process substantially different than the 'give-and-take' one described above, hence by a process led by the maxim of simplicity and maximal order—the maxim that is operative in any science. In particular, we do not believe that we can get from the surface form to the logical form by some process that has nothing to do with the considerations described above under the heading of *reflective equilibrium* and we don't therefore think that logic could be left with the task of pulling out the ready-made structure and lending it a perceptible form. We are convinced that the way from the surface to the so-called logical form involves considerations largely constitutive of logic, so that the resulting logical form is not what logic merely describes or reports, but rather what logic helps bring into being.

According to this picture, logical formalisms basically generalize and systematize the inferential and semantic features of natural language and so they are liable to criticism as other empirical generalizations. However, due to the fact that natural language is vague and open-ended, formalization also does the job of sharpening, explicating and removing inconsistencies; and, as a consequence of this, the result gains a certain normative authority over the use of means of natural language.

7 Conclusion

Accepting a certain logical system, we typically proceed by *regimenting* a natural language sentence into a hybrid sentence/formula, from which we then *abstract away* the (extralogical) remnants of natural language thus reaching formulas that represent what is traditionally called the *logical form* of the sentence (in the language of the given system of logic). The most basic of the criteria governing this enterprise can be termed the criterion of reliability; it is supplemented by the criteria of ambitiousness, transparency and parsimony. The criteria do not guarantee that there is anything like a unique logic form

to be found. Especially the latter three operate on a give-and-take basis, but even the first is not essential in the sense that it would be absolutely non-negotiable.

Logic aims at bringing order to our argumentative practices, by means of achieving the *reflective equilibrium*. Thus, logic has a certain descriptive aspect in the sense that it has to reflect the basic inferential structures of natural language, but it also has a normative aspect in the sense that once established, it has a (limited) authorization to brand natural language arguments as correct or incorrect.

Jaroslav Peregrin and Vladimír Svoboda
Department of Logic, Institute of Philosophy
Academy of Sciences of the Czech Republic
Jilská 1, 110 00 Praha 1
Czech Republic
e-mail: jarda@peregrin.cz, svoboda@site.cas.cz