Volume 6 Number 1
2010
ISSN 1807-9792

abstracta

Linguagem, Mente & Ação

ABSTRACTA

Linguagem, Mente e Ação

ISSN 1807-9792

Volume 6 Number 1
2010

TABLE OF CONTENTS

Abstracta 6 : 1, pp. 3 – 17, 2010

ON THE POSSIBILITY OF CONTINGENTLY DISPOSITIONAL PROPERTIES

Vassilios Livanios

Abstract
Metaphysicians who hold that there is an ontological distinction between two kinds of fundamental natural properties assume that properties are dispositional or non-dispositional *necessarily*. In contrast to this, I suggest that one can admit the existence of fundamental *contingently* dispositional properties. After some clarifications concerning the content of the suggested view, I respond to several objections regarding its intelligibility and viability and outline two of its important consequences.

The dispositional/non-dispositional[1] distinction is (as Armstrong (2005) recently suggested) one of the main disputes about properties.[2] Some philosophers, however, deny the ontological character of the distinction. For instance, according to Mumford's earlier (1998) view properties can be *characterised* either dispositionally or structurally (categorically), relative to a particular causal role. Others (C.B. Martin (1997), J. Heil (2003) and G. Strawson (2008)) defend an identity theory; for instance, on Martin/Heil's view, dispositionality and qualitativity are the self-same property differently considered and there are no ontological features that ground (or simply are) the dispositionality and qualitativity of any property. In contrast to the above thinkers, a number of metaphysicians hold that there is an ontological distinction between dispositional and non-dispositional properties; one that is not about predicates but rather about specific ontological features of properties in question. On the one hand, *property dualists* think that there are irreducible and ineliminable differences between two *kinds* of property (dispositional and non-dispositional), which both exist in the world. On the other hand, *property monists* argue that *only one* kind of property exists. *Categorical monists* defend the view that no genuine

[1] Instead of the term "categorical" (which for many – including myself – has unpleasant connotations), I shall use the expression "non-dispositional".
[2] The following discussion does not concern mathematical, logical and mere Cambridge properties. It is about the fundamental natural properties that carve nature at its joints.

property is dispositional, whereas *dispositional monists* claim that *every* genuine property is dispositional.

Those metaphysicians who hold that there is an ontological distinction between kinds of *fundamental* natural properties assume that properties are dispositional or non-dispositional necessarily. Namely, they suppose that a fundamental natural property is dispositional (or non-dispositional) in all possible worlds in which it exists. Furthermore, in order to give a verdict on the dispositional character of any property they assume the *rigid* application of any adequate criterion of dispositionality that can be used in the actual world (from now on, an @-criterion).[3] In other words, they examine whether the property in question has in all possible worlds (in which it exists) the appropriate ontological features in order to be dispositional according to the @-criterion. For instance, according to dispositionalists the ontological mark of dispositionality of any fundamental natural property is the necessity of its causal/nomological/metaphysical roles.

In general, it is routinely assumed that ontological distinctions (including, besides the dispositional/non-dispositional distinction, the universal/particular distinction and the abstract/concrete distinction), refer to features that entities possess necessarily. Particulars are particulars in each world in which they exist and there is no world in which a concrete entity exists and it is abstract. There are, however, a few dissenting voices suggesting that certain entities might possess at least some of the aforementioned features contingently. Linsky and Zalta (1994), in their attempt to defend an actualist interpretation of the simplest quantified modal logic, suggest the existence of contingently non-concrete entities. And Fraser MacBride (1999), in the context of his examination of the prospects of modal reductionism, argues that if there are no necessary de re connections in nature, then we may have to countenance the possibility that any particular entity (such as Armstrong himself!) could have been a universal.

Having these remarks in mind, let us now return to the issue of dispositionality of fundamental natural properties. Recall that each metaphysician applies her own @-criterion for dispositionality in all possible worlds. Notice, however, that the implementation of the @-criterion in every possible world does not *in general* guarantee a unique upshot. It may

[3] The relevant criterion is ontological. It aims to demarcate the distinction in terms of ontological features.

well be that a fundamental natural property satisfies the @-criterion of dispositionality in the actual world and fails to satisfy it in another possible world. In other words, it might possess the ontological marks of dispositionality only contingently. So, following the unorthodox route, I suggest that in this case one is entitled to say that the property in question is dispositional in the actual world and non-dispositional in the possible world. More generally, I suggest, one can admit the existence of fundamental *contingently* dispositional properties, provided they fail to satisfy the @-criterion of dispositionality in at least one possible world (whether it is the actual world or not).[4]

Two preliminary clarifications; first, one should clearly distinguish the aforementioned view from a kind of dual-sided theory once being defended by C.B. Martin (1993). According to Martin's view, each natural property has (in each world in which it exists) two distinct but inseparable ontological 'sides', its dispositionality and qualitativity, which are necessarily co-existent. On my view, there are fundamental natural properties which, in each possible world in which they exist, *and in their entirety*, are either dispositional or non-dispositional. Second, I must emphasise that I am not suggesting the common Humean view according to which the contingent laws of nature impose upon any fundamental natural property a *particular* dispositional character. The thesis of Contingent Dispositionality (which, henceforth, I will abbreviate as CD) is not a mere reformulation of the orthodox Humean view. To see that, consider, for instance, the property of being electrically charged. I am not arguing that 'attracts other opposite charged bodies' is a contingent second order feature of the first order property in question. Rather, my claim is that *dispositionality* itself contingently characterises this fundamental natural property.

I am sure that a number of philosophers may find the very idea concerning the alleged contingency of the dispositionality implausible (or even a bit bizarre). There are various reasons that might justify this belief. Consider first the case of natural dispositional properties, such as solubility, which are *by definition* related to a specific causal role. According to CD, the property of solubility, which is dispositional in the actual world, may be non-dispositional in another possible world. In that world, which may have radically

[4] The rigid application of the @-criterion does not entail that all possible worlds are populated only by properties of the actual world. They may exist alien properties which in *their* worlds are dispositional (or not) according to the @-criterion.

different laws from the actual, there is no assurance that solubility confers to its bearers the same causal powers that it actually confers. Since, by hypothesis, solubility is non-dispositional in that possible world, it is the alien laws of *that* world that determine the behaviour of objects that instantiate solubility. Hence, in the world in question, an actual soluble object may fail to dissolve (in the appropriate circumstances) despite the fact that it instantiates solubility. But is it not absurd to claim that objects may instantiate solubility and not dissolve in the appropriate circumstances?

To meet this prima facie strong objection we must notice that dispositional properties which are by definition related to specific causal roles are not fundamental natural properties.[5] Consider, for instance, electric charge which is one of those fundamental properties that belong to the reduction (or supervenience) base of solubility. Charge actually confers to its bearers the power "attracts other opposite charged bodies"; but it is not by definition related to this specific causal behaviour. What holds for charge does also hold for all fundamental natural properties that physical science acknowledges. So, to the extent that CD concerns the fundamental properties, the above objection raises no difficulties to it.[6] Of course, one may insist and raise an objection for the fundamental properties themselves: how can charged bodies fail to attract other opposite charged bodies in other possible worlds? The answer is that they can, provided that we hold that the transworld identity of fundamental natural properties is independent of their causal roles.[7] If there is a genuine difficulty here, it concerns the issue of transworld identity of properties (for which we have some things to say in the sequel) and not the intelligibility of CD. It is important to make clear that the issue of the acceptance of CD is orthogonal to the issue of the transworld stability of property's roles. One may hold that any fundamental natural property *must* be dispositional in *all* possible worlds (in which it exists) and, nonetheless,

[5] They are either supervenient upon or can be reduced to a web of fundamental natural properties.

[6] I would like to thank an anonymous referee for pressing me to make clear CD's range of validity.

[7] Returning to the case of solubility, it is plausible to assume that the 'weird' behaviour of charged bodies in the world in question may prevent any actually soluble object to dissolve in the appropriate circumstances, and since solubility is by definition related to this specific behaviour, the aforementioned object may be no longer soluble. Since we cannot exclude this possibility, it seems that solubility cannot exist in a world and fail to be dispositional in it. (Recall that assuming a non-dispositional character for solubility leaves open the possibility that objects instantiating it do not dissolve; an absurdity that gave rise to the objection in the first place.) But, as we have already remarked, that is not a problem for CD; it is not supposed in the first place that, according to CD, solubility (*qua* non-fundamental property) is a contingently dispositional property.

believe that it may have different roles in different possible worlds. For instance, Hendry and Rowbottom (2009) defend the thesis of dispositional contextualism, according to which to have a dispositional property is to have a single set of actual and possible dispositions, rather than just a set of actual dispositions. Dispositional contextualism is a position that respects the orthodox view according to which dispositional properties possess dispositionality necessarily; it simultaneously allows, however, a kind of transworld variation in a property's dispositional profile. Vice versa, a fundamental natural property may have the same causal roles in all possible worlds in which it exists and, nonetheless, be dispositional in some worlds and non-dispositional in others. In the former worlds, those roles are grounded in the dispositional nature of the property itself; in the latter, they are imposed upon it by the contingent laws of nature.

The claim of the unintelligibility of CD can be also expressed in a different manner. Consider the view according to which properties are ways objects are; by analogy, dispositionality must be a way a property is. The objection is that though we can conceptually discern an object from a way that object is (or could be), the supposed distinction between ways a property is and the property itself is obscure. Saying that there are contingent ways a property is is tantamount to saying that the property itself might have been different. And considering the case of the property that could have been different we just (the objector continues) contemplate a different property. So, once again, how can one and the same property be dispositional in one possible world and non-dispositional in another?

In a sense, the above objection simply begs the question against CD. For, under the perspective of CD, a fundamental natural property can be characterised differently (in other words, there are contingent ways a property could be) though remaining the same. Nonetheless, as we have remarked earlier, it may be that the whole objection rests upon the issue of the transworld identity of contingently characterised properties. Prima facie, it seems that CD is indispensably committed to the controversial thesis of quidditism.[8] There

[8] One might be tempted to think that quidditism is tantamount to the acceptance of quiddities as second order non-qualitative properties which distinguish particular properties. He might say that when two worlds differ quiddistically they disagree about which quiddities are instantiated by which qualitative first order properties.

are various formulations that aim to capture the core content of quidditism, but the most serious objections attack a view (which can be called *extreme quidditism*) characterised by the following three theses: firstly, transworld identification of properties is a matter of strict identity. Secondly, there exist non-qualitative determinants[9] of the transworld identification of properties which just are the identities between the 'inhabitants' of each possible world. And thirdly, transworld identification does not depend at all on qualitative characters; any property can have any qualitative character.[10] As dispositionalists have already argued, extreme quidditism faces several difficulties which, of course, also beset any proposal that is committed to it. The problem has also another aspect; if quidditism is presupposed, CD seems to be almost trivial. If any property can have any qualitative character whatsoever, then *of course* an actually dispositional property may turn out to be non-dispositional and vice versa.[11]

In order to reply to this objection, it is crucial to notice that CD is *not* indispensably committed to quidditism. It is in fact compatible with a kind of transworld identification that differs both from the one that quidditists have embraced and the one that dispositionalists hold. In my (2010) I argue that fundamental natural properties, such as rest mass, electric charge and spin, can be identified by conceptual means which are entirely independent of the powers that those properties confer to their bearers. More precisely, they can be identified as invariants under the action of fundamental symmetry transformations. Consider, for example, rest mass and spin. It is standard (within the mathematical framework) to discuss symmetries using group theory. Various transformations of physical interest form groups that can be analysed mathematically. One of them, the Poincare group, is associated with a symmetry concerning Lorentz boosts, rotations and space-time

---

Yet, this is not necessarily so; a quidditist need not believe in second order non-qualitative properties, or even reject the view that there are any second order properties in general.

[9] The qualitative determinants of the transworld identification of a property are related to its causal, nomological and metaphysical roles.

[10] Modest versions of quidditism that allow qualitative constraints on transworld identification of properties spring from some minimal-essential causal, nomological and metaphysical roles of those properties (in other words, there exist limits on how different a property could have been from the way it actually is). Furthermore, there is also a counterpart-version of extreme quidditism, according to which a) each possible world has its own properties, and b) transworld 'identification' of property counterparts is completely independent of their causal, nomological and metaphysical roles.

[11] I would like to thank an anonymous referee for revealing this problematic aspect.

translations. Wigner (1939) showed that rest mass (m) and spin (s) are two properties that characterise the elementary physical particles and are intimately associated with the action of the transformations of the Poincare group. To be more specific, according to group theory, in any irreducible representation of a continuous group, there are operators (called Casimir operators) that commute with all operators of the representation, and they are multiples of the unit operator. These operators have fixed numerical values in a given irreducible representation, which can be used as labels to characterise the irreducible representation (Hamermesh 1989: 318). Wigner (1939) computed all the irreducible representations of the Poincare group and found that the representations can be labelled by the parameters *m* and *s*, where *m* can be identified with the rest mass of the particle and *s* can be identified with its spin. Mass is the property that appears (as a parameter) in the first Casimir operator of the Poincare group; that is, the one which is formalised with the aid of only one parameter (which, of course, represents mass). Having identified mass, we can then identify spin as the property represented by the second parameter that appears in the second Casimir operator (the one which is formalised with the aid of two different parameters).[12]

Given that fundamental natural properties can be identified by means which are entirely independent of the powers that those properties confer to their bearers, it is not absurd – given the similar move that dispositionalists take about dispositional essences – to assume the existence of non-dispositional determinants of the transworld identification of those properties about which, in contrast to quiddities, something substantive can be said. Since the transworld identificatory elements do not involve features that ground @-criteria of dispositionality, the aforementioned suggestion is perfectly compatible with CD.[13]

Furthermore, even holding that CD is indispensably committed to quidditism does not turn it into a trivial position. To illustrate that, suppose that you are a categorical monist (like Armstrong) and you accept quidditism as the most plausible position about the transworld identity of properties. Nevertheless, in line with all categorical monists, you do

[12] The Casimir operators of the Poincare group are $c_1 = -m^2$ and $c_2 = -m^2 s(s+1)$, where *m* is the rest mass and *s* the spin.

[13] I assume that being invariant under the action of fundamental symmetry transformations is not a plausible candidate for being an ontological mark of dispositionality.

not think that natural properties possess their categorical character contingently. You implicitly assume the necessity of the categorical nature and so you accept that natural properties can have different causal roles in different worlds but cannot have dispositional nature in any of these worlds. This shows that in the case of categorical properties quidditism does not imply CD. Now, do we have a cogent reason to suppose the contrary, as far as the dispositional properties and their dispositional character is concerned? I think not and so I conclude that CD is not a trivial consequence of quidditism.

It might also be objected that I offer no explanation of *how* an ontological feature, such as the dispositional character, can be contingent. But what kind of *explanation* is the objector asking for? I suppose that a description of a (causal?) mechanism which explains how a property can acquire or lose its dispositionality in different possible worlds would be enough. In this case, however, I can show that the objector's demand is exaggerated. Consider, for instance, the analogous case of the contingent features of concrete particulars. Humeans and non-Humeans alike accept the fact that there are properties which characterise concrete particulars only at some possible worlds in which they exist; most often the existence of possible worlds at which those particulars do not instantiate the aforementioned properties is grounded only on intuitions regarding what is possible or not. Humeans, for instance, may try to ground these intuitions by invoking a principle of recombination, the application of which 'generates' all possibilities. (Some of these possibilities are compatible with the fact of non-instantiation of the contingent properties.) Following this way, Humeans are able to tell a story about how the actual world (in which the instantiation takes place) differs from possible worlds in which the instantiation of the contingent features does not take place. They merely insist that the difference is due to the different global distributions of properties to the same concrete particulars (or to their counterparts, if they are modal realists). But, in any case, they do not posit a mechanism which (causally) explains the difference by generating the different distributions. I think that something analogous can be said in the case of the contingent dispositionality. It is too excessive to ask Humeans or anyone else to describe a mechanism which generates the relevant differences. Similarly to the previous case, philosophers can rely on intuitions in order to ground the relevant possibilities. Humeans may even try to reply to the objector's

explanatory demand by telling a story about how worlds at which a property is dispositional differ from worlds at which the same property is not dispositional. They may appeal to a recombination principle concerning the ontological features of properties themselves and argue that the relevant difference is due to different global distributions of these features to the same properties.

Finally, some philosophers may object that CD, though conceivable, is not viable. To illustrate that, consider the @-criterion of dispositionality. For CD to be viable, the @-criterion should not prejudge the result of its application in all possible worlds. But how can the @-criterion not prejudge the upshot of its application in other possible worlds, given that, till now, all the suggested @-criteria of dispositionality are modal in character (so, by definition, they involve in their application other possible worlds)?

To this I can reply that the fact that all the suggested @-criteria of dispositionality are modal in character does not imply that the required (for the application of each criterion) range of possible worlds must be universal or be the same in every possible world. The criterion must be the same in all worlds, but the family of worlds used for its implementation may differ; hence, insofar as the @-criterion does not refer to *essential* ontological features (in which case the range of relevant worlds could be universal), a property which satisfies it in the actual world may fail to satisfy it in a possible world that does not belong to the family of worlds relevant for the application of the @-criterion. Hence, the result of the application of the @-criterion is not in general predetermined.

Of course the adequacy of the above reply depends crucially on the assumption that the @-criterion does not refer to *essential* ontological features. It is exactly this assumption, however, that dispositionalists deny. They crucially entangle the transworld identificatory elements of any fundamental natural property with the features that ground its @-criterion of dispositionality. They think that the mark of dispositionality of any fundamental natural property is the necessity of its causal/nomological/metaphysical roles, while the latter are also essential features of the property which *exclusively* constitute its identity in every possible world in which it exists. So they hold that if a property satisfies their criterion in the actual world, it must satisfy it in every possible world (in which it exists), and hence it is necessarily dispositional. In other words, under the perspective of dispositionalists, the

expected result of the application of the @-criterion is unique and fixed in advance. Given that the metaphysical thesis of dispositionalism is independently plausible the aforementioned objection raises (at least prima facie) a serious threat to CD.

The power, however, of the dispositionalist's objection depends upon the adequacy of her @-criterion of dispositionality which, in turn, depends upon the plausibility of her criterion of properties' transworld identity. It is the latter criterion, I argue, that faces the most serious difficulties. It is not only that we can easily imagine possible worlds in which actual properties confer to their bearers different causal powers from the ones they actually confer. Perhaps this intuition can be seriously undermined by embracing the view that imaginability does not entail metaphysical possibility. We have, in addition, cogent reasons to question the universality and the correctness of the criterion. First of all, there is the case of spatiotemporal relations for which it can be plausibly claimed that they confer no causal powers to their bearers. Alexander Bird (2007) recently tried to defend the contrary in the context of General Theory of Relativity wherein the spacetime metrical structure incorporates (in a sense) the set of all spatiotemporal relations. He argues that metrical structure (and, as a result, spatiotemporal relations themselves) confers causal powers because it possesses a dispositional essence. But, as I have argued in my (2008), his argument fails. Firstly, spatiotemporal relations are not active dispositions, because spacetime metrical structure does not *causally* affect material bodies; it just determines which paths are *available* to bodies when moving inertially. And they are not passive dispositions either, because, under a cautious interpretation, Einstein equations do not show that spacetime metrical structure depends causally on matter; they just give a law-like consistency constraint upon the joint features (space-time structure and mass-energy distribution) of any (physically) possible world.

Secondly, there are fundamental properties of the elementary particles (the so called, quantum numbers) that do not only confer no causal powers to their bearers, but, in addition, exclude the latter from possessing certain causal powers. Consider, for example, protons, the well known elementary particles that belong to a kind of entities that experience the strong interaction. Protons (like all baryons) instantiate a fundamental property, the baryon number, the conservation of which prohibits their decay. In other

words, having the property of baryon number exclude protons from possessing the power to decay.

The abovementioned examples show that even assuming the truth of the criterion, we have reasons to question its universality. Finally, we may plausibly doubt the correctness of the dispositionalist's criterion even in its allegedly uncontroversial range of application. For it can be claimed that the identity of fundamental natural properties that dispositionalists themselves often invoke (such as mass, electric charge and spin) can be determined independently of any causal roles that those properties may confer to their bearers. In particular, as we have already remarked, they can be identified as invariants under the action of fundamental symmetry transformations. If the identity of the above fundamental physical properties can be provided via symmetry considerations, why can't we claim that being invariant under the action of fundamental symmetries is an essential feature of the fundamental physical properties? Even if we suppose that this kind of alleged invariance essence would not exhaust the essence of those properties, it would certainly be a constituent of it. Granted that, we have a direct refutation of the dispositionalists' claim that the identity of any fundamental physical property is *exhaustively* constituted by its powers.

I would like to conclude by outlining two interesting consequences of the thesis of Contingent Dispositionality. The first of them is related to Hume's dictum (i.e., the rejection of necessary connections between contingent, wholly distinct, existents). Embracing the suggestion about the contingency of the dispositionality can help the supporters of the dictum to avoid troublesome necessary de re connections. In order to see that, consider the case where one follows the spirit of the dispositionalists' criterion of dispositionality and associate the possession of the latter with the 'necessary' conferment of specific causal powers.[14] Even supposing that, the most the dispositionalist can assume is

[14] It is true that, recently, some philosophers have challenged the orthodox view that there exist metaphysically necessary de re connections between dispositional properties and their manifestations. For instance, Markus Schrenk (2008) argues for a view which arguably may make room for the compatibility between Humeanism and genuine dispositional properties. While examining the antidote cases of the dispositions literature, he posits a dynamic, *intraworld*, de re link between dispositions and their manifestations that has nothing to do with metaphysical necessity. Yet, besides those few dissenting voices, the orthodox view is still prevailing among property metaphysicians. Some philosophers take a step further and suggest that the link between dispositions and necessity is more intimate than that. They embrace what Antony Eagle (2009) calls *Dispositional Actualism*, according to which the metaphysical necessity *itself* is

that an instantiation of a natural dispositional property confers the same causal powers to its bearer only in those possible worlds at which the property in question is dispositional. Given the contingency of dispositionality, there exist worlds in which the property in question is non-dispositional; in those worlds, there is no reason to suppose that the instantiation (by its bearer) of the property is associated with the (actual) specific behaviour characteristic of the possession of the aforementioned causal powers. Therefore, there are no necessary connections between the instantiation of natural properties, the presence of specific activating conditions, the absence of certain disturbing factors (such as finks and antidotes) and the proper manifestations of the aforementioned properties.

Contemporary Humeans (which claim that all fundamental natural properties are non-dispositional) are typically followers of Hume's dictum and so they may take advantage of the above consequence of CD. One of their main problems is to explain the unmanifested dispositions that we often ascribe to particulars (and regularly associate with their properties) without assuming any kind of physical de re necessity. Humeans do not deny that we associate dispositional *truths* with the fundamental properties, but they insist that the truthmakers for such truths are not dispositional properties having an essential dispositional character. Rather, the truthmakers are the fundamental *non-dispositional* properties plus the totality of the (contingent) laws of nature. However, for many metaphysicians the suggested truthmakers are insufficient. For those that do not share Humean intuitions, no distribution of intrinsically inert properties (supported by a 'thin' conception of laws of nature which hardly govern world's events) can serve as an adequate explanation basis for all we observe in a world which is full of 'threats and promises'. So, non-Humeans insist that Humeans are wrong in holding an *eliminativist* view about genuine unmanifested dispositional *properties*. In my opinion, non-Humeans are right in their criticism, but they are too hasty in rejecting Humean position for that reason. For under the perspective of CD, Humeans *can* accept the existence of fundamental dispositional properties *provided that these properties have a modally restricted dispositional character.*

---

grounded upon the constraints that the essentiality of actual causal profiles of properties place on the space of possibilities.

Following this strategy they can adequately explain the ascriptions of unmanifested dispositions while avoiding any kind of physical de re necessity.

The second consequence of CD is related to the controversial issue of the contingency of laws of nature. There is a strong intuition that laws are contingent which stands in contrast to the thesis (advocated by dispositionalists) according to which they are necessary (because natural properties confer essentially causal powers to their bearers). Of course, even in the context of dispositionalism (and setting aside the arguments for the 'illusion' of the metaphysical contingency of laws), a kind of contingency can be restored, provided that the fundamental natural properties are *contingent* beings. For in that case, the only thing that dispositionalists can prove is that possible worlds with the same fundamental properties as the actual must be governed by the laws of our world. Strictly speaking, and given that not all worlds are inhabited by the actual properties, the laws of nature are no longer necessary.[15] CD, however, supports a more *robust* kind of contingency for laws of nature, which holds, even granted that actual fundamental natural properties are dispositional *and* necessary beings. To illustrate that, let us first notice an important point about laws of nature. Most metaphysical accounts agree that laws of nature express relations between natural properties but disagree on the nature of these relations. Categorical monists hold that laws of nature are contingent and express *external* relations between natural properties. The latter are non-dispositional and either have no intrinsic nature or, alternatively, have a nature independent of their causal roles. In each possible world, laws express external relations that determine the role of each natural property in that world. In contrast, dispositional monists hold that laws of nature express *internal* relations between natural properties. The intrinsic dispositional nature of properties determines their roles in each possible world; it also determines completely and necessarily the instantiation of internal relations that laws of nature express. Having said that, let us now suppose that all fundamental natural properties are *necessary* beings; a hypothesis that entails that all fundamental natural properties inhabiting the actual world also exist in all other possible worlds. Consider the case of two actual fundamental dispositional properties and a natural

[15] Bird calls this thesis weak necessitarianism. For a defence of strong necessitarianism (according to which all worlds have the same laws), see Bostock (2003) and Bird (2007: 50–59).

law that involves only them. If (as dispositionalists claim) these properties have a dispositional nature in all possible worlds, the aforementioned law would express an internal relation holding between them in all worlds, since any internal relation is determined by the nature of its terms in each possible world. According to CD, however, properties can have different natures in different worlds. Hence, even if the above law expresses an internal relation in the actual world, it would still express an internal relation between those properties only in those worlds where the latter are still dispositional. In any possible world where both properties are non-dispositional, the law could express an external relation, *or could express no relation at all, since in that case nothing determines the existence of the aforementioned relation.* So, according to CD, laws of nature are doubly contingent. A law may express a relation that exists only in some worlds and in those worlds in which it exists it may have different natures.

Vassilios Livanios
University of Patras

vlivan@phs.uoa.gr

References

Armstrong, D.M. (2005) 'Four Disputes About Properties', *Synthese* 144, 309–320.

Bird, A. (2007) *Nature's Metaphysics: Laws and Properties*, New York: Oxford University Press.

Bostock, S. (2003) 'Are All Possible Laws Actual Laws?', *Australasian Journal of Philosophy* Vol. 81, No. 4, 517–533.

Eagle, A. (2009) 'Causal Structuralism, Dispositional Actualism, and Counterfactual Conditionals', IN T. Handfield (ed.), *Dispositions and Causes*, Oxford: Oxford University Press.

Hamermesh, M. (1989*) Group Theory and its Applications to Physical Problems*, New York: Dover Publications.

Heil, J. (2003) *From an Ontological Point of View*, New York: Oxford University Press.

Hendry, R.F. & Rowbottom, D.P. (2009) 'Dispositional Essentialism and the Necessity of Laws', *Analysis* Vol. 69, No 4, 668–677.

Linsky, B. & Zalta, E.N. (1994) 'In Defence of the Simplest Quantified Modal Logic', IN J. Tomberlin (ed.), *Philosophical Perspectives: Logic and Language,* Atascadero: Ridgeview Press, pp. 431-458.

Livanios, V. (2008) 'Bird and the Dispositional Essentialist Account of Spatiotemporal Relations', *Journal for General Philosophy of Science*, Vol.39, No.2, 383-394.

———. (2010) 'Symmetries, Dispositions and Essences', *Philosophical Studies*, Vol.148, No.2, 295-305.

MacBride, F. (1999) 'Could Armstrong Have Been a Universal?', *Mind*, Vol. 108.431, 471–501.

Martin, C.B. (1993) 'Power for Realists', IN J.Bacon, K.Campbell and L.Reinhardt (eds.), *Ontology, Causality and Mind: Essays in Honour of D.M.Arsmstrong*, Cambridge: Cambridge University Press, pp. 175-186.

———. (1997) 'On the Need for Properties: The Road to Pythagoreanism and Back', *Synthese* 112, 193–231.

Mumford, S. (1998) *Dispositions*, New York: Oxford University Press.

Schrenk, M. (2008) 'Hic Rhodos, Hic Salta: From Reductionist Semantics to a Realist Ontology of Forceful Dispositions', IN G. Damschen, R. Schnepf, and K. Stueber (eds.), *Debating Dispositions: Issues in Metaphysics, Epistemology and Philosophy of Mind*, New York, Berlin: De Gruyter.

Strawson, G. (2008) 'The Identity of the Categorical and the Dispositional', *Analysis* 68.4, 271–282.

Wigner, E. (1939) 'On Unitary Representations of the Inhomogeneous Lorentz Group', *Annals of Mathematics* 40, 149–204.

Abstracta 6 : 1, pp. 18 – 41, 2010

THERE IS NO SUCH THING AS REFERENCE FAILURE[1]

Xiaoqiang Han

Abstract

I argue that the idea of reference failure which is frequently mentioned and occasionally argued for in the recent philosophy of language literature is a misnomer at best and incoherent when taken seriously. In the first place, there is no such thing as an empty name or name that fails to name anything, where names are understood as not replaceable by descriptions. In the case of demonstrative reference, because the speaker's perception fixes the referent and the speaker's referential intention is not formed prior to the fixation of the referent, reference is guaranteed. My argument is based on an analysis of the alleged cases of reference failure.

I. Proper Names and Reference Failure

One of the things and perhaps the most important thing we do with language is to use it to talk about the world. We can do so because some words we use are able to be used in such a way that they, as Marga Reimer picturesquely describes, somehow "hook on to things in the world" or "attach to bits of reality" (Reimer 2010). Proper names such as "Marcus Tullius Cicero" and "Barack Obama" are such words; they are often believed to be paradigmatic referring expressions, as they refer to particular objects or individuals in the world.[2]

[1] I wish to thank an anonymous referee for his/her time and consideration and for his/her valid and pertinent comments and suggestions that I examined with great care and that have lead to an improvement in the paper.

[2] I subscribe to Strawson's view that referring is not something an expression does, but something that someone can use an expression to do. (Strawson 1950: 320.) Thus it is the speaker who uses certain expressions, rather than expressions themselves, that refer. That is, it is the language user that "hooks words on to things in the world" or "attaches them to bits of reality". "Marcus Tullius Cicero" and "Barack Obama" are called referring expressions because they are used to refer, not because they themselves refer. But I will continue to speak of "expressions that refer" as a shorthand for "expressions that are used to refer".

But other words like "Pegasus" and "Zeus" are also deemed proper names, hence referring expressions, because they at least purport to refer, although there is nothing in the world to which they actually refer. What "reference failure" describes is presumably this kind of situation: Referring expressions such as "Pegasus" and "Zeus" *fail* to refer to anything. Of course, not all philosophers consider uses of "Pegasus" and "Zeus" as cases of reference failure. Some think that such expressions have bearers which really exist, albeit as abstract objects (van Inwagen 1979: 299-308, Zalta 1983: 277-319). Others also hold that they have bearers, but deny that the bearers of such expressions exist (Reimer, 2001: 491-506). Now if reference is understood so broadly, as it is often suggested in ordinary discourse[3], then there is no such thing as reference failure, for any expression that has meaning or a semantic value refers, that is, refers to whatever the expression means.[4] So in order for the idea of reference failure makes any good sense at all, reference has to be defined as only to things that exist in space and time. It is precisely in this restricted sense of reference that "Pegasus" and "Zeus" may be thought as cases of reference failure, as such words do not "hook on to things in the world" or "attach to bits of reality" in the way "Marcus Tullius Cicero" and "Barack Obama" do.

The concept of reference failure is a teleological one; it contains the sense of a discrepancy between the purpose of a speech act to refer to something and its outcome. "Pegasus" and "Zeus" can perhaps be said to fail to refer, because they are referring expressions, not in the sense of actually referring to some particular objects or individuals, but in virtue of purporting or being intended to refer. It is tempting to think of a referring expression as analogous to an arrow which purports or is intended to hit a certain target,

[3] For instance, one can say that *Sola Fide* refers to the doctrine of justification by faith alone or the term butterfly effect refers to the concept of sensitive dependence on initial conditions in chaos theory.

[4] Kent Bach (2008: 16, note 2) summarizes the view thus, "there is a broad sense in which every expression refers (or at least every expression that has a semantic value that contributes to the propositional content of sentences in which it occurs)."

and which may or may not succeed in doing so. It seems clear that for any meaningful talk of the arrow's success or failure, there must be a target, i.e., something the arrow is aimed at in the first place. One cannot assert with any good sense that the arrow fails to hit anything, just because there is nothing to hit. The arrow's failure to hit its target is a discrepancy between the purpose for the act of shooting the arrow and the outcome of that act. If the analogy holds, it would appear that any meaningful talk of the success or failure of a use of referring expression presupposes the existence of the expression's "target", namely, the referent, which, however, contradicts the above characterization of reference failure: It is the absence of referents that uses of expressions such as "Pegasus" and "Zeus" are said to fail to refer. There is indeed a sense that unlike the well perceived dichotomy of the arrow's hitting a target and its purporting to hit it, an expression's referring to something and its purporting to refer to it collapse into one. In other words, an expression cannot fail to refer to its referent if it has a referent at all; to have a referent is precisely to refer to it.

Thus the very idea of reference failure may just be one of self-contradiction. This is perhaps what Russell meant when he said, "If it [an expression] were really a name the question of existence could not arise, because a name has got to name something or it is not a name, …" (Russell 1956: 243). Calling an expression that refers to nothing but only purports to refer to something (or an expression that names nothing but only purports to name something) a referring expression (or a name) is nothing more than calling a fake passport a passport or a snowman a man. The apparent self-contradiction is really a result of some metaphoric use of the words or a sheer confusion of the different meanings of the words.

However if it makes sense to say that a fake passport fails to be a (genuine) passport, one should also be allowed to say that a non-referring expression such as "Pegasus" or "Zeus" fails to be a "genuine" referring expression, that is, fails to refer. After all,

"Pegasus" or "Zeus", unlike "but", "therefore" or "since", at least seems to purport (or be intended) to refer, despite the fact that it does not actually. One can use a non-referring expression as a referring one, mistakenly of course.[5] Now it should be kept in mind that the idea of reference failure induced by the use of expressions such as "Pegasus" and "Zeus", namely expressions with the appearance of proper names, assumes a direct theory of reference, which holds that a proper name refers to whatever is linked to it in a way that does not require speakers to associate any identifying descriptive content with the name. According to Kripke's version of direct reference theory, reference is initially fixed at a dubbing, after which, the name is passed on from speaker to speaker through communicative exchanges. Subsequent speakers are said to succeed in referring by using a proper name, if they simply "borrow" its reference from the speaker who performed the dubbing, that is, with the intention to use the name to refer to whatever the initial speaker used it to refer to, without having to identify it.[6] Presumably reference failure occurs when a speaker who has "borrowed" an expression intends to use it to refer to whatever it was initially used to refer to, whereas, unbeknownst to her, it was never used by anyone to refer to anything.

---

[5] One can also intend to use an expression to refer deceitfully, that is, to try to deceive others into believing that one is referring. But in order to do that, one must not intend to use that expression to refer.

[6] Kripke (1977) also distinguishes between semantic reference and the speaker's reference. The speaker's reference of an expression is a property of a dated, particular use of the expression, whereas the semantic reference is not tied to a particular use. If reference is understood not as something an expression does, but as something someone does by using an expression, even semantic reference is not independent of its use by speakers, although unlike the speaker's reference, it is independent of particular uses. Michael Luntley (1999: 53) suggests that semantic reference is a property of an expression as *standardly* used to make a judgement, whereas the speaker's reference is a property of an expression as used on particular occasions where the speaker's intention override standard use. The distinction was largely made to solve the problems involving descriptions. Now if the direct reference theory holds, such a distinction does not apply to names, because any particular use of a name is a standard use and the initial dubbing is when the standard is set. To "borrow" an expression is simply to "borrow" the standard of using the expression and one can standardly use a name one "borrows" without having to know what the standard is. Of course I may name my dog "Nietzsche", but that is not a particular use of this expression that overrides its standard use (referring to a certain philosopher). My first use of the expression to refer to the dog sets an entirely different standard.

To be sure, this is not how we usually take such non-referring expressions or "empty names" to be. When we use "Pegasus", "Zeus", "Santa Clause" and "Hamlet", we know or at least suppose that they do not refer to anything. It would be utterly irrational for someone to intend to use "Pegasus" to refer, while at the same time believing that "Pegasus" does not refer. Now if we do not intend to use such expressions to refer in the first place, how would there be reference failure, which requires the speaker believing the existence of Pegasus in order to form an intension to refer to it? Of course there are cases which do seem to fit the above description of reference failure. Suppose that someone, not knowing that "Vulcan" is empty, intends to use it to refer to something. It would seem that there is indeed a discrepancy between the intention to use the expression and the undesired outcome, hence a reference failure. However, to construe the use of such expressions in this way misses the essence of the intention involved. The speaker's intention to use the expression to refer to whatever it was used to refer to by the initial user is part of her intention to use the expression in whatever way it was initially used. If the expression was never used to refer, then the current use of it does not refer either, as the speaker's intention cannot override the reference of an expression which she "borrows". Consider the name "Lao-tzu" which had long been believed to be the name of an ancient Chinese philosopher until the mid-twentieth century when doubt emerged as to whether there was ever a person bearing the name. Suppose that the historicity of Lao-tzu can never be proven. What happens to my use of the expression "Lao-tzu"? Surely it refers if there was indeed a person by that name, or it does not if otherwise. I can always defer my decision as to whether it is a referring expression or not to the future when some conclusive evidence is available. But that decision is no part of my current intention to use the expression.

Since the speaker who uses a "borrowed" expression, name or otherwise, cannot form an intention relevant to the reference of expression, for whether or not it refers is

determined by its initial use, there is no question of whether there is a discrepancy between the speaker's intention to use an expression and the outcome of such a use, and the initial use, therefore, is the only place where reference failure can possibly occur. According to Kripke, reference is initially fixed at a dubbing either by perception or by description. Reference-fixing by perception is, for instance, when a speaker says of a perceived object, "this is to be called 'Lao-tzu'", whereas reference-fixing by description is when a speaker stipulates, "whoever as a single person wrote *Tao Te Ching*" is to be called 'Lao-tzu'", should there existed such a person.[7] If no reference was ever fixed by using "Lao-tzu" either by perception or by description, "Lao-tzu" can only be thought to have been initially associated with a description or a set of descriptions, which identify nothing.

We are often told that one of the major problems for the direct reference theory is how to account for the fact that sentences containing empty names such as "Pegasus" and "Zeus" seem to be meaningful, if the meaning of a declarative sentence is determined by the meaning of its constituents of which such an expression is one, whose meaning is exclusively its referent. Now if it is true that such expressions were only associated with descriptions, the Fregean-Russellian descriptive analysis provides not only the neatest, but also the most reasonable explanation for the meaningfulness and truth valuability of sentences that contain such expressions. According to Russell's theory of descriptions, "Pegasus" is a disguised description and hence replaceable by something like "the winged horse". Under the descriptive analysis, expressions such as "Pegasus" and "Zeus" cause no reference failure, as they have nothing to do with reference, which is exactly how they were created, and the

[7] The idea that the reference of a name is fixed by description is not equivalent to the idea that the reference of a name is mediated by a description.

sentences that contain such expressions express only some general propositions.[8] The term "empty names" is really a misnomer, as the so-called empty names are not names at all.

II. Can There Be Demonstrative Reference Failure?

Reference fixing of a name by perception involves the use of a demonstrative. A person may say to others "let's call this 'Fido'", pointing at a dog in front of her whom she and her audience both perceive. Should there be no dog or anything, why would the person ever use a demonstrative to refer? If no attempt to refer is made, how can there be reference failure? When Russell says that if an expression were really a name the question of existence could not arise, what he calls "name" is a shorthand for his "logically proper name". A "logically proper name" is really a demonstrative, which refers to an object of perception or object of immediate acquaintance. Contrary to Russell whose above remarks may well be taken as a categorical denial of reference failure, contemporary direct reference theorists readily acknowledge its possibility. David Kaplan, for one, claims that there can be empty or vacuous demonstratives (Kaplan 1989a: 490). Although Kaplan maintains that "pure indexicals" (e.g., "I", "here" and "now") are immune to reference failure, such that "I am here now" is true *a priori*, he allows reference failure involving the use of demonstratives (e.g., "this", "that" and "he"). Kaplan defines reference failure in terms of the difference between an expression's character and its content. According to him, all expressions must have two sorts of meaning: character and content. The character of an expression is its linguistic meaning or the rules that govern the use of the expression, and its content is the proposition or propositional component such

[8] Contemporary direct reference theorists or Millianists have proposed ingenious solutions to the problem. While the solutions may well be feasible alternatives to the Fregean-Russellian descriptive account of such expressions, they seem to insist, needlessly, on the coherence of the idea of reference failure: such expressions are referring expressions, albeit of a special kind. See David Braun (1993), Nathan Salmon (1998), etc. However, if they are not referring expressions, which they certainly are not, given the fact that they originate from an association with some descriptions, and are therefore replaceable by them, they do not pose any problem for the direct reference theory.

as a referent expressed by an expression in a context. While for non-indexicals content and character are identical, for indexicals they are not, as the character of indexicals are constant, but their content varies from context to context. Thus, the character of "this" may be defined tentatively as "the object the user of the word is currently pointing at", which is always the same, but its content, the object, may be different, as the same expression can be used in different contexts, such as the time, location, the gesture of the speaker and her intentions. It is Kaplan's contention that unlike pure indexicals, which, whenever used, always have content, demonstratives may be empty or vacuous, that is, without content, if they are used in certain contexts.

What exactly are the contexts in which demonstratives are empty or vacuous, that is, the use of demonstratives fail to refer? In his early piece "Demonstratives", Kaplan mentions three kinds of using an empty or vacuous demonstrative: (1) hallucination; (2) wrong demonstratum (referent of a demonstrative), which is, for instance, when the speaker is pointing to a flower and saying "he" in the belief that one is pointing at a man disguised as a flower; (3) too many "demonstrata", as in the case where the subject is pointing to two intertwined vines and saying "that vine." (Kaplan 1989a: 490-491) It is well known that for the early Kaplan, what distinguishes demonstratives from pure indexicals, in addition to the former being possibly empty, is that a use of the former is accompanied by demonstration, which is "typically, though not invariably, a (visual) presentation of a local object discriminated by a pointing" (Kaplan 1989a: 491), whereas the character or linguistic rules of the latter which govern their use fully determine the referent for each context. Therefore, "A demonstrative without an associated demonstration is incomplete. The linguistic rules which govern the use of the true demonstratives 'that', 'he', etc., are not sufficient to determine their referent in all contexts of use. Something else—an associated demonstration—must be provided" (Kaplan 1989a: 490). An incomplete demonstrative is

not vacuous much as an incomplete definite description ("the prime minister") is not vacuous. A demonstrative is vacuous when the associated demonstration is vacuous.[9]

Now if there are limits to what even the best of intentions can do and it is demonstration that ultimately fixes the referent, as the early Kaplan suggests, an appropriate analogy for demonstrative reference would be a gambler throwing a dice, rather than an archer shooting an arrow. The gambler can be said to succeed in the game insofar as there is always a number turning out after each throw, where success is defined as having a number, any number, not as having the particular number the gambler wishes for. He fails only when, for instance, the dice rolls away and disappears, that is, no number turns out. A classic illustration of this "dice throw" theory of demonstrative reference is provided by Kaplan in another early piece of his, "Dthat", where he imagines a scenario in which he, without turning and looking, points at a picture of Spiro Agnew, which is hanging on a wall behind him, where there used to be a picture of Rudolf Carnap, and says, "dthat is a picture of one of the greatest philosophers of the twentieth century." (Kaplan 1979)[10] By the aforementioned arrow shooting analogy, the Carnap-Agnew example should be read as an instance of reference failure, as the speaker's act of demonstration does not "hit" the intended "target", namely, the picture of Carnap. On the early Kaplan's

[9] It is not obvious that (2) is an instance of reference failure even on the early Kaplan's criterion. It is an instance of reference failure only if the demonstrative "he" is defined as primitive and irreducible to other demonstrative, which Kaplan seems to hold, but is not convincing. For one thing, since "he" has some descriptive content (indicating the object being human and male), it may be said that the speaker fails not in referring, but in describing or predicating. Calling a flower "he" is simply another way of saying "that *man*" or "that which *is a man*". There is no significant difference between referring to a flower as "he" and referring to a cross dressing man as "she", the latter of which, many would acknowledge, is clearly a case of misidentification. As Donnellan has shown, even descriptions can be used referentially such that the sense of descriptions may sometimes be overridden. If "he" is treated as replaceable by a complex demonstrative "that man", the sense of the descriptive component "man" may as well be overridden by the referent of the demonstrative "that". By calling a flower "he", the speaker fails to correctly predicate of the object she is pointing at, but she does not fail to refer to it, because the expression "he" she uses does single out an object, a flower, which she merely wrongly thinks of as a man disguised as a flower.

[10] "Dthat" is a term invented by Kaplan for the demonstrative use of "that".

view, however, since it is the demonstration, and not intention, that fixes the referent, and in this case, the demonstration does single something out, the speaker succeeds in referring.

Despite vigorous defenses by philosophers such as Reimer (Reimer 1991b) and Wettstein (Wettstein 1984), this "dice throw" theory of demonstrative reference, as it stands, seems to encounter the difficulty of reconciling its elimination or minimization of the teleological elements (intentions) in reference with the idea of reference success or failure which presupposes intentions. If the referent of the use of a demonstrative is not what the speaker intends, but only what the demonstration demonstrates, which is not intentional, how can we consider such a reference success without contradicting the basic notion about a successful act that it is the fulfillment of an intention? This difficulty might have been a reason for Kaplan to change his mind.[11] In his "Afterthoughts", Kaplan abandons the "dice throw" theory altogether and argues instead that the demonstration has no bearing on the determination of the referent, which is fixed by what he calls the "directing intention" of the speaker, the speaker's intention to demonstrate a perceived object on which his attention has focused. This move, however, leaves the Carnap-Agnew case largely unexplained—the later Kaplan seems to avoid it on the grounds of its being complex and atypical (Kaplan, 1989b: 582, no. 34). It is at least in part because of the later Kaplan's reticence on the case that his newly adopted intentionism became the target of the early Kaplan's followers such as Reimer and Wettstein, who insist that demonstration plays an essential semantic role and that the speaker's intention is marginal at best.

III. Bach's Solution

From the later Kaplan's point of view, one may think it natural to construe the Carnap-Agnew case as an instance of reference failure, as the speaker's intention to refer to

[11] As Reimer observes and others agree, Kaplan never gives reasons for his newly adopted view.

the picture of Carnap is not fulfilled by his pointing gesture, which instead demonstrates an unintended object, the picture of Agnew. But according to the later Kaplan, the speaker's intention to refer is specifically an intention to "to point at a perceived individual on whom he has focused" (Kaplan, 1989b: 582). Since, as Reimer points out, there is no "perceived" object or individual on whom the speaker has "focused," and therefore, nothing can count as an intention to refer to the picture of Carnap, the Carnap-Agnew case is not an instance of reference failure. Between the "dice throw" theory which is vulnerable to the charge of contradicting the teleological conception of reference and the later Kaplan's inability to handle this "complex" and "atypical" case, Kent Bach proposes a "middle way" to the effect that the speaker intends to demonstrate and actually demonstrates the picture of Agnew (Bach 1992: 297b). Bach contends that demonstration itself is intentional or that what the demonstration demonstrates is always what the speaker intends, although what he intends may not be what he demonstrates. In essence, the intention of the speaker can include the intention to refer to what one is demonstrating (Bach 1992a: 140). To modify the "dice throw" analogy, one can say that the gambler may wish for a particular number, but by committing himself to chance, he also intends whatever comes out of the dice throw. The speaker in the Carnap-Agnew case may have the picture of Carnap in mind when he uses "dthat" to refer. But given his commitment to the rules of communication, he also intends to refer to the picture of Agnew which he is pointing to behind him. The Carnap-Agnew case is an instance of reference success, not because the speaker's act of demonstration accidentally singles something out, but because his intention to refer to something he is demonstrating is fulfilled by the fact that he is actually demonstrating it.

Since demonstration itself is intentional, the use of a demonstrative unaccompanied by a demonstration will then be construed as a misuse of such an expression for its lack of referential intention. In an example provided by Reimer, a speaker, intending to refer to

Fido, says to someone nearby, "that dog is Fido," but fails to accompany her utterance with any sort of ostensive gesture (pointing, nodding, glancing, etc.), due to some sort of sudden, momentary, paralysis (Reimer 1991b: 194). According to Reimer, since no demonstration is made, no referent is fixed. There is a reference failure, simply because nothing is demonstrated. On Bach's account, however, the intention relevant to reference success or failure is not the speaker's intention to refer to Fido, but her intention to refer to the dog she is pointing at (Bach 1992b: 297). Since she failed to point at Fido, the relevant intention is empty. Since the intention is empty, there is no question of whether or not it is fulfilled. This case is similar to Kaplan's intertwined vines example, which on Bach's account, would no longer be an instance of reference failure as well, as it too involves no relevant referential intention: the speaker is unable to distinguish one vine from the other by means of his demonstration.

But how to account for cases in which the speaker intends to demonstrate and refer to an object which he perceives, but ends up demonstrating another? The following is an example also provided by Reimer. Suppose that two dogs, Spot and Fido, have been racing about. "The speaker, focusing on Fido, comes out with an utterance of 'That dog is Fido'. But because the dogs are moving about so quickly, the gesture which accompanies the speaker's utterance, ends up discriminating Spot—rather than the intended Fido, who is by now just out of the range of the speaker's pointing finger." (Reimer 1991a: 181) Instead of intending to refer to something the speaker does not perceive as in the Carnap-Agnew case, the speaker in Reimer's Fido-Spot example intends to refer to and demonstrate what she clearly sees, Fido, which, however, is not what she actually demonstrates. This may appear *prima facie* an instance of reference failure, as the speaker's intention to refer to Fido is not fulfilled. Both Reimer and Bach, however, predict that it is a reference success, yet for different reasons. For Reimer, it is an instance of reference success, because the speaker's

act of demonstration actually picks something out regardless of his intention. For Bach, the speaker can be said to have both failed and succeeded to refer to the dog she intended to refer to, in the sense that she failed with regard to her intention to refer to and demonstrate Fido, as Fido is the dog she intended to refer to, and succeeded with regard to her intention to refer to the dog she is demonstrating. Because only the latter, not the former intention, according to Bach, is the specifically referential intention, the speaker succeeded in referring (Bach 1992a:143).

A specifically referential intention is one which the speaker intends and expects her audience to recognize and rely on in order to identify a certain dog as the referent (Bach 1992a:143). "Such an intention is not fulfilled if the audience fails to identify the right individual in the right way, that is, the one intended in the way intended" (Bach 1992b: 296). So Bach's notion of referential intention, as he himself acknowledges, is not exactly the same as that of later Kaplan (Bach 1992b: 296), which does not take the audience's successful identification of the object intended by the speaker as the determinant of reference success. Demonstrative reference, according to Bach, is speaker's reference, which, as opposed to semantic reference, should be understood in terms of a four-place relation, between a speaker, an expression, an audience, and a referent: A speaker uses an expression to refer his audience to an individual, because "communication is essentially interpersonal affair, and reference by a speaker is part and parcel of an act of communication" (Bach 2008: 16). A demonstrative reference failure is the speaker's failure to demonstrate anything, and the speaker's failure to demonstrate anything is specifically the audience's failure to identify or recognize what the speaker intends them to identify or recognize by means of his demonstration.

It is clear that Bach separates questions of reference success and failure from questions of reference fixing, which are treated by Kaplan and Reimer among others as

overlapping. For Bach, reference fixing concerns how the speaker picks out an object as his intended referent, whereas reference success and failure concerns whether the audience identifies or recognizes the referent the speaker has picked out. This "audience-oriented" conception of reference success and failure, however, is not without problems. For one thing, according to Bach, a reference failure is ultimately the audience's failure to identify or recognize the object the speaker intends them to identify or recognize by means of his demonstration. This, of course, presupposes the prior existence of the speaker's intention which may or may not be fulfilled by the forthcoming audience's effort to identify the intended referent. Since the relevant intention of the speaker is always the intention to refer to the object he is demonstrating, and therefore cannot be overridden by his act of demonstration, there is a sense that his demonstration which is intentional fixes the referent. But what counts as a demonstration that demonstrates an object? If, for instance, the speaker intends to refer to a dog, Spot, which he is said to be pointing at, for what reasons his pointing gesture qualifies as a demonstration of Spot, rather than Fido? Of course he thinks that he is pointing at him. But from the viewpoint of his audience, under some circumstances, he may well appear to be pointing at Fido. Which dog is he pointing at? Perhaps there are criteria based on which objective judgments can be made, criteria akin to those for an arrow's hitting a target. Given Bach's view that the speaker's intention is "audience-oriented", it is Fido that should be the demonstrated dog, because Fido is the one which the audience takes to be what the speaker is pointing at. The Carnap-Agnew case, as Bach construes it, can be read as suggesting that the audience ultimately determines the object of a demonstration. The speaker, not perceiving the picture behind him, is in no position to know by himself whether he is pointing at anything at all or what, if anything, he is pointing at. He takes the picture of Agnew as the one he is pointing at and forms the intention to refer to it, precisely because it is the picture that the audience takes to be what

he is pointing at. In short, if what is demonstrated by the speaker is determined by the audience, it is the audience's decision as to what is demonstrated that fixes the referent. At the end of the day, when a speaker refers demonstratively, he refers his audience to something which is in the mind of the audience. So understood, there can be no such thing as reference failure.

I agree with Bach that communication is essentially an interpersonal affair, and reference by a speaker is part and parcel of an act of communication. But this does entail that a successful demonstrative reference depends on a particular audience's successful identification of the referent, which as I have shown leads the conclusion that the audience is the ultimate authority in fixing the referent. When a speaker uses a demonstrative to refer to an object, he intends his audience to identify the object. It is quite possible that only some in the audience actually identify the object, but some others for some reason, say vision impairment or some particular viewpoint at which the audience is located, do not. It does not seem right to claim that the same reference made by the speaker both succeeds and fails. It is also possible that no one in the audience actually identifies the object, and the speaker's intention to refer to the audience to the object is therefore unfulfilled. However, it is conceivable that had those who did not identify the object normal vision or had been located differently, they would have identified it. The point is, even in the absence of successful identification by the audience, the use of a demonstrative is still an act of communication, and not part of a soliloquy.

So I do not think that Bach's revision of the later Kaplan's intentionism yields promising results. A demonstrative reference failure is not the audience's failure to identify the referent, instead it is the speaker's failure to fix the referent, or to use Kaplan's scheme, a demonstrative expression's lack of content when used in a particular context. The fact that the speaker fails to refer his audience to a certain object by using an expression in no

way implies that the expression itself is empty (lack of content). The expression obtains its content, when the speaker focuses his attention on a certain object which he perceives and to which he intends to refer by a forthcoming act of demonstration, which itself is a mere externalization of the inner directing intention.

IV. There Is No Such Thing as Demonstrative Reference Failure

Now if the speaker's referential intention fixes the referent, what exactly is involved in the speaker's having that intention? First of all, the speaker must perceive the object he intends to refer to, which I take as most obvious. The idea that a successful use of a demonstrative such as "this" can be determined solely by the demonstration gesture is not quite intelligible. Certainly one can point out distant objects with eyes shut and ears plugged, and blind people can point out the moon. But with such "demonstrations" the subject lacks the understanding or knowledge of the thing being "demonstrated", which is an integral part of his intention to refer (Evans 1982: 143). One of the reasons for treating the Carnap-Agnew case as complex and atypical by the later Kaplan is presumably that the speaker, when uttering the demonstrative, is perceiving neither the picture of Carnap nor that of Agnew, which does not fit the descriptions of the typical cases, or cases involving what he calls perceptual demonstratives, where the speaker is perceiving the object he intends to refer to or the speaker's directing intention is aimed at a perceived object (Kaplan 1989b: 583). On the other hand, the attempted reference in the Carnap-Agnew case is not entirely without perceptual basis. It is most likely that the speaker did previously see the picture of Carnap, and in this sense his intention to refer to it can be said to be perceptually based. But unlike "that" in "that was so bright" which is intended to refer to something in the past, say a flash of lightening that just burst through the clouds a second ago, the demonstrative used in the Carnap-Agnew case is intended to refer to an object that currently exists. The problem is

precisely that the intention to refer to something in the "updated" surroundings is solely based on an "outdated" perception, which conveniently assumes that nothing has ever changed. Simply put, the speaker is in no position to form an intention to refer to the picture of Carnap in the first place. Nor, of course, is he in a position to form an intention to refer to the picture of Agnew, even though he manages to point to it, because he is not perceiving and perhaps has never perceived it. Given that the intention in question has no perceptual basis, it does not qualify as a referential intention and the question of reference success or failure, namely whether the intention is fulfilled, does not arise.

Perhaps the case most friendly to the defenders of the notion of reference failure is hallucination, where the speaker intends to refer to something which she perceives or at least thinks she perceives, but which does not exist. When Macbeth utters "this is a dagger", the demonstrative "this" he uses to refer fails to pick out anything—he does not merely mistake something which is really there for something else. Clearly Macbeth's intention to use the word "this" to refer is unfulfilled. However, it must be noted that even in hallucination, there is something that the speaker actually refers to, namely, a certain sense datum. According to Russell, demonstratives such as "this" and "that" are logically proper names and therefore are not subject to descriptive analysis, because they refer directly to sense data, which are objects of immediate acquaintance. Sense data are ontologically neutral in the sense that both ordinary common-sense objects and hallucination images may be constructed from them. Because reference involves only sense data, no reference failure can ever occur. This is precisely what Russell means, when he says that if an expression were really a name the question of existence could not arise. One cannot be mistaken about the existence of a sense datum, because that a sense datum appears to exist is no different from that it exists.

Now if we accept the sense data theory as an explanation for hallucination, "this" used in this context must be understood as referring to a sense datum, such that when Macbeth says "this is a dagger", what Macbeth really states is either (1) "this represents a real dagger", where "this" stands for a sense datum, or (2) "the object represented by this is a dagger", where "this" stands for a sense datum and "the object represented by this", a description or quasi-description (for the ineliminable "this" packed in the phrase), replaces the "this" in the original statement "this is a dagger". Either way, the initial attempted reference to a physical object disappears, and as a result, reference failure in cases of hallucination is analyzed away.

A merit of introducing the sense data theory is that it accounts for the fact that in such a situation the speaker does perceive something, namely, a sense datum, and not sheer nothing, and he intends to use the expression to refer to it, despite the fact he does not know that what he perceives and intends to refer to is merely a sense datum. It may be objected, however, that such a stipulation of the character of demonstratives such as "this" ("expression used to refer to a sense datum the user perceives") and the sense data theory generally seem too far away from how we understand such words of ordinary language, as it requires that whenever we use demonstratives to refer, we always end up with referring to some sense data, which is too much theory laden and counter-intuitive. A possible response to this objection is that the supposed reference failure in hallucination can be explained away in terms of sense data without invoking the sense data theory itself. In other words, that someone in hallucination uses "this" to refer to a sense datum need not entail that he does so in normal circumstances. It is perfectly coherent to take "this" to be an expression that is used to refer to a sense datum in hallucination, but to a physical object in normal

circumstances.[12] In essence, there is no need to retain the restriction required by the sense data theory on the character of the demonstratives. Perhaps an adequate formulation of "this" should be something like "expression used to refer to the thing, whatever it is, the user of the expression perceives and intends to refer to", as such a formulation does not commit the user to the type of things to which it applies and their ontological status.

It would appear that the demonstrative "this" is guaranteed to refer, as long as there is something going on within the perceptual field of the speaker, with or without external stimuli and the speaker intends to refer to it. The problem is, the referent of the use of a demonstrative in the case of hallucination is purely private. Linguists and philosophers generally distinguish the demonstrative use of "this" or "that" from its anaphoric use. While the use of "this" in hallucination is surely not anaphoric, it is not demonstrative either in the sense in which a demonstrative use is generally understood: When one uses a demonstrative to refer, one is able to demonstrate for others what one is referring to by an act of demonstration, whenever it is required. The intention to refer to a sense datum in hallucination, being an intention to refer to a private sensation, cannot be externalized by an act of demonstration serving as an aid to communication. I may say "this really hurts" referring to a pain I am experiencing. But there is no way I can demonstrate the pain for anyone else. All I can do is to indicate it indirectly through grimace or to describe it. Demonstrative reference in this sense must take place in a public space and the referent of a demonstrative must in principle be accessible perceptually to the audience, although they may not be able to successfully identify it in a given occasion.

The question is whether a speaker under hallucination can form an intention that qualifies as an intention to demonstratively refer to something external in the first place.

[12] Some believe that the possibility of hallucinations shows that even normal perception always involves sense data (Robinson 1994: 151-62, Jackson 1977: 50ff.). But a sense data explanation of hallucination by no means entails a sense data explanation of perception in general.

The speaker may sincerely believe that he is perceiving something actually there in front of him, and may intend to use a demonstrative to refer to a real thing. What exactly is his intention? I made the point in the preceding that in order for the speaker to form an intention to refer to an object, he must perceive the object he intends to refer to. This means that if the speaker does not perceive the object, there is no basis for him to intend to refer to it. The perception I talked about is of course normal perception. With normal perception the speaker would surely not intend to refer to things he does not perceive. But when he is under hallucination, he does not know he is, and may therefore indeed intend to refer to a real thing. Now if the thesis that demonstrative referential intention is (normal) perception based holds, the intention formed by the speaker under hallucination cannot be a demonstrative referential intention. Normal perception is something that is presupposed by the intention to refer demonstratively. If the speaker's perception is abnormal as in the case of hallucination, that is, the presupposition is false, his intention (with regard to demonstrative reference) is empty or irrelevant. And if his intention is empty, there will be no reference failure, which requires a non-empty intention. It may be objected that a false presupposition does not boil down to an empty intention. Think of a description for example, one can use "the present king of France" with the non-empty intention of referring to someone even though the presupposition that there is a king of France is false. Now the question we should ask here is what exactly is presupposed by the intention to refer by using a description. If it is the existence of the present king of France, then the intention is surely not empty. However, what is presupposed by the intention to refer to something by using a description is not the existence of a present king of France, but the speaker's correct understanding of the phrase which constitutes the description he is using. If the speaker does not know the meaning of the phrase, his intention to refer has no basis, and is therefore empty. Having a normal perception for the speaker to form an intention to

refer demonstratively is like having an understanding of the phrase for him to form an intention to use a description to refer. If the speaker does not have normal perception, which is presupposed by his intention to demonstratively refer, his intention, if anything at all, is empty.

Finally let's consider another situation which may appear as a better candidate for reference failure: A speaker perceives a rapid succession of many different things, each of which lasts for a period of time short enough to not only disallow an utterance of "this" to complete, but also escape the perceptually focused attention of the speaker. Unlike in hallucination, the speaker in such a situation has normal perception and has things publicly displayed within her perceptual field. It does seem that a reference failure always obtains whenever an attempt is made to use a demonstrative "this" to refer to one of the things in rapid succession. Such a situation was in fact discussed quite extensively by Plato and was treated by him as a dramatization of what he took to be the phenomenal world. In a number of occasions (Plato *Timaeus* 49a6-c7, *Cratylus* 439d and *Theaetetus* 182c1-183b5), Plato describes it as one in which demonstratives such as "this" (*tode*) or "that" (*touto*) cannot be used to refer anything in flux. The constant incessant transformation between the phenomenal stuffs, fire, water, earth and air, makes it impossible to say that any one of them is really one thing (e.g., water) rather than some other. In a well-known passage in the *Timaeus*, Plato claims, "since none of these [fire, water, and etc.] appears ever to remain the same, which one of them can one categorically assert, without engrossment, to be some particular thing, *this* one, and not something else? One can't" (Plato *Timaeus*, 49c7-50a4). That is, in order for someone to use an expression such as "this" or "that" to refer to anything at all, whatever is intended to be referred to must have some sort of stability. Since nothing in flux is stable, it appears, any such reference always fails.

Russell once noted that anything referred to by "this" (which for him is a particular sense datum, not a physical object) should last for at least a minute or two, long enough for anyone who uses "this" to finish talking about it (Russell 1956: 203). Quite certainly, the time needed can be much shorter as far as the utterance of "this" is concerned, as there is no difficulty to utter "this" to demonstratively refer to a flash or a bang that lasts as briefly as only a second or two. However it is not so much the utterance of a demonstrative as the speaker's perception of the intended referent that requires the minimal stability of the intended referent. As I have argued in the foregoing, a successful demonstrative reference is such that the speaker must perceive what she intends to refer to. Nothing in flux can be demonstratively referred to precisely because first and foremost nothing in flux can last long enough to be perceived by the speaker or be focused on by the speaker's perceptual attention. Consider the static images projected successively on the movie screen at the speed of 24 frames per second. Now if the speaker cannot perceptually attend to any of the images, she would not form an intention to refer to it in the first place, and her utterance of "this" is little more than a noise. No reference failure can occur in this situation because of the absence of an intention.

Given that the speaker's perception fixes the referent and that the speaker's referential intention is not formed prior to the fixation of the referent, demonstrative reference is guaranteed. The alleged reference failure in the contexts of hallucination and flux can be analyzed away, when in each case the intention to refer is shown to be empty.

Xiaoqiang Han

University of Toronto

hanxiaoqiang@gmail.com

References

Bach, Kent. (1987) *Thought and Reference*, Oxford: Clarendon Press.

———.(1992a) 'Intentions and Demonstrations', *Analysis* 52, 140-146.

———.(1992b) 'Paving the Road to Reference', *Philosophical Studies* 67 : 295-300

———. (2008) 'On Referring and Not-Referring', IN Jeanette K. Gundel, Nancy Hedberg and Nancy Ann Hedberg (eds.), *Reference: Interdisciplinary Perspectives*, Oxford: Oxford University Press.

Braun, David. (1993) 'Empty Names', *Nous* 27:4, 449-469.

Evans, Gareth. (1982) *The Varieties of Reference*, Oxford: Oxford University Press.

Jackson, Frank. (1977) *Perception: A Representative Theory*, Cambridge: Cambridge University Press.

Kaplan, David. (1979) 'Dthat', IN P. French, T. Uehling and H. Wettstein (eds.), *Contemporary Perspectives in the Philosophy of Language*, Minnesota: University of Minnesota Press, 383-400.

———. (1989a) 'Demonstratives', IN Joseph Almog, John Perry and Howard Wettstein (eds.), *Themes from Kaplan*, New York and Oxford: Oxford University Press, 481-563.

———. (1989b). 'Afterthoughts', IN Joseph Almog, John Perry and Howard Wettstein (eds.), *Themes from Kaplan*, New York and Oxford: Oxford University Press, 582-584.

Kripke, Saul A. (1977) 'Speaker's Reference and Semantic Reference', *Midwest Studies in Philosophy* 2, 255–276.

Luntley, Michael. (1999) *Contemporary Philosophy of Thought: Truth, World, Content*, Oxford and Malden: Wiley-Blackwell.

Plato. (1997) *Timaeus*, *Cratylus* and *Theaetetus*, IN John M. Cooper (ed.), *Plato's Complete Works*, Indianapolis and Cambridge: Hackett.

Reimer, Marga. (1991a) 'Do Demonstrations Have Semantic Significance?', *Analysis* 51: 177–83.

——— . (1991b) 'Demonstratives, Demonstrations and Demonstrata', *Philosophical Studies* 63: 187-202.

——— . (2001) 'The Problem of Empty Names', *Australian Journal of Philosophy* 79, 491-506.

——— . (2010) 'Reference', *The Stanford Encyclopedia of Philosophy*, forthcoming URL = http://plato.stanford.edu/archives/spr2010/entries/reference/.

Robinson, Howard. (1994) *Perception*, London: Routledge.

Russell, Bertrand. (1956) *Logic and Knowledge*, London: George Allen and Unwin Ltd.

Salmon, Nathan. (1998) 'Nonexistence', *Nous* 32, 277-319.

Strawson, Peter. (1950) 'On Referring', *Mind* 59, 320-344.

van Inwagen, Peter. (1979) 'Creatures of Fiction', *American Philosophical Quarterly*, 14 : 4, 299-308.

Wettstein, Howard. (1984) 'How to Bridge the Gap Between Meaning and Reference', *Synthese* 58 : 63-84.

Zalta, Edward. (1983) *Abstract Objects*, Dordrecht: Reidel.

Abstracta 6 : 1, pp. 42 – 64, 2010

UM EXAME DE OBJEÇÕES A RYLE SOBRE O FUNCIONAMENTO DOS TERMOS PSICOLÓGICOS INTENCIONAIS

Filipe Lazzeri

Jorge M. Oliveira-Castro

Resumo

Este artigo apresenta, brevemente, uma perspectiva, baseada em parte na abordagem de Ryle, acerca das funções dos termos psicológicos intencionais, tais como empregados na linguagem ordinária. De acordo com esta perspectiva, termos psicológicos intencionais descrevem padrões conhecidos de comportamento, que são determinados por mecanismos seletivos de causação. Isto é, esses termos descrevem relações entre certas respostas, selecionadas a partir das consequências que elas produzem no ambiente, e contextos de sua ocorrência, aos quais elas se tornam associadas. Não se trata de termos que designem causas internas de um dado comportamento, antes o explicando apenas no sentido de enunciarem que ele era de se esperar, se pudermos identificar seu padrão comportamental maior e o contexto em que ocorre. Passamos, então, a examinar três objeções principais que foram levantadas contra a posição de Ryle, nomeadamente: (a) o desafio de Davidson a perspectivas não causais sobre as explicações em termos de razões; (b) as preocupações de Armstrong em "deixar-se os contrafactuais suspensos no ar"; e (c) a objeção holista ao (equivocadamente) presumido atomismo de Ryle. Procuramos mostrar que nenhuma dessas objeções coloca problemas sérios à perspectiva sugerida.

Abstract

This paper briefly presents an account, partially based upon Ryle's approach, of the functions of intentional psychological terms as they are used in ordinary language. According to this account, intentional psychological terms describe known patterns of behavior that are determined by selective mechanisms of causation. That is, these terms describe relations between certain responses, selected on the basis of the consequences they produce in the environment, and contexts of their occurrence, to which they become associated. Intentional psychological terms do not point to inner causes of a given behavior, but can explain it only in the sense of stating that it could be expected to occur, if we can identify its behavior pattern and the context in which it occurs. We proceed then to examine three main objections that have been raised against Ryle's position, namely: (a) Davidson's challenge to the non-causal accounts of reason-explanations; (b) Armstrong's worries about "leaving the counterfactuals hanging in the air"; and (c) the holistic objection to the (wrongly) presumed Ryle's atomism. We aim at showing that none of these objections pose serious problems to the proposed account.

Introdução

A abordagem de Ryle (1949) sobre o funcionamento dos termos psicológicos comuns, como é sabido, vem, geralmente, sendo considerada pouco plausível, em livros de introdução à filosofia da mente e outros textos que examinam alguns de seus aspectos. Costuma-se alegar contra a abordagem (*inter alia*) ser inaceitável conferir um estatuto não-causal a tais termos[1] e que o caráter holista dos enunciados que eles compõem vai-lhe de encontro[2].

Presentemente, com efeito, abordagens de caráter mentalista sobre os termos psicológicos comuns são predominantes. Segundo esta perspectiva, tais termos, pelo menos em suas funções predicativas (ou atributivas), têm como característica básica de seu funcionamento designar entidades internas (localizadas em uma parte do corpo, geralmente considerada o cérebro) que determinam causalmente os comportamentos que supõem explicar ou predizer. O mentalismo é assumido, de modo explícito ou tácito, como uma premissa básica, diante da qual qualquer perspectiva de tipo não-causal sobre o vocabulário psicológico comum "é" reduzida ao absurdo e desqualificada – como se a premissa fosse óbvia e indiscutível.

Em contraposição à tendência mentalista, este trabalho sustenta uma abordagem sobre o funcionamento de parte dos termos psicológicos comuns que se baseia, fundamentalmente, naquela de Ryle (1949) e no behaviorismo operante. Trata-se de uma combinação de algumas análises dos termos psicológicos intencionais delineadas por Ryle e das linhas gerais da concepção selecionista do comportamento intencional ou proposital defendida pelo behaviorismo operante, respondendo a algumas das objeções principais levantadas a Ryle. Como tal combinação assemelha-se ao behaviorismo teleológico de Rachlin (1994), valemo-nos de alguns raciocínios deste autor e cremos não ser incorreto julgar-se que estejamos simultaneamente apoiando alguns aspectos centrais de sua abordagem.

O trabalho cinge-se aos termos psicológicos intencionais, tais como, por exemplo, 'achar' (no sentido de opinião), 'querer' e 'tencionar', ou seja, àqueles que, conforme se

[1] Cf., por exemplo, Armstrong (1968: 56), Fodor (1975: 2ss) e Braddon-Mitchell e Jackson (2007 [1996]: 42).
[2] Cf., por exemplo, Heil (2004 [1998]: 61-62), Putnam (1964: 673) e Braddon-Mitchell e Jackson (2007 [1996]: 44-45).

tem usualmente reputado, estão associados a fenômenos que exibem a propriedade da intencionalidade (no sentido de Brentano) e são analisáveis sob a forma '*s V* que *p*' (em que *s* está para um organismo ou sistema qualquer, *V* é um verbo intencional e *p* é o complemento deste verbo expressando um "conteúdo proposicional")[3]. Embora a abordagem e as respostas aqui delineadas às objeções estendam-se não só a tais termos, mas também àqueles a respeito de traços de personalidade ou de caráter (como, por exemplo, 'ser inteligente', 'ser organizado', 'ser agressivo', etc.), optamos por fazer tal recorte, porque pelo menos duas das três objeções consideradas não são geralmente levantadas incluindo-se a categoria dos termos para tais traços. Observe-se que não estão em questão predicados para sensações e emoções (como, por exemplo, 'sentir raiva', 'sentir dor' e 'estar com frio'), o que ocorre porque consideramos que possuem algumas feições peculiares; dentre as quais, uma feição não inteiramente disposicional (sem que isso implique que mereçam um entendimento mentalista).

Está em foco primário o emprego predicativo e ordinário dos termos intencionais. A observação de que apenas o emprego predicativo o está deve-se a que, por certo, os termos intencionais nem sempre figuram em predicações. Isto ocorre, marcadamente, em certos casos de enunciados em primeira pessoa que Ryle e Wittgenstein chamam de *manifestações* ('avowals', 'Äusserungen') e que se qualificam como formas daquilo que Austin (1975 [1962]) denomina *proferimentos performativos*. Nossa opção baseia-se no fato de que o foco das objeções que avaliamos está no uso destes termos para se explicar e predizer comportamentos. A observação de que apenas o emprego ordinário está em foco maior deve-se a que tais termos, como é sabido, têm um sentido técnico em algumas discussões em filosofia e pesquisas empíricas.

Uma motivação natural deste trabalho, além de a de considerarmos equivocada a tendência mentalista (em razão de argumentos como aqueles que ressaltamos na seção 1), é a relevância da questão. Tais termos são centrais nas práticas linguísticas ordinárias, e,

[3] Ao longo do texto, a partir daqui, empregamos a expressão 'termos intencionais' e expressões análogas (como 'atribuições intencionais'), ao invés de 'termos psicológicos intencionais' e análogas, sem, no entanto, pressupor que todo termo intencional seja de caráter psicológico. Autores como Dennett (1987) e Millikan (1984) mostram bem, a nosso ver, que a intencionalidade está dispersa na natureza, para além de fenômenos psicológicos.

como salienta Strawson (1992), é de interesse filosófico uma compreensão explícita do funcionamento de categorias assim centrais, para além do domínio simplesmente tácito que delas se tem. Além disso, é sabido que se trata de uma categoria conceitual que permeia várias questões em filosofia da mente e em outras áreas da filosofia. Ademais, em algumas ciências (tal como a psicologia e as ciências sociais), há decisões metodológicas que pressupõem uma posição com respeito ao seu funcionamento. Em outras palavras, é correto dizer-se (nisso seguindo parte de um raciocínio de Ryle[4]) que um mapeamento correto de tal funcionamento, inclusive distinguindo maneiras equivocadas de operar estes predicados, é potencialmente útil para aguçar-se a percepção dos fenômenos a eles verdadeiramente relacionados, evitar-se erros de categoria e ser-se bem guiado em certas decisões metodológicas.

Na primeira seção do trabalho, delineiam-se alguns elementos da abordagem de Ryle, coadunada com a concepção operante, selecionista do comportamento dito intencional ou proposital. É possível que a interpretação que é desta maneira conferida à abordagem de Ryle não lhe seja inteiramente fidedigna; porém, note-se que a preocupação aqui não é de mantê-la estritamente em seus termos e sem modificações. Antes, é a de mostrar que, assim complementada (uma combinação que reputamos ser correta como compreensão da categoria conceitual referida), a abordagem não sofre as objeções provavelmente principais que lhe foram feitas. Nas seções subsequentes, procura-se responder a elas: (a) o desafio de Davidson ao não-causalismo sobre as explicações intencionais; (b) a objeção de Armstrong de que Ryle estaria deixando os contrafactuais das predicações intencionais "suspensos no ar"; e (c) a objeção holista, que é frequentemente feita ao autor. Se nossa argumentação estiver correta, mostramos que, ao contrário do consenso que parece ter sido criado em filosofia da mente, nenhuma destas objeções constitui real dificuldade à sua proposta, ao menos quando assim complementada.

O trabalho não entra em pormenores sobre a concepção operante mencionada, mesmo porque nos é relevante, nele, assumirmos apenas os aspectos mais gerais desta

[4] Cf. Ryle (1949: 7-8).

concepção[5]. A saber, a ideia de que se trata de um comportamento regido por causação seletiva (como o são as espécies biológicas). Assumimos que o comportamento deste tipo é uma entidade funcional e histórica, que se define em termos das consequências ambientais que as respostas emitidas pelo organismo produzem (e não da topografia destas). As respostas variam em suas propriedades, algumas das quais, fazendo frente às constrições ambientais, são selecionadas, passando a ter uma ocorrência mais frequente, enquanto que, outras, destituídas das propriedades relevantes ou produzindo consequências aversivas, passam a ter uma ocorrência menos frequente.

Essa forma de processo de *seleção pelas consequências*, como Skinner (1988 [1981]) o denomina, é responsável pela modelagem, determinação e manutenção dos padrões comportamentais constituídos por tais respostas. Assim, estas são entendidas como produtos da história de interação de respostas passadas com o ambiente maior (e da filogênese que as enraízam). Segundo este modelo, estímulos ambientais presentes (ou contextos), que antecedem a ocorrência delas, não são seus determinantes mais fundamentais, embora eventualmente exerçam controle sobre elas. Aquelas selecionadas tornam-se associadas aos contextos de sua ocorrência, quando a produção das consequências que as selecionaram depende deles. Desta maneira, os contextos que compartilham a característica relevante, seja física ou funcional, passam a constituir ocasião para a futura ocorrência de tais respostas. Às relações entre consequências, respostas e contextos (que, segundo a perspectiva molar que adotamos[6], não são necessariamente contíguas temporalmente), chamamos de *contingências de reforço*, quando as primeiras aumentam a probabilidade de ocorrência das segundas (ou reforçam-nas), e *de punição*, quando, ao invés, diminuam sua probabilidade (ou punem-nas).

1. Sobre o funcionamento das predicações intencionais

> [Q]uando descrevemos as pessoas como estando exercendo qualidades mentais, não estamos nos referindo a episódios ocultos dos quais seus atos e proferimentos manifestos

[5] Sobre ela, cf., por exemplo, Baum (2005 [1994]: cap. 4), Chiesa (1994: cap. 5), Skinner (1969) e Glenn *et al.* (1992).
[6] A respeito da visão molar, cf., por exemplo, Rachlin (1989).

> são efeitos; estamo-nos referindo àqueles atos e proferimentos mesmos. Há, claro, diferenças, cruciais para nossa investigação, entre descrever-se uma ação como sendo realizada de modo distraído e descrever-se uma ação fisiologicamente similar como sendo com propósito, cuidado ou astúcia. Mas tais diferenças de descrição não consistem na ausência ou presença de uma referência implícita a alguma ação-sombra prefaciando encobertamente a ação manifesta. Consistem, ao contrário, na ausência ou presença de certos tipos de asserções explicativas-preditivas testáveis.[7]

> [Q]uando falamos da mente de alguém, não estamos falando de um segundo teatro de acontecimentos de estatuto especial, mas de certas maneiras em que alguns destes acontecimentos de sua vida estão ordenados.[8]

O objetivo desta seção é sumarizar *alguns* traços dos termos intencionais com base na abordagem de Ryle (1949) e no modelo supramencionado, a fim de preparar o terreno para nossa avaliação das objeções à abordagem[9]. Um primeiro traço destacável é o de que se tratam de predicados da categoria do – são aplicados com sentido apenas ao – *organismo* (ou *sistema*) *como um todo*, e não da categoria de partes dele[10]. Diz-se que a pessoa, o cachorro e, de modo geral, certos organismos inteiros (ou, eventualmente, por exemplo, um robô que satisfaça os critérios relevantes) têm este ou aquele atributo intencional, e não (salvo metaforicamente) que mentes, cérebros ou partes de cérebro tenham atributos intencionais. Por exemplo, não se diz que o cérebro de um pombo queira pousar em determinada árvore, mas, antes, que o pombo o queira; e não se diz que algo em uma pessoa pretenda escrever uma obra filosófica extensa, mas, sim, que a pessoa o pretenda. Os predicados em questão e suas negações, simplesmente, por uma questão de gramática, não se aplicam a partes dos sistemas, tal como 'caminhar', 'acordar' e suas respectivas negações não se aplicam a rochas e planetas.

Uma segunda característica é a de que possuem um caráter disposicional. Ou seja, eles são aplicáveis em um momento *t* a um sistema mesmo sem correspondentes ocorrências em *t* que constituem critérios para sua aplicação. Por exemplo, uma pessoa pode ter vários propósitos e opiniões, mas não estar realizando qualquer coisa diretamente

---

[7] Ryle (1949: 25).
[8] Ryle (1949: 167).
[9] Embora falemos dos predicados em questão de uma maneira geral, admitimos que nem todos necessariamente apresentam todos os traços seguintes, na medida em que, sendo parte da linguagem comum, informal, possuem, por vezes, algumas nuanças que escapam a uma arregimentação maior.
[10] Este aspecto conceitual é, mais recentemente, enfatizado e desenvolvido por Bennett e Hacker (2003).

relacionada a tais propósitos e opiniões (como acontece quando está dormindo). Além disso, sua lógica é diferente daquela de expressões para episódios (ou ocorrências), não fazendo sentido se dizer, como destes, que crenças, saberes, propósitos e similares tenham propriedades como a de poderem ser coisas paradas por um momento e continuadas em seguida, poderem ser apontadas ostensivamente como estando em certo lugar e poderem ocorrer sincronicamente[11]. Não se diz que alguém tenha tido duas opiniões e duas expectativas durante dez minutos, depois suspendidas e retomadas uma hora depois, e não se as pode apontar com o dedo dizendo-se "Aqui estão, veja-as!".

Este caráter disposicional de enunciados intencionais não significa que estejam para disposições como entidades de algum tipo. Fazer isso seria deixar-se de levar em conta seus critérios de aplicação, ou tomá-las como expressões referenciais que elas não são[12]. Podemos entendê-lo em termos daquilo que Tanney (2009) qualifica como sendo o funcionamento *contextualizador* de tais enunciados. Eles abreviam disjunções de enunciados hipotético-subjuntivos, da forma (*grosso modo*) "Se a circunstância *S* fosse o caso, então ocorreria (provavelmente) a ação *A*", e servem como "bilhetes para inferência" (*inference-tickets*). Ou seja, são atribuições que explicam ou predizem comportamentos ao sinalizar que determinados contextos são ocasião para certos comportamentos, assim legitimando inferências sobre correlações deste tipo existentes ao longo do tempo. Encaixando comportamentos a contextos aos quais estão associados, supondo-se que a associação é em certa medida familiar, elas tornam inteligível e previsível sua ocorrência.

O critério de aplicação das atribuições intencionais são relações entre comportamentos e contextos. Quando predicamos propósitos, expectativas, etc., olhamos para comportamentos ocorridos nos contextos que lhe dão ocasião, e não para causas (ou supostas causas) interiores ao organismo. São tais relações que decidem a aplicabilidade ou não de determinada atribuição intencional[13]. Por exemplo, se Pedro diz que tem o propósito de escrever uma obra filosófica extensa, julgamos pelo que ele faz. A pergunta que

[11] Cf. também Wittgenstein (1967).
[12] Este equívoco conceitual por vezes é decorrência, dentre outras coisas, da forma lógica enganadora dos enunciados intencionais, semelhante à de expressões referenciais. Sobre tal tipo de influência enganadora, cf. Ryle (1932).
[13] Cf. também Melden (1961).

normalmente seria feita para se saber sobre a veracidade disso averiguaria se sua conduta condiz com determinadas práticas de alguns filósofos. Não condissesse com elas, evidentemente não se diria que Pedro teria tal propósito. Além disso, é comum sermos levado a dizer que alguém quer fazer determinada coisa e tem certa opinião apenas por tê-lo asseverado (às vezes de modo bastante direto, em *avowals*), mas, no futuro, virmos a constatar, com base em sua conduta ao longo do tempo, que, na verdade, suas intenções e crenças eram outras, e, às vezes, a própria pessoa vir a corrigir-se.

Esse ponto fica particularmente claro quando há alguma razão para a pessoa esconder o real propósito de suas condutas, como frequentemente ocorre em um tribunal criminal, por exemplo. Quando o réu se declara inocente de um determinado crime, como o júri decide sobre suas verdadeiras intenções? A decisão baseia-se em fatos sobre as condutas do réu que se encaixam, mais ou menos, em padrões conhecidos de fluxos meio-fim[14]. Se a pessoa comprou uma arma de fogo dois dias antes do crime, havia discutido fervorosamente com a vítima alguns dias antes a respeito de um negócio que envolvia milhões de reais, foi vista no local do crime no horário estimado da morte, não foi capaz de apresentar qualquer álibi para tal horário, a balística confirmou que os indícios são compatíveis com a arma adquirida pelo réu, e assim por diante, não haverá muito espaço para dúvidas quanto à autoria e propósito do crime. E, quando isso ocorre, não tem muita relevância o fato do réu alegar inocência. São suas condutas que são analisadas e avaliadas, com referência aos padrões que conhecemos, de forma geral e não sistemática, sobre como as pessoas costumam agir em certos contextos. Em momento algum é necessário, para se decidir sobre o dolo e a culpabilidade do acusado, investigações relacionadas ao interior dele.

Segundo esta perspectiva, então, é redundante tomar-se os predicados intencionais como remissões a coisas para além das relações entre comportamento e o ambiente maior envolvendo o sistema como um todo. Inferir-se que alguém interage, interagiu e/ou vai interagir de uma determinada maneira, exibindo tais e tais relações comportamentais, *e* tais e tais atributos intencionais, é análogo a proceder como alguém que conhece os prédios e atividades de uma universidade, mas ainda se pergunta por onde está a universidade. O

[14] Cf. também Peters (1958).

termo 'universidade' não diz respeito a algo além destas relações, e, analogamente, os enunciados intencionais não estão para algo além de relações apropriadas entre ações e circunstâncias ambientais. A conjunção de enunciados sobre relações comportamentais e contextuais relevantes estendidas no tempo, levando-se em conta aquelas às quais, no passado, o sistema foi submetido – ou seja, conforme sugerimos, enunciados das contingências de reforço ou de punição –, não dá margem para acréscimos de elementos paralelos como *denotata* das atribuições intencionais.

O fato de alguém exibir comportamentos que costumam constituir critério para determinadas atribuições intencionais sem que verdadeiramente as satisfaçam, como acontece nos filmes, não constitui contraexemplo à abordagem. Deve-se considerar que a diferença entre um ator e a pessoa que ele interpreta reside no fato de os atos de um compõem os padrões de comportamento relevantes, enquanto que, os do outro (o ator), não[15]. Os atos imitativos do ator não estão em harmonia com os demais que realiza em contextos fora de cena, se ele efetivamente não compartilha as opiniões e intenções interpretadas. Em outras palavras, sua conduta estendida no tempo não satisfaria a um teste de condicionais hipotéticos estabelecendo correlações comportamentais e contextuais que a pessoa interpretada satisfaria.

Outra característica das atribuições em questão é a de que elas, pelo menos em um sentido importante, não são causais, ainda que os comportamentos que por vezes explicam, certamente, sejam causados e admitam explicações causais. Elas não são causais no sentido de que, com elas, não estamos fazendo referência a coisas internas que respondem a por que o comportamento ocorre. Elas explicam comportamentos apenas no sentido de que os contextualizam no âmbito do padrão molar de atividades, sinalizando prováveis relações em que entram ao longo do tempo, excluindo outras. (São causais apenas no sentido muito amplo de que sinalizam *indiretamente* para processos históricos de seleção pelas consequências.)

Por fim, naturalmente, não se nega a importância dos estados e processos estruturais subjacentes aos padrões comportamentais. O que se sugere é que o significado destas atribuições, em seu emprego ordinário (ou ainda, o que significa ter um propósito, uma

[15] Cf. Rachlin (1994).

opinião, etc.), não diz respeito às entidades estruturais, mas sim aos próprios padrões. Estes, certamente, têm precondições estruturais – no caso de organismos, neurofisiológicas, herdadas da filogênese, e, além disso, aspectos estruturais moldados pela seleção ontogênica (que passam a atuar como causas próximas). No entanto, não há equivalências, sejam tipo-tipo ou exemplar-exemplar, entre predicações intencionais e enunciados correspondentes aos fatores estruturais. O fato de que haja certos estados e processos cerebrais no momento de ocorrência de uma ação pode ser indicativo de que tais entidades concomitantes sejam precondições dela e, portanto, precondições para a verdade de certas predicações intencionais; mas uma coisa é ser condição para sua verdade, e, outra, ser seus fazedores (ou fatores) de verdade. Ocorre algo análogo aqui à relação entre os átomos do corpo, que são condições para que ele realize certos processos, e os processos mesmos, que se definem em razão de outros fatores.

2. O caráter racionalizador das predicações intencionais requer que sejam causais?

Uma das objeções à abordagem de Ryle é feita por Davidson, dirigida à perspectiva não-causal das "razões":

> Quando perguntamos por que alguém agiu como o fez, queremos ser providos com uma interpretação. [...] Quando aprendemos sua razão, temos uma interpretação, uma nova descrição da ação que realizou, que a ajusta a uma imagem familiar. A imagem inclui algumas de suas crenças e atitudes [...]. Para além disso, a redescrição de uma ação proporcionada por uma razão pode colocar a ação em um contexto social, econômico, linguístico ou avaliativo maior. [...]
> [É] um erro pensar-se que, porque colocar a ação em um padrão mais abrangente a explica, disso entendamos o tipo de explicação envolvida. Falar-se em padrões e contextos não responde à questão de como razões explicam as ações, já que o padrão ou contexto relevante contém tanto a razão como a ação. Uma maneira de explicar um evento é colocá-lo no contexto de suas causas; causa e efeito formam o tipo de padrão que explica o efeito, em um sentido de 'explicar' que entendemos tão bem quanto qualquer outro. Se razão e ação ilustram um padrão diferente de explicação, este padrão deve ser identificado.[16]

Cremos que a objeção pode ser representada, de modo aproximado, da seguinte maneira, restringida às explicações intencionais:

[16] Davidson (1980 [1963]: 9-10).

(1) Se a explicação intencional não fosse de tipo causa-efeito, então, se ela dissesse que uma ação ocorre por causa de atributos intencionais e o contexto de uma ação incluísse os atributos intencionais que o agente tem com relação a esta (a ação), então o tipo de padrão explicativo da explicação intencional não seria inteligível.
(2) Suponha-se que a explicação intencional não seja de tipo causa-efeito. (Hipótese)
(3) Logo, se ela dissesse que uma ação ocorre por causa de atributos intencionais e o contexto de uma ação incluísse os atributos intencionais que o agente tem com relação a esta, então o tipo de padrão explicativo da explicação intencional não seria inteligível. (De (1) e (2), *modus ponens*)
(4) Ora, a explicação intencional diz que uma ação ocorre por causa de atributos intencionais e o contexto de uma ação inclui os atributos intencionais que o agente tem com relação a esta.
(5) Logo, o tipo (não-mentalista) de padrão explicativo da explicação intencional não seria inteligível. (De (3) e (4), *modus ponens*)
(6) Logo, se a explicação intencional não fosse de tipo causa-efeito, então seu tipo de padrão explicativo não seria inteligível. (De (2) e (5), introdução do condicional)

O argumento de Davidson, assim entendido, não é uma tentativa de inferir conclusivamente que as explicações intencionais sejam causais, mas, antes, um desafio à perspectiva geral exemplificada pela abordagem de Ryle de mostrar como enunciados intencionais podem ser explicativos sem que exibam a forma de enunciado para atributo intencional como causa e ação como efeito.

Deixando-se de lado, por um momento, a pressuposição problemática (que se figura na premissa (4), particularmente na segunda parte da conjunção) de que há atributos correspondentes presentes no contexto do comportamento, julgamos que o desafio de demonstrar que explicações intencionais não se constituem em explicações de causa-efeito pode ser enfrentado. Do fato de elas terem uma forma segundo a qual um comportamento ocorre "por causa" de determinados atributos intencionais do agente, não se segue que estes sejam suas causas. Segundo a abordagem aqui proposta, um comportamento é efetivamente explicado desta maneira no sentido de ser contextualizado no âmbito de um padrão (molar

operante) que ele compõe e de uma circunstância que constitui ocasião para resultados tipicamente produzidos sob ela por comportamentos similares. Trata-se de uma forma de explicação teleológica, em que o porquê do comportamento particular é remetido às consequências que comportamentos similares tendem a produzir, sob os contextos apropriados. Silogismos práticos e atribuições "racionalizadoras" análogas têm alguma força explicativa (para fins das práticas comuns), ainda que sem ser propriamente causal, então, na mesma medida em que o comportamento referido é direcionado a consequências típicas e que participam na determinação histórica do padrão maior do qual é parte constitutiva[17].

No caso das explicações intencionais, ao afirmarmos, por exemplo, que "João adquiriu um cavalo árabe porque tem a intenção de participar de enduros equestres", a expressão 'porque' indica, simplesmente, que "comprar um cavalo árabe" é um dos comportamentos típicos de quem pretenda participar de enduros, pois a pessoa precisa possuir um animal de montaria e a raça árabe é considerada a mais adequada para esse tipo de atividade. Desta maneira, indica-se que o comportamento em questão faz parte de um determinado padrão comportamental, e não de outros padrões comportamentais. Está-se negando, por exemplo, que João tenha comprado o animal para presentear sua filha, ou para iniciar um criatório de animais árabes, ou para agradar o vendedor que é um político importante de quem precisa de favores. Tais expressões explicam no sentido de contextualizar uma ação no âmbito de um padrão conhecido de atividades e relações ambientais, excluindo outros, o que faz com que o comportamento apresentado passe a ser esperado ou previsível[18].

A objeção em questão alega não haver inteligibilidade em uma maneira não-causal de enunciados intencionais desdobrarem um papel explicativo provavelmente em razão do

[17] Adotamos, aqui, uma interpretação da forma de explicação teleológica proposta por Wright (1972). Em Rachlin (1994), encontramos esta forma de explicação delineada de maneira um pouco diferente, a partir da noção aristotélica de causa final. Os padrões de comportamento que as predicações intencionais conotam são entendidos por ele como "causas" finais de padrões menos molares e das atividades singulares que os constituem. Cf. também Dutra (2006).

[18] Outros níveis de explicação podem ser necessários se houver interesse em indagar-se a respeito do modo em que João veio a aprender que cavalos árabes são os mais aptos para a prática de enduro equestre ou de como João veio a tomar gosto pelo esporte. Esses outros níveis podem envolver explicações selecionistas para o comportamento de João.

pressuposto, correspondente a uma parte da premissa (4), de que os termos intencionais estão para coisas presentes. Embora tal premissa possa estar subjacente a algumas das análises não-causalistas indagadas por Davidson, não o está na abordagem de Ryle, ou pelo menos na perspectiva aqui sugerida nela baseada. Como salientado na seção precedente, tais termos (em seu emprego predicativo) dizem respeito a padrões de comportamento estendidos no tempo, e, assim, não é correto dizer-se que estejam para entidades ou propriedades presentes em um momento particular.

3. Os fazedores de verdade das predicações intencionais têm de ser estados internos?

Considere-se agora a objeção de Armstrong. O autor assim a expressa:

> Ora, não há dúvida de que enunciados verdadeiros que atribuem disposições a objetos na ausência de uma manifestação [...] regularmente envolvem a verdade de contrafactuais adequados. Mas Ryle, por assim dizer, deixa seus contrafactuais suspensos no ar ['leaves his counterfactuals hanging in the air']. Quero dizer com isso que ele parece pensar que não precisa dizer o que há no mundo que faz tais contrafactuais serem verdadeiros. No lugar de disposições não-manifestas, ele conecta ['he plugs in'] contrafactuais, mas não diz nada sobre aquilo que faz com que tais contrafactuais sejam verdadeiros.
> [...] Qual é o fundamento na realidade – o que alguns filósofos hoje em dia chamam de *fazedor de verdade* – para a verdade deles? Pareceria que uma parte essencial do fazedor de verdade deve ser um estado interno apropriado da coisa que está na disposição. Mas em tal caso, por que não identificar a disposição com o estado interno?[19]

Armstrong está de acordo com Ryle que as predicações intencionais deixam-se analisar em termos de condicionais contrafactuais que relacionam contextos e comportamentos. No entanto, o autor pensa que Ryle não estabelece os fazedores de verdade destes enunciados, por não fazer com que estes se refiram a estados internos. O fazedor de verdade de um contrafactual constituinte de uma predicação intencional teria de ser um estado interno porque, do contrário, o contrafactual ficaria "suspenso no ar". Esta objeção pode ser entendida, então, como uma forma de desafio de mostrar-se como tais predicações podem ter fatores para sua veracidade que não sejam estados internos.

Como dito acima, o caráter disposicional dos enunciados em questão, segundo a abordagem, não significa que eles estejam para disposições entendidas como algum tipo de

[19] Armstrong (1999: 64; grifo do autor).

entidade, mas que desempenham suas funções de uma maneira contextualizadora. Trata-se de uma abreviação de certas relações históricas e indução de outras similares, assim permitindo inferências e cobranças de comportamentos que nelas se encaixem, e não de uma maneira de reportar fatos específicos. Por certo, é característica da lógica dos enunciados intencionais serem eles verdadeiros ou falsos quando não há comportamentos relevantes exibidos em um momento particular. Sugerimos que isso se deve, justamente, por aquilo que determina sua verdade serem padrões de comportamento, estendidos no tempo. (Nessa medida, pode-se considerar que a abordagem estabelece os fazedores de verdade destes enunciados, no sentido amplo de um "fundamento na realidade" para sua verdade.)

Um possível diagnóstico da suposição de Armstrong é o de que não atenta suficientemente para os critérios de aplicação dos termos em questão. Ao avaliarmos a aplicabilidade ou não de determinado enunciado intencional, são basicamente os comportamentos que o decidem. Fundamentamos nossas decisões a esse respeito com base nas condutas, as quais devem se encaixar em determinados padrões mais ou menos conhecidos, frequentemente em termos daquilo que funciona como meio para que fins sejam alcançados em determinada sociedade. Se Pedro diz (de maneira verídica) que tem o propósito de escrever uma obra filosófica extensa, pode-se esperar que exiba um certo padrão de ações que funcionam como meio para atingir esse objetivo, tais como ler muito e escrever vários textos de filosofia. Se essas atividades não ocorrerem, dificilmente poder-se-ia afirmar que Pedro tenha realmente essa intenção, e em momento algum seria necessário investigar o que acontece no interior de Pedro, neurológica ou psicologicamente.

Armstrong (1999: 62)[20], em sua argumentação, evoca certos casos extremos, nos quais um organismo satisfaria uma predicação intencional mesmo que em nenhum contexto apropriado o organismo realizasse os comportamentos esperados. Isso supostamente refutaria a tese de que correlações comportamentais-ambientais são o amparo dos enunciados subjuntivos referidos, pois, dessa maneira, a veracidade deles ficaria sem ser explicada. Por exemplo, dado o argumento de Armstrong, seria, em princípio, possível

[20] Cf. também Armstrong (1968: 71-72).

aplicar corretamente uma atribuição intencional a uma pessoa totalmente paralítica. Assim entendida, a objeção pode ser representada, aproximadamente, da seguinte maneira:

(1) Se predicações intencionais pudessem ser verdadeiras quando não houvesse qualquer comportamento sendo realizado em quaisquer contextos que lhe são ocasião, então elas não teriam como fazedores de verdade correlações entre comportamentos e contextos.
(2) Ora, elas podem ser verdadeiras quando não há qualquer comportamento sendo realizado em quaisquer contextos que lhe são ocasião.
(3) Portanto, elas não têm como fazedores de verdade correlações entre comportamentos e contextos. (De (1) e (2), *modus ponens*)

Consideramos que esta objeção não atinge à abordagem de Ryle porque a premissa (2) é implausível. É preciso considerar-se as ações em sua dimensão temporal, em uma escala possivelmente larga de tempo. Para um organismo paralítico, a veracidade de uma predicação do tipo em questão é, ainda e efetivamente, uma questão de realização de comportamentos. Se o organismo estivesse nos contextos apropriados, ele provavelmente realizaria aqueles que estaríamos legitimados a esperar. Se, por exemplo, uma pessoa quisesse alguma coisa, ela provavelmente pediria a alguém que lha trouxesse, ou faria algum gesto ou alguma outra coisa que sinalizasse para isso. Se se tratasse de um caso em que, por anos e anos, a pessoa estivesse totalmente paralisada, sem poder executar *qualquer* ato, a legitimidade das atribuições intencionais à pessoa seria questionável. Não há a possibilidade de um organismo ter, por exemplo, o desejo de beber água e não se comportar de modo a obter água em pelo menos algumas das ocasiões propícias, pois é constitutivo do significado de ter-se o desejo de beber água (é critério de aplicação das atribuições respectivas) haver condutas que conduzem a tal consequência (reforçadora). Dado que não há qualquer sinal dessa conduta (sob contextos que lhe seriam ocasião), não há desejo de beber água.

Em alguns casos a pessoa pode, apesar de ter desejo de beber água, não emitir os comportamentos típicos em um dos contextos. Por exemplo, quando ela está se preparando

para fazer um exame médico que exija jejum total. O que ocorre neste tipo de caso é que o desejo de beber água e a intenção de fazer o exame corretamente "estão se contrapondo". Mas o fato de dois ou mais atributos intencionais oporem-se não faz com que se tornem ocorrências internas. Mesmo nesse tipo de caso, comportamentos relacionados ao desejo de beber água provavelmente ocorreriam, pois a pessoa provavelmente comentaria que sente sede e não pode beber. No caso de um animal não-humano, por exemplo, quando se afirma que "o rato deseja beber água, mas teme levar um choque", não se está a inferir estados internos, mas a descrever uma história de interação do organismo com seu ambiente, a qual observamos ou conhecemos de outra forma.

No entanto, se, mesmo na ausência de obstáculos ou tendências contrárias identificáveis no ambiente, o rato não mais beber água, o que diríamos? Comprovaríamos que é possível haver uma intenção interna, não formada por comportamentos, ou, como diria Austin (1946), não saberíamos o que dizer? Parece-nos que, nessa situação hipotética, tenderíamos a modificar os nossos conceitos e distinguir ratos que bebem água de ratos que não bebem. Os exemplos extremos de Armstrong se assemelham a esse caso, pois com o passar do tempo e com a ausência total de atos, as pessoas não saberiam mais o que dizer a respeito de expectativas e desejos de uma pessoa completamente paralítica. A objeção do autor, portanto, deixa de levar em conta os critérios de aplicação dos predicados em questão em seu emprego ordinário.

4. A cadeia inferencial de predicações intencionais implica o mentalismo?

Por fim, há a conhecida objeção holista, a mais comumente mencionada como barreira à abordagem. A objeção é levantada nos termos seguintes:

> Mas há realmente qualquer comportamento característico de uma dada crença? Pode uma ação ser descrita como "agindo como se você tivesse uma tal e tal crença", a não ser que tomemos por certo, ou estivemos de alguma maneira especialmente informados sobre, as necessidades e quereres do agente?[21]
>
> Qualquer tentativa de dizer qual comportamento segue-se de um dado estado mental pode ser mostrada como falsa pela invenção de um exemplo no qual o estado mental está

[21] Geach (1957: 8).

> presente, mas, devido à adição de novas crenças ou desejos, o comportamento não se segue. Não adiantará tentar impedir tais casos através de uma cláusula geral: [fixar-se que] se você acha que há um urso no caminho, acha que ursos são perigosos e deseja evitar animais perigosos, estará disposto a fugir. O problema aqui é que se reintroduz menção a estados mentais na cláusula. Estas são precisamente as coisas que se estava tentando reduzir pela análise.[22]

A objeção afirma que os enunciados intencionais andam implicitamente sempre juntos, não isoladamente, de modo que não é possível caracterizar um deles sem introduzir outros. Um seria verdadeiro sobre um organismo ou sistema apenas se vários outros o fossem ao mesmo tempo sobre ele. Por exemplo, uma pessoa querer fugir de ursos constitui uma predicação verdadeira apenas se for verdade também que suspeita que haja ursos no caminho e teme serem-lhe danosos, e, por sua vez, tais suspeita e temor pressupõem que a pessoa tenha certas crenças, desejos, etc. e não tenha contrários. No entanto, a abordagem de Ryle estaria tentando reduzir tais enunciados a outros que não os introduzissem, assim indo de encontro com esta feição holista. A objeção é de que, pelo fato destas predicações andarem implicitamente juntas (ou em massa), seguir-se-ia que uma específica poderia ser analisada em termos de um determinado conjunto de condicionais subjuntivos apenas dada a presença de certas entidades mentais correspondentes.

Poder-se-ia, então, reconstruir assim:

(1) Se as predicações intencionais se caracterizassem (holisticamente) em conjunção umas com outras, então elas designariam entidades mentais correspondentes.
(2) Se elas designassem entidades mentais correspondentes, então não poderiam ser analisadas em termos de condicionais subjuntivos relacionando contextos e comportamentos sem fazer menção a entidades mentais internas correspondentes.
(3) Ora, elas caracterizam-se (holisticamente) em conjunção umas com outras.
(4) Logo, elas não podem ser analisadas em termos de condicionais subjuntivos relacionando contextos e comportamentos sem fazer menção a entidades mentais internas correspondentes. (De (1)-(3), sorites)

[22] Heil (2004 [1998]: 61-62).

Esta objeção revela-se equivocada por pelo menos duas razões. Em primeiro lugar, ela comete petição de princípio. Há nela a pressuposição, como se figura no consequente do condicional representado na premissa (1), de que predicações intencionais designem (supostas) causas internas correspondentes. Ou seja, tal objeção pressupõe de antemão o mentalismo a respeito delas (que questionamos). Por si só, isso releva que, apesar de ser muito frequentemente apontada à abordagem Ryle, ela não é bem colocada.

Em segundo lugar (e isso é também um diagnóstico de tal petição de princípio), é preciso dizer-se que a abordagem, diferentemente de um "behaviorismo lógico" como aquele de Hempel (1980 [1935]), não tem uma motivação reducionista, no sentido de que abraçasse como uma finalidade reduzir as atribuições intencionais a atribuições que não as incluísse. O que Ryle defende, e que está sendo aqui defendido, é uma tentativa de clarificação do funcionamento da linguagem intencional ordinária, e não reduzir o que supostamente fosse mental e interno a algo público. Ora, não há problema com a ideia de que a linguagem intencional possua um traço holista – as cadeias inferenciais de predicações intencionais fazem parte das regras de operação dela –, na medida em que, sendo isso um fato a ela inerente, é algo que se pretende capturar.

O traço holista deste "jogo de linguagem" é um fato, mas não se segue disso, entretanto, o consequente do condicional expresso em (1). Em outras palavras, do fato de estarmos legitimados a inferir outras predicações intencionais a partir de uma, não se segue que elas digam respeito a algo além de correlações entre comportamentos, contextos e consequências. Normalmente, não é preciso ir-se muito além em predicações tais na explicação ou predição de um comportamento, mas, em todo caso, ir-se adiante em tais inferências é ir-se adiante em contextualizar a conduta do organismo em seu padrão maior. A imagem holista sugerida por autores como Dennett (1978, 1987) pode ser admitida livre do elemento mentalista que lhe infundem.

Tomando o exemplo de Heil, dizer-se que uma pessoa queria fugir de ursos e que suspeitava que houvesse ursos no caminho pode significar, por exemplo, que ela se comportaria de modo a fugir de ursos, se estivesse em circunstâncias que sinalizassem a provável presença de ursos, e que, efetivamente, ela estava em uma circunstância deste tipo. Se fosse perguntado à pessoa por que tinha suspeitado que houvesse ursos no

caminho, é bastante provável que diria o suspeitar por ter visto certos sinais que eram indícios disso, dado que passou por isso ou casos similares no passado. Podemos imaginar também que diria temer ursos por que achava que eles eram perigosos, já os tendo visto atacar presas ou ouvido falar que eles eram perigosos. Desta maneira ("racionalizando" o comportamento da pessoa), estar-se-ia, indiretamente, remetendo mais e mais às contingências de reforço e de punição de seu repertório comportamental.

5. Conclusão

Este trabalho apóia a tese de que os termos intencionais, em seu uso ordinário, desempenham suas funções de explicar e predizer comportamentos de uma maneira diferente daquela que as abordagens mentalistas defendem ou assumem: formam enunciados que não funcionam pela designação de entidades internas causadoras das ações. De modo positivo, apóia-se a tese de que o fazem sinalizando padrões de comportamento de sistemas inteiros – a qual encontramos, sob certa interpretação, em Ryle (1949), e, mais recentemente, de modo semelhante, em autores como Rachlin (1994). Os enunciados intencionais, em seu uso predicativo, incidem sobre o sistema como um todo, no contexto molar (estendido no tempo) de correlações entre sua conduta e determinadas circunstâncias das quais efeitos controladores são contingentes e lhes constituem ocasião. Desta maneira, tais enunciados servem como bilhetes que nos dão licença para inferir e aguardar ocorrências similares sob circunstâncias similares.

Um dos possíveis problemas maiores com o mentalismo, em muitas de suas versões usuais, tratando-se de uma ideia de seu bojo, é a concepção de comportamento dito intencional ou proposital como comportamento determinado, fundamentalmente, por causas endógenas. Mesmo autores que dão alguma proeminência a aspectos da seleção do comportamento pelas consequências, como Dretske (1988) e Millikan (1993), situam a agência em tais causas, às quais correlacionam (cada um ao seu modo) as predicações intencionais. Nisso, obliteram a redundância que tal correlação implica, e, em alguns casos, deixam de levar em conta a questão de se há coerência conceitual na correlação (coerência com as regras de operação da linguagem intencional; por exemplo, com a mereologia destas predicações). De um ponto de vista operante, como bem enfatiza Rachlin (1994), as causas

endógenas dão-nos apenas respostas a perguntas sobre *como* tais comportamentos ocorrem, e não a perguntas sobre seu *porquê*, que requerem visualização de (por vezes complexas) dinâmicas históricas de *feedback* entre o organismo e seu ambiente, em conjunção com circunstâncias presentes.

Procurou-se mostrar, neste trabalho, que nenhuma das objeções consideradas se revela cogente[23]. O desafio colocado por Davidson, perguntando pela inteligibilidade de as explicações intencionais não se referirem a causas, tem subjacente a suposição de que os predicados intencionais estão para atributos presentes no contexto de uma ação, o que a abordagem que defendemos justamente coloca em questão. Sugerimos que o desafio pode ser superado à luz de uma premissa selecionista apropriada sobre as causas do comportamento, tal como enfatizada pela tradição operante em psicologia. As predicações intencionais explicam teleologicamente, contextualizando o comportamento em seu padrão estendido. Sua força explicativa deriva do fato de que as consequências às quais o comportamento se direciona são fatores causais históricos dele e de que os contextos de sua ocorrência nos permitem induzir eventos familiares deste tipo.

A objeção de Armstrong de que os contrafactuais das predicações intencionais têm de ter como fazedores de verdade estados internos deixa de levar em consideração os critérios de aplicação dos termos intencionais em seu emprego ordinário. Ademais, a objeção assume que tais enunciados funcionam referencialmente, como qualquer outro enunciado de estado de coisas. Então, no caso de nenhuma ação estar sendo realizada em determinado momento, eles teriam de estar para algum fato recôndito, essencialmente neurofisiológico ou similar.

Por sua vez, a objeção holista presume que Ryle teria proposto caracterizações reducionistas destas predicações, o que é implausível. Em todo caso, sua abordagem é plenamente compatível com a feição holista delas. É verdade que, a partir desta perspectiva, a linguagem em questão diz respeito a formas de interação comportamental e ambiental, mas isso não implica que seja redutível a uma linguagem mais básica. (A teoria dos sistemas intencionais de Dennett contribui para fazer ver ambas as coisas, com

[23] Uma quarta objeção principal – que encontramos formulada, dentre outros lugares, em Davidson (1984) – questiona o entendimento que a abordagem propõe do estatuto do autoconhecimento. Este é um tema que pretendemos tratar em outro trabalho.

independência do elemento mentalista que acresce à abordagem de Ryle[24].) Ademais, o argumento pressupõe tacitamente que, desta feição, se seguiria que elas funcionam pela designação de entidades mentais internas. Explicitando esta premissa, mostramos que a objeção comete petição de princípio e, logo, que não é uma objeção bem colocada.

Filipe Lazzeri
Universidade de Brasília
filipelazzeri@gmail.com

Jorge M. Oliveira-Castro
Universidade de Brasília
jocastro@unb.br

Referências

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, London: Routledge.

Armstrong, D. M. (1999) *The Mind-Body Problem: An Opinionated Introduction*, Boulder, CO: Westview Press.

Austin, J. L. (1946) 'Other Minds', *Proceedings of the Aristotelian Society*, Supplementary Volume 20: 148-187.

Austin, J. L. (1975 [1962]) *How to Do Things with Words*, 2nd ed, J. O. Urmson & M. Sbisà (eds.), Cambridge, MA: Harvard University Press.

Baum, W. M. (2005 [1994]) *Understanding Behaviorism: Behavior, Culture, and Evolution*, 2nd ed., Malden, MA: Blackwell.

Bennett, M. R. & Hacker, P. M. S. (2003) *Philosophical Foundations of Neuroscience*, Oxford: Blackwell.

[24] Cf. Lazzeri (*forthcoming*).

Braddon-Mitchell, D. & Jackson, F. (2007 [1996]) *Philosophy of Mind and Cognition: An Introduction*, 2nd ed., Oxford: Blackwell.

Chiesa, M. (1994) *Radical Behaviorism: The Philosophy and the Science*, Boston: Authors Cooperative.

Davidson, D. (1980 [1963]) 'Actions, Reasons, and Causes', IN *Essays on Actions and Events* (1980), Oxford: Oxford University Press, 3-19.

Davidson, D. (1984) 'First Person Authority', *Dialectica* 38 (2-3): 101-11.

Dennett, D. (1978) *Brainstorms: Philosophical Essays on Mind and Psychology*, Cambridge, MA: MIT Press.

Dennett, D. (1987) *The Intentional Stance*, Cambridge, MA: MIT Press.

Dretske, F. (1988) *Explaining Behavior: Reasons in a World of Causes*, Cambridge, MA: MIT Press.

Dutra, L. H. de A. (2006) 'Comportamento Intencional e Contextos Sociais: Uma Abordagem Nomológica', *Abstracta* 2 (2): 102-128.

Fodor, J. A. (1975) *The Language of Thought*, New York: Crowell.

Geach, P. T. (1957) *Mental Acts: Their Content and their Objects*, London: Routledge.

Glenn, S. S., Ellis, J. & Greenspoon, J. (1992) 'On the Revolutionary Nature of the Operant as a Unit of Behavioral Selection', *American Psychologist* 47 (11): 1329-1336.

Heil, J. (2004 [1998]) *Philosophy of Mind: A Contemporary Introduction*, 2nd ed., New York: Routledge.

Hempel, C. G. (1980 [1935]) 'The Logical Analysis of Psychology', W. Sellars (trans.), IN N. Block (ed.) (1980), *Readings in Philosophy of Psychology*, Vol. 1, Cambridge, MA: Harvard University Press, 14-23.

Lazzeri, F. (*forthcoming*) 'Um Balanço de Parte da Teoria dos Sistemas Intencionais de Dennett', *Psicologia: Teoria e Pesquisa*.

Melden, A. I. (1961) *Free Action*, London: Routledge.

Millikan, R. G. (1984) *Language, Thought, and Other Biological Categories: New Foundations for Realism*, Cambridge, MA: MIT Press.

Millikan, R. G. (1993) *White Queen Psychology and Other Essays for Alice*, Cambridge, MA: MIT Press.

Peters, R. S. (1958) *The Concept of Motivation*, London: Routledge.

Putnam, H. (1964) 'Robots: Machines or Artificially Created Life?', *Journal of Philosophy* 61 (21): 668-691.

Rachlin, H. (1989) *Judgement, Decision, and Choice: A Cognitive/Behavioral Synthesis*, New York: Freeman.

Rachlin, H. (1994) *Behavior and Mind: The Roots of Modern Psychology*, New York: Oxford University Press.

Ryle, G. (1932) 'Systematically Misleading Expressions', *Proceedings of the Aristotelian Society* 32: 139-170.

Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson.

Skinner, B. F. (1969) *Contingencies of Reinforcement: A Theoretical Analysis*, New York: Appleton-Century-Crofts.

Skinner, B. F. (1988 [1981]) 'Selection by Consequences', IN A. C. Catania & S. Harnad (eds.) (1988) *The Selection of Behavior*, New York: Cambridge University Press, 11-20.

Strawson, P. F. (1992) *Analysis and Metaphysics*, Oxford: Oxford University Press.

Tanney, J. (2009) 'Reasons as Non-Causal, Context-Placing Explanations', IN C. Sandis (ed.) (2009), *New Essays on the Explanation of Action*, Hampshire, UK: Macmillan, 94-111.

Wittgenstein, L. (1967) *Zettel*, G. E. Anscombe & G. H. von Wright (eds.), G. E. Anscombe (trans.), Oxford: Blackwell.

Wright, L. (1972) 'Explanation and Teleology', *Philosophy of Science* 39 (2): 204-218.

Abstracta 6 : 1, pp. 65 – 70, 2010

CONTRA A NECESSIDADE METAFÍSICA DA LEI "O SAL SE DISSOLVE EM ÁGUA"

Rodrigo Reis Lastra Cid

Resumo

Neste artigo, pretendo argumentar contra a tese de Alexander Bird (2001) de que a lei *o sal se dissolve na água* é metafisicamente necessária. Indico brevemente qual é o argumento de Bird a favor da necessidade de tal lei e, posteriormente, provejo um contra-argumento à sua tese. Bird, que de modo geral, quer mostrar que a existência de certas substâncias depende da veracidade de certas leis e que, por isso, a existência de tais substâncias implica a verdade de tais leis. Isso faria que as leis existissem sempre que existissem as substâncias que elas regulam; o que, segundo Bird, faria tais leis metafisicamente necessárias. Meu contra-argumento a Bird é que tal concepção apreende apenas o que chamamos de "necessidade fraca", e não a necessidade forte que esperaríamos de uma lei metafisicamente necessária.

Abstract

In this paper, I intend to argue against Alexander Bird's thesis (2001) that the law *salt dissolves in water* is metaphysically necessary. I briefly indicate Bird's argument for the necessity of such law, and then I provide a counter-argument to his thesis. In a general way, Bird wants to show that the existence of certain substances depends on the truth of certain laws, and that because of this the existence of such substances implies the existence of such laws. That would make the laws existing at least while the substance it rules exists; what, for Bird, makes such laws metaphysically necessary. My counter-argument to Bird is that such conception apprehends just what we call "weak necessity", and not the strong necessity we would like a metaphysically necessary law to have.

Alexander Bird (2001) sustenta que a lei "o sal se dissolve em água" é metafisicamente necessária, ou seja, existe em todos os mundos metafisicamente possíveis. No entanto, ele faz uma importante ressalva: a lei "o sal se dissolve em água" é metafisicamente necessária quando vige em todos os mundos metafisicamente possíveis nos quais sal e água existem. Daí, ele nos diz que o que permite a ligação eletrostática que forma as moléculas de sal e de água é a lei de Coulomb, e que o que permite que o sal se dissolva na água é também a lei de Coulomb – já que é ela que regula as ligações eletrostáticas em cada uma dessas moléculas e que a dissolução é um processo eletrostático.

O argumento para sustentar a necessidade da lei *o sal se dissolve na água* procede por redução ao absurdo: parte-se da premissa de que essa lei é contingente e chega-se a uma conclusão contraditória. O argumento é o seguinte: se a lei *o sal se dissolve na água* é contingente, então há um mundo onde, embora as condições estejam presentes[1], o sal não se dissolve na água; se há um tal mundo, então nesse mundo a Lei de Coulomb é falsa, pois a Lei de Coulomb é o que, dadas as condições adequadas, permite a dissolução do sal na água. Mas para a existência do sal é necessário que a Lei de Coulomb seja verdadeira, pois, além da dissolução, é ela que regula a magnitude da atração eletrostática entre os átomos num íon como o sal – pois, no geral, a Lei de Coulomb regula a atração eletrostática entre corpos carregados –; logo, um mundo no qual sal e água existem, mas o sal não se dissolve na água, embora as condições estejam presentes, é um mundo no qual a Lei de Coulomb é falsa (porque a dissolução não ocorre) e verdadeira (porque o íon de sal existe). Assim, diz Bird (2001, p. 271), "não há tal mundo; e, com isso, a pressuposição de que é contingente que o sal se dissolve na água é refutada" (p. 271).

A principal objeção contra este argumento é dizer que um composto iônico como o sal não requer a existência da Lei de Coulomb: pode ser o caso que outra lei, semelhante à de Coulomb, mas com valores diferentes para as constantes, permita a existência do sal e de comportamentos similares aos atuais para aos objetos eletricamente carregados. A resposta de Bird a essa objeção dá um passo além em seu argumento, pois o generaliza para quaisquer conjuntos de leis que regulamentem o comportamento de certo tipo de substâncias. A resposta é a seguinte: se houvesse uma lei semelhante à de Coulomb, ela regeria o comportamento dos objetos eletricamente carregados, ou seja, regulamentaria tanto o processo de dissolução do sal na água, como a existência do próprio íon de sal (pela ligação entre os átomos de cloro e sódio) e, portanto, seria necessária: em qualquer mundo onde os objetos por ela relacionados existissem, ela seria verdadeira, e onde não existissem, ela seria vacuamente verdadeira.

[1] Aqui cabe explicitar uma certa noção, indicada pelo próprio Bird (2001) em seu artigo, a saber, a de leis *ceteris paribus*. Quando falamos que é uma lei que "o sal se dissolve em água", é claro que não queremos dizer que isso sempre é verdade, afinal há casos em que a água está por demais saturada e o sal não se dissolve mais nela. Assim, o sal se dissolve em água, *dadas certas condições*. Quando falamos sobre leis neste artigo, gostaria que entendêssemos as leis como restringindo sua atuação a condições constantes, ou seja, como leis *ceteris paribus*.

Para imaginar bem isso, Bird nos pede para pensar numa lei L de ordem elevada e relativa a um conjunto de substâncias S, onde a existência das substâncias S depende de um certo conjunto C de leis mais básicas que L. Assim, se as substâncias S existem, então também existe o conjunto C de leis que permitem a existência de S. É possível que muitos conjuntos de leis diferentes permitam a existência de S, de modo que a existência de S implique $<C_1 \vee C_2 \vee C_3...>$. Se L regulamenta a relação entre substâncias S, então $<C_1 \vee C_2 \vee C_3...>$ implica L, pois L é dedutível ou sobreveniente ao conjunto de leis mais básicas C. Assim, se S existe, isso implica a verdade de $<C_1 \vee C_2 \vee C_3...>$ e de L. O que quer dizer que “a existência de substâncias (que existem no nosso mundo) implica a verdade das leis que as relacionam” (p. 273) e, conseqüentemente, sua necessidade, segundo Bird.

Esse argumento não responde inteiramente a nossa questão. Quando nos perguntamos se uma determinada lei é metafisicamente necessária, o que queremos saber é se ela vige em todos os mundos metafisicamente possíveis (o que é chamado de “necessidade forte”), e não apenas se ela vige em todos os mundos em que seus objetos existem (o que é chamado de “necessidade fraca”). Não podemos *pressupor* que uma lei seja metafisicamente necessária se ela se for verdadeira apenas naqueles mundos nos quais os próprios objetos que ela regula também existirem, porque haverá mundos onde não existem os objetos por ela regulados, se estes não forem excluídos por argumento posterior; e, se houver tais mundos, uma pergunta central para falarmos sobre a necessidade metafísica das leis naturais será se tal lei é verdadeira nesse mundo. Se pressupusermos sua necessidade, estaremos cometendo petição de princípio. É claro que Bird poderia argumentar que se as leis são conexões nomológicas entre entidades universais, elas podem ser verdadeiras em mundos em que não existem instâncias dos universais. Mas isso não o ajudaria, pois não queremos saber se elas podem ser verdadeiras, mas se elas são verdadeiras em todos esses mundos, ou melhor, em todos os mundos metafisicamente possíveis. Assim, uma questão fundamental, que temos que responder sem pressupor – à pena de cometer petição de princípio –, é se uma lei é verdadeira em mundos onde não há instâncias dos universais por ela regulamentados.

De acordo com a concepção de “lei” a ser aceita, isso implicará em diferentes respostas à questão acima – como apresentarei a seguir. O que é uma lei ser verdadeira em

um mundo? Qualquer resposta a essa pergunta depende fundamentalmente de como entendemos uma lei: (I) como meras regularidades observadas, ou (II) como uma conexão nomológica entre propriedades que torna compreensível as regularidades observadas.

Se acreditarmos que (I) é o caso, então trataremos a lei como uma regularidade observada nesse mundo e, por conseguinte, não poderia haver lei onde não houvesse instâncias dos universais expressos por ela. Se não há os objetos, não há regularidade a ser observada entre eles, e se não há tal regularidade, não há lei, pois, segundo essa concepção, a lei nada mais seria do que um tipo de regularidade observada. O problema crucial dessa concepção é o seguinte: ela não nos permite distinguir leis de regularidades meramente acidentais, suprimindo, ademais, a função explicativa que conferimos às leis naturais.

Quem sustenta (II), concebe a lei natural como uma conexão entre entidades abstratas (propriedades) que poderiam existir sem serem instanciadas. A lei natural nada mais seria do que a conexão nômica de propriedades que tornaria compreensível a observação da conexão regular dos objetos e eventos que instanciam tais propriedades. Essa concepção alternativa admite a vigência de leis naturais mesmo naqueles mundos possíveis nos quais as propriedades em questão não estão sendo instanciadas. Na verdade, como essa concepção se utiliza das leis para *explicar* suas instâncias, nela a lei deve estar em vigor previamente às suas instâncias a fim de que estas possam ocorrer.

Ao ponderarmos sobre as diferentes concepções, vemos que ambas deixam em aberto a questão que nos perguntamos. E, assim, para fugir da petição de princípio, não podemos pressupor – como Bird – que uma lei seria verdadeira em mundos nos quais as propriedades que ela conecta não estariam sendo instanciadas. Como Bird assinala, em mundos em que sal e água existissem, a lei “o sal se dissolve na água” seria válida. Não há problemas em concordar com isso. Mas sal e água são objetos contingentes e tudo que é contingente é, por definição, não-necessário. E, assim, há mundos em que eles não existem; e a questão sobre a vigência da lei nesses mundos não é trivial. Por exemplo, em todo mundo onde há humanos, existem leis que regulam o funcionamento da mente humana; mas em mundos onde humanos são impossíveis de surgir, não há tais leis. É claro que vivemos num mundo onde há humanos e, por isso, para a maioria dos nossos objetivos, as possibilidades que nos importarão serão aquelas que levam em consideração leis sobre

humanos; o que também é verdadeiro para o sal e para a água. Mas para alguns objetivos, como, por exemplo, quando nos perguntamos se uma lei é verdadeira em todos os mundos metafisicamente possíveis, não será satisfatório dizer que ela existe em todos os mundos metafisicamente possíveis em que seus objetos existem, pois tais objetos são supostamente contingentes e, portanto, poderia ser o caso que eles fossem metafisicamente impossíveis de surgir a partir de um certo mundo; e, assim, haveria mundos metafisicamente possíveis em que a lei não é verdadeira.[2]

Como salientamos, se leis se reduzirem a conexões regulares entre objetos, não haverá lei onde não houver os objetos. O problema dessa concepção é que não resgata a necessidade metafísica que intuitivamente imputamos às leis naturais. Se Bird sustentasse essa concepção de lei, o regularista, já perderia o debate sobre a necessidade da lei “o sal se dissolve em água” ao admitir que sal e água são objetos contingentes, pois se a lei é uma regularidade, ela não existe nos mundos em que os objetos contingentes que ela regula não existem. Se não houver certa lei em um mundo metafisicamente possível – por haver um mundo metafisicamente possível em que não há os objetos que ela relaciona – então essa lei não existe em pelo menos um mundo metafisicamente possível e, portanto, não é metafisicamente necessária, embora exista em todos os mundos metafisicamente possíveis em que seus objetos existem – ou seja, é apenas fracamente metafisicamente necessária. Ora, para ser fortemente metafisicamente necessária, ela teria que valer mesmo naqueles mundos em que seus objetos não existem. A garantia de que a lei de Coulomb vige pelo menos enquanto sal e água existem não é uma garantia de que sal, água e a lei de Coulomb existem em todos os mundos possíveis. E, para valer em mundos em que seus objetos não existem, ela não pode ser uma mera apreensão da regularidade entre objetos, ela deve constituir um substrato metafísico independente. Ou seja, para Bird conseguir falar sobre a existência de uma lei em todos os mundos metafisicamente possíveis, ele não poderá ser regularista; terá de tomar a lei como um substrato metafísico independente.

[2] É claro que isso não seria verdadeiro em sistemas de lógica modal S5; mas a discussão sobre qual sistema apreende melhor tais relações entre as modalidades é por demais extensa para ser tratada aqui. É suficiente que vejamos que não é truísmo que todos os mundos são acessíveis a todos os mundos, mas que é uma tese um tanto controversa.

Mostrar que uma lei é metafisicamente necessária consiste em assinalar que ela existe em todos os mundos metafisicamente possíveis ou que ela é dedutível de leis mais básicas, que em última instância são elas mesmas metafisicamente necessárias. Por exemplo, Bird deveria provar que a lei de Coulomb é dedutível de leis mais básicas que regulam cada mundo metafisicamente possível. E isso ele não prova; seus argumentos não nos dão razões suficientes para pensar que não é metafisicamente possível um mundo completamente diferente do nosso, inclusive nos objetos e leis básicas, e onde é impossível, dada a natureza daquele mundo, que surjam os objetos ou relações indicados pelas leis básicas de nosso mundo. Ou melhor: dada a resposta de Bird, a questão sobre a contingência ou necessidade das leis da natureza (inclusive a lei de Coulomb) ainda fica em aberto.

Rodrigo Reis Lastra Cid
Universidade Federal do Rio de Janeiro

rodrigorlcid@hotmail.com

Referências

Bird, A. (2001) ‘Necessarily salt dissolves in water’, *Analysis*: vol. 61, n. 4, pp. 267-274.

abstracta
Linguagem, Mente & Ação

Editorial

This is the first time *Abstracta* publishes articles on philosophical logic and philosophy of logic. It is important to say that using logical tools to examine philosophical concepts is not a sufficient condition for good philosophy, but sometimes logical tools allow us to formalize philosophical theories in order to determine whether they are sound or not. This dimension is mirrored in the articles collected in this special session which was originaly developed in Brazil during *Seminário Newton da Costa* at the *Federal University of Paraiba.* Although sometimes the articles here are not related to his theoretical work, this special session is in his honor, given the central role he is playing in Brazilian philosophy and logic.

Béziau's article has logic as the main character and this is an article on the philosophy of logic. This text contains a study on many polemical aspects of the concept of *logic* as well about what this subject really is, and, as it could not be different, it is in the scope of the author's project called *universal logic*, which tries to find general properties shared by all logical systems. Costa-Leite's article deals with the relations between imagination, conception and logical possibility. It is showed how to understand Cartesian and Humean theses using combinations of modal logics. It proposes a logic in which imagination and conception are kinds of diamond operator. This is an example of a work in philosophical logic. Koslow's text is a genuine application of the structuralist theory of logic, also in the scope of universal logic. Studying very general properties of implication structures, Koslow approaches semantical aspects of propositional logic showing how to understand Carnap's problem according to which there are non-normal interprations of classical operators producing weird properties. Silvestre's article is another example of philosophical logic and it shows how to explore inductive reasoning using tools from paraconsistent multimodal logic in order to model concepts like *plausibility* and *certainty*.

In general, in this special issue, we have two articles exploring the very nature of logical system and two articles showing how to use these logics to investigate philosophical concepts. We are happy to make this new experience in *Abstracta*, considering that, in fact, philosophy recently is almost impenetrable without understanding basic logical constructions.

THE EDITORS.

Abstracta 6 : 1, pp. 73 – 102, 2010

LOGIC IS NOT LOGIC

Jean-Yves Béziau

Abstract
In this paper we discuss the difference between *logic as reasoning* and *logic as a theory about reasoning*. In the light of this distinction we examine central questions about history, philosophy and the very nature of logic. We study in which sense we can consider Aristotle as the first logician, Descartes's rejection of syllogistic as logical, Boole rather than Frege as the initiator of modern logic. We examine also in this perspective the unfolding of logic into logic and metalogic, the proliferations of logic systems, the questions of relativity and universality of logic and the position and interaction of logic with regards to other sciences such as physics, biology, mathematics and computer science.

1. Anatomy of logic

11. Logic in the shadow

The word "logic" is a common word part of ordinary language, but the adjective "logical" is more frequently used. So the meaning of the word "logic" is generally understood through the adjective "logical". But what is logical and what logic is, the layman doesn't exactly know. As for many words, there is a semantic wavering: the same word can mean different things more or less contradictory; "logic" is no exception, even in the mouth of logicians. There is vagueness, ambiguity and confusion.

We don't want here just to discuss the meaning of the word "logic". We want to clarify what logic is. We want to put the logic room in order. This is a crucial point because if there is confusion at the level of logic, where shall we meet clarity and understanding? Should we accept global confusion, fall into soft relativism or look for another messiah such as cognitive science?

At some point the world was becoming a logical world. People like Wittgenstein, Carnap, Tarski were architects of such a world. Carnap wrote a book with the suggestive title *The logical structure of the world* (1928). Renewing the Aristotelian project these people had the idea that logic is essential, that it is the basis of science and rationality. In

this spirit, the Poles were using the expression "methodology of deductive sciences" as synonymous to "logic".

About 100 years later this logical world has vanished. We are living neither in a Fregean conceptual paradise, nor in a Russellian type heaven. Who is guilty? Gödel with his incompleteness theorems? Church with his undecidabilty result? Or Steve Jobs shamelessly exposing his apple in supermarkets all around the world?

Gödel's and Church's results can be seen as the failure of Leibniz's program[12] to build a big system that we can use to think without thinking, in the same way that we can clean our clothes using a washing machine without degrading our hands. But this failure is a happy end to the best-of-the-worlds story. These results are good news: human mind cannot be reduced to an algorithm and in showing that logicians have developed the theory of computation. They gave birth to machines, not replacing human beings, but releasing their brain from an activity that only supporters of Deep Blue may consider as the reflect of intelligence.

Maybe the vanishing of logic is due to the success of computers. Logic is now in the shadow of computer science. Technology is prevailing over science. But we cannot forget the root of computer science: its marvelous fruits are by-products of logic. However, even if logic does not reduce to computer science, this latter has changed the face of logic in a positive lifting. At the dawn of modern logic there was a tendency to try to construct big architectonic logic systems describing everything, solving all the problems. Computer scientists have broken this prehistoric trend being guided by efficiency rather than by megalomania. This has led to many different complementary logical systems. But we have to be careful not to get lost in such a jungle and to keep in mind what logic is. Logic has to do with rationality, it is not only a bouquet of efficient but limited tools.

[1] About Leibniz's project, see the excellent book by Louis Couturat, *La Logique de Leibniz* (1901).

[2] Gödel's comments are the following: "In I678 Leibniz made a claim of the universal characteristic. In essence it does not exist: any systematic procedure for solving problems of all kinds must be nonmechanical." (Wang, 1996, 6.3.16); "My incompleteness theorem makes it likely that mind is not mechanical" (Wang, 1996, 6.1.9).

12. Logic medley

To point out the ambiguity surrounding the very nature of logic and the way the word "logic" is used, let us have a look at what some famous thinkers have written about it.

Logic is the anatomy of thought.	John Locke	ca 1700 *Unsourced*
That *logic* has advanced in this sure course, even from the earliest times, is apparent from the fact that, since Aristotle, it has been unable to advance a step, and thus to all appearance has reached its completion.	Immanuel Kant	1787 Preface of the Second Edition of *Critique of Pure Reason*
The design of the following treatise is to investigate the fundamental laws of those operations of the mind by which reasoning is performed; to give expression to them in the symbolical language of a calculus, and upon this foundation to establish the science of *logic*.	George Boole	1853 *An Investigation of the laws of thought*
Man has such a predilection for systems and abstract deductions that he is ready to distort the truth intentionally, he is ready to deny the evidence of his senses only to justify his *logic.*	Fyodor Dostoyevsky	1864 *Notes from the underground*
Contrariwise, if it was so, it might be; and if it were so, it would be; as it isn't, it ain't. That's *logic*.	Lewis Carroll	1871 *Through the looking glass*
Bad reasoning as well as good reasoning is possible; and this fact is the foundation of the practical side of *logic.*	Charles Sanders Peirce	1877 *The fixation of belief*
Logic takes care of itself and all what we have to do is to look and to see how it does it.	Ludwig Wittgenstein	1914 *Journal*
Pure *logic* is the ruin of the spirit	Antoine de Saint-Exupéry	1942 *Flight to Arras*
With the discovery of the conventional and relative character of *logic*, human spirit has burned his last idol.	Louis Rougier	1955 *Traité de la connaissance*
If one wishes to speak about the atomic particles themselves one must either use the mathematical scheme as the only supplement to natural language or one must combine it with a language that makes use of a modified *logic* or of no well-defined logic at all.	Werner Heisenberg	1958 *Physics and philosophy*

It is obvious that in these quotations the word "logic", independently of style and personal views, is used with different meanings. In the following sections we will try to disentangle the logic meanings of this table.

13. Logic and logic

Beyond the paradoxical claim "Logic is not logic", there is an important distinction: *logic as reasoning* and *logic as the study of reasoning*. This distinction is quite similar to the distinction between History as the series of events and history as the science which studies these events, *History* being the object of study of *history* (see e.g. Woolf, 2011). To keep this parallel in mind we can use the word "Logic" for reasoning and "logic" for the science which studies reasoning, *Logic* being the object of study of *logic*. This is a nice "differance", pointing the close connection between the two sides of the logical coin.

For many sciences the two sides of the coin are generally clearly linguistically separated, although the distinction ranges from few letters to different words. A radical difference of words is sometimes due to a language shift as in the case of biology and physics. This is also the case of logic: when we say that logic is the science of reasoning, the name of the object of study - "reasoning" - is a word completely different from the name of the science of it - "logic" -, "reasoning" being the Latin word for logic as reasoning.

Here is a table showing differences and variations:

SCIENCE	**OF**
Biology	Living organisms
Physics	Matter
Linguistics	Languages
Sociology	Society
Poetics	Poetry
Anthropology	Human beings

Why in the case of history people are using the same word, making only and not always a graphical differentiation? And why in logic people are making no difference, graphical or not, favouring the confusion between a science and its object of study. Such a

mess looks similar to what happens with cooking. But in this case the confusion can be justified by a theory/practice rave mix leading to delicious omelettes. Is logic an art like cooking? This is in fact the French conception of it according to the title of the famous book by Arnauld and Nicole: *Logic or the art of thinking* (1662). To go on sambing with examples, when someone is studying dance, it is in general not in a contemplative mood, it is for dancing. Dance is an art and the practice of it, the difference here is a substantive/verb difference not necessarily graphically expressed; we can say: "Mary is studying dance to dance". What about John? We will say: "John is studying logic to reason". In this formulation the distinction is explicit because there is no "logic" verb, the closer we can get to a redundant formulation is: "John is studying logic to be logical". But is it really the goal of John when studying logic?

In history, dance and logic, there is on the one hand an activity performed by human beings (we are ignoring here dance history of logical ducks), on the other hand a theory about this activity produced by the same mammals. The connection is strong because there are human beings on both sides of the coin; this is a characteristic of human sciences (but is logic a human science?). A connection that can be understood as an interaction, considering a theory/practice duality.

In history the difference is stronger because rarely the historian will work in order to practice History. But the reason why the difference is not so strong and nearly the same word is used is because the object of historical science is not so objective. There are many different stories, in style and focus. Sometimes we may wonder if they are referring to the same History. In particular History is lost in a "once upon the time". When did History start? Not right at the beginning, because there is also pre-History with pre-historical men and women, generally not confused with specialists of pre-History, the pre-Historians. [3] Historians are also not confused with the first historical human beings, but maybe there is here a secret connection. When and where historical human beings did emerge? There is no clear answer to the question, this reflects the unreality of History. Settling, housing, agriculture, painting, writing, all these activities may be considered as activities surrounding the birth of historical human beings, but maybe they are too much bourgeois,

[3] Notions such as *protohistory* (Otte 2008) and *deep history* (Smail 2008) have also been introduced.

emphasizing immobility and comfort rather than the strong stream of History with conquests, revolutions and crises. According to a quite different view, History started in the year 0 or 1, and will never end, a mathematical vision of the world ... A more rational view has been promoted by Heidegger. He would rather say that History started with historical science, with Herodotus and Thucydides, the first historians (see e.g. Shanske 2007). Their objective stories did change History by contrast to mythologies perpetuating karmic circles. Can we similarly claim that Logic started with Aristotle? This is a question we will examine in the next section after comments about the Magritteean character of the title of our paper.

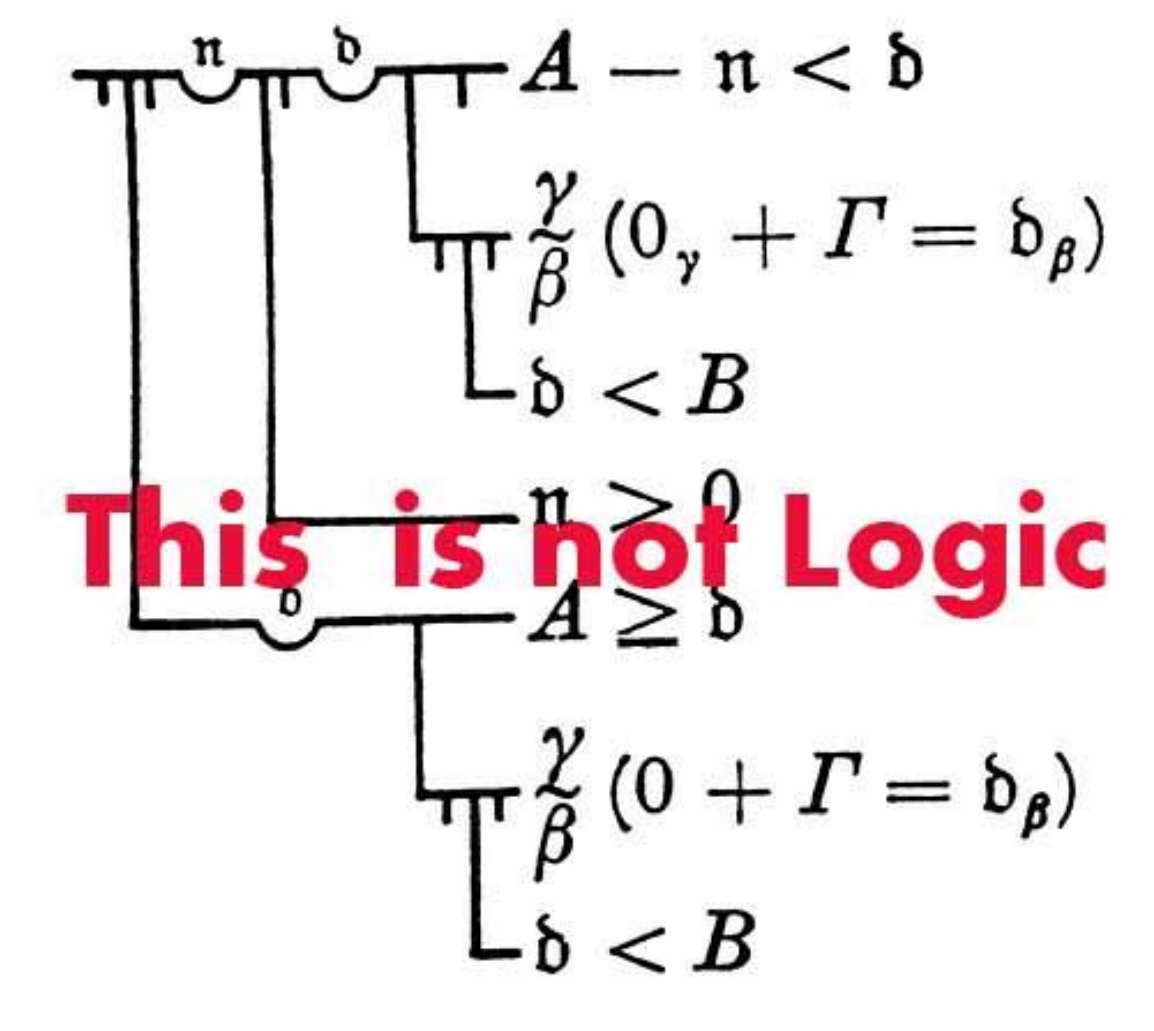
The paradoxical claim "Logic is not logic" remembers Magritte's paradoxical claim: *Ceci n'est pas une pipe* (*This is not a pipe*). This claim appears in his most famous painting. The name of this painting is not *This is not a pipe* but *La trahison des images*, meaning *The treachery of images*. Magritte explicitly deals with the ambiguity of pictorial representation. Someone may think at first sight that the pipe in the painting of Magritte is more similar to a real pipe than logic science[4] is similar to reasoning. But pictorial similarity is really a treachery; it is one of the most powerful illusions - visual illusion. The painted pipe is in fact very different from a real pipe, as one can easily understand if he tries to use it for smoking. Of course the word "banana" has not the same taste as a real banana, but the contrast between the thing and its representation is not so strong, since there is no resemblance between the two.

By claiming that "Logic is not logic" we want to stress both the similarity and the difference between logic as reasoning and logic as a science. At first, not paying attention to the scriptural difference or thinking that the capital "L" is due to the beginning of the sentence, this claim may sound like a real contradiction such as "life is not life", the converse of the declamation "life is life", a successful song. Such declamation, with many variations, such as "black is black", is not a tautological claim, the idea is to emphasize the

[4] We are using here the expression "logic science" as synonymous to "science of logic", in the same way that "logic theorem" is used as synonymous to "theorem of logic". "Logic science" therefore means logic as a science. The expression "science of logic" is ambiguous because it can be interpreted as "science of *Logic*" or "science of *logic*". The latter should be interpreted similarly to "science of biology", which is synonymous to "biology" tout court, just emphasizing that biology is a science. It is in this way that the expression "science of logic" is used by Boole in the table of our logic medley.

very nature of the thing, which may be indefinable. The famous historian of logic Jean van Heijenoort used to claim: “Life is not first-order, life is not second order: Life is life.” (Anellis, 1994, p.45). But what is the meaning of the antilogical claim “life is not life”? This may stress that life is not what we usually think it is.

We finish this flowering of our discourse with a picture of a Magrittean flavour. The picture below anticipates our discussion in the next sections pointing out that the *Begriffsschrift* is not logic as reasoning – difficult to practice it – but a (pictorial) theory of reasoning.



2. Evolution of logic
21. Logical animals

Many people would say that Aristotle is the first logician. But do they think that he was the first man to reason? No, generally they think he was the first to develop a science of reasoning. And they are right,[5] but we have to understand the full story. For this it is useful to have in mind the distinction *Logic/logic* and also to merge in classical Greece. The word "logic" derives from a word typical of the Greek culture, the word "logos", which has no equivalent in other languages. There are four main meanings in its semantic network: *relation*, *language*, *reason*, *science* (Later on, in the Bible, logos became God – cf. John 1:1[6]). The table below describes the situation with examples.

LOGOS 4 MEANINGS	**EXAMPLES**
relation	irrational numbers (not a *relation* between natural numbers)
language	neologism (new *word*)
reason	*rational* animals λογικό ον
science	anthropology (*science* of human beings)

Maybe one could claim that modern first-order logic is the full realization of the *logos*: it is the science of reasoning describing relations with a language. But let us come back more than two thousand years ago. Some people argue that mathematics started with the proof of the irrationality of square root of two by the Pythagoreans (see e.g. Dieudonné, 1987). They consider that this was the first mathematical proof. A central feature in this proof is the use of the reduction to the absurd. We can say that with the reduction to the absurd we have a new way of reasoning, a new *Logic*, maybe the birth of *Logic*, tracing here the difference between rational and non-rational human beings. This would be contrary to Aristotle's definition of human beings as *rational animals* (literal translation:

[5] Bochenski who has extensively studied Indian logic (see Bochenski 1956) says that Indian logicians can be credited to inventions as valuable as Aristotle's ones, the difference being that it took them several centuries (see Bochenski 1990).

[6] The standard translation of John 1:1 is: "In the beginning was the Word, and the Word was with God, and the Word was God", where "Word" is the translation of "logos": Ἐν ἀρχῇ ἦν ὁ λόγος, καὶ ὁ λόγος ἦν πρὸς τὸν θεόν, καὶ θεὸς ἦν ὁ λόγος. Gödel speaks of a "rational principle behind the world" (Wang, 1996, 8.4.10).

logical animals). This definition implies that rationality is an essential feature of human beings and that consequently they have always been rational.

According to Szabó (1969, 1994), the reduction to the absurd was first used by the philosophers of the Eleatic school, Parmenides and Zeno. Szabó's thesis based on a detail historiographic study is a confirmation of an idea defended more than one century before by Schopenhauer, emphasizing that rationalism was a philosophical attitude based on the rejection of sense data. But Schopenhauer thinks that rationalism, in particular mathematical rationalism supported by the reduction to the absurd, is wrong. Eighty percent of the proofs in Euclid's *Elements* are based on the reduction to the absurd and Schopenhauer is not afraid to say that this method is properly absurd. In his first book *On the fourfold root of the principle of sufficient reason* (1813), he presents some new proofs of theorems of geometry, based on pictures, not on "absurd reasonings". Schopenhauer's approach has influenced Wittgenstein and the Intuitionists. Schopenhauer's philosophy of mathematics is not something new. For him the way to escape the empirism/rationalism dichotomy is the Kantian theory of pure intuitions of space and time: we can reason directly and safely about space using some intuitive pictorial proofs, this is much better than kilometers of reduction to the absurd proofs. Schopenhauer has been inspired by Kant, but he has developed much more the theory, elaborating the distinction and relation between logic and mathematics, being probably the first to introduce the terminology "metalogical". He uses this word to qualify the fourth class of truths, corresponding to the fourth root of the principle of sufficient reason (see Béziau 1993).

Let us come back to good old Greece. Greeks were rational animals. But why did these rational animals introduced *irrational* numbers? This is one of the mysteries of the *logos*. Pythagoreans had the belief that everything can be explained with natural numbers or *relations* between such numbers, *rational* numbers. But their belief was dismissed by the logos through a *reasoning* based on the reduction to the absurd, the proof of the *irrationality* of square root of two. So they were rationally led to irrationality or better: the reduction to the absurd became the key of the logos, opening the kingdom of rationality to

irrational numbers and much more. Summarizing: Logic was born in the Pythagorean boat and its true face is the reduction to the absurd.[7]

Now what is the relation between Aristotle and such *Logic*? Aristotle didn't practice it and Aristotelian *logic* is not a science of it. Moreover mathematicians never used syllogistic to practice mathematics. It is interesting to compare the situation with tragedy: Aristotle elaborated a theory of tragedy in his *Poetics* based on the great tragedies of his time. He didn't write tragedies but his theory has been used - in Hollywood it is still a basis for screenwriting (see e.g. Tierno, 2002).

Aristotle was nevertheless a great promoter of logic, as a tool and as a science. He had the idea that *proof* is the central characteristic of science. He was interested to develop logic as a general methodology of science but also to avoid sophisms. Modern logic is in continuity with the Aristotelian perspective: logic appears as methodology of science and as critical thinking. What has been rejected is Aristotelian logic as a given theory describing reasoning: syllogistic. The main reason of this rejection is that it is not giving an accurate description of mathematical reasoning.

First-order logic is a better description, but one may argue that first-order logic, like syllogistic, is more a theory of reasoning than an effective way of reasoning. Nevertheless in modern times there had been a better interaction between *Logic* and *logic*. Despite the rejection of the new logic science by some mathematicians, logic has changed mathematics: looking closer we see that modern mathematics is directly connected to modern logic. If we consider that mathematics is (part of) reasoning, we can say that *logic* has changed *Logic*. And vice versa *Logic* has changed *logic*, because logic as mathematical reasoning has been applied to develop logic science – that was not the case of Aristotle's syllogistic, not based on mathematics. Algebra, topology, category theory have been applied to develop logic, as the science of reasoning.

The highest development of the first stage of modern logic, the mathematical foundations wave, is model theory. Model theory is a beautiful interaction between *Logic* and *logic* establishing a vital link between mathematical structures and the way we reason

[7] About the relation between irrational numbers and irrationality, see the very interesting book by G.-G.Granger, *L'irrationnel* (1998).

about them. The second stage of modern logic connected with the proliferation of non-classical logics is also an interaction between *Logic* and *logic*: people are constructing systems of logic and applying them, we have here a theory/practice duality like in dance and other arts and techniques - this is *techno-logic*.

More than ever we are logical animals.

22. Logical cuisine

An important philosopher who was against the Aristotelian trend was Descartes. We can see Descartes as the father of modern philosophy, in particular breaking the Aristotelian tradition. Descartes is not anti-rationalist, but he is again the rationalism of Aristotle and neo-Aristotelian philosophy (scholastic). Descartes is promoting a new kind of rationalism. The distinction between logic as reasoning and logic as science is useful to understand this shift of rationalism. For many people, Cartesian means logical. So if we say that Descartes didn't like logic, people may be surprised. This Cartesian paradox is clear up if we explain that Descartes didn't like logic as a science, in particular syllogistic. He thought that to be logical it is not necessary to use syllogistic, it can even disturb our reasoning in the same way that if are trying to apply a theory to walk, we will not walk in a better way, but maybe fall.

So what is the Cartesian way? Descartes believes in a natural disposition: "Good sense is, of all things among men, the most equally distributed". To reason we don't need a theory of reasoning, *Logic* doesn't need *logic*. Descartes wrote two books with suggestive titles: *Rules for the direction of mind* (1628) and *A Discourse of a method - For a method for the well guiding of reason, and the discovery of truth in the sciences* (1637), but these books don't develop a science of reasoning.

Descartes emphasizes that we must think clearly and distinctly and summarizes his methodology in four principles he presents in contrast with logic as syllogistic: "Instead of the great number of precepts of which logic is composed, I believed that the four following would prove perfectly sufficient for me, provided I took the firm and unwavering resolution never in a single instance to fail in observing them" (Descartes, 1637). Here is a table presenting Descartes's four precepts (the nicknames on the left column are ours):

DESCARTES 4 PRECEPTS	
Clarity	Never to accept anything for true which I did not clearly know to be such; that is to say, carefully to avoid precipitancy and prejudice, and to comprise nothing more in my judgment than what was presented to my mind so clearly and distinctly as to exclude all ground of doubt.
Division	To divide each of the difficulties under examination into as many parts as possible, and as might be necessary for its adequate solution.
Ascension	To conduct my thoughts in such order that, by commencing with objects the simplest and easiest to know, I might ascend by little and little, and, as it were, step by step, to the knowledge of the more complex; assigning in thought a certain order even to those objects which in their own nature do not stand in a relation of antecedence and sequence.
Exhaustivity	To make enumerations so complete, and reviews so general, that I might be assured that nothing was omitted.

We can consider these precepts as promoting logic, considered as good rational thinking. The aim of Descartes is not to theorize, developing a science of reasoning, but to practice: *The Discourse of a method* includes applications of the method to Dioptrics, Meteors and Geometry. Descartes also applies his methodology to philosophical issues: *proving* his own existence and also the existence of God. Such proofs are not chains of syllogisms, Descartes clearly states that *cogito ergo sum* is not the conclusion of a syllogism. Descartes is promoting Logic as rational thinking, free of Barbarian syllogistic.[8]

Blaise Pascal has on this respect a position quite similar to Descartes. For Pascal the highest way of reasoning is the one we find in geometry, based on an obvious natural methodology and we can say bye bye to Barbara, Celarent and all their syllogistic friends, which are of no use to develop right thinking and avoid sophisms. Pascal wrote in *The Art of Persuasion* (1656): "To discover all the sophistries and equivocations of captious reasonings, they have invented barbarous names that astonish those who hear them … It is not Barbara and Baralipton that constitute reasoning. The mind must not be forced; artificial and constrained manners fill it with foolish presumption, through unnatural elevation and vain and ridiculous inflation, instead of solid and vigorous nutriment. And one of the principal reasons that diverts those who are entering upon this knowledge so much from the true path which they should follow, is the fancy that they take at the outset

[8] A good presentation of Descartes's views on logic is (Gaukroger 1989).

that good things are inaccessible, giving them the name of great, lofty, elevated, sublime. This destroys everything. I would call them low, common, familiar: these names suit them better; I hate such inflated expressions."

By opposition to the sophistry of syllogistic, Pascal defends 8 rules, "the true ones", that are "simple, artless, and natural". We present them in the following table:

PASCAL 8 RULES	
Rules for Definitions	Not to undertake to define any of the things so well known of themselves that clearer terms cannot be had to explain them.
	Not to leave any terms that are at all obscure or ambiguous without definition.
	Not to employ in the definition of terms any words but such as are perfectly known or already explained.
Rules for Axioms	Not to omit any necessary principle without asking whether it is admitted, however clear and evident it may be.
	Not to demand, in axioms, any but things that are perfectly evident of themselves.
Rules for Proofs	Not to undertake to demonstrate any thing that is so evident of itself that nothing can be given that is clearer to prove it.
	To prove all propositions at all obscure, and to employ in their proof only very evident maxims or propositions already admitted or demonstrated.
	To always mentally substitute definitions in the place of things defined, in order not to be misled by the ambiguity of terms which have been restricted by definitions.

Can we say that Descartes/Pascal position is typically French? Is this the quintessence of a French vision of logic extending up to Poincaré (1905-06) not afraid to qualify the new logistic as *pipi de chat*? Maybe clarity of thought is, with champagne and Roquefort, a typical French specialty - a good mix indeed.

But this French logic cuisine does not necessarily reduce to a regional delicacy. Tarski presented a conference at the *9th International Congress of Philosophy* organized in Paris in 1937 entitled "Sur la méthode deductive" and published in French in the Annals of this event under the same title. This paper is the same as Chapter 6 of Tarski's bestseller *Introduction to Logic and to the Methodology of the Deductive Sciences* (1936) entitled "On the Deductive Method".

Alfred Tarski was a great admirer of Blaise Pascal. He considers that modern logic, as methodology of deductive science, is very similar to Pascal's methodology of *The Art of persuasion*. The first section of the chapter "On the Deductive Method" is entitled "Fundamental constituents of a deductive theory—primitive and defined terms, axioms" and the following footnote is attached to the first sentence of this section (p.109): "Ideas which are closely related to those presented in this section can be found in earlier literature. See, for instance, the opusculum (posthumously published), *De l'esprit géométrique et de I'art de persuader*, of the great French philosopher and mathematician B. PASCAL (1623-1662)."

Like Pascal, Tarski thinks that the model of reasoning has to be found in Euclid's geometry. Tarski sees modern logic as a renewal of this method. For Tarski there is continuity between Euclid, Pascal and Hilbert. The deductive method as the trinity *Definition-Axiom-Proof*, promoted by Pascal and considered as the central architecture of the deductive method by Tarski, is not due to Aristotle. It was developed by mathematicians and it is rather a methodology than a science of reasoning. But Tarski, like Aristotle,[9] considers that reasoning does not reduce to mathematical reasoning, modern logic "arose originally from the somewhat limited task of stabilizing the foundations of mathematics. In its present phase, however, it has much wider aims. For it aspires to relate to the whole of human knowledge. In particular, one of its goals is to perfect and to sharpen the deductive method, which not only has a central place in mathematics, but in addition, in just about every domain of intellectual endeavor, serves as an indispensable tool for deriving conclusions from accepted assumptions." (Tarski, 2004, p.IX) However Tarksi, differently to Aristotle, considers that this methodology is fundamentally based on mathematics: "Logic (the deductive method) applies to every science and in particular to itself, which should for this reason be regarded as a mathematical discipline" (p.112).

We will in the next section study more this *reflexive character of logic*.

[9] Tarski like Aristotle had a strong interest for biology - this was his first love, and he encouraged his friend J.H.Woodger to develop the methodology of biology, see (Woodger, 1937)

23. Logical buildings

Hilbert has coined the word "metamathematics". This monstruous name is not so much in use nowadays, people prefer to use the expression "proof theory" that Hilbert was using synonymously. Hilbert's follower, Nicolas Bourbaki started his famous multivolume treatise *Eléments de Mathématique* by claiming "qui dit mathématique, dit démonstration" (see Bourbaki 1970). For Hilbert the substance of mathematics are proofs. So for him the study of what mathematics is, is the study of what a proof is. To perform such study we have to go above mathematics, hence the bigname "metamathematics", where the prefix "meta" is understood as "above".

During some years the word "metamathematics" had an extensive use, in fact it was used as a synonymous to "logic". It was the time of Hilbert's reign. In 1952 Kleene wrote a book called *Introduction to metamathematics* which became a fundamantal textbook of logic. Eleven years later was published a book with a punny name: *The mathematics of metamathematics* (1963). Is this punny title the beginning of the end of metamathematics? Surely there is a shift of perspective. This title is provocative because the idea of Hilbert's program was to develop metamathematics using a different methology as *within* mathematics, in particular in view of proving consistency results. The standard view is that proofs are made of strings of symbols and that we stay in the denumerable to manipulate them. But Gentzen, member of the Hilbert's school, was himself changing things, at the time of Gödel's second incompleteness result he proved the relative consistency of arithmetic, using transfinite induction.

In Poland there was a different perspective right at the start. The Polish school was much influenced by the work of Schröder on algebra of logic (see Woleński, 1992). In Poland grew a tradition of using mathematics to develop logic as any branch of mathematics. As pointed out by Tarski, there is a certain reflexive character in logic, but this is not necessarily seen as a vicious circle, it can be viewed as an elevating spiral. Logic is not the foundation on which the building of mathematics or science is erected; it is rather its architecture. A different perspective that may be understood through the definition by Tarski of logical notions are those invariant under any transformation (Tarski, 1986).

"Metamathematics" is related with Hilbert; in Poland a connected word has been much popular, the word "metalogic". What is metalogic? If logic is synonymous to metamathematics, metalogic is a nickname for metametamathematics. So we are facing here a three story building, that can be represented by the following table, placing our differance between *Logic* and *logic*:

	Name	**What it is**
3rd Floor	Metalogic	theory of the theory of reasoning
2nd Floor	logic	theory of reasoning
1st Floor	Logic	reasoning

THE 3-STORY BUILDING OF LOGIC

Generally people consider rather a two story building. This is not without ambiguity. Are they collapsing the 2nd and the 3rd floors? In a proof system, say LK, a theorem is something proven within the system LK, it is an object of the 2nd floor. A metatheorem such as the cut-elimination theorem is an object of the 3rd floor, it is a result about LK, not a result proved in LK. One may want to study the system where cut-elimination is proved. This can lead to a 4th floor, but logicians will rather try to go down than up, trying to reduce the 3rd floor to the second floor. Gödel's work is typically on this direction, with the arithmetization of syntax and the notion of proof turning into a modal operator within a logical system.

But the reduction to a two story building can be based on a different view. Sometimes people use the word "metalogic" to speak of a theory whose object is logic, not clearly making the distinction between reasoning (*Logic*) and a system describing this reasoning (*logic*). In this case they are collapsing the 1st and 2nd floors.

The prefix "meta" has become popular in philosophy, maybe due to a complex of superiority. Philosophers are talking about "metaethics", "metaphilosophy" and even "metametaphysics". On the other hand nowadays logicians are not using so much metawords. Maybe things are clear enough without climbing above, paradise in on earth: if John says to Mary he is studying logic, she will generally understand that he is studying some logical systems describing reasoning. Maybe she would have been more impressed if

he had said he was studying metalogic. Then she would have invited him to the hell of a metadance.

3. Cosmology of logic

31. Logic, logic and logics

The XXth century has been very fruitful for logic, in quality and in quantity. Uncountable logical systems were born during this baby boom period. And this fertility boost is going on. Is it a multiplicity of *logics* or of *Logics*? To properly answer this question we must examine the relation between a (system of) logic, and the Logic it describes.

To have examples at hands and to better understand the problem, let us first try to classify the multiplicity of logical systems. Our objective here is not to present an exhaustive classification of all existing and possible logics but to show that there are different ways of slicing the logic cake, that we will illustrate with typical specimens.

We can start with the distinctions between *deviation* and *expansion* of classical propositional logic, deviation means that the properties of standard classical connectives are modified, expansion means there are some additional connectives.[10]

DEVIATION/EXPANSION	
Deviations	Intuitionistic logic Relevant logic
Expansions	Modal logic Causal logic

Let us note that here "modal logic" is not the name of a logical system but the name of a class of logical systems: there are thousands of systems of modal logic. Moreover the dichotomy deviation/expansion is not necessarily exclusive: we can have a relevant modal logic. A variation that does not appear explicitly in this table is a rather vertical variation, that can be understood through the distinction propositional logic / first-order logic and that is more fully expressed by the following table based on the example of classical logic but that can be applied to intuitionistic logic and other logics.

[10] Haack (1974) uses the terminology "extensions", which is quite ambiguous, we use here "expansions" by reference to the use of this word in model theory.

GRADES	
Subsystems	Positive classical propositional logic Full classical propositional logic
Supersystems	Many-sorted classical first-order logic Second order classical logic

There is not only one way to generate a system of logic. And we can consider that in some sense two techniques correspond to two different systems. We can classify logic systems according to the way they have been generated. *Substructural logics* are defined in this perspective: they are logics constructed by modifying the structural rules of sequents systems. Here is a simple table describing classification by techniques:

TECHNIQUES	
Proof	Hilbert systems Sequents systems
Semantics	Logical matrices Kripke structures

It is important to note that these different ways of slicing the cake don't lead to the same results. For example if we call Kripke logics, logic systems generated by the technique of Kripke structures, this class of logics is not the same as the class of modal logics, because on the one hand there are some modal systems that cannot be characterized by a Kripke structure, and on the other hand Kripke structures can be used to develop logics without modalities, such as a logic of implication.

Now let us examine if the multiplicity of *logics* (logic systems) corresponds to a multiplicity of *Logics* (ways of reasoning).

One may argue that there is only one Logic and that the multiplicity of logics is not against this oneness: multiple logic systems can be viewed as descriptions of the many aspects of this big Logic. In physics, there are many theories, this does not necessarily mean that there are different physical realities. These theories can be viewed as describing different aspects of the same reality. In logic, when we have expansions, such as a modal logic, or grade variations such as a many-sorted logic, we can argue that these logics are all

describing the same Logic. And also in the case of deviations, as pointed out by da Costa (1980) in the case of paraconsistent logics: a paraconsistent negation can be seen as an additional operator.

In physics we may have different concurrent theories, for example Ptolemy's theory is not the same as Newton's theory. They are about the same reality but one seems closer to reality than the other one. We can say the same of first-order logic comparing it to syllogistic. Here again this does not mean that we have different physical realities or different Logics, just different ways to look at the same thing.

But the case in logic is more difficult than the case of physics, because logic is both a normative and a descriptive theory.[11] The normative/descriptive distinction is useful to understand different positions non-classical logicians may have. When someone says that classical logic is not "real reasoning", this may be understood in two different ways. On the one hand one may argue that classical logic is not the right description of reasoning as it is, on the other hand one may argue that classical logic is not the right way of reasoning. In the first case one is speaking of classical *logic*, in the second case of classical *Logic*.

A typical example of the second case is Brouwer. He thought that classical *Logic* was wrong: this is not the way we should reason in mathematics, for him the right way is intuitionistic *Logic*. Brouwer, like other mathematicians, had no interest in logic as a science of reasoning, he was not interest to develop intuitionistic *logic*. This was done by his student Heyting and people generally think that Heyting's *logic* is a good description of Brouwer's *Logic* (see van Stigt, 1990 and Moschovakis, 2009).

32. The relativity of logic

The answer to this question depends on the two sides of the logic coin and the relation between them, but here we shall also talk about a hidden third dimension: the logic of reality.

[11] Lewis Carroll defines logic as the "science of reasoning rightly", see (Moktefi, 2008). This classical definition shows the normative character of logic. But this formulation is quite ambiguous: its meaning may range from "the art of reasoning (rightly)" to "a theory of what correct reasoning is".

Without being a post-modernist for whom the theory of relativity has the same value as Tupi-guarana's cosmology, one may think that science is relative because it is always changing, an idea which seems quite natural nowadays. The main idea is that science is progressing. Theories are slowly improving, or sometimes there are some breaks: a theory is rejected as false and replaced by a totally different one. Syllogistic was rejected, but some people see first-order logic as an improvement of syllogistic.

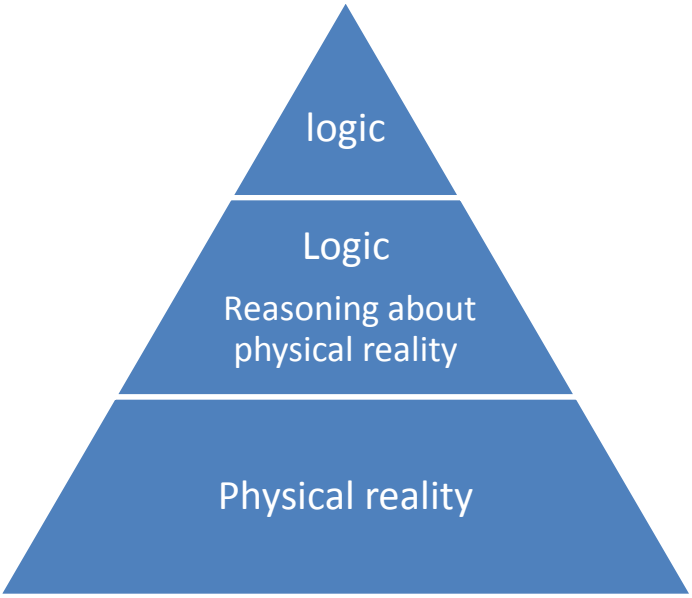
It is true that there are still nowadays people revering the religious scientificity of the enlightenment period, a time when a scientific theory could be seen as absolutely true. This was in particular the case of Newton's theory, a truth absoluteness in harmony with the intrinsic absoluteness of Newtonian physics: absoluteness of space, time and laws of nature. Some people may similarly think that first-order logic is an absolutely true theory perfectly describing perfect reasoning. Newtonian absoluteness was seriously challenged by the changes of modern physics: both at the microscopic level, with quantum physics, and at the macroscopic level, with relativity theory. And some physicists are arguing that not only time and space are relative but also that the laws of nature are changing (see e.g. Barrow 2002).

We have to face the very nature of objective reality. The physicists from the Copenhagen school have not necessarily rejected the idea of an objective reality; in fact Bohr's complementary theory is a way to save such reality. But due to the results of quantum physics there is the idea that reality cannot be known as it is, independently of some human experimentations which modify it. Since physics is a major science defining the house we are living in, the universe we are merged in, all these astonishing changes of physics have an impact in our way to consider science and reality in general.

And in fact since the start there was an interaction between modern physics and modern logic. Heisenberg has argued that our reasoning based on classical logic cannot describe the phenomena of the microscopic world (see e.g. Heisenberg 1958). We may try to change our logic and new logics have been proposed.[12] The situation can be interpreted in different

[12] The universe of quantum logics is in continuous expansion. An interesting logic system dealing directly with Heisenberg's uncertainty principle was proposed by Paulette Février in 1937 in Paris at the same congress where Tarski was talking about the deductive method (see Février, 1937). Tarski was a good friend of Paulette Février during many years (see Feferman and Feferman, 2004). In this book the authors also

ways: one may think that the logic of physical reality is not classical logic. If we consider that logic is reasoning, how can we speak of "the logic of physical reality"? Can we say that a stone is reasoning? For Aristotle the principle of contradiction was a law of reality, it was the structure of the world. Nowadays we have a less anthropomorphic view of reality, we don't see negation as part of reality. Reality is not black or white. It is not tricolor either. Negation, classical or neo-classical, is a tool to conceive reality. We can speak of the logic of physical reality in this sense, this logic of reality is a way of reasoning about reality, and we can develop a system of logic describing it, indirectly describing physical reality.[13] Here is a picture summarizing the stratification:



Physics is not the only science with theories radically changing our vision of science and the world. Changes started in fact before, in biology, with the theory of evolution, rejecting the idea of living beings of a permanent type, human beings included. If our brain is evolving, and if we think that the brain is an organ strongly connected with reasoning, it is

explain how Tarski encouraged Patrick Suppes to organize in 1957 at Berkeley a big event on the axiomatic method with special reference to geometry and physics (Paulette Février was there). After this successful event, Tarski decided to launch the series of congresses LMPS (Logic, Methodology and Philosophy of Science) still going on.

[13] Compare with the views of Joe Brenner in his book *Logic in reality* (2008).

natural to think that reasoning changes, that *Logic* is relative. And also *logic* as a science: cognitive scientists may rightly think we have to replace the study of the Aristotelian organon by the study of another organ: the brain; not only human brain but also the brain of other animals, to see the resemblance and the difference, to eventually find what characterizes the logicality of human brain. Recent discoveries tend to show that the logicality of human beings is not so different from the logicality of other animals – it was proved many years ago that dogs can perform disjunctive syllogisms, without reading Aristotle (see Aberdein, 2008).

We are nowadays in a situation completely different from the time when people like Kant had the idea that Logic was a set of fixed laws of thought perfectly described by Aristotelian logic. Boole also had the idea that Logic was made of laws of thought but he didn't think that syllogistic was a good description of these laws. He started to use mathematics to describe them, and then everything started to change by an interaction between the object of study and the theory, an interaction between *Logic* and *logic*.

For this reason we can consider that Boole is the true generator (rather than creator) of modern logic by opposition to Frege. Frege with his *Begriffsschrift* produced a static picture of reasoning, it is a beautiful cliché but it cannot be used. Human beings are not Boolean, human brains are not necessarily working according to the laws of a Boolean algebra. But a science of reasoning based on mathematics has led us to conceive new systems of logic, for example with a negation obeying neither the principle of contradiction, nor the law of excluded middle. A logic that can be implemented in computers (see e.g. Belnap, 1997) but also that can be used by human beings to reason in a different way about reality.

The first section of Chapter 6 of *Philosophy of logic* by Quine (1979) is entitled, "change of logic, change of subject*"*. Quine, against deviant logics, claims: "If sheer logic is not conclusive, what is? What higher tribunal could abrogate the logic of truth functions or of quantification?". We may wonder if Quine is talking here about *Logic* or about *logic*. Quine says that people dealing with non-classical negations are making a linguistic confusion; they don't know what they are talking about. Now can we say that

people by claiming that the earth was spherical rather than plane were not denying the doctrine, but just changing the subject?

The logic of truth functions is a system of logic dated from the beginning of the XXth century. It has some quality and it has some defects. To think that it describes the true *Logic* because it is a nice system would be like arguing that everything in the universe is spherical because spherical astronomy is a beautiful theory. Luckily enough our reasoning is not a slave of some laws of thought described by binary truth functions. Using our reasoning we can develop some new logical systems changing *Logic*.

33. The universality of logic

Let us first examine the question of universality of logic, as a science. It is not necessarily contradictory to argue that science is both relative and universal.[14]

We can say that universality is a fundamental and characteristic feature of science, in the sense that: (1) science is not a private business, it is objective, not subjective, not a question of taste; (2) science explains not the idiosyncrasies of a particular phenomenon, but some general patterns of phenomena. We can see that logic, since Aristotle, has these two universal features. The first feature is explicitly manifested through a theory like syllogistic which is a system with a set of rules. The second feature is also clear since syllogistic is concerned with all kinds of reasonings. On the other hand Aristotle thought it was impossible to develop a science of history, because for him there was no universality beyond the particulars turbulences of human societies – easier to develop logic, biology and meteorology.

Science is concerned with a double ALL, *ALL* minds and *ALL* objects. It is interesting to make a connection with the universal quantifier in logic. Chuaqui and Suppes (1995) have shown that classical physics can be described with a first-order logic theory with only universal quantifiers.

But a given science is not purely universal. A science like biology does not apply to all phenomena, it applies to a *portion* of reality, or better an *aspect* of reality. Biology is the science of life, so tautologically it does not apply to non-living things like a stone, but also

[14] Rougier was defending the relativity of logic and at the same time the unity of science (see Marion 2011).

it does not necessarily apply to all aspects of living beings, for example the basic concepts of biology are generally not used to explain the mathematical activity of a living being. So biology is not completely universal from outside. Neither it is from inside. There are different branches of biology, using different methods and having different objects of study. Studying whales is not the same as studying mushrooms, different also from studying trees. There are things in common and differences. Someone who is studying the origin of life is concerned with all living beings, but not directly with all aspects of all living beings. Within biology there are different levels of universality.

In the logic science, similar things happen: one may be concerned with some particular kinds of reasoning. There are many different kinds of reasoning, each having its specificity. When study legal reasoning, it is interesting to develop a deontic system of logic with an operator of obligation. This operator may have some common properties with other logical operators like necessity, but someone studying mathematical reasoning doesn't need to work with a system of logic with an obligation operator. To develop a big "universal" system of logic encompassing all the varieties of reasoning would be quite monstrous. But, as a biologist concerned with the very nature of living beings, one may be interested in the very nature of reasoning beyond all particular kinds of reasoning. In this case one may study some general concepts, like the notion of consequence relation. This goes in the direction of "universal logic" (see Béziau 2006) understood similarly to universal algebra (see Birkhoff 1987).

Such universal logic is not a universal system of logic, in the same way that general linguistics is not a universal language. General linguistics is the study of the common features of all languages. It is universal in this sense. Human beings are using thousands of different languages. We can ask the question "is language universal?" meaning "is there something in common beyond all particular human languages?" We can have a positive answer to this question, but this does not mean that what is beyond all particular human languages is itself a language. General linguistics may be considered as universal because having as object of study something which is universal, an object which is not a universal language, but the universal features of all languages. Because of the universal character of

its object of study and because of its methodology, general linguistics has the universal characteristic of a science, but without being itself a universal language.

Similarly it is possible to develop logic as a science whose object of study are the general features of all kinds of reasonings, not a universal reasoning. This logic science is universal because its object of study is universal, but it is neither a universal logical system nor a universal way of reasoning (for more details see Béziau 2010).

Jean-Yves Béziau
Universidade Federal do Rio de Janeiro

jyb@ifcs.ufrj.br

4. Bibliography[15]

Aberdein, A. (2008) 'Logic for dogs', IN S.D.Hales (ed), *What philosophy can tell you about your dog*, Chicago: Open Court, pp. 167-181.

Anellis, I.H. (1994) *Logic and its history in the work and writings of J. van Heijenoort*, Ames: Modern Logic Publishing.

Arnauld, A. and Nicole, P. (1662) *La Logique ou l'art de penser* (*Logic or the art of thinking*), Paris: Savreux.

Barrow, J.D. (2002) *The Constants of nature: From alpha to omega - The Numbers that encode the deepest secrets of the universe*, New York: Pantheon Books.

Belnap, N. (1977) 'A useful four-valued logic', IN M.Dunn (ed), *Modern uses of multiple-valued logic*, Boston: Reidel, pp.8-37.

Béziau, J.-Y. (1993**)** 'La critique Schopenhauerienne de l'usage de la logique en mathématiques', *O Que Nos Faz Pensar*, 7, pp.81-88.

———. (2006) '13 questions about universal logic', *Bulletin of the Section of Logic*, 35, pp.133-150.

———. (ed) (2010) Special Issue of *Logica Universalis* on the topic "Is logic universal?", number 2, volume 4.

———. (ed) (2011) *An Anthology of universal logic*, Basel: Birkhäuser.

Birkhoff, G. (1987) 'Universal algebra', IN Rota, G.-C. and Oliveira, J. S. (eds.), *Selected Papers on Algebra and Topology by Garret Birkhoff*, Birkhäuser, Basel, pp. 111-115.

Bochenski, J.M. (1956) *Formale Logik*, Karl Alber, Fribourg (English translation *A History of formal logic*, Indiana: Notre Dame, 1961).

———. (1990) *Entre la logique et la foi*, Editions noir sur blanc, Montricher.

Boole, G.: (1854) *An Investigation of the laws of thought, on which are founded the mathematical theories of logic and probabilities*, London, and Cambridge: Macmillan & Co.

Bourbaki, N. (1970) *Théorie des ensembles*, Paris: Hermann.

[15] Thank you to Arthur Ronald de Vallauris Buchsbaum for his support in finding references with the help of electronic devices and also to Amirouche Moktefi.

Brenner, J.E. (2009) *Logic in reality*, Berlin: Springer.

Carnap, R. (1928) Der *Logische Aufbau der Welt*, Felix Meiner, Leipzig (English translation *The Logical Structure of the World - Pseudoproblems in Philosophy*, University of California Press, 1967).

Carroll, L. (1871) *Through the Looking-Glass, and What Alice Found There*, London: Macmillan.

da Costa, N.C.A. (1980) *Ensaio sobre os fundamentos da lógica*, São Paulo: Hucitec.

Chuaqui, R. and Suppes, P. (1995) 'Free-variable axiomatic foundations of infinitesimal analysis: A fragment with finitary consistency proof', *The Journal of Symbolic Logic*, 60, 122–159.

Couturat, L. (1901) *La Logique de Leibniz – D'après des documents inédits*, Paris: Félix Alcan.

Descartes, R. (1637) *Discours de la méthode* (*Discourse on the method)*, Leyde.

———. (1628) *Regulae ad directionem ingenii* (*Rules for the direction of mind)*, published in Amsterdam, 1701.

Dieudonné, J. (1987) *Pour l'honneur de l'esprit humain*, Paris: Hachette. (English translation: *Mathematics – The music of reason*, Berlin: Springer, 1992).

Dostoyevski, F. (1864) *Zapiski iz podpolya* (Записки из подполья) (English translation: *Notes from Underground*).

Feferman, S. and Feferman, A.B. (2004) *Tarski: Life and Logic*, Cambridge: Cambridge University Press.

Février, P. (1937) 'Les relations d'incertitude d'Heisenberg et la logique', IN *Travaux du IXème Congrès International de Philosophie*, VI, Paris : Hermann, pp. 88-94.

Frege, G. (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens* (*Concept Notation, the Formal Language of the Pure Thought like that of Arithmetics*), Louis Nebert, Halle a. S.

Gaukroger, S. (1989) *Cartesian logic – An essay of Descartes's conception of inference*, Oxford: Oxford University Press.

Granger, G.-G. (1998) *L'Irrationnel*, Paris: Odile Jacob.

Haack, S. (1974) *Philosophy of logics*, Cambridge: Cambridge University Press.

Heidegger, M. (1927) *Sein und Zeit (Being and Time),* Tübingen: Niemeyer.

Heisenberg, W. (1958) *Physics and philosophy - The revolution in modern science*, New York: Harper.

Hilbert, D. (1965) *Gesammelte Abhandlungen*, 3 volumes, New York: Chelsea.

Kant, I. (1787) *Kritik der reinen Vernunft (Critique of pure reason,* second edition).

Kleene, S.C. (1952) *Introduction to metamathematics*, Amsterdam: North-Holland.

Kneale, W. and Kneale, M. (1962) *The Development of logic*, Oxford: Oxford University Press.

Marion, M. (2011) 'Louis Rougier on the relativity of logic – an early defence of logical pluralism', IN (Béziau 2011).

Moktefi, A. (2008) 'That's logic!', *Daresbury Lewis Carroll Society Journal*, 8, pp.23-22.

Moschovakis, J.R. (2009) 'The logic of Brouwer and Heyting', IN D.M.Gabbay and J.Woods (eds), *Handbook of the History of Logic – Volume 5,* Amsterdam: Elsevier.

Otte, M. (2008) *La Protohistoire*, Bruxelles: De-Boeck.

Pascal, B. (1657) *De l'esprit géométrique et de l'art de persuader* (*The art of persuasion*).

Peirce, C.S. (1877) 'The fixation of belief', *Popular Science Monthly*, pp.1-5.

Poincaré, H. (1905-06) 'Les mathématiques et la logique', *Revue de Métaphysique et de Morale*, 1905, p. 815-835, 1906, p. 17-38, et p. 294-317.

Quine, W.V.O. (1970) *Philosophy of logic*, New Jersey: Prentice Hall.

Rasiowa, H. and Sikorski, R. (1963) *The Mathematics of Metamathematics*, Warsaw: Polish Academy of Sciences.

Rougier, L. (1955) *Traité de la connaissance*, Paris: Gauthiers-Villars.

———. (1941) 'The Relativity of logic', *Philosophy and Phenomenological Research,* 2, pp.137-158, reprinted in (Béziau 2011).

de Saint-Exupéry, A. (1942) *Pilote de guerre* (*Flight to Arras*), New York: Editions de la maison française.

Schopenhauer, A. (1813) *Über die vierfache Wurzel des Satzes vom zureichenden Grunde*, Rudolstadt (*On the fourfold root of the principle of sufficient reason*).

Schanske, D. (2007) *Thucydides and the philosophical origin of history*, Cambridge: Cambridge University Press.

Smail, D.L. (2008) *On Deep history and the brain*, Berkeley: University of California Press.

van Stigt, W.P. (1990) *Brouwer's Intuitionism*, Amsterdam: North Holland.

Szabó, A. (1969) *Anfänge der griechischen Mathematik*, Budapest: Akademiai Kiádo (English translation: *The Beginnings of Greek mathematics*, Dordrecht: Kluwer, 1978).

———. (1994) *Die Entfaltung der griechischen Mathematik*, Mannheim: B.I. Wissenschaftsverlag.

Tarski, A. (1936) 0 *logice matematycznej i metodzie dedukcyjnej*, Lwów and Warsaw: Ksiaznica-Atlas (English translation; we refer here to the 4th edition by Jan Tarski: *Introduction to Logic and to the Methodology of the Deductive Sciences*, Oxford: Oxford University Press, 1994).

———. (1937) 'Sur la méthode déductive', in *Travaux du IXe Congrès International de Philosophie*, VI, Paris: Hermann, pp.95-103.

———. (1986) 'What are logical notions?', *History and Philosophy of Logic*, 7, pp.143-154.

Tierno, M. (2002) *Aristotle's Poetics for screenwriters: Storytelling secrets from the greatest mind in Western civilization*, New York: Hyperion.

Wang, H. (2006) *A Logical journey - From Gödel to philosophy*, Cambridge: MIT Press.

Wittgenstein, L. (1961) *Notebooks 1914-1916*, New York: Harper.

Woleński, J. (1989) *Logic and Philosophy in the Lvov-Warsaw School*, Dordrecht: Kluwer.

Woodger, J.H. (1937) *The Axiomatic method in biology, with appendices by Alfred Tarski and W.J.Floyd*, Cambridge: Cambridge University Press.

Woolf, D. (2011) *A Global history of history*, Cambridge: Cambridge University Press.

5. Acknowledgments
I don't remember when exactly I started to discuss the distinction between logic as reasoning and logic as the science of reasoning using the calligraphic difference "Logic" and "logic", but I can objectively trace it back via slides as far as a talk I presented at the *Santa Fe Institute* in New Mexico, September 29, 2005. Thank you to David Krakauer my host at the Santa Fe Institute.

I have good souvenirs of this trip to New Mexico followed by a trip from San Francisco to Los Angeles via Highway Number 1. My thanks go also to the late Herb Enderton who invited me to give a talk at UCLA on the same topic on October, 7, 2005, and who was also my host during a Fulbright stay at UCLA ten years before, in 1995.

Abstracta 6:1, pp.103-116, 2010

LOGICAL PROPERTIES OF IMAGINATION *

Alexandre Costa-Leite

Abstract

Inspired by Niiniluoto's account of the logic of imagination, this work proposes a combined logic able to deal with interactions of imagination, conception and possibility. It combines Descartes' view according to which imagination implies conception with Hume's view according to which both imagination and conception imply possibility.

1 Introduction

This study argues that imagination and conception are two weak kinds of possibility, although they are intrinsically connected. For this purpose, we have constructed a logic in which the relations between them is clearly defined. This system characterizes the minimal logic of imagination and related notions.

In order to understand the relations between imagination, conception and possibility, it is important to note that R.Descartes in [2] already proposed a distinction between them: for him, imagination implies conception, while conception does not imply imagination.[1] In this sense, we have two distinct levels of mental acts: imagining and conceiving. D.Hume in [6] defends an empiricist notion of

*Work supported by the *Swiss National Science Foundation (CH)* from 2007 to 2008 at the *City University of New York (USA)*, *FAPESP-Brazil* from 2008 to 2009 at the *State University of Campinas (BR)* and by *FINATEC (BR)* at the *University of Brasilia (BR)* (2010).

[1] We can recognize this same distinction in N. Vasiliev (see [11]): he states that we can conceive an n-dimensional logic but we cannot imagine it.

imagination, which we accept without restriction. Gendler and Hawthorne in [4], as well as R. Sorensen in [10], recognized in Hume a reduction of conceiving to imagining.

We accept that imagination and conception are distinct concepts in the sense defended by R. Descartes. But we also accept that D. Hume is correct while announcing that both imagination and conceiving imply possibility. Indeed, we have proposed a sound and complete combined logic showing that both Descartes' view and Hume's view are compatible.

Philosophical studies on imagination, conception and possibility (as well as their interactions) are frequent in the history of philosophy: from Descartes, Hume and Vasiliev to very recent studies as those which can be found in the book edited by Gendler and Hawthorne (see [4]) and also in the book edited by Nichols (see [8]), where many contemporary philosophers study the subject (Chalmers, Yablo, Fine, Stalnaker, Sorensen etc). However, none of the mentioned philosophers has proposed a logic of imagination and conception. Indeed, R. Sorensen vaguely proposed a "logic" of meta-conception. He even considered a conceivability operator and formalized it as C. Moreover, the author studied an interaction of conceivability and possibility, but he has reduced conceivability to conception.

Any attempt to formalize the concepts of imagination and conception should take into consideration the first and unique proposal to elaborate a logic of imagination developed by I. Niiniluoto in [9]. His approach has many merits, but also some gaps. His idea consists in exploring imagination as a modal operator in the same sense that J. Hintikka in [5] studied the notions of knowledge and belief. There are many kinds of epistemic notions which are usually called *propositional attitudes*. The first philosopher who developed a logical and formal account of epistemic notions is J. Hintikka in [5]. His work on epistemic logic has been important for all later work on the subject.

Considering imagination as an operator, Niiniluoto was able to investigate properties of it. He introduces the imagination operator I in order to formalize sentences of the form *an agent imagines that* φ. Therefore, he proposes the fol-

lowing set of axioms in Hilbert-style presentation plus one inference rule (where $\vdash$ means the standard notion of syntactical logical consequence - the basis of the system is classical logic):

1. $I(\varphi \rightarrow \psi) \rightarrow (I\varphi \rightarrow I\psi)$;
2. $I(\varphi \wedge \psi) \leftrightarrow (I\varphi \wedge I\psi)$;
3. From $\vdash \varphi$, we derive $\vdash I\varphi$;

Niiniluoto states that the above axioms are consequences of the following semantical condition: an agent a imagines φ in w if and only if φ is true in all possible worlds compatible with what a imagines in w. The author also argues that $I\varphi \rightarrow \varphi$ does not hold while $I\varphi \rightarrow \Diamond\varphi$ holds. So, imagination is viewed semantically as a kind of $\Box$-operator. Niiniluoto's approach to logical aspects of imagination is insufficient, considering that:

1. It examines a case of modal interaction without appealing to combining logics;
2. Metalogical properties of the logic are not examined;
3. It does not distinguish between imagination and conception;
4. Not intuitive inference rule.

Thus, given the insufficiency of Niiniluoto's approach, one has to search for a plausible logic of imagination. In the same way Niiniluoto has introduced a new operator to reason about imagination, we can go on and introduce another operator to reason about conception. So, we introduce in the object language an operator C formalizing sentences of the type *an agent* a *conceives* φ as $C_a\varphi$. Thus, we are able to construct an adequate environment to discuss about the distinctions and similarities between imagination and conception.

2 Imagination and conception as diamonds

Imagination is a faculty of minds able to generate images of objects (be they real or not). Whenever an agent imagines something (in this case the *something* is a particular proposition), we say that there is an *act of imagining*. We take imagination in an empirical fashion, following Hume's approach. This means that acts of imagination are connected to previous sense data. For theoretical reasons, we assume that the content of a given act of imagination is a proposition. Then, we speak about *propositional imagination* in the sense of [8].

Conception (or pure intellection, in Descartes' terminology) is a faculty of minds able to generate understanding of concepts and/or propositions. It is not necessarily related to images, but to comprehension. Whenever an agent conceives something (a proposition), we say that there is an *act of conceiving*.

Evidently, both imagination and conception are ways of representing things (representation mechanisms, acts of thought), and any act of imagining is an act of conceiving, but conception cannot be reduced to imagination. Our favorite example to elucidate this topic is the Cartesian one. Descartes in [2] has argued that although the mind cannot imagine a geometrical structure with a thousand sides, it can conceive it. In this sense, conception is a kind of understanding, a notion much more general than imagination.

We assume that agents imagine propositions, but they can imagine more (or less). Take these examples:

1. John imagines that it is raining in Manhattan;

2. John imagines Manhattan;

In (1) the content of imagination is the proposition *it is raining in Manhattan*, while in (2) the content of imagination is only *Manhattan*. Both can be understood as propositional imagination because although the first one is a proposition and the second a mere object, *Manhattan* can be viewed as a collection of properties and can be defined, therefore, as a collection of propositions. So, each object

corresponds in some sense to a given proposition. Given any object, for instance, *Manhattan* we can associate to it a proposition: *There exists Manhattan.*[2] This lead us to the view according to which all kinds of imagination can be reduced to propositional imagination.

Generally speaking, imagination is a weaker concept than that of conception which, in its turn, is a weaker concept than that of logical possibility. Comparisons between imagination, conception and possibility can find a good environment in modal logics.

There are many notions and philosophical distinctions concerning the concept of possibility. Basically, there are two kinds of possibility: empirical and logical. Empirical possibility depends of a given context: given a context X (a scientific area for instance), one can define the X-possibility. In this sense, some authors say there are things physically-possible, biologically-possible and so on. All these kinds of possibility can be reduced to what we call here empirical possibility. In this sense, a proposition is empirically-possible if and only if it does not contradict the underlying empirical theory. Logical possibility is something more general and it is our favorite notion of possibility. But there are indeed many ways one could define logical possibility. Consider a standard interpretation of logical possibility using the symbol $\Diamond$. Thus we can formalize sentences of the form "φ is possible" by $\Diamond\varphi$. Take also a Kripke frame. The notion of imagination is studied considering its relation with the notion of logical possibility. In order to define imagination one needs to use a notion of possibility able to capture in some sense the content of imaginative acts. In this sense we have (where R is an accessibility relation without restrictions):

$$\text{(MODAL)} w \vDash \Diamond\varphi \text{ if and only if } \exists w' \text{ such that } wRw', w' \vDash \varphi$$

This notion has been made clear by the developments of modal logic. It contains the key idea of possible worlds and given that possible worlds are important

[2] We can add quantifiers to the logic of imagination in order to prove this fact. Thanks to Niiniluoto for this remark.

for a part of the constitution of what is imagination, it follows that it is the choice in a logical theory trying to model the concept of imagination. The modal criterion of possibility can be applied to model imagination and conception. At the same time, these can be used to determine whether something is logically possible or not, playing a role of guides to possibility (See discussions on whether conceivability/imaginability are guides to possibility in [4]).

3 The logics of imagination, conception and possibility

The first thing to be said concerning a logic of imagination is that imagination can be treated as a modal notion. In this sense, it has some connections with different kinds of modality. It has a diamond-like truth condition. A logic of imagination is constructed using I (imagination), C (conception) and $\Diamond$ (possibility). Combined they can give rise to interesting philosophical interactions: $\Diamond I$ (imaginability) and $\Diamond C$ (conceivability), for instance.

Many formulas containing interactions of these notions can be presented. Studying how these operators behave is one of the motivations for a logic of imagination and related notions. While it is very difficult to compare the concept of imagination with necessity, it is very easy to compare it with possibility, given that imagination implies possibility seems to be plausible, but possibility does not imply imagination. Moreover, imagination does not imply necessity and vice-versa.

As we said, here we have to introduce in the language of modal logic for possibility a new modal operator in the same style of Niiniluoto. We represent this new operator by I and call it the *imagination operator*. We want our logic of imagination to respect some basic and most important properties of imagination, and we also want that it denies strange properties as for instance the property according to which imagination implies truth and that possibility implies imagination. So we have to build a formal system taking all these facts into consideration.

Consider the language L of classical propositional logic (**CPL**) defined by the

structure $L =< \wedge, \vee \rightarrow, \neg >$. Adding to this language the imagination operator, we generate the minimal language to describe imagination, let's call it $L_I =< \wedge, \vee \rightarrow, \neg, I >$. We repeat the procedure, taking now C for conception operator and $\Diamond$ for possibility in order to get languages L_C and $L_\Diamond$, respectively. So, we have three languages: one for imagination, other for conception and another for possibility.

For each language, consider $\sharp \in \{I, C, \Diamond\}$. Then, we define three axiomatic systems using a diamond-based presentation of **K** as the one proposed by Blackburn, De Rijke and Venema in [1] in order to guarantee normality:

1. $\sharp\bot \leftrightarrow \bot$;
2. $\sharp(\varphi \vee \psi) \leftrightarrow (\sharp\varphi \vee \sharp\psi)$;
3. $\vdash \varphi \rightarrow \psi$ then $\vdash \sharp\varphi \rightarrow \sharp\psi$.

Replacing uniformly each occurrence of $\sharp$ by I, C or $\Diamond$, we have three axiomatic systems. Therefore, from the syntactical viewpoint, we do not have any criteria to distinguish between imagination, conception and possibility. For each operator, we can build the related dual. In this sense, we have the dual of imagination $\Box_I\varphi$ (This dual is exactly Niiniluoto's imagination operator). This operator satisfies all standard axioms for $\Box$ and it is very useful in the completeness proof. The same holds for duals of conception and possibility.

For each axiomatic system, we have a respective frame such that for each $\sharp \in \{I, C, \Diamond\}$ we define $F_\sharp =< W, R_\sharp >$. Thus, imagination, conception and possibility have the following truth-condition:

$$w \Vdash \sharp\varphi \text{ if and only if } \exists w' \text{ such that } wR_\sharp w', w' \Vdash \varphi.$$

In the same way, semantically, we do not have elements to distinguish between imagination, conception and possibility. For each instantiantion of $\sharp$, we have a sound and complete logic with respect to its class of all frames. Let's call these

logics K_I, K_C and $K_\Diamond$. So, up to now, there are no tools to interact and reason about each operator in connection with another. However, the situation can change extending our logics by fusions and adding interaction axioms.

In order to define interactions of imagination, conception and possibility, let's take the fusion of the languages, axiomatic systems and frames. In this sense we have a logic

$$K_I \oplus K_C \oplus K_\Diamond$$

which is sound and complete with respect to the class of frames of the form

$$F =< W, R_I, R_C, R_\Diamond >$$

The proof of this could be constructed by canonical models or by preservation of completeness by fusions as developed by Fine and Schurz in [3] and Wolter and Kracht in [7]. However, even in the fusion, we cannot distinguish between imagination, conception and possibility.

We need to add interaction axioms in order to reason about the distinctions mentioned above. We use basic philosophical intuitions to determine which are the interesting axioms to be added in the fusion. In this sense, considering that imagination implies conception, we define Descartes-Vasiliev law:

$$I\varphi \rightarrow C\varphi$$

Considering that conception implies possibility, and imagination implies possibility, we define the so-called laws of Hume:

$$LH = \begin{cases} C\varphi \rightarrow \Diamond\varphi; \\ I\varphi \rightarrow \Diamond\varphi. \end{cases}$$

Obviously, the second law of Hume is a consequence of the law of Descartes-Vasiliev and the first law of Hume. Thus, we express the relations between these concepts, expanding the fusion in the following way:

$$K_I \oplus K_C \oplus K_\Diamond \oplus (I\varphi \rightarrow C\varphi) \oplus (C\varphi \rightarrow \Diamond\varphi)$$

Let's call this logic **IMAG**. This system is sound and complete with respect to the class of all frames F such that $R_I \subseteq R_C \subseteq R_\Diamond$. We denote this class of frames as $F^{\subseteq}$. **IMAG** has many very interesting properties. Before checking them, let's see that **IMAG** is sound and complete with respect to $F^{\subseteq}$.

For soundness, we need to verify that both Descartes-Vasiliev law and the first law of Hume - interaction axioms added to the fusion - are valid. That the inference rules and other parts of the logic preserve validity is evident. To check that the law of Descartes-Vasiliev is valid, take $w \vDash I\varphi$ but $w \nvDash C\varphi$. Thus:

1. $w \vDash I\varphi \Longleftrightarrow \exists w'$ such that $wR_I w'$, $w' \vDash \varphi$;

2. $w \nvDash C\varphi \Longleftrightarrow \forall w'$ such that $wR_C w'$, $w' \nvDash \varphi$;

Given that $R_I \subseteq R_C$, it follows the desired result. The same argument applies to the first law of Hume.

For completeness, we need to show that all **IMAG** valid formulae are theorems. We can proceed by canonical models method adapting standard proofs. While proving completeness we need to consider dual operators of I, C and $\Diamond$ as well their properties, which behave like $\Box$ operators.

The result presented here can be shown to be a special case of a general result on interaction axioms. Consider a hierarchy of diamond operators:

$$\Diamond_1, \Diamond_2, ..., \Diamond_n$$

such that each $\Diamond_i$ is weaker than a $\Diamond_j$ if $i \leq j$.

Each language containing a $\Diamond_i$ generates a logic which only modal operator is $\Diamond_i$, for some i. Thus, we define a fusion

$$K_{\Diamond_1} \oplus K_{\Diamond_2} \oplus ... \oplus K_{\Diamond_n}$$

This fusion can be expanded by the addition of finite many interaction axioms in the following way:

$$K_{\Diamond_1} \oplus K_{\Diamond_2} \oplus ... \oplus K_{\Diamond_n} \oplus (\Diamond_1 \varphi \rightarrow \Diamond_2 \varphi) \oplus ... \oplus (\Diamond_{n-1} \varphi \rightarrow \Diamond_n \varphi)$$

The above fusion is sound and complete with respect to the class of frames $F^{\subseteq}_{\Diamond} =< W, R_{\Diamond_1}, R_{\Diamond_2}, ..., R_{\Diamond_n} >$ such that $R_{\Diamond_1} \subseteq R_{\Diamond_2} \subseteq ... \subseteq R_{\Diamond_n}$.

4 Properties of IMAG

Now we answer questions posed in the literature on the relations between imagination, conception and possibility. We note that derived notions such as imaginability and conceivability cannot be reduced to imagination and conception, respectively, if some restrictions are not added to the accessibility relations. The following are valid in **IMAG**:

Interactions	Distributions	Connections
$I\varphi \rightarrow C\varphi$	$I(\varphi \wedge \psi) \rightarrow (I\varphi \wedge I\psi)$	$I\Diamond\varphi \leftrightarrow \Diamond I\varphi$
$C\varphi \rightarrow \Diamond\varphi$	$C(\varphi \wedge \psi) \rightarrow (C\varphi \wedge C\psi)$	$IC\varphi \leftrightarrow CI\varphi$
$I\varphi \rightarrow \Diamond\varphi$	$\Diamond(\varphi \wedge \psi) \rightarrow (\Diamond\varphi \wedge \Diamond\psi)$	$C\Diamond\varphi \leftrightarrow \Diamond C\varphi$

These would be valid if **IMAG**-frames were reflexive and transitive, respectively:

Reflexive	Transitive
$\varphi \rightarrow I\varphi$	$\Diamond C\varphi \rightarrow \Diamond\varphi$
$\varphi \rightarrow C\varphi$	$\Diamond I\varphi \rightarrow \Diamond\varphi$
$\varphi \rightarrow \Diamond\varphi$	$\Diamond\Diamond\varphi \rightarrow \Diamond\varphi$

None of the formulae below is valid in **IMAG**:

$I\Diamond\varphi \rightarrow \Diamond\varphi$	$I\Diamond\varphi \rightarrow \varphi$	$C\varphi \leftrightarrow \varphi$
$I\Diamond\varphi \rightarrow I\varphi$	$IC\varphi \rightarrow C\varphi$	$\Diamond\varphi \leftrightarrow \varphi$
$I\Diamond\varphi \rightarrow \varphi$	$IC\varphi \rightarrow I\varphi$	$\Diamond\varphi \rightarrow C\varphi$
$C\Diamond\varphi \rightarrow \Diamond\varphi$	$IC\varphi \rightarrow \varphi$	$C\varphi \rightarrow I\varphi$
$C\Diamond\varphi \rightarrow C\varphi$	$I\varphi \leftrightarrow \varphi$	$\Diamond\varphi \rightarrow I\varphi$

Considering that **IMAG** has standard metalogical properties, it can be used to settle disputes on the properties of imagination, conception and possibility, as well its interactions. It can be useful to the philosopher lost in the plurality of debates founded in the literature, as for instance those in the books [4] and [8]. Thus, using basic properties of **IMAG**, let's discuss what we consider to be the most interesting properties of it, approaching problems we can find in the literature.

4.1 Descartes-Vasiliev law

Descartes in [2] proposed a distinction between imagination and pure intellection (conception), using the very intuitive example that we can imagine a triangle but we cannot imagine a chiliagon. We can conceive it: understand that it is a figure composed by a thousand sides. Considering this example, it seems very plausible to accept that *imagination implies conception*, if we take imagination as a faculty of generating images while conception as a faculty of understanding a concept (or proposition), even without images. Thus, what we have called Descartes-Vasiliev law is a plausible principle which all logics of imagination should satisfy.

Moreover, Vasiliev's imaginary logic is not a logic of imagination, but this one is the logic of the imaginary worlds. Vasiliev is obviously also concerned with conception.

4.2 Hume's laws

D. Hume collapses the notions of imagination and conception, using both in the same empirical sense. This collapse we cannot accept. However, he is right to state in [6] that both concepts imply possibility. Thus, if we are trying to determine whether a given proposition is logically possible, the best thing to do is to check whether the proposition can imagined or conceived. In this sense, to be able to imagine or conceive φ is a clue to the possibility of φ. It seems impossible to find an intuitive counter-example to these laws.

4.3 Conceivability and imaginability

The reader is now able to distinguish between concepts. Conceivability is an hybrid notion, while conception is a primitive, non-interactive concept. Inside the environment of **IMAG**, it is quite natural to find a counter-example to $\Diamond C\varphi \leftrightarrow C\varphi$. However, if we are in transitive frames, the equivalence holds. The same argument applies to imaginability and imagination. So, the best answer to the question posed in [4] of whether conceivability is a guide to possibility or not, is to state that it depends in what kind of frame our concepts are used. Conceivability and imaginability are good guides to possibility if and only if our frames are transitive. Otherwise, we can find useful counter-models.

5 Conclusion

This text has proposed some new ideas concerning the logics of imagination presented by Niiniluoto. One of the claims of this paper is that Niiniluoto's account is insufficient to deal with imagination and related notions. Other plausible claim of this paper consists in showing that to each object we can associate a proposition and, then, we use this fact to show that any kind of imagination can be reduced to propositional imagination.

The main conclusion of this paper is that imagination and conception are two kinds of possibility. Without interaction axioms relating these notions, they remain the same from the syntactical and semantical viewpoint. This article combined Descartes and Hume's position showing that they are compatible. Using tools from combining logics, we have proposed a combined logic of imagination, conception and possibility showing that the resulting system is sound and complete with respect to combined Kripke frames with special properties in the accessibility relations.

As a very interesting open question to be studied in the future, we can point out how would it be a version of **IMAG** able to deal with contradictions? In this system, we would be able to formalize contradictory conception and contradictory

imagination. Thus, we would need to change the underlying logic.

6 Acknowledgements

Thanks to Arnold Koslow, Walter Carnielli, Ilkka Niiniluoto and Fabien Schang for comments on this text. Thanks also to Jean-Yves Béziau and Gillman Payette for discussions on the nature of imagination and its logic.

Alexandre Costa-Leite
University of Brasilia
costaleite@unb.br

References

[1] BLACKBURN, DE RIJKE, VENEMA. (2001). Modal Logic. Cambridge University Press.

[2] DESCARTES, R. (1641). Méditations métaphysiques. In *Oeuvres complètes*: Vrin.

[3] FINE, K; SCHURZ, G. (1996). *Transfer theorems for stratified modal logics*. In Logic and Reality. In memory of Arthur Prior, pages 169-213. Oxford University Press.

[4] GENDLER, T. S; HAWTHORNE, J.(eds.) (2002). Conceivability and Possibility. Oxford University Press.

[5] HINTIKKA, J. (1962). Knowledge and Belief: An Introduction to the Logic of the Two Notions. Cornell: Cornell University Press.

[6] HUME, D. (1748). A Treatise of Human Nature. Edited by D. Norton and M. Norton. Oxford University Press.

[7] KRACHT, M.; WOLTER, F. (1991). Properties of independently axiomatizable bimodal logics. Journal of Symbolic Logic 56, 14691485

[8] NICHOLS, S.(2006). The Archicteture of the Imagination. New Essays on Pretence, Possibility, and Fiction. Oxford University Press.

[9] NIINILUOTO, I. (1985). *Imagination and Fiction*. Journal of Semantics 4(3), p. 209-222.

[10] SORENSEN, R. (2006). *Meta-conceivability and Thought Experiments* in [8], 257-272.

[11] VASILIEV, N. (1912). *Lógica Imaginária*. In N.A.Vasiliev e a Lógica Paraconsistente. Org. by Ayda Ignez Arruda, Campinas: Coleção CLE, v.07, 1990.

Abstracta 6 : 1, pp. 117 – 135, 2010

CARNAP'S PROBLEM: WHAT IS IT LIKE TO BE A NORMAL INTERPRETATION OF CLASSICAL LOGIC?*

Arnold Koslow

Abstract
Carnap in the 1930s discovered that there were non-normal interpretations of classical logic - ones for which negation and conjunction are not truth-functional so that a statement and its negation could have the same truth value, and a disjunction of two false sentences could be true. Church argued that this did not call for a revision of classical logic. More recent writers seem to disagree. We provide a definition of "non-normal interpretation" and argue that Church was right, and in fact, the existence of non-normal interpretations tells us something important about the conditions of extensionality of the classical logical operators.

1. Carnap's Problem

In the decade from the early thirties to the mid fifties, there was a brief and scattered discussion of a problem raised by B.A. Bernstein (1932), R.Carnap (1943), and A. Church (1944, 1956) of what has now been referred to referred to as "Carnap's Problem". Carnap discovered the existence of what Church later called "non-normal interpretations" of sentential classical logic, and first –order logic. Church's major criticism of Carnap's reformulation of sentential logic was that it essentially incorporated semantical assumptions into what was supposed to be a syntactically presented formulation of the logic.

In what follows, I shall consider only sentential logics. Roughly speaking Carnap took interpretations to be truth-value assignments (He called them "interpretations.") which assigned truth to all theorems, and which respected deducibility—that is, if some collection of sentences is true under an interpretation τ, then any sentence deducible from those sentences is also true under τ. What Carnap discovered was that there were interpretations of the classical sentential calculus which assigned the same truth value to statements as well as

*A paper read to the Logic Seminar, Cambridge University April 23, 2009. Many thanks to Peter Smith, Luca Incurvati, and the other members of that seminar.

their negations, and interpretations which assigned “true” to a disjunction while also assigning “false” to all of its disjuncts.

Church thought was that there was no need to correct the formulation of the logic. There was no “deficiency” in the formalization, and no need to seek a “fuller formalization”. Presumably if one just avoided the use of these non-normal interpretations, there would not be any mismatch between the proof-theoretic (syntactic) deductive presentation of the system, and the usual (semantic) truth-tables for the logical connectives.

In the more recent literature devoted to Carnap’s discovery, the issue takes on a more serious cast. If one thought that the truth-tables provided the meaning of the logical connectives (I do not), then if the proof-theoretic formulation does not match up with the tables then one might put the significance of Carnap’s discovery as showing that the (deductive) rules of inference do not determine the meanings of logical constants (Raatikainen (2008)). Other writers (Murzi & Hjortland (2010)) have taken the moral to be one that concerns inferentialism, and a problem concerning a special kind of categoricity. Shoesmith and Smiley (1978) have explored specific examples of non-normal interpretations from the vantage of multiple -conclusion logics, using essentially a four-element Boolean algebra, and Smiley (1996), Incurvati & Smith (2009) have considered embedding the sentential calculus in a system with rules of rejection and acceptance (a “fuller fomalization”?) to eliminate the mismatch. Rumfitt (1997 & 2000) and a host of other logicians have made their own case for understanding the import of these strange truth-value assignments.

In addition to all of these, I now wish to reconsider “Carnap’s challenge” and its import from a more general “structuralist” vantage. This kind of approach has been explained at length in Koslow (1992), and somewhat differently but in lesser length in Koslow (1999). The following discussion however is intended to be self contained.

2. The more general structuralist background

To indicate the generality of "Carnap's Problem" we shall use the notion of an *implication structure* $\mathfrak{I} = \langle S, \Rightarrow \rangle$, where S is any non-empty set, and “$\Rightarrow$” is an *implication relation.* That is, any relation on S satisfying the following six slightly redundant conditions:

(1) ***Refexivity***: $A \Rightarrow A$, for all A in S.

(2) ***Projection***: $A_1,A_2,...,A_n \Rightarrow A_k$, for any k= 1,...,n.

(3) ***Simplification*** (sometimes called Contraction): If $A_1,A_1,A_2,...,A_n \Rightarrow B$, then $A_1,A_2,...,A_n \Rightarrow B$, for all A_i and B in S.

(4) ***Permutation***: If $A_1,A_2,...,A_n \Rightarrow B$, then $A_{f(1)},A_{f(2)},...,A_{f(n)} \Rightarrow B$, for any permutation f of {1,2,...,n}.

(5) *Dilution*: If $A_1,A_2,...,A_n \Rightarrow B$, then $A_1,A_2,...,A_n,C \Rightarrow B$, for all A_i,B, and C in S.

(6) ***Cut***: If$A_1,A_2,...,A_n \Rightarrow B$, and $B, B_1,B_2,...,B_m \Rightarrow C$, then $A_1,A_2,...,A_n,B_1,B_2,...,B_m \Rightarrow C$.

These conditions are of course those which G.Gentzen put forward as the structural conditions for implication. We understand them in a very general sense as giving a story about implication that does not appeal to truth or any other familiar semantical concept. And it is conspicuous that the story is told without any appeal to the logical operators. They (that is, the operators of conjunction, disjunction, the conditional, negation, and universal and existential quantification), as it turns out can all be defined in terms of implication. It is also a feature of the structuralist story that the set S is not restricted to syntactically presented elements. In that lies the generality of this way of looking at things. Nevertheless it is a generality that will not be used in the following.

Here are a few definitions that we shall need in order to show how, despite the absence of apparently semantical concepts in this story, we can in fact define the concept of truth-value assignments (valuations), and obtain with them, a remarkable completeness theorem for the theory of implications given by (1) – (6). This will allow us to introduce the semantic notion of a valuation in the structuralist setting.

(i) A ***bisection*** on S is any ordered pair T = <K, L> where K and L are non-empty subsets of S that are disjoint , and whose union is S.

(ii) Let T = <K, L> be a bisection on S. Then $\Rightarrow^T$ is the corresponding bisection implication relation defined as follows:

$A_1,A_2,...,A_n \Rightarrow^T B$ if and only if some A_i is in K or B is in L.

(iii) Let $\mathfrak{I}$ = <S, $\Rightarrow$ > be an implication structure, then any subset R of S is ***strongly closed*** under the implication relation if and only if whenever several members of R together imply A then A also belongs to R. We shall say that any bisection T = <K, L> on an implication structure is a strong bisection on $\mathfrak{I}$ if and only if L is strongly closed under the implication relation $\Rightarrow$ of the structure.

3. Truth-value assignments (valuations)

We can now show several interesting facts about strong bisections, on the basis of which we can define the notion of a ***truth-value interpretation*** on arbitrary implication structures.

(iv) Let $\mathfrak{I}$ = <S, $\Rightarrow$ > be an implication structure. Then for any strong bisection implication relation $\Rightarrow^T$ on $\mathfrak{I}$, we have $\Rightarrow \subseteq \Rightarrow^T$. That is, for any structure, every strong bisection implication $\Rightarrow^T$ on it extends $\Rightarrow$.

(v) The strong bisection implications are maximal. That is, if $\Rightarrow^T$ and $\Rightarrow^{T*}$ are strong bisection implication relations on $\mathfrak{I}$, and $\Rightarrow^T \subseteq \Rightarrow^{T*}$, then $\Rightarrow^T = \Rightarrow^{T*}$.

We now come to the basic result that allows the introduction of truth-value interpretations on implication structures:

(vi) **Lindenbaum-Scott Completeness I (**Scott, 1974). Let $\mathfrak{I}$ be any non-trivial implication structure (there are at least two elements of it neither of which implies the other). Then $A_1,A_2,...,A_n \Rightarrow B$ if and only if $A_1,A_2,...,A_n \Rightarrow^T B$, for all strong bisection relation $\Rightarrow^T$ on the structure.

Another way of stating this result is that if we define an implication structure as complete if and only every member of it is either a thesis (it is implied by everything in the structure, or

it is an antithesis, it implies everything in the structure, then the Lindenbaum-Scott Theorem says that every non-trivial implication relation is the intersection of all complete implication relations that extend it.

> Simple proof: For any (non-trivial) implication structure $\mathfrak{I}$ = <S, $\Rightarrow$ >, clearly every bisection implication relation $\Rightarrow^{T}$ extends the implication relation $\Rightarrow$, because L is strongly closed under the implication relation $\Rightarrow$. The converse is fairly straightforward: Suppose that $A_1,A_2,...,A_n \Rightarrow^{T} B$ for all bisection implication relations $\Rightarrow^{T}$, but $A_1,A_2,...,A_n \Rightarrow B$ fails. We define a strong bisection T* = <K, L> as follows: Let L be the set of all members C of S such that $A_1,A_2,...,A_n \Rightarrow C$. L is strongly closed and is also non-empty since it contains all the A_i. Let K be the rest of S, that is, all C such that $A_1,A_2,...,A_n \Rightarrow C$ fails. It is non-empty since B is in it. So T is a strong bisection on the structure. Therefore, by hypothesis, $A_1,A_2,...,A_n \Rightarrow^{T*} B$. Therefore some A_i is in K or B is in L. But none of the A_i are in K so B is in L. But that is impossible. Consequently, $A_1,A_2,...,A_n \Rightarrow B$.

We can now define the notion of a truth-value assignment for arbitrary (non-trivial) implication structures. Simply stated, truth-value assignments (valuations) on implication structures are uniquely associated with strong bisections on those structures. That is

If $\mathfrak{I}$ = <S, $\Rightarrow$ > is an implication structure, then any truth-value assignment on it is a function τ associated with a strong bisection T = <K, L> on it such that for any A in S,

$\tau(A) = t$, if A is in L, and

$\tau(A) = f$, if A is in K.

With this notion of a valuation in place, the Lindenbaum-Scott theorem can be stated in a way that is truly a completeness result:

> **Lindenbaum-Scott Completeness II.** Let $\mathfrak{I}$ be any non-trivial implication structure. Then $A_1,A_2,...,A_n \Rightarrow B$ if and only if for all strong bisection relations τ on the structure, if $\tau(A_i) = t$ for all A_i, then $\tau(B_i) = t$. That is the implication relation on the structure holds if and only if every valuation that makes each of the premises true, also makes the conclusion true.

A few observations are in order. In some of the recent literature on Carnap's Problem, much has rested on the truism that if the premises imply a conclusion, then if the premises are true under a valuation, then the conclusion has got to be true as well. Some distinguished logicians like J. Myhill, and J. Corcoran have had their doubts whether this is even correct, but the simple proof shows that something like it is correct. However, a glance at the simple proof of completeness shows it to be a result which holds without assuming how the valuations behave with respect to the logical operators. Thus on the present story, it is not at all plausible, that without some additional assumptions such matters as how valuations distribute over conjunctions, disjunctions, negations and conditionals will be settled, or indeed whether they can be settled in any but special cases.

Nevertheless, it is this theorem that motivates our taking these valuations as truth-value assignments. It is what one sees in the usual classical case, only instead of the usual truth values, we take truth (falsity) of an interpretation to be just membership in the sets L and K of the associated strong bisection (this is an insight which is due to D. Scott (1974).

It is worth recalling that this notion of a truth-value assignment relies only on the notion of an implication relation as we described it using Gentzen Structural conditions. Those conditions for implication made no appeal to any notion of truth, or truth-value assignments. So the definition of these valuations does not rely on some hidden semantic devices.

Nevertheless, there is a fair amount of semantic information that can be gleaned. It can be shown that

For any conjunction [A $\wedge$B is in L if and only if A is in L and B is in L].

For negation, [If A in in L then $\neg$A is in K] (but not conversely.

For disjunctions, [If A is in L or B is in L, then (A $\vee$ B) is in L] (but not conversely), and

For conditionals [If (A $\rightarrow$ B) is in L, then either A is in K or B is in L] (but not conversely).

So if we want more, we need to require more than just truth-preservation under all valuations.

4. Carnap's Problem, non-normal truth-value interpretations, and the logical operators

With the Lindenbaum-Scott theorem in place we are in a position to reconsider Carnap's Problem. It involves the observation that there are certain truth-value valuations which satisfy the condition that any sentence that is implied by sentences that are all assigned the value "true", will also be assigned the value "true", which, nevertheless, lead to unwanted consequences for the logical operators. It will turn out that there are assignments that will assign the value "f" both to a sentence and it's negation, and some that will also assign "t" to a disjunction of sentences each of which has been assigned "f". There's something peculiar about the examples of non-normal valuations that Carnap, and Church provided. But we shall see with the help of other examples that there are such non-normal valuations on implication structures even when the notion of a truth-value assignment is given the clear foundation provided by the Lindenbaum-Scott theorem. The "Carnap" phenomenon is real.

It is instructive to note that there are cases of implication structures with implication relations that do not give rise to a Carnap Problem; all truth-value assignments for the logical operators behave in the expected way. Let S be the sentences of the classical sentential calculus, and let the implication relation be given by the bisection implication relation $\Rightarrow^{T}$, where T is a strong bisection <K, L> (that is, L is closed under the implication relation $\Rightarrow^{T}$). It is easy to see that the following holds:

(1) $(A \wedge B)$ is in L (t) if and only if A is in L (t) and B is in L (t).

(2) $\neg A$ is in L (t) if and only if A is in K (f).

(3) $(A \vee B)$ is in L (t) if and only if A is in L (t) or B is in L (t).

(4) $(A \rightarrow B)$ is in L (t) if and only if A is in K (f) or B is in L (t).

That is, all the operators behave in the familiar extensional pattern for this particular truth-value assignment. Later, we shall see that this is exactly what extensionality with respect to a valuation requires. There is no "Carnap Problem" here. This is however, a case of a special implication relation on the sentences of classical sentential logic. The proof theory and this semantics are perfectly matched. Things don't always go this smoothly.

Consider the following structure: $\Im_{CSC}$, where S is the set of sentences of the Classical Sentential Calculus (CSC), and the implication relation $\Rightarrow$ is one given by say the standard deductive rules for (CSC). One would have thought that for such a familiar classical system, it would be obvious that the logical operators would all exhibit the same extensional pattern for every truth-value assignment. That is not so. We know that this structure has theses (those members of S which are implied by everything in the structure), that it has antitheses (those members of S which imply every member of the structure), and it is also incomplete in the sense that there are sentences in S such that neither they nor their negations are theses (this is sometimes call *syntactic incompleteness*, and sometimes *incompleteness with respect to negation*). Consider the following valuation: Let T = <K, L>, where L is the set of all theses, and K is the rest of S:

$\tau(A) = t$, if A is a thesis (A is in L), and
$\tau(A) = f$, if A is not a thesis (A is in K).

Let A_0 be a member of S such that neither it nor its negation is a thesis. Since the structure is classical, it follows that

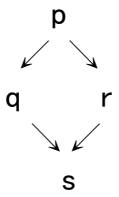
$\tau(A_0 \vee \neg A_0) = t$, and since neither A_0 nor $\neg A_0$ are theses,
$\tau(A_0) = \tau(\neg A_0) = f$.

Evidently under this assignment, a disjunction is assigned "t" although its disjuncts are both assigned "f". Furthermore there is a statement, A_0, such that it and its negation get assigned the same value. This is of course not your standard extensional distribution of truth-values.

We mentioned in passing that Carnap offered the example of a non-normal valuation that consisted in the assignment of "truth" to all sentences in a structure. In that case we get the bizarre distribution according to which every sentence and its negation are assigned the same value. Both these assignments satisfy the requirement that for any implication, if all the premises are "true" then so too is the conclusion. That looks like a "cheap shot". Nevertheless the response to such an example cannot be that we exclude such distributions, since that looks arbitrary. We have provided a less contentious example which supports Carnap's essential point.

In any case we cannot make use of Carnap's version since all the valuations provided by the Lindenbaum-Scott Completeness Theorem are based on (strong) bisections. If all sentences were assigned "t", then they would all belong to the set L, so that the set K would be empty. And that is impossible.

There are more exotic examples of which I shall consider just one. D. J. Shoesmith and T.J. Smiley give an example (*Multiple-Conclusion Logic*, p.3) in which there are four truth-values. It is possible to see their example as a case where the implication structure is given by a set S of four elements {p, q, r, s} where p implies q and implies r, q implies s, r implies s, and neither r nor s imply each other. In this case consider L to be the set {s}, and K to be the set {p,q,r}. Membership in L is "t", membership in K is "f". The structure looks like this:



In this four-membered Boolean algebra, s is the negation of p (and conversely), q is the negation of r (and conversely), the disjunction of q and r is s, so that the disjunction of two elements (q and r) that are "f", is "t", and q and its negation r are both "f". Negation is classical. Church, in his review of Carnap's *Formalization of Logic*, had already indicated that non-normal truth-value assignments could be based on a four-element Boolean algebra.

This example involves cases where there are various different kinds of falsity {p, q, r, }, and in such cases, one might be inclined to dismiss the example as a case of a generalization to multiple truth-values, and one might have expected with such a generalization, that distribution patterns of truth and falsity would lead to these peculiar results. That reaction however is not warranted. In order to see whether these non-normal interpretations are a serious problem requiring serious modification in the presentation of some of our standard logics, we have to have to have a better working definition of normal and non-normal assignments. That is the next task.

5. The Normal, and the Non-Normal

We can distinguish the normal truth-value valuations from those which are non-normal in a very simple systematic way, rather than appeal to a non-homogeneous collection of various clever, but strange constructions. Recall that a valuation on a structure $\mathfrak{I}$ = <S, $\Rightarrow$ > is a function which for any *strong* bisection <K, L> assigns "t" or "f" to a member of S according as it belongs to L or to K. And the bisection is strong just in case L is strongly closed under the implication relation $\Rightarrow$. Recall as well, that by the Lindenbaum-Scott theorem, it is guaranteed that an implication relation holds between premises and a conclusion if and only if whenever any valuation is true for all of the premises, then it is also true for the conclusion.

Among all the valuations we shall single out those which are *normal*, for which we shall need the notion of the ***dual*** of an implication relation. It was R. Wojcicki (1973) who first defined the important notion of the dual of an implication relation ("Dual Counterparts of Consequence Relations", in *Bulletin of the Section of Logic*, v.2, n.1, 1973, pp. 54-57, and they have been studied at some length in Koslow (1992). Although Wojcicki seems not to have made any use of his discovery in his later writings, it is a seminal notion. Implication relations which are duals of other implication relations play a powerful role in defining a general notion of the duality of logical operators, and provide some insight into multiple conclusion logic as well. They also show that there are implication relations and consequence relations which are falsity-preserving rather than truth-preserving. [See also Koslow, *A Structuralist Theory of Logic* (1992).]

The dual of any implication relation is also an implication relation satisfying our (6) conditions, and it can be shown to satisfy two conditions: (1) A dually implies B (A $\Rightarrow^{\wedge}$ B) if and only if B implies A (B $\Rightarrow$ A), and although the dual is defined in all structures, in those structures where disjunctions exist, it reduces to this: (2) A finite number of premises dually implies B if and only if B implies their disjunction.

We now can introduce the notion of a *normal bisection*, which is a strong bisection with an additional condition: Let T = <K, L> be a bisection of an implication structure $\mathfrak{I}$ = <S, $\Rightarrow$ >. Then T is a *normal bisection* on $\mathfrak{I}$ if and only if:

> (1) L is strongly closed under the implication relation $\Rightarrow$ of the structure, and
>
> (2) K is strongly closed under $\Rightarrow^{\wedge}$, the dual of the implication relation on the structure.

In the usual way, a *normal valuation* on the structure is one which for any normal bisection assigns "t" to the members of L, and "f" to those of K.

> A **non-normal valuation** on a structure with implication relation $\Rightarrow$ is a valuation based on a strong bisection <L, K> for which (1) L is strongly closed under the implication relation, but (2) K is not strongly closed under the dual of that implication relation.

The restriction of valuations to normal ones introduces a nice symmetry in their construction: L is strongly closed under implication, and K is strongly closed under the dual implication. But there is more than just a symmetry that is reflected here. Normalcy requires that the concepts of true under a valuation and false under a valuation be duals of each other.

What we have in mind is the following: In the case of the logical operators on implication structures, the duals of operators can be obtained by taking their definitions which are framed in terms of implication, and simply replace the implication relation everywhere in that definition by it's dual. Thus for example, if in the characterization of conjunction we replace the implication relation by its dual, the result is the characterization of disjunction.

Similarly, we suggest, consider the characterizations we gave for truth (and falsity) in a normal valuation:

> (1) For the normal bisection T = <K, L>, A is "t" if and only if (L is strongly closed under $\Rightarrow$) and (A is in L).
>
> And the dual would be given by
>
> (2) For the normal bisection T = <K, L>, A is "f" if and only if (K is strongly closed under $\Rightarrow\wedge$) and (A is in K).

In other words, the assignment of falsity to the members of K is what the assignment of truth to the members of L becomes if we replace implication of the structure with its dual. Another way of seeing the connection with duality is to consider the simpler case: Let T = <K, L> be a strong bisection on a structure. The Ls are the truths, and the Ks are the falsehoods (for this bisection of course). For any member of L, anything which it implies (using $\Rightarrow$) is true, and so in L. Now the Ks are false, and anything which implies them (using $\Rightarrow$) is false, and so in K. However, any A which implies (using $\Rightarrow$) some B in K is such that B dually implies A (B $\Rightarrow\wedge$ A). So K is closed under the dual implication. In effect, the same thing is going on, only in the one case it is by implication, and in the other it is by its dual.

Thus we see that the motivation to consider the normal valuations as the ones to use, is not to preserve some core logical truths like disjunctions being true if and only if at least one disjunct is true. There is the other possibility that the restriction to normal valuations respects a feature of truth and falsity: that they are dual concepts.

In certain recently studied logical systems, that duality has not been preserved. That doesn't mean that logic has been left in dire straits, and needs to be rescued. It only means that there are other paths that logicians can study and even advocate.

We can now see with this notion of normality in place, that ***if*** there are non-normal valuations on an implication structure, then there are going to be deviations from the usual distributional patterns for some of the logical operators.

Here is why we say "if": There are some implication structures on which all the valuations are normal, and there are some implication structures on which some valuations are normal and some are not. Here are some examples of these possibilities:

(1) If $\Im$ = <S, $\Rightarrow^T$ >, where the implication relation on S is a bisection implication relation on S, then it is easily shown that in that case where T = <K, L> is a bisection, then the valuation on it has to be normal. All the operators as a consequence have the familiar extensional distribution features. This was the example we already discussed on pp.123-4.

(2) Suppose that there is a non-normal valuation on an implication structure $\Im$ = <S, $\Rightarrow$ > in which disjunctions exist. Then there will be a disjunction such that the non-normal valuation will assign "f" to each of the disjuncts, but assign "t" to the disjunction. The proof is straightforward

> There is a strong bisection <K, L> in which L is closed under implication, but K is not closed under the dual implication. Then there will be some $A_1,A_2,...,A_n$ and B in S, such $A_1,A_2,...,A_n \Rightarrow\!\wedge$ B, all the A_i are in K, but B is not. Then $B \Rightarrow (A_1 \vee A_2 \vee ... \vee A_n)$. B is in L, so $(A_1 \vee A_2 \vee ... \vee A_n)$ is also in L. Consequently we have a disjunction of members all assigned "f" by the non-normal valuation, but their disjunction is assigned "t". This shows that given our notion of a non-normal valuation, then in a very broad variety of cases, there will be Carnapian style examples of a non-standard distribution of truth-values.

(3) Here is a specific example of a non-normal valuation. Consider the classical implication structure (CSC) that we referred to earlier. Let <K, L> be a bisection where L is the set of all theses of (CSC), and K is the set of the remaining sentences of S (all the non-theses). (CSC) is incomplete (with respect to negation), so there is some sentence A_0 such that neither it nor its negation are theses of the structure. This is a non-normal bisection on the classical sentential calculus: L is certainly closed under the usual classical implication relation (say) $\Rightarrow$, but K is not closed under its dual. The reason is that $A_0, \neg A_0 \Rightarrow\!\wedge (A_0 \vee \neg A_0)$ (because $(A_0 \vee \neg A_0) \Rightarrow (A_0 \vee \neg A_0)$). So we have A_0 is in K, and $\neg A_0$ is in K, but $(A_0 \vee \neg A_0)$ is in L. Thus with this non-normal valuation we have two statements each assigned

"f", whose disjunction is assigned "t", and a statement (i.e. A_0) such that it and its negation are both assigned "f".

Thus in classical implication structures, the distributional patterns of non-normal valuations for some of the logical operators deviate from the usual (extensional) patterns. This we have just seen is true for negation, and disjunction.

Before we turn to a way of getting some perspective on these observations, and try to understand the significance of the difference that non-normal valuations make, it is important to note that in the classical case, if we consider only the behavior of the normal valuations, then there is no departure from the familiar patterns. For this we need a brief discussion of the extensionality of the logical operators.

6. Extensionality and the Logical Operators

For any implication structure $\mathfrak{I}$ = <S, $\Rightarrow$ > , we think of the logical operators as functions that map members of S, or pairs of members of S to S. The full story of how to define the logical operators using only the implication relation of the structure is a story told elsewhere in Koslow (1992). Suppose that one has an operator O(A) on the structure. Let T = <K, L> be any strong bisection on the structure. And let τ be the valuation based on that structure. We shall say that O(A) *is extensional with respect to the valuation* τ (for short, "**O[ext, τ]**"), if an only if ,

> For any A and A* in S, if τ(A) and τ(A*) are in the same set of the bisection (K, or L), then τ(O(A)) and τ(O(A*)) are in the same set of the bisection (K, or L).

That is, if A and A* have the same truth value, then O(A) and O(A*) also have the same truth-value. This definition covers the case of operators on single arguments. There is the obvious natural generalization for operators of two or more arguments.

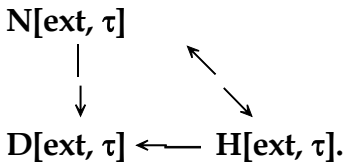
It can be shown that for any valuation τ, normal or not, and any A and B, (where "N", "D", "C", and "H", stand for the negation, disjunction, conjunction and conditional operators on a structure) that

(1) **N[ext, τ]** if and only if: ¬ A is in K if and only if A is in L.
(2) **D[ext, τ]** if and only if:(A ∨ B) is in L if and only if A is in L or B is in L.
(3) **C[ext, τ]** if and only if: (A ∧ B) is in L if and only if A is in L and B is in L.
(4) **H[ext, τ]** if and only if: A is in K or B is in L.

So (1) says that the negation operator ('¬") is extensional with respect to the valuation **τ,** if and only if [the negation of any A is assigned f if and only if A is assigned t]. Similar readings for (2) – (4).

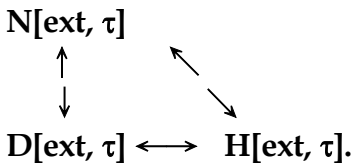
Therefore if a valuation departs from the customary distribution of truth-values for a logical operator, then that operator will fail to be extensional with respect to that valuation.

It is not difficult to show, for any implication structure, how the extensionality of the various logical operators with respect to any valuation (normal or not), are related. The result can be summed up this way:



So for example one can show that there is some valuation such that disjunction is extensional with respect to it, but negation is not.

The connection of the normality of a valuation and the extensionality of the various logical operators with respect to it is a matter of some delicacy. If we assume that the negation operator on an implication structure is classical, then it can be shown that all the logical operators on that structure are extensional with respect to any normal valuation—if any one of them is. That is:



Since it is straightforward to show that for any valuation τ, disjunction is extensional with respect to it, if and only if it is normal, it follows that in any classical implication structure, all the logical operators are extensional with respect to any normal valuation, since disjunction is extensional with respect to any normal valuation.

The matter is different if the implication structure is non-classical. Let $\mathfrak{I}_{ISC}$ be the structure that is associated with the Intuitionistic Sentential Calculus. It is easy to prove the following simple theorem:

> If $\mathfrak{I}$ is any implication structure such that (1) negation is non-classical, (2) it has the disjunctive property [$(A \vee B)$ is a thesis if and only if either A is a thesis or B is a thesis], and (3) $\mathfrak{I}$ is incomplete with respect to negation (some member of the structure is neither a thesis nor is its negation), then there exists a normal valuation on the structure such that negation is not extensional with respect to it.

This shows immediately that for the Intuitionistic implication structure, negation is not extensional with respect to some normal valuation, and by the first triangle diagram, the conditional is not extensional either. So the restriction to normal valuations, unlike the case of classical structures, doesn't help restore extensionality. The negation and conditional operators in the Intuitionistic structure are not -extensional even with respect to normal valuations.

> The proof is direct. Let T = <K, L>, where L is the set of Intuitionistic theses (all A such that it is provable in ISC that A). Clearly, L is closed under Intuitionistic implication $\Rightarrow^{ISC}$. To see that it is a normal bisection, consider A, B, C such that A, B $(\Rightarrow^{ISC})\wedge$ C , where A and B are in K. We want to show that C is in K. Suppose it is in L, then since C $\Rightarrow^{ISC}$ (A v B) it follows that (A v B) is in L. Therefore by the disjunctive property, either A is in L or B is in L. But by hypothesis, neither of them is in L. Therefore C has to be in K. So T is a normal bisection. However, since ISC is incomplete with respect to negation, there is some Z such that neither it nor it's negation are theses. So neither it nor its negation are in L – both are in K. Therefore the valuation based on T assigns "f" to both Z and to its negation.

This is a nice way of incorporating what some philosophers think should be the Intuitionistic version of "truth" simpliciter. Here we take the appeal to the theses of Intuitionism to

be one way of giving a truth-value interpretation—that is, a particular valuation. We have seen that taking the set of Intuitionistic theses as the L's of a strong bisection gives a nice example of a normal bisection. We do not confuse a particular truth-value assignment with "truth" (say some notion satisfying the Tarski T-Schema), no more than we would make the mistake of confusing a valuation in a two-valued sentential logic or a truth-value assignments for a possible world, with "truth".

So we see that a normal valuation on an Intuitionistic structure can give rise to deviations from the usual extensional distribution—in this case negation (and by the first triangle extensionality diagram the conditional will also fail to be extensional with respect to this normal valuation). The disjunction and conjunction operators, however, will be extensional with respect to this normal valuation.

It is also worthwhile mentioning the well known fact that just as there are non-normal valuations on classical structures, there are also non-normal valuations on non-classical structures. And just as in the classical case, they give rise to strange behavior for some logical operators. To see this, let $\Im_{ISC} = \langle S, \Rightarrow^{ISC} \rangle$ be an Intuitionistic implication structure. Let $T = \langle K, L \rangle$ be a bisection where L is the set of *classical* theses (*sic*), and let K be the rest (S – L). L is closed under $\Rightarrow^{ISC}$, because if A is a classical thesis and $A \Rightarrow^{ISC} B$, then B is also a classical thesis. So T is at least a strong bisection. The question is whether it is normal. Is K closed under the dual of the Intuitionistic implication relation $\Rightarrow^{ISC}$? The answer is negative.

> A proof: Since (CSC) is incomplete with respect to negation, there is some A_0 such that neither it nor its negation is a thesis of (CSC). Consequently neither of them is a thesis of ISC. Now since neither of them is a classical thesis, they are both in K. Now we have $A_0 , \neg A_0 \Rightarrow^{ISC} \wedge(A_0 \vee \neg A_0)$, because that is equivalent to the condition that $(A_0 \vee \neg A_0) \Rightarrow^{ISC} (A_0 \vee \neg A_0)$. Therefore, if K is closed under the dual of $\Rightarrow^{ISC}$, then $(A_0 \vee \neg A_0)$ is in K. That is impossible since K contains only sentences which are not classical theses. Consequently, K is not closed under the dual, and this bisection is non-normal.

Therefore the valuation that is based on this strong bisection is non-normal. Since $\tau(A)$ is "t" if A is in L and "f" otherwise, we have the result that $\tau(A_0)$ and $\tau(\neg A_0)$ are both 'f". It is

also clear from this proof, that disjunction and the conditional are also non-extensional with respect to this valuation. In any case, the restriction to only normal valuations may restore extensionality in the classical case, but it certainly doesn't achieve that in the Intuitionistic case, nor should it, since it is very clear that the negation and the conditional operators in the Intuitionistic case are easily seen to be non-extensional.

I agree with Church's reaction to the existence of non-normal valuations. He thought that Carnap had discovered that you could have valuations that assigned "t" to (say) the tautologies of the classical sentential calculus, but deviated elsewhere from the normal assignments on the logical operators. Van Fraassen's supervaluations are another very different way of showing that possibility.

In this note I have tried to indicate that the non-normal valuations have more interest than that. Their existence doesn't show that there is something wrong with the usual presentations of some simple sentential logical systems, nor does it show that the claim that implication is truth preserving is inadequate since it by itself doesn't guarantee the familiar extensional distribution of valuations on the logical operators. Nor does it guarantee that truth and falsity are duals. The motivation for removing non-normal valuations from the logical scene is cosmetic. There's no need to treat them like the lepers of logic.

Arnold Koslow

The Graduate Center, The City University of New York

akoslow@gc.cuny.edu

References

Bernstein, B.A. (1932) *Bulletin of the American Mathematical Society*, vol.38, pp.390, 592.

Carnap, R. (1943) *Formalization of Logic*, Cambridge, Mass: Harvard University Press.

Church, A. (1944) 'Review of Carnap 1943', *The Philosophical Review*, 53, pp.493-498.

———. (1956) *Introduction to Mathematical Logic I*, Princeton, N.J.: Princeton University Press.

Incurvati, L. and Smith, P. (2009) 'Rejections and valuations', *Analysis* 70 (1), pp.3-10.

Koslow, A. (1992) *A Structuralist Theory of Logic*, Cambridge: Cambridge University Press.

———. (1999) 'The Implicational Nature of Logic: A Structuralist Account', *European Review of Philosophy: The Nature of Logic*, vol.4 (ed. A.C. Varzi, CSLI Publications, Stanford California) pp.111 -155.

Murzi, J. and Hjortland, O.T. (2009), 'Inferentialism and the Categoricity Problem: Reply to Raatikainen', *Analysis* 69 (3), pp.480-488.

Raatikainen, P. (2008) 'On the Rules of Inference and the Meanings of Logical Constants', *Analysis* 68 (4), pp. 282-87.

Rumfitt, I. (1997) 'The Categoricity Problem and Truth-value Gaps', *Analysis* 57 (4), pp.223-36.

———. (2000) '"Yes" and "No"', *Mind* 109, pp.781-823.

Scott, D. (1974) "Completeness and Axiomatizabiity in Many-Valued Logic.", in L. Henkin et al (eds.), *Proceedings of the Tarski Symposium*, Providence: American Mathematical Society, pp. 411 – 35.

Smiley, T. (2005) 'Rejection', *Analysis* 56 (1), pp.1-9.

Shoesmith, D. J. and Smiley,T. (1978) *Multiple-conclusion Logic*, Cambridge: Cambridge University Press.

Wojcicki, R. (1973) 'Dual Counterparts of Consequence Relations', *Bulletin of the Section of Logic*, 2 (1) , pp. 54-57.

Abstracta 6 : 1, pp. 136 – 155, 2010

AN INDUCTIVE MODAL APPROACH FOR THE LOGIC OF EPISTEMIC INCONSISTENCY[1]

Ricardo Silvestre

Abstract
The purpose of this paper is twofold. First we want to extent a specific paranormal modal logic in such a way as obtain a paraconsistent and paracomplete multimodal logic able to formalize the notions of plausibility and certainty. With this logic at hand, and this is our second purpose, we shall use a modified version of Reiter's default logic to build a sort of inductive logic of plausibility and certainty able to represent some basic principles of epistemic inductive reasoning, such as a negative autoepistemic principle, an 'error-prone feature of induction' principle and a confirmation by enumeration principle.

Some things make the combination of modal logic and paraconsistent logic (da Costa 1974) a very interesting enterprise (Fuhrmann 1990) (da Costa and Carnielli 1986) (Goble 2006). First of all, many knowledge representation problems involving modalities seem to require a paraconsistent reasoning mechanism. An agent able to represent its beliefs and doxastic states, for example, may have evidences both to belief and not belief something; or its normative component might both require and prohibit something. Second, some have defended the idea that normal modal logic already embodies a kind of paraconsistency (Béziau 2002) (Marcos 2005) (Silvestre 2006). For instance, defining in S5 the derivated operator ~ as $\neg\Box$, we have a unary operator that does not satisfy the principle of explosion and has enough properties to be called a negation, entitling us then to classify S5 as a paraconsistent logic (Béziau 2002).

In (Silvestre 2011) and (Silvestre 2006) we presented a combination of modal and paraconsistent logic called paranormal modal logic. The motivation for this logic lies on the concept of inductive plausibility. By inductive plausibility we mean the same as Carnap's pragmatical probability (Carnap 1946), that is, a qualitative label we attach to the

[1] Work partially supported by CNPq (National Counsel of Technological and Scientific Development of Brazil), public notice MCT/CNPq N° 03/2009.

conclusion of inductive inferences. The novelty here is that when we seriously take into consideration the contradictions that are sure to arise from the use of such inferences (Perlis 1987) (Pequeno and Buchsbaum 1991), we see that there are not one, but two authentic approaches to deal with the problem (Silvestre 2007). These skeptical and credulous approaches to induction, as we have named them, give rise to two different plausibility notions which bear important relations to the field of paraconsistent and paracomplete logic (Loparíc and da Costa 1984): while the skeptical plausibility is a paracomplete notion, the credulous plausibility is a paraconsistent one. The idea of paranormal modal logic then is to analyze these two notions inside a modal framework.

First of all, we have a modal operator ? (used in a post-fixed notation) meant to represent the notion of credulous plausibility. Alike to $\Diamond$, $\alpha?$ is true iff α is true in at least one world (which we call plausible world). In addition to ?, there is the $\Box$-like operator ! meant to represent the notion of skeptical plausibility or acceptance: $\alpha!$ is true iff α is true in all plausible worlds. While the primitive negation $\neg$ is, in connection with ?, paraconsistent – we might have both $\alpha?$ and $\neg(\alpha?)$ –, in connection with ! it is paracomplete – it might be that neither $\alpha!$ nor $\neg(\alpha!)$ are true. Being its paraconsistency and paracompleteness dependent on the modality attached to the formula, we call $\neg$ a modality-dependent paranormal negation. Alike to normal modal logic, there is a family of paranormal modal logics related both axiomatically and semantically to each other. For instance, add the axiom $\alpha!\rightarrow\alpha$ ($T_?$) to $K_?$, which is the most basic paranormal modal logic, and you have the system $T_?$; add $\alpha!\rightarrow\alpha!!$ ($4_?$) to $T_?$ and you have $S4_?$; add $\alpha\rightarrow\alpha?!$ ($B_?$) to $S4_?$ and you have $S5_?$, and so on and so forth.

Considering some key aspects of the philosophical framework behind paranormal modal logic, two related combination developments can be thought. First, following the original motivation of the very first versions of paranormal modal logic (Pequeno and Buchsbaum 1991), we might think of using the notions of plausibility along with an inductive reasoning mechanism, therefore giving rise to an inductive and consequently nonmonotonic paraconsistent logic. Second, since ? and ! represent epistemic notions, it might be useful to investigate the relation between these plausibility notions and other

epistemic notions. This is significant, for when we look deep at the epistemic nature of inductive inferences we see that in the same way that the conclusions of such inferences must be marked with a plausibility operator, their premises should also be referred to with the help of some epistemic notion (Silvestre 2010).

Our purpose in this paper is to advance these two combination developments. For the sake of simplicity, we shall consider only the propositional case[2]. In the next section we briefly present paranormal modal logic $K_?$. In Section 2 we introduce a multimodal logic meant to function as a logic of plausibility and certainty. In Section 3 we use this multimodal logic along with a nonmonotonic inferential mechanism to obtain a sort of logic of inductive. Finally, in the last section, we lay down some conclusive remarks.

1. Paranormal Modal Logic

As we have said, the intended meaning for the modal operators ? and ! of paranormal modal logic are the notions of credulous plausibility and skeptical plausibility or acceptability. If α is a formula, then the $\alpha?$ and $\alpha!$ mean, respectively, "α is credulously plausible" and "α is skeptically plausible or accepted". While ? is, we might say, a paraconsistent modal operator, ! is a paracomplete one: there is a model M such that both $\alpha?$ and $\neg(\alpha?)$ are satisfied in M and there is a model M such that neither $\alpha!$ nor $\neg(\alpha!)$ are satisfied in M. Both ? and ! are introduced as primitive symbols of the language. We have below the axiomatics of $K_?$, which is the most basic paranormal modal logic:

Positive Classical Axioms

P1: $\alpha\rightarrow(\beta\rightarrow\alpha)$
P2: $(\alpha\rightarrow(\beta\rightarrow\varphi))\rightarrow((\alpha\rightarrow\beta)\rightarrow(\alpha\rightarrow\varphi))$
P3: $\alpha\wedge\beta\rightarrow\alpha$
P4: $\alpha\wedge\beta\rightarrow\beta$
P5: $\alpha\rightarrow(\beta\rightarrow\alpha\wedge\beta)$
P6: $\alpha\rightarrow\alpha\vee\beta$

[2] For an account of the first order case of the class of logics to be introduced here see Silvestre (2010) and (2011).

P7: $\beta\rightarrow\alpha\vee\beta$
P8: $(\alpha\rightarrow\beta)\rightarrow((\varphi\rightarrow\beta)\rightarrow(\alpha\vee\varphi\rightarrow\beta))$

Paranormal Classical Axioms

A1: $(\alpha\rightarrow\beta)\rightarrow((\alpha\rightarrow\neg\beta)\rightarrow\neg\alpha)$, wherein β is ?-free and α is !-free
A2: $\neg\alpha\rightarrow(\alpha\rightarrow\beta)$, wherein α is ?-free
A3: $\alpha\vee\neg\alpha$, wherein α is !-free

Non-Positive Additional Classical Axioms

N1: $\neg(\alpha\rightarrow\beta)\leftrightarrow(\alpha\wedge\neg\beta)$
N2:$\neg(\alpha\wedge\beta)\leftrightarrow(\neg\alpha\vee\neg\beta)$
N3:$\neg(\alpha\vee\beta)\leftrightarrow(\neg\alpha\wedge\neg\beta)$
N4:$\neg\neg\alpha\leftrightarrow\alpha$
N5:$((\alpha\rightarrow\beta)\rightarrow\alpha)\rightarrow\alpha$

Paranormal Modal Axioms

K1: $\alpha?\leftrightarrow\sim((\sim\alpha)!)$
K2: $(\neg\alpha)!\leftrightarrow\neg(\alpha!)$
K3: $(\neg\alpha)?\leftrightarrow\neg(\alpha?)$

Modal Axioms

$K_?$: $(\alpha\rightarrow\beta)!\rightarrow(\alpha!\rightarrow\beta!)$

Rules of Inference

MP: α, $\alpha\rightarrow\beta$ / β
$N_?$: α / $\alpha!$

Axioms A1-A3 are restricted in such a way as to guarantee the paraconsistent and paracomplete behavior of ? and !, respectively. Axioms N1-N5 are there to restore the deductive power awakened by the restrictions of A1-A3. K1 sets ? and ! as the dual of each other. ~ is a derived operator meant to play the role of classical negation: $\sim\alpha =_{def} \alpha\rightarrow p\wedge\neg p$, there *p* is an arbitrary propositional symbol. Along with A1-A3, axioms K2 and K3 are the key of $K_?$'s non-classical behavior. While K2 allows us to go from the skeptical implausibility of α ($\neg(\alpha!)$) to the skeptical plausibility of $\neg\alpha$ ($(\neg\alpha)!$), K3 allows us to go from the credulous plausibility of $\neg\alpha$ ($(\neg\alpha)?$) to the credulous implausibility of α ($\neg(\alpha?)$)[3].

[3] For an explanation of the philosophical reasons behind these axioms see Silvestre (2011).

Finally, $K_?$ and $N_?$ are paranormal modal logic equivalents to axiom K and rule N, respectively.

Regarding the notion of deduction, following Fitting (Fitting 1993) we make use of the distinction between global and local premises. From a proof-theoretical point of view, the difference is that only those formulas obtained exclusively with the help of the global premises are able use the necessitation rule. In symbols we have $A \cdot\!\!\cdot B \vdash \varphi$ as meaning that φ is deducted from A and B, A being the set of global premises and B the set of local premises. The same distinction shall be used in our definition of the notion of logical consequence.

A frame in paranormal modal logic is a pair <W,R> where W is a non-empty set of entities called worlds (or plausible worlds) and R is a binary relation on W called accessibility relation. A model then is a triple <W,R,ν> where F = <W,R> is a frame and ν is a function mapping elements of P and W to truth-values 0 and 1. We say that the model M is based on F and that $w \in W$ is a world of M. Bellow you have the semantics of $K_?$:

$\Omega_{M,w}(p) = \mho_{M,w}(p) = 1$	iff	$\nu_w(p) = 1$;
$\Omega_{M,w}(\neg\alpha) = 1$	iff	$\mho_{M,w}(\alpha) = 0$;
$\mho_{M,w}(\neg\alpha) = 1$	iff	$\Omega_{M,w}(\alpha) = 0$;
$\Omega_{M,w}(\alpha\rightarrow\beta) = 1$	iff	$\Omega_{M,w}(\alpha) = 0$ or $\Omega_{M,w}(\beta) = 1$;
$\mho_{M,w}(\alpha\rightarrow\beta) = 1$	iff	$\Omega_{M,w}(\alpha) = 0$ or $\mho_{M,w}(\beta) = 1$;
$\Omega_{M,w}(\alpha\wedge\beta) = 1$	iff	$\Omega_{M,w}(\alpha) = 1$ and $\Omega_{M,w}(\beta) = 1$;
$\mho_{M,w}(\alpha\wedge\beta) = 1$	iff	$\mho_{M,w}(\alpha) = 1$ and $\mho_{M,w}(\beta) = 1$;
$\Omega_{M,w}(\alpha\vee\beta) = 1$	iff	$\Omega_{M,w}(\alpha) = 1$ or $\Omega_{M,w}(\beta) = 1$;
$\mho_{M,w}(\alpha\vee\beta) = 1$	iff	$\mho_{M,w}(\alpha) = 1$ or $\mho_{M,w}(\beta) = 1$;
$\Omega_{M,w}(\alpha?) = 1$	iff	for some $w' \in W$ such that wRw', $\Omega_{M,w'}(\alpha) = 1$;
$\mho_{M,w}(\alpha?) = 1$	iff	for all $w' \in W$ such that wRw', $\mho_{M,w'}(\alpha) = 1$;
$\Omega_{M,w}(\alpha!) = 1$	iff	for all $w' \in W$ such that wRw', $\Omega_{M,w'}(\alpha) = 1$;
$\mho_{M,w}(\alpha!) = 1$	iff	for some $w' \in W$ such that wRw', $\mho_{M,w'}(\alpha) = 1$.

Formula α is satisfied in model M and world *w* (in symbols: $M,w\Vdash\alpha$) iff $\Omega_{M,w}(\alpha)=1$; if α is satisfied in all worlds *w* of M we say that M satisfies α (in symbols: $M\Vdash\alpha$). We then say that α is a logical consequence of A and B, A being the global premises and B the local ones (in symbols: $A\therefore B\vDash\varphi$) iff, given a specific set of frames F (which in $K_?$ is the set of all frames), for every model M based on F, if M satisfies all members of A, then for every world *w* of M such that $M,w\Vdash\beta$, for every $\beta\in B$, $M,w\Vdash\varphi$[4].

Ω and $\mho$ are evaluation functions which, depending on the modal operator at hand, maximize or minimize the truth-value of formulas: while Ω minimizes and $\mho$ maximizes !-marked formulas, Ω maximizes and $\mho$ minimizes ?-marked ones. As we have shown above, it is Ω which is used in the definition of the notion of satisfaction. The need of these two functions lies on the interpretation of the negation symbol $\neg$: the result of Ω applied to $\neg\alpha$ is defined in function of $\mho$, and vice-versa. This in fact is the semantic key of paranormal modal logic's non-classical behavior. $K_?$ is sound and complete (Silvestre 2011).

As we have said, exactly in the same way as it happens with normal modal logic, there is a semantic and axiomatic relation between the several paranormal modal systems. If, for instance, we restrict ourselves to the class of serial frames we obtain system $D_?$, which is syntactically obtained by adding $\alpha!\rightarrow\alpha?$ to the axiomatics of $K_?$; considering the class of all reflexive frames we have the logic $T_?$, which is syntactically obtained by adding $\alpha!\rightarrow\alpha$ to the axioms of $K_?$; taking into account the class of all reflexive and symmetric frames we obtain the system $B_?$, which is the same as $T_?$ plus axiom $\alpha\rightarrow\alpha?!$; and so on and so forth.

[4] For more on the formalization of the notions of deduction and logical consequence inside a global-local premises framework see Fitting (1993) and Silvestre (2011).

2. A Logic of Plausibility and Certainty

What we call the logic of plausibility and certainty is a multimodal logic with two sets of operators. On the one hand we have the operators ? and ! (which as we have seen behave paraconsistently and paracompletly, respectively); on the other we have the classically behaved operators $\Box$ and $\Diamond$ meant to represent the notions of certainty and epistemological possibility: while $\Box\alpha$ means "α is certain", $\Diamond\alpha$ means "α is epistemologically possible". Alike to ! and ?, $\Box$ and $\Diamond$ are primitive symbols of the language.

An important point related to the meaning of formulas in general and non-modal formulas in particular concerns the place they appear in the relation of deductibility or logical consequence. Suppose that $A \therefore B \models \varphi$ (or $A \therefore B \vdash \varphi$). While an arbitrary formula α belonging to the set of global premises A can be said to mean "α is true" or "α is a true hypothesis", a formulae β belonging to the set B of local premises means simply "β is a hypothesis." This is why we can apply the N rules ($\alpha/\alpha!$ and $\alpha/\Box\alpha$) only to the global premises: since α is a true hypothesis, we sure can claim it to be skeptically plausible ($\alpha!$) as well as to be certain about its truth ($\Box\alpha$). Looking at the other way round, the fact that we can semantically conclude $\Box\alpha$ and $\alpha!$ from α (which is due to all models taken into account being exactly those in which α is true in all of its worlds) reflects the idea that α is being taken as a true hypothesis and not just as a certain or accepted one. In its turn, β helps to select, out of the multitude of worlds belonging to some of these models, the individual worlds that will be used to evaluate the conclusion φ. It therefore functions like a local, hypothetical premise whose truth is guaranteed not in all, but only in a few possible worlds of the models in question.

In addition to the axioms and inference rules of $K_?$, the logic of plausibility and certainty has the following axioms and inference rules:

Paranormal Modal Axioms

$D_?$: $\alpha! \rightarrow \alpha?$

$B_?$: $\alpha \rightarrow \alpha?!$

Normal Modal Axioms

NP: $\Diamond\alpha \leftrightarrow \neg\Box\neg\alpha$
K: $\Box(\alpha \rightarrow \beta) \rightarrow (\Box\alpha \rightarrow \Box\beta)$
NN: $\sim\Box\sim\alpha \leftrightarrow \neg\Box\neg\alpha$
D: $\Box\alpha \rightarrow \Diamond\alpha$
B: $\alpha \rightarrow \Box\Diamond\alpha$
4: $\Box\alpha \rightarrow \Box\Box\alpha$

Multimodal Axioms

PC: $\Box\alpha \rightarrow \alpha!$

Rules of Inference

N: α / $\Box\alpha$

K is system K's axiom. While NP is there to guarantee $\Box$ and $\Diamond$ as the dual of each other (recall that both are primitive symbols), NN is needed in order to set the normal and classical behavior of $\Box$ (and, consequently, of $\Diamond$.) D and $D_?$ guarantee, respectively, that what is certain is also epistemically possible and what is skeptically plausible is also credulous plausible. B and $B_?$ say, respectively, that if α is a true then it is certain that α is epistemologically possible and it is skeptically plausible that α is credulously plausible. The reasonableness of these principles is self-evident in the case where α is a true hypothesis. Concerning the local, unqualified hypothesis case, B and $B_?$ state a sort of minimal rationality principle about the hypotheses we are allowed to consider: even though they may be neither plausible nor epistemologically possible, they must be so from a second-order point of view. 4 is a sort of principle of positive introspection: if we know that α, then we know that we know that α. From B and 4 we deduce 5, $\neg\Box\alpha \rightarrow \Box\neg\Box\alpha$, which is a principle of negative introspection: if we are not certain about α, then we are certain that we are not certain about α. PC or the plausibility-certainty axiom states that if α is certain then it is also an accepted hypothesis. From it, along with MP and K1, we obtain $\alpha? \rightarrow \Diamond\alpha$, that is to say, that if α is (credulously) plausible then it is epistemically possible.

The reason why we have excluded axioms T ($\Box\alpha\rightarrow\alpha$) and $T_?$ ($\alpha!\rightarrow\alpha$) is that they represent a kind of principle of epistemic arrogance undesirable in the case of both certainty and skeptical plausibility. Taking α as meaning "α is true," while T means that if we are certain that α is true then it is true, $T_?$ means that accepting α as true entails that it is true. On similar grounds, $T_?$ and T cannot be accepted if we take α as representing an unqualified hypothesis. While from $T_?$ along with K1 we conclude $\alpha\rightarrow\alpha?$, which means that every conceivable hypothesis is automatically a plausible one, from T we derive $\alpha\rightarrow\neg\Box\neg\alpha$, which means that every conceivable hypothesis is an irrevocable one. $4_?$ ($\alpha! \rightarrow \alpha!!$) was not included on account of the desirableness of allowing gradations of credulous plausibility ($T_?$ along with K1 entails $\alpha??\rightarrow\alpha?$), from which it is possible to develop, as we shall see below, a quantitative theory of plausibility.

About the relation between our modal operators, we have that the following axioms are valid in the logic of plausibility and certainty: $\Box\alpha\rightarrow\alpha!$, that is, from certainty we obtain acceptance, $\alpha!\rightarrow\alpha?$, that is, from acceptance we obtain (credulous) plausibility, and $\alpha?\rightarrow\Diamond\alpha$, that is, from plausibility we get epistemic possibility.

A frame in the logic of plausibility and certainty is a triple $<W,R_?,R_\Diamond>$ where W is a non-empty set of worlds, $R_?$ is a binary relation on W called plausibility accessibility relation and $R_\Diamond$ is a binary relation on W called certainty accessibility relation. $R_?$ and $R_\Diamond$ satisfy the following conditions: (i) for every $w,w'\in W$ if $wR_\Diamond w'$ then $wR_?w'$, (ii) for every $w\in W$ there is at least one $w'\in W$ and at least one $w''\in W$ such that $wR_\Diamond w'$ and $wR_?w''$, (iii) for every $w,w'\in W$ if $wR_\Diamond w'$ then $w'R_\Diamond w$ and if $wR_?w'$ then $w'R_?w$, and (iv) for every $w,w',w''\in W$, if $wR_\Diamond w'$ and $w'R_\Diamond w''$ then $wR_\Diamond w''$. A *model* then is a quadruple $<W,R_?,R_\Diamond,\nu>$ where $F = <W,R_?,R_\Diamond>$ is a frame and ν is function mapping elements of P and W to truth-values 0 and 1. For the evaluation functions Ω and $\mho$ we have the following modification on what has been shown above:

$\Omega_{M,w}(\alpha?) = 1$ iff for some $w' \in W$ such that $wR_?w'$, $\Omega_{M,w'}(\alpha) = 1$;
$\mho_{M,w}(\alpha?) = 1$ iff for all $w' \in W$ such that $wR_?w'$, $\mho_{M,w'}(\alpha) = 1$;
$\Omega_{M,w}(\alpha!) = 1$ iff for all $w' \in W$ such that $wR_?w'$, $\Omega_{M,w'}(\alpha) = 1$;
$\mho_{M,w}(\alpha!) = 1$ iff for some $w' \in W$ such that $wR_?w'$, $\mho_{M,w'}(\alpha) = 1$;
$\Omega_{M,w}(\Diamond\alpha) = 1$ iff for some $w' \in W$ such that $wR_\Diamond w'$, $\Omega_{M,w'}(\alpha) = 1$;
$\mho_{M,w}(\Diamond\alpha) = 1$ iff for some $w' \in W$ such that $wR_\Diamond w'$, $\mho_{M,w'}(\alpha) = 1$;
$\Omega_{M,w}(\Box\alpha) = 1$ iff for all $w' \in W$ such that $wR_\Diamond w'$, $\Omega_{M,w'}(\alpha) = 1$;
$\mho_{M,w}(\Box\alpha) = 1$ iff for all $w' \in W$ such that $wR_\Diamond w'$, $\mho_{M,w'}(\alpha) = 1$.

The definitions of satisfatibility and logical consequence are the same as $K_?$'s. About the peculiarities of the semantics of the logic of plausibility and certainty we first note that given a frame $<W,R_\Diamond,R_?>$ and a world $w \in W$, the sets $R_\Diamond(w) = \{w'|wR_\Diamond w'\}$ and $R_?(w) = \{w'|wR_?w'\}$ represent, respectively, what we may call the epistemically possible worlds of w and the plausible worlds of w. Second, every plausible world is also an epistemic possible world (in symbols: $R_?(w) \subseteq R_\Diamond(w)$); this is restriction (i) of the frame structure, which from a proof-theoretical point of view corresponds to axiom PC. Third, all frames considered are serial frames; this is restriction (ii), which in the axiomatics corresponds to axioms D and $D_?$. Fourth, while $R_\Diamond$ is a symmetric and transitive relation, $R_?$ is only a symmetric one; this, which is stated in restrictions (iii) and (iv), corresponds, respectively, to axioms B and 4 and axiom $B_?$.

3. A Logic of Inductive Implication

Traditionally the purpose of a logic of induction is one of confirmation: given a piece of evidence *e* and a hypothesis *h*, it should say whether (and possibility to what extent) *e* confirms or gives evidential support to *h* (Carnap 1950) (Hempel 1945). About the status of hypothesis *h* when *e* confirms *h* and *e* is true, despite the diversity of approaches, all theorists agree on one basic point: given that *e* confirms *h* and that *e* is true, whatever we conclude about *h* it should reflect the uncertainty inherent to inductive inferences. Almost invariably some probability notion has been chosen to do the job: even though from "*e*

confirms *h*" and "*e* is true" we cannot conclude that *h* is true, we can conclude that it is probable.

This notion of probability should not be confounded with Carnap's logical probability (Carnap 1950): while the later is supposed to be a purely logical notion connecting two sentences, the former must be seen as an epistemic label we attach to inductive conclusions in order to make explicit their defeasible character. Carnap calls this non-logical notion of probability pragmatical probability (Carnap 1946); we shall prefer the qualitative and hopefully less problematic term "inductive plausibility" or simply "plausibility".

This characterization of induction in terms of pragmatical probability or plausibility is significant, first because considering that the truth of *e* warrants us to inductively conclude not the truth but the plausibility of *h*, we can trivially say that what *e* confirms or evidentially supports is not the truth of *h*, but its plausibility. Therefore, rather than saying that *e* confirms or inductively supports *h*, we should say that *e* confirms or inductively supports the plausibility of *h*. And given that "*h* is plausible" will possibly be inferred, the whole thing might be read as "*e* inductively implies the plausibility of *h*." We shall call such sort of statements inductive implications.

Second, as we have mentioned, the contradictions that are sure to arise from the use inductive inferences force us to consider two different but complementary approaches to induction. A consequence of that is that sentences like "*e* confirms or evidentially supports *h*" shall necessarily mention the approach according to which the confirmation is being made. This is easily done by qualifying the plausibility notion appearing in the consequent of inductive implications: while "*e* inductively implies the credulous plausibility of *h*" characterizes a credulous approach, "*e* inductively implies the skeptical plausibility of *h*" characterizes a skeptical approach.

Third, attaching an epistemic label to the conclusions of inductive inferences leaves the door open to taking the whole notion of induction as an epistemic one. In the same way that what is confirmed or evidentially supported is not the truth of *h* but its plausibility, we may

say that what confirms the plausibility of *h* is not the truth of *e*, but the certainty or plausibility of *h*[5].

As far as our formalization of these points is concerned, we shall use a version of Reiter's default logic (Reiter 1980) to represent the notion of inductive implication. The rationale behind this choice is that default logic incorporates the inferential and non-truth preserving aspects of inductive logic (Silvestre and Pequeno 2005). For example, we may quite naturally read default $\alpha:\varphi/\beta$ as "α inductively implies β unless $\neg\varphi$". We shall represent this by $\alpha\succ\beta\precsim\neg\varphi$. Second, the monotonic basis of this default logic shall be exactly the logic of certainty and plausibility just introduced in the previous section. Third, in order to capture the epistemological nature of inductive implications just mentioned, we shall force the components of our defaults to be marked with the correspondent modal operators. For instance, an inductive inference made according to a credulous approach might be represented as $\Box\alpha\succ\beta?\precsim\varphi$, which shall be read as "the certainty of α inductively implies the plausibility of β, unless φ".

Let $\Im$ be the language of the logic of certainty and plausibility. The inductive language $\Im_{\succ}$ built over $\Im$ is defined as follows: (i) If $\alpha\in\Im$ then $\alpha\in\Im_{\succ}$; (ii) If $\alpha,\beta,\varphi\in\Im$ then $\alpha\succ\beta\precsim\varphi\in\Im_{\succ}$; (iii) Nothing else belongs to $\Im_{\succ}$. We call $\alpha\succ\beta\precsim\varphi$ and inductive implication, being α its antecedent, β its consequent and φ its exception. $\alpha\succ\beta$ is an abbreviation of $\alpha\succ\beta\precsim\bot$, $\beta\precsim\varphi$ an abbreviation for $\top\succ\beta\precsim\varphi$ and β° an abbreviation for $\top\succ\beta\precsim\bot$, where $\bot$ is an abbreviation for $p\wedge\neg p$ and $\top$ is an abbreviation for $p\vee\neg p$, where *p* is an arbitrary propositional symbol. Any formula that is not an inductive implication is called an ordinary formula. With the help of $\Im_{\succ}$ we can define the notion of extension:

Let $A\subseteq\Im_{\succ}$ be a set of our inductive language and $S\in\Im$ a set of formulas of our multimodal language. $\Gamma(S)\subseteq\Im$ is the smallest set satisfying the following conditions: (i) $A\subseteq\Gamma(S)$; (ii) If $\Gamma(S)\dot{\cup}\varnothing\vdash\alpha$ then $\alpha\in\Gamma(S)$; (iii) If $\alpha\succ\beta\precsim\varphi\in A$, $\alpha\in\Gamma(S)$, $\varphi\notin S$ and $\sim\beta\notin S$ then

[5] For a full description of the theory of induction sketched here see (Silvestre 2007) and (Silvestre 2010).

$\beta \in \Gamma(S)$. A set of formulas E is an *extension* of A iff $\Gamma(E) = E$, that is, iff E is a fixed point of the operator Γ.

We first note that this language $\mathfrak{I}_{\succ}$ is a mixed language containing ordinary formulas as well as inductive implications. Therefore the set used as parameter in the definition of P-extension plays the role of both components of a default theory: it contains both a set of ordinary formulas as well as a set of inductive implications. Second, in mentioning the deduction relation of the logic of certainty and plausibility $\vdash$ we make use exclusively of global premises, the reason for that being that we want our notion of extension to incorporate the autoepistemic principle according to which we are aware of whatever our inductive mechanism infers (see below)[6]. Finally, we make the test of consistency of the consequent (in terms of ~) inside the very definition of extension, turning then $\alpha \succ \beta \not\sim \varphi$ into an equivalent of default $\alpha:\beta \wedge \neg\varphi/\beta$. This has the advantage of preventing so-called abnormal defaults (Morris 1988).

As one might have concluded, this inductive language does not incorporate yet the epistemological considerations we have made above about inductive inferences. As we have advanced, one way to incorporate the theory of induction we are sketching here is to require the antecedent of inductive implications to be marked with the $\Box$ symbol and the consequent with the ? symbol. We thus have what we call the epistemic inductive language $\mathfrak{I}_{E\succ}$: (i) If $\alpha \in \mathfrak{I}$ then $\alpha \in \mathfrak{I}_{E\succ}$; (ii) If $\alpha,\beta,\varphi \in \mathfrak{I}$ then $\Box\alpha \succ \beta ? \not\sim \varphi \in \mathfrak{I}_{E\succ}$; (iii) Nothing else belongs to $\mathfrak{I}_{E\succ}$. Trivially $\mathfrak{I}_{E\succ} \subset \mathfrak{I}_{\succ}$.

In order to use this $\mathfrak{I}_{E\succ}$ language, we have to slightly change our definition of extension and introduce what we shall call a Δ-extension: Let $\Delta \in \mathfrak{I}_{\succ} - \mathfrak{I}$ be a set of inductive implications, $A \subseteq \mathfrak{I}_{E\succ}$ a set of formulas of the epistemic inductive language and $S \in \mathfrak{I}$ a set of formulas of our multimodal language. $\Gamma(S) \subseteq \mathfrak{I}$ is the smallest set satisfying the following conditions: (i) $A \subseteq \Gamma(S)$; (ii) If $\Gamma(S) \dot{-} \varnothing \vdash \alpha$ then $\alpha \in \Gamma(S)$; (iii) If

[6] To see a formulation in terms of both global and local premises see Silvestre (2010).

$\alpha \succ \beta \not\precsim \varphi \in A \cup \Delta$, $\alpha \in \Gamma(S)$, $\varphi \notin S$ and $\sim\beta \notin S$ then $\beta \in \Gamma(S)$. A set of formulas E is a Δ-*extension* of A iff $\Gamma(E) = E$, that is, iff E is a fixed point of the operator Γ.

The idea here is that while A behaves like a default theory where its defaults satisfy the above motioned epistemic restrictions, Δ is a set of inductive implications meant to function like axioms able to nonmonotonically extend the inferential power of our logic of plausibility and certainty. About which inductive inferences compose Δ we have as follows.

First of all, there is the serious limitation of the logic of plausibility and certainty that we cannot conjunct plausible formulas: from $\alpha?$ and $\beta?$ we cannot conclude $(\alpha \wedge \beta)?$. The reason for that is obvious: it might be that α and β contradict each other in a strong sense ($\sim(\alpha \wedge \beta)$ or $\alpha \wedge \beta \rightarrow \bot$), in which case $(\alpha \wedge \beta)?$ also trivializes the theory ($\sim((\alpha \wedge \beta)?)$ or $(\alpha \wedge \beta)? \rightarrow \bot$). However, for cases where there is no contradiction between α and β it is desirable to be able to conclude $(\alpha \wedge \beta)?$ from $\alpha?$ and $\beta?$. In order to deal with that we introduce the schema of inductive implications below

$C_{?\wedge}$: $\alpha? \wedge \beta? \succ (\alpha \wedge \beta)?$

and set all instances of $C_{?\wedge}$ as belonging to Δ. See that if we have $\alpha? \wedge \beta?$ as belonging to $\Gamma(S)$ and α and β contradict each other (that is to say, $\sim((\alpha \wedge \beta)?) \in S$) then $(\alpha \wedge \beta)?$ shall not be included in $\Gamma(S)$.

Second, axiom 4, theorem 5 and rule N embody a sort of autoepistemic principle: while 4 and 5 says that we are aware of the facts we known as well as of the facts we do not know , respectively, N says that we are aware of all those propositions we take as true. But how about those statements whose truth we have no hint about? Suppose that Th(A) is all we can conclude from knowledge situation A. By N, for each $\alpha \in Th(A)$ we will have that we know that α ($\Box\alpha$.) But how about those statements which do not belong to Th(A)? It seems reasonable that for all β such that $\beta \notin Th(A)$ we conclude $\neg\Box\beta$. This is what we could call a *negative autoepistemic principle*. It is trivially a nonmonotonic rule: if from A we infer $\neg\Box\beta$, from $A \cup \{\beta\}$ the same inference cannot be done. It therefore might formalized only with the help of an inductive implication:

NA: $((\neg\Box\alpha)?)^{\circ}$

NA, all instances of which belong to Δ, is the axiom which transform our system into a truly autoepistemic logic. Note that $((\neg\Box\alpha)?)^{\circ}$ is an abbreviation for $\top \mathrel{\sim\!\!\succ} (\neg\Box\alpha)? \not\sim \bot$. Therefore, independently of the knowledge situation at hand, if it does not contain $\sim((\neg\Box\alpha)?)$ we will be able to infer nonmonotonically that $\neg\Box\alpha$ is plausible. The purpose of this is of course to make explicit that our agent does not know about the truthfulness of those formulas whose certainty cannot be inferred from his knowledge base: in the cases where $\Box\alpha$ does not belong to the logical theory, that is to say, α is not known, $(\neg\Box\alpha)?$ will be the case. One may think that because what we conclude through NA is $(\neg\Box\alpha)?$ and not $\neg\Box\alpha$, NA does not in fact perform the task we are claiming it performs. Not quite so. Since $\alpha?\rightarrow\Diamond\alpha$ (which is obtained from PP, K1 and $\Diamond\alpha\leftrightarrow\sim\Box\sim\alpha$), from $(\neg\Box\alpha)?$ we get $\Diamond\neg\Box\alpha$. From that, along with NP, we get $\neg\Box\neg\neg\Box\alpha$, which is equivalent to $\neg\Box\Box\alpha$. Since $\neg\Box\Box\alpha\rightarrow\neg\Box\alpha$, we have then that $\neg\Box\alpha$.

Third, some have defended what might be called the error-prone feature of inductive reasoning (Perlis 1987): since inductive conclusions may be mistaken even when its premises are true (something the very past use of such sort of inference has shown), any fair account of inductive reasoning should have an axiom saying that, independently of the circumstances we are working on, it is plausible that one of the beliefs we now take as rational is false. This can be formalized by the following axiom:

$I_?$: $\alpha_1?\wedge\ldots\wedge\alpha_n? \mathrel{\sim\!\!\succ} (\neg(\alpha_1\wedge\ldots\wedge\alpha_n))? \not\sim (\neg(\alpha_1\wedge\ldots\wedge\alpha_n\wedge\beta))?$,

wherein $\alpha_1,\ldots,\alpha_n$ and β are *different* basic formulas

A basic formula is an atomic formula (a propositional formula) or the negation of an atomic formula. All instances of $I_?$ belong to Δ. $I_?$ says that if *n* basic formulae are plausible, then it is also plausible that some of them is false (or, as we wrote, that the negation of their conjunction is plausible.) The exception part of $I_?$ is meant to guarantee that no plausible

atomic formula will be out of the conjunction $\alpha_1?\wedge\ldots\wedge\alpha_n?$: if this is the case, then the induction implication at hand cannot be used.

Finally, we have not spoken about skeptically plausible formulas. First, if we are allowed to use inductive implications only in connection to credulously plausible formulas (that is to say, inductive implications belonging to the epistemic inductive language $\mathfrak{I}_{E\succ}$), how are we to nonmonotonically introduce skeptically plausible formulas? Second, how are we to deal, in terms of inductive implications, with the relation we know there is between ?-marked formulas and !-marked ones?

One way to answer these questions is to use a very simple sort of confirmation by enumeration philosophy according to which α will be taken as accepted ($\alpha!$) only after it has got enough credulous confirmation. It is as if, by observing one black raven we turn the hypothesis "all ravens are black" into a very weakly plausible one; by observing another one we increase a little bit its degree of plausibility; and so and so forth, until that, after we have observed a certain number of black ravens, say *n*, we raise the hypothesis in question to the status of an accepted or skeptically plausible statement. In order to formalize that, we need of course to quantify how much a hypothesis was weakly confirmed or, in the context of taking weak confirmation and credulous plausibility as the same, how weakly plausible a hypothesis is.

The most straightforward way to do that is to count in how many plausible worlds a hypothesis is true. If α is true in at least one plausible world we write $\alpha?_1$; if it is true in at least two plausible worlds we write $\alpha?_2$ … until it is true in at least *n* plausible worlds, in the case we write $\alpha?_n$ or $\alpha!$. This can be done by defining the following abbreviations:

(i) $\alpha?_1 =_{def} \alpha?$;
(ii) $\alpha?_2 =_{def} (\alpha\wedge q)?\wedge(\alpha\wedge\neg q)?$, where *q* is an arbitrary atomic formula of $\mathfrak{I}$;
(i) $\alpha?_n =_{def} (\alpha\wedge p_1\wedge q)?\wedge\ldots\wedge(\alpha\wedge p_m\wedge q)?\wedge(\alpha\wedge p_1\wedge\neg q)?\wedge\ldots\wedge(\alpha\wedge p_m\wedge\neg q)?$, where $n = 2^{k+1}$, $m = 2^k$, $k>0$, $\alpha?_m \equiv (\alpha\wedge p_1)?\wedge\ldots\wedge(\alpha\wedge p_m)?$ and *q* is an arbitrary atomic formula of $\mathfrak{I}$ which do not occur in p_1;
(ii) $\alpha?_n =_{def} (\alpha\wedge p_1)?\wedge\ldots\wedge(\alpha\wedge p_n)?$, where $2^{k+1}>n>2^k$ and $\alpha?_{n+1} \equiv (\alpha\wedge p_1)?\wedge\ldots\wedge(\alpha\wedge p_n)?\wedge(\alpha\wedge p_{n+1})?$.

$\alpha?_n$ may be understood as meaning "the degree of plausibility of α is *n*." As we have mentioned above, such meaning is achieved by counting in how many plausible worlds α is true, which is performed with the help of the classical feature of worlds. Given an atomic formula *q*, we know that q and $\neg$q cannot be true at the same time in world *w*. Therefore, if $(\alpha\wedge q)?$ and $(\alpha\wedge\neg q)?$ are true, then the plausible worlds which make these two formulae true cannot be the same. Consequently, α is true in at least two worlds. Similarly, given an atomic formula *p* distinct from *q*, $(\alpha\wedge q\wedge p)?\wedge(\alpha\wedge\neg q\wedge p)?\wedge(\alpha\wedge q\wedge\neg p)?$ means that α is true in at least three worlds, $(\alpha\wedge q\wedge p)?\wedge(\alpha\wedge\neg q\wedge p)?\wedge(\alpha\wedge q\wedge\neg p)?\wedge(\alpha\wedge\neg q)?$ that α is true in at least four worlds, and so on and so forth. With the help of this abbreviation we can nonmonotonically obtain skeptically plausible formulas thought credulously plausible ones according to the confirmation by enumeration philosophy mentioned above:

$!_n$: $\alpha?_n \succ \alpha! \precsim (\neg\alpha)?$

All instances of $!_n$, for some specific *n*, belong to Δ. Note that, according to $!_n$, even if $\alpha?_n$ is true (that is, α is true in at least *n* plausible worlds) two situations might prevent $\alpha!$ from being inferred: if $\alpha!$ implies a contradiction or if $(\neg\alpha)?$ is the case. This second situation is significant, for it illustrates how the exception part can be used to set priority between inductive implications. For instance, imagine that we somehow have got $\alpha?_n$ but there is belonging to A the inductive implication $\beta \succ (\neg\alpha)?$. Suppose further that we have got β. In this case, because of the exception part of $!_n$, $\alpha!$ shall not be inferred: $\beta \succ (\neg\alpha)?$ has priority over $\alpha?_n \succ \alpha! \precsim (\neg\alpha)?$.

4. Conclusion

We have in this paper elaborated on how one might extend paranormal modal logic in such a way as to use the notions of plausibility along with an inductive reasoning mechanism which takes seriously into consideration the epistemic nature of inductive reasoning. More specifically, we introduced a non-classical multimodal logic of plausibility and certainty in

which, on the one hand, the operators of plausibility ? and ! behave paraconsistently and paracompletly, respectively, and on the other hand the operators of certainty and epistemic possibility behave classically. Along with a version of Reiter's default logic, we were able to use this logic of plausibility and certainty to formalize a very simple theory of induction. It should be noted that this formalization is just one among the several possibilities we can to use the logic of plausibility and certainty along with a nonmonotonic reasoning mechanism to formalize a theory of induction. For an illustration of some of these possibilities along with the formalization of less naïve theories of induction see (Silvestre 2010).

Ricardo Silvestre
Universidade Federal de Campina Grande

ricardoss@ufcg.edu.br

References

Béziau, J. (2002) 'S5 is a paraconsistent logic and so is first-order classical logic', *Logical Studies*, 9, 301-309.

Carnap, R (1946) 'Remarks on Induction and Truth', *Philosophy and Phenomenological Research*, 6, 590-602.

———. (1950), *Logical Foundations of Probability*, Chicago: University of Chicago Press.

da Costa, N. C. (1974) 'On the theory of inconsistent formal system', *Notre Dame Journal of Formal Logic*, 15, 497-510.

da Costa, N. C., Carnielli, W. (1986) 'On paraconsistent deontic logic', *Philosophia*,

16, 293-305.

Fitting, M. (1993) 'Basic modal logic', IN Gabbay, D., Hogger, D., Robinson, J. (eds.), *Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. 1, Logical Foundations*, Oxford: Oxford University Press, pp. 368-448.

Fuhrmann, A (1990) 'Models for relevant modal logics', *Studia Logica*, 49, 501-514.

Goble, L. (2006) 'Paraconsistent modal logic', *Logique et Analyse*,193, 3-29.

Hempel, C. (1945) 'Studies in the Logic of Confirmation', *Mind*, 54, pp. 1-26, 97-121.

Loparíc, A. da Costa, N. C. (1984), 'Paraconsistency, paracompletenes and valuations', *Logique et Analyse*, 106, 119-131.

Marcos, J. (2005) 'Nearly every normal modal logic is paranormal', *Logique et Analyse*, 48, 279-300.

Morris, P. (1988) 'The Anomalous Problem in Default Reasoning', *Artificial Intelligence*, 35, 383-399.

Pequeno, T., Buchsbaum, A. (1991) 'The logic of epistemic inconsistency', IN Allen, J., Fikes, R., Sandewall, E. (eds.) *Principles of Knowledge Representation and Reasoning: Proceedings of Second International Conference* (1991) San Mateo: Morgan Kaufmann, pp. 453-460.

Perlis, D. (1987) 'On the Consistency of Commonsense Reasoning', *Computational Intelligence*, 2, 180-190.

Reiter, R. (1980) 'A Logic for Default Reasoning', *Artificial Intelligence*, 13, 81–132.

Silvestre, R. (2006) 'Modality, paraconsistency and paracompleteness', IN Governatori, G., Hodkinson, I., Venema, Y. (eds.), *Advances in Modal Logic 6* (2006) London: College Publications, pp. 459-477.

———. (2007) 'Ambigüidades Indutivas, Paraconsistência, Paracompletude e as Duas Abordagens da Indução', *Manuscrito*, 30, 101-134.

———. (2010) *Induction and Plausibility: A Conceptual Analysis from the Standpoint of Nonmonotonicity, Paraconsistency and Modal Logic*, Berlim: Lambert Academic Publishing.

———. (2011) Paranormal Modal Logic and the Logic of Skeptical and Credulous Plausibility, Parts 1 and 2, *Logic and Logical Philosophy*, 20, 3 (to appear).

Silvestre, R., Pequeno, T. (2005) 'A Logic of Inductive Implication or AI Meets Philosophy of Science II', IN Kégl, B., Lapalme, G. (eds.) *Advances in Artificial Intelligence – AI 2005 (LNAI 3501)* (2005) Berlin-Heidelberg: Springer-Verlag, pp. 232-243.