

BALKAN JOURNAL OF PHILOSOPHY

Vol. 12, Issue 2, 2020

TABLE OF CONTENTS

Articles

Invited article

Platonism And Mathematical Explanations: Some Critical Remarks On Explanatory Proofs
And Debunking Arguments (I) / 63

Fabrice Pataut

Infinite Lifespans, Terraforming Planets, And Intergenerational Justice / 75

Adelin-Costin Dumitru

The Role Of Truth In Explanatory Understanding / 87

Stefan Petkov

On A Regularity Of The Intellectual Evolution Of Mankind / 99

Rovshan Sabir Hajiyeu

How to Assign Weights to Values? Some Challenges in Justifying Artificial Explicit Ethical
Agents / 111

Silviya Serafimova

“Paper Wave”, “Paper Pressing” And “Paper Genocide” As Applicable To The Russian
Education System / 119

Alexander M. Osipov

Healing the Black Sea’s Patrimony: Hospitals, Health and Heritage. An Argument In Favor
Of Creating A Network Of Hospitals-Monument Belonging To The States Across The Black
Sea / 125

Oana Șerban

Muslims In Europe Between “European” And “Islamic” Values: Identity Crisis / 133

Gyulnara Gadzhimuradova

Book reviews

A Coherent Temporalist Metaphysics / 139

Vesselin Petrov

Rudolph Hermann Lotze: Mikrokosmos: Ideen zur Naturgeschichte und Geschichte der Menschheit. Versuch einer Anthropologie. 3 vols., edited by **Nikolay Milkov**, Hamburg: Felix Meiner, 2017. ISBN 978-3-7873-3180-2 / 143

Michele Vagnetti

Articles

Invited article

PLATONISM AND MATHEMATICAL EXPLANATIONS: SOME CRITICAL REMARKS ON EXPLANATORY PROOFS AND DEBUNKING ARGUMENTS (I)*

Fabrice Pataut

Sciences, Normes, Démocratie, UMR 8011,
CNRS, Université Paris Sorbonne,
email: patautfabrice@gmail.com

Contents

Short abstract

Key words

1. Some widely held assumptions about existence, indispensability and best explanation
 2. Two nominalization cases: Field and Rizza
 3. Knowledge (I)
 4. Knowledge (II)
 5. Debunking
 6. Conclusions
- References

Short abstract

Ontological parsimony requires that if we can dispense with *A* when best explaining *B*, or when deducing a nominalistically statable conclusion *B* from nominalistically statable premises, we must indeed dispense with *A*. When *A* is a mathematical theory and it has been established that its conservativeness undermines the platonistic force of mathematical derivations (Field), or that a nonnumerical formulation of some explanans may be obtained so that the platonistic force of the best numerical-based account of the explanandum is also undermined (Rizza), the parsimony principle has been respected.

Since both derivations resorting to conservative mathematics and nonnumerical best explanations also require abstract objects, concepts and principles, ontological parsimony must also be required of nominalistic accounts. One then might of course complain that such accounts turn out to be as metaphysically loaded as their platonistic counterparts. However, it might prove more fruitful to leave this particular worry to one side, to free oneself, as it were, from parsimony thus construed and to look at other important aspects of the defeating or undermining strategies that have been lavished on the disposal of platonism.

Two aspects are worthy of our attention: epistemic cost and debunking arguments. Our knowledge that good mathematics is conservative is established at a cost, and so is our knowledge that nominalistic proofs play a theoretical role in best explanations. I will suggest that the knowledge one must acquire to show that nominalistic deductions and explanations do play their respective theoretical role involves some question-begging assumptions regarding the nature of proofs. As for debunking, even if the face value content of either conservative or platonistic mathematical claims didn't figure in our explanation of why we hold the mathematical beliefs that we do, we could still be justified in holding them so that the distinc-

* I was the respondent to Mark Colyvan's paper "The Ins and Outs of Mathematical Explanation" at the Indispensability and Explanation workshop I co-organized with Marco Panza at the Institut d'Histoire et de Philosophie des Sciences et des Techniques (IHPST) in Paris in November 2012. Section 1 of this paper builds on my critical remarks. The proceedings of the workshop have since then been published, alas without Colyvan's contribution (see *Molinini, Pataut, and Sereni (eds.) 2016*). At the time the present paper was in print, Colyvan's paper was included in a list of Selected Recent Papers as a piece forthcoming in *Mathematical Intelligencer* (see Colyvan's website at <http://www.colyvan.com>).

A substantially different version of the paper was presented in May 2019 at the Seminar in the philosophy of applied mathematics (MathsApp) at the Maison de la Recherche in Paris. Many thanks are due to Mark Colyvan, Marina Imocrante, Daniele Molinini, Marco Panza, Andrea Sereni and Julien Tricart for their criticisms and comments on these two occasions.

The paper is published in two installments. The second part, containing sections 4 to 6, will appear in the next issue of this journal.

tion between nominalistic deductions and explanations and platonistic ones turns out to be invidious with respect to the relevant propositional attitude, i.e., with respect to belief.

Key words: Abstract Objects; Baker (Alan); Clarke-Doane (Justin); Colyvan (Mark); Debunking Arguments; Explanatory Proof; Field (Hartry); Indispensability Argument; Inference to the Best Explanation; Lange (Mark); Mathematical Explanation; Mathematical Realism; Mathematical Truth; Nominalism; Nominalization; Numbers; Ontological Economy; Platonism; Proof; Quantification (objectual and substitutional); Realism; Rizza (Davide); Safety; Sensitivity.

1. Some widely held assumptions about existence, indispensability and best explanation

Ever since philosophers of mathematics have been facing the so-called Quine-Putnam challenge of the late sixties and early seventies (see, e.g., Quine [1968] 1969 and Putnam [1971] 1979), it has been taken for granted by many that if it is shown that mathematics is indispensable in the best explanations of natural phenomena, existence claims of applied mathematics thereby acquire a bona fide philosophical justification.¹ The leading idea is that if

¹ Liggins argues that the phrase “Quine-Putnam challenge” is a misnomer because Quine has another argument for platonism than the one mistakenly attributed to him, and because Putnam isn’t arguing at all in favor of abstract objects. He detects in both Quine and Putnam a stronger argument for platonism than the standard indispensability argument. I shall not discuss this important point here, although, as far the Putnamian point is concerned, the last paragraphs of this section strongly suggest that abstractness, i.e., the noncausal, nonspatial and non-temporal character of the disputed objects, is *not* the issue.

Let me however stress the following point. There is a claim about commitment to objects via truth, which comes directly from Quine, and another one about translatability which comes directly from Putnam. These are strictly speaking distinct, and the very idea of there being a Quine-Putnam challenge obviously plays down the distinction. Field, for instance, conflates the two in *Field 1980a*: 107n4. For Quine, we must conclude that classical mathematics constitutes a body of truths and this implies that we must “countenance,” as Quine often says, i.e., tolerate, its existential assertions. For Putnam, we must accept the so-called realist reading if only because there is still no translation of the language of science — a language which does indeed involve reference to and quantification over mathematical objects — into a language free of such reference and quantification. Or, if there is one, it disrupts logical relations (Putnam [1971] 1979 : sect. V). See Liggins 2008 for his rejoinder to the prevalent understanding of Quine and Putnam’s stand on the familiar epistemological challenge to platonism.

mathematics plays a role in these explanations, it is hard to resist a realist reading of it.² I understand realism, here, as consisting in an existence claim *simpliciter* with respect to mathematical objects, e.g., numbers, involving *prima facie* no commitment to a further independence claim to the effect that such objects exist *ante rem*, independently of mathematical practice, thinking, structures, principles and language. This isn’t a matter of bias with respect to the meaning of the word “realism” which would have one dismiss the independence claim as superfluous, as not being constitutive of the platonist picture, or perhaps as not even being part of it, no matter how small the part. A realist or platonist (I shall be using these terms interchangeably) will, of course, defend the view that mathematical theories relate to an external realm of abstracta, “cut off from all links with the reflecting subject,” as Bernays nicely puts it when contrasting Euclid’s postulates, where figures *must* be “constructed,” with Hilbert’s axioms *stating from the outset the existence, in the system of straight lines, of points, straight lines and planes* (Bernays [1935] 1983: 259, emphasis mine).³

The point, rather, of insisting on existence *simpliciter*, is that the argument Colyvan is considering when defending platonism rests on the idea that if you need *A* to explain *B*, you should take *A* at face value, and that since what you so take in best explanations of natural phenomena is mathematics, the referents of the mathematical terms involved in the explanation and the values of the variables bound by the objectual existential quantifier of the applied regimented mathematical theory are also to be taken at face value.⁴ So, to resort to Baker’s celebrated

² Colyvan argued verbatim in favor of this view in his presentation at the 2012 workshop.

³ See also Bernays *op. cit.*: 267.

For a nice and simple example of construction in Euclid, see *Euclid [ca. 300 BC] 2008*: Book 1, Proposition 1, on how to construct an equilateral triangle on a given finite straight line. There are, of course, many other relevant examples of this kind throughout the *Elements*.

⁴ If we were to favor a substitutional reading of the existential quantifier, or introduce a new symbol for a substitutional existential quantifier in the regimented language of applied mathematics, noted Σ , we’d be claiming that a sentence of the form $(\exists x) Fx$, or of the form $(\Sigma x) Fx$, is true if and only if there is a closed term *t* of the language such that *Ft* is true. If, on the contrary, we were to stick to the standard objectual reading and refrain from such an introduction, we’d be claiming that a sentence of the form $(\exists x) Fx$ is true if and only if there is some object *x* in the relevant universe of discourse such that *F* is true of that object. In the first case, we would be substituting expres-

magicicada exemple (Baker 2005), if you need prime numbers to explain why periodical magicicadas crawl out of their subterranean hideouts every 13 or 17 years, depending on the kind they belong to, you should also hold that primes exist.⁵

You might furthermore believe that the mathematical theory that asserts the existence of 13 and 17 and of their properties such as that of being prime, i.e., number theory, describes structures, say à la Shapiro (Shapiro 1997) or à la Resnik (Resnik 1997), so that these numbers are not truly entities or objects in any (perhaps) naïve, non-Fregean sense, but places that stand in structural relations to one another, i.e., places in the structure of the natural numbers. Were that option chosen, places will be referred to by some appropriate term of the philosophical theory (structuralism, in the case at hand) that discloses what number theory, as we're sometimes fond of saying, is *really* about (not just numbers *tout court* but, more specifically, *numbers qua places*), i.e., quite naturally, "place." Places will then be among the values of the bound variables in expressions of the form $(\exists x) Fx$, insofar as the genuine content of our assertions of the existence of numbers and of their properties is disclosed by the aforementioned philosophy of arithmetic, so that F will be true of places, or of numbers *qua places*, *under the structuralist reading of the theory*.

But we shouldn't be concerned here with any particular take on the nature of numbers or of any other mathematical object needed in the application of mathematical theories. The point, rather, is that if the justification we have for the existence assertions, beliefs in the entities and so on and so forth, comes from the success of the application of the mathematics we need, we *must* take the mathematical existence claims *at face value*, i.e., as claims about whatever it is they are quite literally, verbatim, about — say, numbers. To be an entity or an object in whose existence we believe is simply to be referred to by a mathematical term of the theory needed in our best

mathematical explanations, or to be among the values of the bound variables of some suitably regimented mathematical theory, given that no paraphrase disposing of such terms and of such values of the bound variables is available, either effectively or in principle.

In other words, the contrasting philosophical mood which might lead you to take mathematical existence assertions and beliefs to be about what a particular philosophical theory claims they are, say places, is *not* available in the present context, either as a solution to the indispensability problem, or as a way out of it. You can't say, "Oh, but I don't quite believe that primes such as 13 or 17 exist, or literally exist, I just believe in places in the structure of $\mathbf{N}$." And if you wish to say it, your belief in *that* must be supported by an explanation of how it is that places, or numbers *qua places*, play an indispensable role in, say, the best mathematical explanation of why periodical magicicadas crawl out of their subterranean hideouts every 13 or 17 years depending on the kind they belong to. However, to repeat, in the present context of investigation, whatever turns out to be metaphysically the case under a particular philosophical reading, interpretation or construal of the applied theory, need not concern us. The face value claim that numbers are needed and useful must indeed be taken at face value — in earnest.

There will also be disagreements about the details of the argument, e.g., typically, on whether it is possible to argue for the realist or platonist reading through confirmation holism, i.e., on whether applied mathematics inherits that reading from whatever evidence or justification is available in favor of the empirical theories in which it plays the role of a theoretical apparatus needed for formulations, computations and predictions.⁶ So if, say, you obtain some appropriate confirmation for Newton's law of gravitation, a law that involves forces and masses, you'll thereby have a basis on which to secure a realist reading of the real numbers that measure them.

The core contention here is that mathematical explanations "are significant for [the] philosophy of mathematics, especially for the realism /anti-realism debate,"⁷ because they justify mathematical existence claims. They do in the sense that it would be *incoherent* to use mathematics *and* to deny the exist-

sions for the bound variables so as to get true sentences; in the second we would assign values to the bound variables so as to get genuine existence claims.

I won't discuss here the moot point of the relevance of the substitutional reading of existential quantification (or of existential quantification of the substitutional kind) to the platonism issue, when platonism is argued for on the basis of an indispensability claim based on an inference to the best explanation.

⁵ I have discussed the so-called "enhanced" version of that argument in a reply to Baker 2016 in Pataut 2016. More on this below in section 3.

⁶ See, e.g., Leng 2010 and Leng 2016 for a defence of the view that applied mathematics does indeed inherit that reading.

⁷ Colyvan's words in his oral presentation.

ence of mathematical objects; say, to explain that the life cycles of cicadas are either 13 or 17 years, *and* to claim that these numerical terms fail to refer, or that mathematical existence assertions that have primes among the values of their bound variables are false (or perhaps that they are neither true nor false). Given that double talk is ruled out from the start as deceptive (and that we really would be in a case of deceptive double talk), inference to the best explanation forces mathematical platonism, or realism, on us.

What is proposed here by the platonist is an inference that relates an indispensability claim with respect to mathematical explanations to a face value claim with respect to mathematical objects. The inference relies on a number of strong assumptions, and it will be useful to make a note of them.

The strong assumptions are the following:

- (A1) If we can dispense with *A* when best explaining *B*, then we must dispense with *A*.
- (A2) If we can dispense with *A* when best explaining *B*, then we must conclude that *A* terms fail to refer and that the *prima facie* (objectual)⁸ existence claims of *A* are false.
- (A3) If we can't dispense with *A* when best explaining *B*, then we must take *A* at face value, conclude that *A* terms refer and that the *prima facie* (objectual) existence claims of *A* are true.

(A1), of course, expresses the philosophical motivation behind any attempt at nominalization; it may itself be motivated by a yet more general belief in the virtues of Ockham's razor so that, in case a mathematical explanation and a non-mathematical explanation of *B* are both on offer, we should prefer the nonmathematical one, out of respect for the parsimony principle inherited from Ockham.

This, I think, tells us something important. It tells us that if we are to judge the respective virtues of mathematical explanations and non-mathematical or nominalized ones by parity and refrain from invidious distinctions between the two kinds, as we certainly must, we should allow ourselves to yet another inference that likewise relates indispensability claims with respect to best *non-mathematical* explanations of *B* to face value claims with respect to whatever will have *replaced* mathematical objects in a nominalistically acceptable explanation of *B*. The inference is this : If a nonmathematical explanation of some natural phenomenon is available, we should

conclude that this particular explanation best explains the phenomenon and that its nonmathematical claims must be taken at face value. If nominalist explanations have any explanatory virtue, then we must dispense with mathematics and embrace a different ontology. Whether that ontology is genuinely lighter is a problem I shall return to.

The relevant distinction here is between a genuinely mathematical explanation and a genuinely nonmathematical one, i.e., between one that applies mathematics and one that applies something else, one that resorts to, say, first order logic and set theory, as in *Field [1980] 2016: ch. 2*, or to axioms of measurement theory as in *Rizza 2011*. Of course, first order logic, set theory and measurement theory are as abstract as can be. In any event, first order logic and set theory are as abstract as arithmetic taken at face value (Field's case), and measurement theory is as abstract as number theory taken at face value (Rizza's case), so that the issue of abstractness and the so-called access problem to abstracta turns out to be pretty much a red herring.⁹

Something else must motivate the choice. Accordingly, parsimony or ontological economy should be construed in such a way that mathematical explanations and nonmathematical explanations cannot fare equally with respect to something *not* related to abstractness. In other words, since (A1)–(A3) applies, *mutatis mutandis*, to nominalized explanations, and since this clearly indicates that the success of an explanation, whether platonistic or otherwise, plays the role of a *defeater*, or *underminer*, with respect to its competitor, we should be wary of any claim to the effect that a particular explanation has an advantage just because it defeats or undermines a metaphysical or ontological view couched in terms

⁹ One might argue, of course, with respect to the Fieldian case, that first order logic is *not* committed to anything. One might object: "there are no objects, in any clear sense of the word 'object' of first order logic." A first order language allows quantifiers to bind only individual variables; or, to put it semantically, first order quantification is quantification over individuals, but in the case of first order *logic*, no ontological commitment is implied via the use of bound individual variables.

Moreover, one might argue that Field's arithmetical example is ill-chosen here because it isn't concerned with what counts as the explanation of a natural phenomenon. Still, we are offered an account of how to reach a conclusion using a mathematical apparatus that's merely conservative so that a dispensability claim is (perhaps only allegedly) established through an argument that relies on substituting logic and set theory for arithmetic. I shall go back to this point in the next section.

⁸ Given what has been said in footnote 5, "objectual" is redundant. I resort to it here as a reminder, both in (A2) and in (A3), and shall not resort to it afterwards.

of access to abstracta. *Something else* must be defeated or undermined. What is the *nonmetaphysical* or ontological benefit of a platonistic or, on the contrary, of a nominalistic explanation?

Suppose, for the sake of argument, that there is nothing in the mathematical world, as we sometimes rather carelessly say, to make mathematical discourse true. Are we thereby losing any justification we might have to look for mathematical explanations? The very same kind of question is in order for competing nominalistic explanations. There is no reason why they should be treated differently in this respect. Suppose there is nothing to make *nonmathematical* explanatory discourse true. Suppose, e.g., that there is nothing to make first order logic and set theoretical discourse true (Field's case) or measurement theory true (Rizza's)?¹⁰ Are we thereby losing any justification we might have to look for nominalistic explanations that resort to these theories, i.e., explanations that will justify a dispensability thesis?

Why bother about justifications, one might complain? For the following reason. I'm thinking here of so-called debunking arguments that have been developed in discussions of the evolutionary standpoint in meta-ethics. This takes us far away indeed from mathematics, but it helps us keep a close watch on (A1)-type strategies, and this is what counts. Suppose there is nothing in the moral world that makes moral judgments true, or that there is no such thing indeed as a moral world. Does it thereby follow that it would be a mistake to either accept or reject moral judgments on the basis of suitable warrants, or to believe them either true or false? Would it be a waste of time to look for justifications and refutations of a distinctly moral kind?

Another connected question concerns the explanatory role of such beliefs, and that question might obviously be asked about mathematical beliefs as well. It might indeed be asked about beliefs in general. Suppose that the face value content of mathematical claims doesn't figure in our best explanation of why we hold the mathematical beliefs we do indeed hold. Would this undermine these beliefs? Would this then in turn undermine the phil-

osophical view that there are mathematical objects?¹¹ In meta-ethics, defenders of debunking arguments have argued that if the face value content of ethical claims doesn't figure in our best explanation of why we hold the moral beliefs we do, this would undermine them in the sense that it would threaten or impair our confidence in them. The very same questions are in order for mathematical beliefs. If we still would have the mathematical beliefs we do as a matter of fact have, had they been false, and if we could not have easily had false mathematical beliefs, using the methods that we do use in current mathematical practice, is there any place left for a debunking argument with respect to mathematical beliefs?

I shall return to these worries in section 4. The crucial question for now is: How are we to assess the respective values of mathematical explanations and non mathematical ones?

2. Two nominalization cases: Field and Rizza

The question may be asked about the two cases I've briefly mentioned so far, from Field (*Field [1980] 2016: ch. 2*) and Rizza (*Rizza 2011*). It will be interesting to look at these, in part because it is often argued, or perhaps merely felt, that they take issue with indispensability and ontological commitment in diverging ways.

In Field's case, the arithmetic of natural numbers supplemented by set theory is replaced by the logic of identity supplemented by existential numerical quantifiers (for ease of exposition), so that a particular empirical conclusion to the effect that there are exactly 63 bugs at a given location in space may be reached by means of a deductively valid argument using the arithmetic of natural numbers supplemented by set theory in a noncommitted way. In Rizza's case, the mathematical results one needs to explain why periodical magicadas crawl out of their subterranean hideouts every 13 or 17 years are paraphrased with the help of axioms of measurement theory so that the whole cicada case is reformulated without any recourse to mathematics, using only betweenness relations on temporal intervals.

It might be thought that what is shown in Field's arithmetical example is that we may go through a derivation by means of a mathematical system that is conservative but not true, either literally or otherwise, but that *no* genuine explanation has thereby been delivered, let alone a best or nomi-

¹⁰ Putnam's argument in *Putnam [1967] 1979: 33-34* is quite different in this respect for he argues that the applicability of number theory requires nothing more than the consistency of mathematics, and consistency involves no commitment whatsoever to existence claims. It involves the claim that nothing logically absurd or self-contradictory is a theorem of mathematics, or that there are no two theorems of mathematics such that one is the negation of the other.

¹¹ See *Clarke-Doane 2016* on the twin issues of safety and sensitivity. More on this, with an unpacking of both notions, in section 5.

nalistic one. I'll try to dispel the fear that this is all we have in that particular instance, and that, in both cases, although for different reasons, we are in a situation where, if the respective platonistic forces of the explanations to be replaced have been undermined, the advantage thereby yielded *isn't* one of ontological economy.

In Field's example, we start with an argument whose premises constitute a body of "nominalistically stated assertions," i.e., of assertions containing no occurrence of terms or of quantifiers for "abstract objects." We then provide an abstract counterpart for each of the premises, apply a mathematical system "that includes the arithmetic of natural numbers plus some set theory" (*Field op. cit.*: 22), so that the conjunction of the abstract counterparts of the premises and the provable results of the mathematical system "entail" the abstract counterpart of the nominalistic conclusion to be reached. Since the equivalence of nominalistic assertions and their abstract counterparts is provable in the logic of identity supplemented by the numerical quantifiers plus the arithmetic of natural numbers supplemented by set theory, "we have proved the nominalistically stated conclusion" (*Field loc. cit.*).

We may reconstruct the full argument in the following way. ("MS1," etc., stand for the numbered assertions of the mathematical system used in the derivation, "Equiv." stands for "logical equivalence," "CP" stands for "conservative principle," [1*] and [1*'] are assertions in which each quantifier has been restricted with a formula " $\neg M(x)$," meaning intuitively " x is not a mathematical entity," so that, as far as ontological commitment is concerned, every quantification of the form " $(\exists x) \dots$ " is replaced by one of the form " $\exists x (\neg Mx)$ and..." and "Nom." stands for "Nominalism.")

(1) There are exactly 21 aardvarks.

[1*] [There are exactly 21 aardvarks that are not mathematical objects.]

(2) On each aardvark there are exactly 3 bugs.

(3) Each bug is on exactly 1 aardvark.

(1') The cardinality of the set of aardvarks is 21.

[1*'] [The cardinality of the set of aardvarks that are not mathematical objects is 21.]

(2') All sets in the range of the function whose domain is the set of aardvarks, and which assigns to each entity in its domain the set of bugs on that entity, have cardinality 3.

(3') The function in 2' is 1-1 and its range forms a partition of the set of all bugs.

MS1. If all members of a partition of a set X have cardinality α , and the cardinality of the set of members of the partition is β , then the cardinality of X is $\alpha \times \beta$.

MS2. The range and domain of a 1-1 function have the same cardinality.

MS3. $3 \times 21 = 63$.

(4') The cardinality of the set of all bugs is 63.

(4) There are exactly 63 bugs.

Equiv. The equivalence of (1)–(4) and (1')–(4') is provable in the logic of identity supplemented by numerical quantifiers plus the arithmetic of natural numbers supplemented by set theory.

CP Let A be any nominalistically storable assertion and N any body of such assertions; let S be any mathematical theory. Then A^* , that results by restricting each quantifier in A with the formula " $\neg M(x)$ " (meaning intuitively " x is not a mathematical entity"), isn't a consequence of $N^* + S + \exists x (\neg Mx)$ unless it is a consequence of N .

Nom. (4) follows from (1)–(3) alone.

What we have in Field's arithmetical case is an example of how a nominalist may apply mathematics for deducing nonmathematical consequences, given that mathematics is conservative (see *Field op. cit.*: 14). What we have in Rizza's take on the Baker example is obviously quite different, namely the application of a system that is *nonmathematical* through and through because it resorts to betweenness relations on temporal intervals, and neither may qualify as mathematical objects.

Baker points out in *Baker 2005* that magicicadas spend either 13 or 17 years in larval form before emerging as adult insects because prime periods uniquely maximize the infrequency of overlap with their periodical predators. The explanation in terms of evolutionary advantage has two salient features that are relevant for the platonism issue as it is discussed here: It is mathematical and it is of the optimization kind.

The explanation counts as mathematical because it draws on the number-theoretic result that any natural number p is coprime with any other natural number q given that $q < p$, and thus maximizes the lowest common multiple with every q just in case p is prime. It is of the optimization kind because the primeness of the periods provides a solution to an optimization problem: The target feature, i.e., the avoidance of periodical predators, is optimized given the constraints imposed by ecological factors deemed relevant and a set of life-cycle

lengths deemed biologically possible given such constraints.

Baker's claim is that since (i) there is no direct empirical evidence for the entities that feature as a solution to the optimization problem, i.e., for periodical predators of magicicadas, and (ii) prime periods of life cycles of magicicadas are uniquely optimal, there are prime numbers (*Baker 2016*). The platonist conclusion is warranted by the fact that (i) and (ii) conjointly amount to the claim that we are in a genuine case of inference to the best mathematical optimization explanation. Since there is no debating that prime periods are uniquely optimal, the existence of numbers is "implied by the number-theoretical result" (*Baker 2016*: 337). To be precise, the controversial existence claim about numbers *simpliciter* is implied *only* because the result is *applied* in order to obtain an optimization explanation that takes it at face value, given that the result, taken at face-value, *asserts* (rather than *implies*) that there are prime and coprime natural numbers.¹²

The platonistic force of the best optimization explanation is entirely a function of mathematical objects (natural numbers, primes, coprimes and lowest common multiples) featuring in a pure number-theoretical result being applied to concreta (periods) in such a way that the application, by assigning a theoretical role to these objects, justifies a conclusion to the effect that there is an evolutionary advantage to the existence of prime numbers of periodical life cycles. Given the theoretical role that such objects play in an explanation in terms of the minimizing of the intersection of prime periods with any shorter periods, the (Hard Road Piecemeal) nominalist's task is to undermine such controversial platonistic force by showing how that role (the minimizing) may be played by some other entities that do not commit one to the platonist's controversial existence claim, but still yield an optimization explanation in terms of evolutionary advantage.¹³

¹² Strictly speaking, the lemma one resorts to in the explanation makes no existential claims. Given Quine's criterion, in order to qualify as a genuine assertion of the existence of mathematical objects that one may furthermore decide to take at face-value, the lemma would have to be rephrased so that natural numbers, prime numbers, coprime numbers and lowest common multiples appear as values of variables bound by the objectual existential quantifier.

¹³ The nominalist is dubbed in this instance a "Hard Road Piecemeal Nominalist" because he "respond[s] to the indispensability argument by attempting to undermine the claim that mathematics is explanatorily indispensable for

We may formulate the argument in favor of the claim that a mathematically-free explanans that carries the same explanatory load as the mathematical or platonistic one is available for the magicicada case in the following way:

1. Mathematical explanation

P1. A cicada subspecies with period p and a predator species with shorter period q will coincide at intervals corresponding to the lowest common multiple (LCM) of p and q .

P2. If p is prime, then its LCM with any smaller q is $p \times q$, which is as large as possible.

C. Prime periods maximize the infrequency of overlap with periodical predators on the assumption that the predators have shorter periods than the cicadas.

Remark on the mathematical explanation. The key property in the explanation *P1-C* is coprimeness. (Two numbers are coprime if they have no common factors, which in turn ensures that their LCM is as large as possible.)

Mathematical result. Let p and q be natural numbers; p is coprime with every $q < p$, and hence maximizes LCM with every $q < p$, iff p is prime.

2. Mathematics- (or, rather, number-) free formulation of the explanans

We may depict life-cycles as segments on a line and then describe divisibility relations between the lengths of sequences of intervals over a segment partitioned into congruent intervals.

(1) The basic results and notions involved in that geometrical description may be given in non-numerical fashion.
(2) The properly numerical formulation of these notions and results may be recovered from non-numerical conditions.

(1) and (2) may be established by using axioms of measurement theory.

Non-committed interpretation of the axioms. Axioms of measurement theory may be interpreted on relations between time-intervals. The axioms have models that are representable on a numerical structure N that allows the reduction of the numerical reasoning to the numerical representation of the relevant facts about a geometry of time intervals, so that the numerical reasoning leading to the derivation of the length of magicicada's life cycles

science [...] by attempting some kind of nominalistic paraphrase for, reconstruction of, or alternative to mathematics-laden scientific explanations." What makes that approach specifically piecemeal as opposed to holistic is that it "take[s] individual putative cases of mathematical explanation in science and come[s] up with alternative explanations which are purely nominalistic" (*Baker 2016*: 339–340). See also *Colyvan 2010*.

depends on the preliminary availability of suitable empirical relations. It is because these empirical relations are in place that it is meaningful to use numbers in order to reason about life-cycles. Once the numerical assignment is in place, it is, of course, possible to introduce the relevant numerical notions and thus describe the relevant divisibility relations. Although this may be done numerically, it isn't necessary to appeal to numbers because divisibility relations between life-cycles, i.e. time intervals, may be represented geometrically.¹⁴

Provided these arguments are conclusive, we have an account of why the mathematics *doesn't have to carry the explanatory load*, either because it is used conservatively, or because it is dispensed with.

The idea that we do have an explanans in Field's arithmetical case is not as preposterous as it might seem at first blush. Of course, one might want to object that the claim that there are exactly 63 bugs at a given location isn't on a par with the claim that magicicadas spend either 13 or 17 years in larval form before emerging as adult insects. One reason is that no regularity is involved in the fact that there are 63 bugs distributed on a limited number of aardvarks, not even an accidental generality. Yet both are empirical facts, facts about the natural world or about natural phenomena, and both claims involved are stated in such a way that the existence of a number of nonmathematical entities (years or life-cycles; bugs) is asserted. The fact that the inference to the best explanation is interesting to biologists – because “there are known predators of cicadas, including some parasitic insect species that target cicadas specifically [although none] of [those] known predators or parasites are themselves periodical” (Baker 2016: 336) – whereas the fortuitous remark about bugs on aardvarks isn't of any interest to anyone, need not worry us.

What should worry us is that the nominalistic proposals play the role of defeaters or underminers with respect to a platonistic competitor, and that the undermining or defeating of mathematical realism, supported either by a derivation or by a best explanation in numerical terms taken at face value, has a cognitive cost.

I've said earlier (towards the end of section 1) that something *not* related to the abstractness, remoteness, acausality or inaccessibility of objects of a certain kind must be defeated or undermined so that, provided that we must prefer nonmathematical explanations to mathematical ones and abide by (A1),

another lesson must be drawn from this preference that is different from (A2) and (A3). I will suggest that the lesson cannot be that explanations with mathematics-free formulations of the explanantia, or explanations with merely conservative mathematics, offer an epistemic advantage.

3. Knowledge (I)

Suppose that the mathematics taken at face value doesn't carry the explanatory load so that we know that there are exactly 63 bugs at a given location and why magicicadas spend either 13 or 17 years in larval form before emerging as adult insects *nonmathematically*. The two cases are strikingly divergent with respect to the kind of nonmathematical knowledge involved. This, of course, is due to the fact that Field's derivation relies on an argument to the effect that mathematics, and arithmetic in particular, is conservative, and thus eschews the prospect of paraphrase and reconstruction, while Rizza's proposal does on the contrary offer a reconstruction of a numerical-based explanation in strictly geometrical fashion so that a numerical reasoning taken at face value may be paraphrased within a non-numerical theory.

Given that we do have an explanans in both cases, we might have two different kinds of worries. To begin with, how do we acquire the knowledge that is required for the undermining of platonism or realism? How do we know that the applied mathematics is conservative, and how do we know, not just the axioms of measurement theory, but that they may be safely interpreted in terms of relations between time-intervals? Why should we believe in their truth in the first place, either when the axioms are taken at face value or when they are suitably interpreted? It seems that quite a lot has to be known, either to secure conservativeness or the explanatory role of the proposed reconstruction. Secondly, it is unclear how the content of that knowledge may in either case benefit from a purely nominalistic account (provided, of course, that one should abide by this further constraint).

In Field's case, we know that a nominalistically stated conclusion follows from nominalistically stated premises *without* having *actually* derived (4) from (1)–(3), or, to be precise, without having derived it directly. It is essential, in that view, that the “intervening” premises one resorts to in the derivation aren't true, either literally or otherwise. It is sufficient that they preserve truth among the nominalistic statements (Field *op. cit.*: 14). How does one

¹⁴ See Rizza 2011: sections 4 and 5, and Sam Baron's criticism in Baron 2016.

know that premises that are not in the market for truth nevertheless preserve truth among nominalistically stated claims?

Field proposes two answers. The weak answer is that “[i]n the conservativeness case, the kind of inductive considerations that are relevant are the knowledge that in the past no one has found counterexamples to conservativeness, and also the knowledge that in many actual cases where platonistic devices are used in proofs of nominalistic conclusions from nominalistic premises [as in the arithmetical case], these devices are eliminable in what seems to be a more or less systematic way” (*Field op. cit.*: 109n10). The stronger answer is that CP follows from another somewhat stronger conservative principle:

(CP’) Let A be any nominalistically statable assertion, let A^* result by restricting each quantifier in A with the formula “ $\neg M(x)$ ” (meaning intuitively “ x is not a mathematical entity”), and let S be any mathematical theory. Then A^* isn’t a consequence of S unless it is logically true.

In Rizza’s case, the explanation that unifies similar instances of periodicity under a common explanatory pattern in terms of betweenness relations on temporal intervals appeals to axioms of measurement theory.

In Baker’s original platonistic account, mathematical objects such as natural numbers, and properties or relations such as being prime and coprime, play a role in the explanation (see, e.g., *Baker 2005*: 234). Rizza denies, however, that such properties and relations *must* be “interpreted as properties and relations on certain distinctive mathematical entities” (*Rizza op. cit.*: 105-106). The key idea is that these properties and relations are properties and relations “of time intervals corresponding to life-cycles which [...] can be studied nonnumerically” (*Rizza loc. cit.*). One may study them geometrically by depicting respective life-cycles as segments on a line. One then asks how often sequences of consecutive life-cycles of predators and cicadas respectively determine time intervals of equal length. When the life-cycles do determine this, a divisibility relation obtains. Rizza concludes that “a form of divisibility can be recovered in non-numerical terms and so a non-numerical version of primality can be recovered too” (*Rizza loc. cit.*).

Of course, as Rizza correctly remarks, although the paraphrase or reconstruction does eliminate reference to numbers, “it does not eliminate the math-

ematical form of the reasoning leading to a derivation of a prime life-cycle length” (*Rizza op. cit.*: 110). The crucial claim in this respect is that “mathematics may play a relevant explanatory role in a non-ontological way, by providing the means to conceptualize the structure of an empirical phenomenon in such a way that its explanation can be represented as a derivation from certain formal conditions, [with non-numerical counterparts in the reasoning]” (*Rizza op. cit.*: 113).¹⁵

In the Fieldian case, CP’ is taken to be obvious, if only because non-conservative mathematics would turn out to be inconsistent and that a proof of the inconsistency of standard applied classical mathematics is highly unlikely: Good, applicable, useful mathematics is both true in all possible worlds and a priori true (*Field op. cit.*: 13). The a priority claim, either couched in terms of truth in all possible worlds or in terms of a knowledge obtained and justified independently of experience, is, of course, a disputable one.¹⁶

What I want to suggest here is that the problem isn’t so much with the view that mathematics is conservative *as such*, or with the mathematical form of the reconstruction or paraphrase *as such*, as with the kind and breadth of knowledge one must acquire that would justify either the claim that mathematics is conservative and a priori, or the claim that one may recover a non-numerical form of deduction that turns out to be essential to some nominalistically respectable paraphrase.

Field’s claim that good mathematics is logically true and Rizza’s claim that the geometrical concepts that are introduced to study the relevant empirical phenomena constrain the selection of a certain length require, respectively, *further* logical and geometrical knowledge, much more in either case than is needed in standard platonistic accounts. One evaluating the epistemic cost might conclude that too much is required to undermine the platonistic force of the explanations to be replaced, so that the cost isn’t really worth our trouble, especially if a nominalization of the content of that knowledge must also be established. Not only does this further requirement sets the bar very high — presumably way too high —, but problems arise with respect to the

¹⁵ Or, again: “the mathematical character of the explanation resides in the form of the reasoning being used and in the concepts invoked, rather than in the reference to [and quantification over] certain particular entities” (*Rizza op. cit.*: 110).

¹⁶ Field doesn’t consider the second, Kantian, construal.

proofs at stake both in the paraphrase of the mathematical apparatus appealed to in nominalizations and in the conservativity argument.

To begin with, is it possible, or indeed advisable, to provide nominalistically acceptable proofs of all the results appealed to in a nominalistic explanation?

Rizza, for one, proposes to paraphrase any additional mathematical apparatus appealed to in the platonistic proof at stake in the magicicada case. Nominalized proofs should presumably be construed as formal derivations (as opposed to, e.g., acts construed à la Brouwer, in orthodox intuitionistic fashion). But how could proofs construed as syntactic objects contribute to a genuinely non-mathematical explanation, whether or not the explanation is of the optimization kind? Proofs, whether construed as acts or derivations, are mathematical objects in their own right. And even if it were established that proofs syntactically construed do contribute to a bona fide nominalistic account or explanation, a further argument would have to be provided that this force is weaker than the platonistic one, in spite of the fact that part of the mathematical apparatus that plays a role in nominalistic explanations, such as axioms and postulates, is just assumed to be true.

Moreover, once it is admitted that mathematical proofs play a theoretical role, it looks as though one is committed to some form or other of justification holism, so that *all* proofs with theoretical roles, along with the mathematical apparatus they require to be carried out, will have to be nominalized for the undermining of the platonistic force of platonistic explanations to be successful. Where should one stop? The platonistic force of an explanation varies with the strength and quantity of mathematics required for a proof of the applied results to be carried out. Not only is it unfair to require that nominalistically acceptable proofs be provided for all the applied number-theoretic results involved in, e.g., the cicada case, but the Hard Road Piecemeal nominalist is now turning into a mathematical holist. By way of a slippery slope argument, once you've started applying mathematics, there is no bound to the amount of theory you have to nominalize in order to get your Hard Road Piecemeal nominalism going.

Baker's idea, of course, is that it is not enough, in order to commit oneself to the existence of mathematical objects that the theories and statements about them be indispensable in the natural sciences: they must be indispensable *in the right way*, i.e., they must themselves offer the best explanation of

some empirical phenomena (Baker 2009). Does this restriction help?

Steiner's take on the notion of explanatory proof turns out to be helpful when discussing this particular point. Steiner argues that proofs are explanatory when they directly refer to a characterizing property of the entity mentioned in a theorem such that from the proof, it is evident that the result *depends* on the property (Steiner 1978). The role of a proof in the case of a mathematical optimization explanation, such as Baker's cicada case, is strong, or truly explanatory, just in case it is "integral" or "essential" or "central" to it. It is weak just in case it provides a "justification" of the (applied) mathematical claim that is not itself part of the explanation.

Should the nominalist offer paraphrases of proofs that are strong or explanatory in Steiner's sense, i.e., paraphrases that offer *more* than a justification for the truth of the nonmathematical claims one must resort to in nominalized optimization explanations? One might think that the most challenging nominalistic approach is one where *both* the mathematical results at stake *and* strong proofs in Steiner's sense are given a nominalistically acceptable construal.

If that is the case, the reasoning centered on the theoretical role puts us in a puzzling situation. It now appears that the platonist accepts proofs as a principal means of justification for useful mathematical claims, so that the business of the nominalist is to paraphrase these in some nominalistic language. This is a strange claim, first of all because platonism, when characterized semantically in terms of truth, must be construed as the claim that mathematical truth is independent from mathematical proof and may indeed go beyond provability in principle (and perhaps even beyond practical or feasible provability);¹⁷ secondly because the nominalist's task isn't so much to establish that one may dispense with the theoretical *role* of mathematical objects, as previously announced, as to paraphrase the definitions of the mathematical terms occurring in the purely mathematical formulations of the proofs and lemmas one resorts to in platonistic explanations (see Tallant 2013 on this point).

The crucial point, then, is that the theoretical terms occurring in the formulation of some nominalized optimization explanation must be such that their referents play a strong explanatory role in the explanation. Tallant's idea is that, most of the time, proofs

¹⁷ See the beginning of section 1 and footnote 4 on Bernays.

only play a weak role, i.e., justify mathematical claims but aren't strictly speaking part of a best explanation. What needs to be assessed here is whether (and if so, why) proofs, especially proofs that explain, as opposed to proofs that merely prove, would necessarily yield optimization explanations that appeal to the relevant aspects of the world's network of causations rather than to strictly mathematical facts, entities or properties.

I shall now briefly turn to Mark Lange's take on the issue (Lange 2013) before going back to the details of the conservativity argument. Lange claims that the magicicada case is as a matter of fact a case of *non* mathematical optimization explanation. Some scientific explanations are distinctively mathematical because they appeal to mathematical "facts." Although it quantifies over mathematical entities, the biological explanation we have been concerned with so far is an ordinary causal one (Lange 2013). What makes it causal and ordinary as opposed to distinctively mathematical is that it relies on a description of the explanandum's causal history and that natural selection is a causal concept *par excellence*. Although the lemmas one uses in the platonistic proof do indeed play an explanatory role, the optimization explanation that relies on them in its explanans *does not* resort to mathematics "in the manner that is exploited by [...] indispensability arguments" (Lange *op. cit.*: 492). Neither the quantification over mathematical entities per se, nor the explanatory role that the entities in the course of values of the bound variables play in an optimization explanation may be used in these arguments.

Even if that were indeed the case, Lange's distinction between mathematical explanations that rely on proofs proper (conceived as acts or otherwise) and mathematical explanations that rely on "facts" alone, i.e., on what proofs establish, as well as on postulates or axioms, doesn't help in the way of an explanation of why that is. Lange goes as far as claiming that distinctively mathematical explanations include "mathematical facts [...] but *not* their proofs — *much less* proofs that explain *why* those mathematical facts hold [emphases mine]" (Lange *op. cit.*: 508). Explanations that include mathematical facts (as opposed to proofs), appeal to something that is modally stronger than what explanations that (purportedly) include proofs of those facts provide. They do because they relate to the framework of *any* causal relation rather than restrict themselves to some specific explanandum's causes and to the particular causal histories that ordinary explanations

take into account in their explanans — e.g., typically in the case at hand, specific ecological factors imposing constraints on the avoidance of predators, or some evolutionary advantage assessed in terms of observed past regularities. Moreover, they help answer why questions.

Notice however that in Baker's original case, what makes the (allegedly ordinary) explanation that resorts to Lemma 2 a better explanation of the maximization of the infrequency of overlap of magicicadas with their periodical predators than the (allegedly ordinary) explanation that resorts to the weaker Lemma 1 is precisely what enables the better explanation to answer why questions, in particular why prime periods would *not* be uniquely optimal in certain counterfactual situations (in case there would be, say, only 2-year predators). In the author's view, the increment in explanatory force is indeed cashed out in counterfactual terms. Now, of course, Lange is looking for something even stronger than that: The explanation must reveal that the explanandum is "more necessary" than the causal laws that govern it, so that optimization explanations in biology may be deemed "distinctively mathematical" only insofar as they are instances of particular mathematical conjectures.¹⁸

Finally, we must reckon with the fact that the syntactic or formal structure of the proofs one resorts to in both platonistic and nominalistic explanations may count more than expected when one defends the idea that the indispensability argument is primarily concerned with "extra-mathematical" explanations, i.e., with explanations which are given "via a 'spill over' from within mathematics itself."¹⁹ After all, explanatoriness is tied to the structure of proofs, if only for the following reason. If you conclude that p from a contradiction derived from $\neg p$, you have a warrant for p , or for the truth of p , but you don't necessarily know *why* p , or why p is true. You haven't always gained what Gowers and Neilson call a "crucial insight" (see Gowers and Neilson 2009).

There is yet another and deeper disagreement here, not about what contradictions amount to — we may agree on their logical form — but about how to proceed to get one, a disagreement between advocates of the classical *reductio* and advocates of the intuitionistic *reductio*. The case of negation and of the introduction and elimination rules for that constant is revealing. In the first case, it is sufficient, in

¹⁸ See my critical comments on this point in Pataut 2016.

¹⁹ Colyvan's words in his oral presentation.

order to assert $\neg p$, that the mere supposition of the truth of p would lead to a contradiction. In the second case, we must possess a construction such that, from the supposition that a construction of p would be obtained, a contradiction may be derived (see Heyting [1956] 1971: sections 2.2.2 and 7.1.1).

[Question: must we leave such a large blank space before the bibliography? Probably not.]

References

- Baker A (2005) Are there genuine mathematical explanations of physical phenomena?. *Mind* 114: 223–238.
- Baker A (2009) Mathematical explanation in science. *The British Journal for the Philosophy of Science* 60: 611–633.
- Baker A (2016) Parsimony and inference to the best mathematical explanation. In Molinini, eds Pataut and Sereni, 333–350.
- Baron S (2016) The explanatory dispensability of idealizations. *Synthese* 193(2): 365–386.
- Bernays P ([1935] 1983) On Platonism in Mathematics. In *Philosophy of mathematics – selected readings*, English translation by C. D. Parsons, eds P. Benacerraf and H. Putnam, 258–271, Cambridge UP, 2nd ed., Cambridge, Mass.
- Clarke-Doane J (2016) What is the Benacerraf Problem?. In *Truth, Objects, Infinity: New Perspectives on the Philosophy of Paul Benacerraf*, ed. F. Pataut, 17–43, Springer, “Logic, Epistemology and the Unity of Science” series.
- Colyvan M (2010) There is no easy road to nominalism, *Mind* 119: 285–306.
- Colyvan M (2012) *An Introduction to the Philosophy of Mathematics*. Cambridge: Cambridge UP.
- Dummett M (1977) *Elements of Intuitionism*, 1st ed. Oxford: Clarendon Press.
- Euclid, *The Elements*, Clay Mathematics Institute Historical Archive; James A. Carlson et al., eds., <http://claymath.org/library/historical/euclid>, [ca. 300 BC] 2008.
- Field H (1980) *Science without Numbers - A Defence of Nominalism*, 1st edition: Princeton UP, Princeton, New Jersey, 1980; 2nd ed. with a Preface to Second edition, a Bibliography for Second Preface, a letter from W. V. O. Quine and an Index (to entire volume): Oxford: Oxford UP, 2016.
- Gowers T, Neilson M (2009) Massively collaborative mathematics, *Nature* 461: 879–881.
- Heyting A ([1956] 1971), *Intuitionism – an Introduction*, 3rd revised edition. Amsterdam: North-Holland Publishing Company.
- Horsten L (2012) Philosophy of Mathematics. In: *Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta, <http://plato.stanford.edu/archives/sum2012/entries/philosophy-mathematics/>.
- Lange M (2013) What makes a scientific explanation distinctively mathematical?. *The British Journal for the Philosophy of Science* 64(3): 485–511.
- Leng M (2010) *Mathematics and Reality*. Oxford: Oxford UP.
- Leng M (2016) Truth, Fiction and Stipulation. In *Truth, Objects, Infinity: New Perspectives on the Philosophy of Paul Benacerraf*, ed. F. Pataut, 147–158. Springer, “Logic, Epistemology and the Unity of Science” series.
- Liggins D (2008) Quine, Putnam and the ‘Quine-Putnam’ indispensability argument. *Erkenntnis* 68(1): 113–127.
- Molinini D, Pataut F, Sereni A eds (2016) Special Issue on Indispensability and Explanation, *Synthese* 193(2): 317–558.
- Pataut F (2016) Comments on ‘Parsimony and inference to the best mathematical explanation’. In Molinini, eds Pataut and Sereni, 351–363.
- Putnam H ([1967] 1979) The Thesis that Mathematics is Logic. *Mathematics, Matter and Method : Philosophical Papers*, vol. 1: 12–42, Cambridge UP, 2nd ed., Cambridge, Mass.
- Putnam H ([1971] 1979) Philosophy of Logic, *Mathematics, Matter and Method : Philosophical Papers*, vol. 1, 323–357. Cambridge UP, 2nd ed., Cambridge, Mass.
- Van Orman Quine W ([1968] 1969) Existence and Quantification. *Ontological Relativity and Other Essays*, Columbia UP, New York 91–113.
- Resnik M (1997) *Mathematics as a Science of Patterns*. Oxford: Clarendon Press.
- Rizza D (2011) Magicicada, mathematical explanation and mathematical realism. *Erkenntnis* 74: 101–114.
- Shapiro S (1997) *Philosophy of Mathematics: Structure and Ontology*. Oxford: Oxford UP.
- Steiner M (Mark) (1978) Mathematics, Explanation, and Scientific Knowledge. *Nous* 12: 17–28.
- Tallant (Jonathan) (2013) Optimus Prime: Paraphrasing Prime Number Talk. *Synthese* 190: 2065–2083.

INFINITE LIFESPANS, TERRAFORMING PLANETS, AND INTERGENERATIONAL JUSTICE

Adelin-Costin Dumitru

National University of Political Studies and Public Administration,
Bucharest, Romania
e-mails: adelin.dumitru1@gmail.com

Abstract:

When it comes to specifying the moral duties we bear towards future generations, most political philosophers position themselves on what could be regarded as a safe ground. A variant of the Lockean proviso is commonplace in the literature on intergenerational justice, taking the form of an obligation to bestow upon future people a minimum of goods necessary for reaching a certain threshold of well-being (Meyer, 2017). Furthermore, even this minimum is often frowned upon, given the non-identity problem and the challenges this presents to the topic of justice between generations. Additional issues are raised at the level of non-ideal theory, the most significant being the problem of non-compliance (Gosseries and Meyer, 2009).

In this paper I intend to probe the limits of “practical political possibility” (Rawls 1999), by inquiring whether embracing the sufficiency view (Frankfurt, 1987; Crisp, 2003; Benbaji, 2005) as a distributive pattern and capabilities as a metric can lead to more burdensome obligations for present generations. More specifically, I try to show that we have a duty to invest in research that aims at prolonging the lifespan of humans (the idea can already be found in the sufficientarian literature, for instance in Farrelly, 2007). Moreover, given the Earth’s limited resources, we ought to encourage the terraforming of other planets in order to make them inhabitable for (future) people.

I argue that these two seemingly far-fetched projects are in fact worthwhile goals to pursue on the one hand, and moral obligations on the other hand. Nonetheless, they are not the only ones we ought to take on; for instance, we must simultaneously pursue them and try to improve the prospects of those who fall under a sufficiency threshold here and now. That is, specifying

these (*prima facie*) duties towards future generations is connected with stronger obligations towards the current generation.

Towards the end of the paper I engage in a discussion regarding the role of the feasibility constraint in a theory of justice, as rationales pertaining to feasibility are perhaps going to be the most recurrent criticisms raised against my proposal. To that end, I defend limitarian policies, which aim at setting an upper limit to how much money individuals are allowed to possess (Robeyns, 2017; Volacu and Dumitru, 2019).

Key words: Intergenerational justice; limitarianism; non-identity problem; sufficiency.

Introduction

For a long time, political philosophy has had a rather limited scope. It is implicit that by “justice” we mean “societal” or “state” justice, and we have to add the “global” adjective if we want to discuss our duties to members of other societies. The same is true when it comes to the temporal dimension of the scope of justice; there seems to be no need to define justice as “intragenerational”: This is the default. Instead, we have to ensure that we add “intergenerational” when we engage in debates regarding distributive processes between generations. The result is that there seems to be a prioritization bias: First, we solve the shortcomings in our society, and then we attempt to provide reparations for historical injustice or to save enough resources for future generations. In this article I intend to challenge this

priority order: The obligations we have towards our descendants are as strong as the obligations we have towards our contemporaries. More specifically, I try to show that we have a moral duty to invest resources in research that aims at prolonging the lifespan of humans. Furthermore, given the paucity of resources on this planet, we ought to encourage the process of terraforming other planets in order to make them inhabitable by future people.

At a cursory glance, this might sound far-fetched. Thought experiments are commonplace in the literature on justice, but to claim that such projects are a requirement of justice seems to go a step beyond that. Nonetheless, I intend to probe the limits of “practical political possibility” (Rawls, 1999) by starting from some plausible and widely acceptable premises. The first of these is that individuals have a right to procreate, which imposes positive and negative obligations on potential parents and on society at large (Gheaus, 2016). The second premise is that one of those obligations is the need to ensure that potential parents can reach their lifelong goals which lead to a flourishing life. One of these lifelong goals is the desire to see one’s child, in turn, lead a fulfilling, flourishing life. A prerequisite for this is prolonging the human lifespan. The problem with this is that it can easily lead to an overcrowded world. Thus, we end up with two horns of a dilemma: On the one hand, people have a right to a flourishing life; on the other, prolonging life can lead to an overcrowded world. In order to find a solution, I shall argue that another ambitious project – the conquest of space – is required.

At this point a question that may arise is why I must add sufficientarianism to the equation. Would not a different distributive pattern better justify such ambitious projects? Utilitarianism, for instance, would have fewer problems explaining why conquering the stars or defeating the ailments brought about by old age are worthwhile goals that we as a society should pursue. Nonetheless, I do not find utilitarianism a plausible view of distributive justice. Are sufficientarians not meant to have their cake and eat it too? Or, more explicitly put, do sufficientarians have to settle on solving the problems manifesting here and now? I believe that sufficientarianism is not only a compelling moral principle, but also capable of accommodating our intuition that prolonging life and exploring outer space are morally appealing ends. The explanation requires, unfortunately, a further controversial claim – that capabilities are the metric of distributive justice. If we endorse a

capabilitarian version of sufficiency, I posit that we can explain why delaying aging and making other planets inhabitable contribute to the realization of some vital capabilities, the achievement of which would be impossible in the absence of these projects.

Given that the main aim of the paper is to show that sponsoring these projects is consistent with obligations of justice, one of the first issues that has to be clarified is the extent to which satisfying them i) interferes with other worthwhile objectives (such as eliminating global poverty, or providing adequate housing to all) and ii) does not represent a violation of the criterion of moral accessibility.

In his analysis of the transition from non-ideal to ideal theory, Allen Buchanan (2004) puts forth several reasonable criteria, one of which is the idea of moral accessibility, *i.e.* that the transition from the present state of affairs to the postulated ideal does not entail too much of a burden for certain individuals. Given the asymmetrical nature of intergenerational relations, this worry becomes even more prominent in the present discussion. In order to show why this will not be the case, I engage in the last section of the paper in a brief discussion regarding limitarianism, the view according to which individuals are not morally allowed to keep more money than they need for flourishing (Robeyns, 2017; Volacu & Dumitru, 2019). Robeyns presents two main arguments for her instrumental version of limitarianism: the unmet urgent needs argument and the democratic argument. In this paper I extend the list with a third objective that has to be accomplished before people could be allowed to keep surplus money (*i.e.*, money that is not necessary for their flourishing): achieving infinite voluntary lifespan (which, taking into account the state of current and foreseeable medicinal progress, is an insatiable principle, that is, it is always theoretically possible to satisfy it to a higher degree (Raz, 1986)).¹

Restating the Sufficiency View

The purpose of the paper is not to defend the sufficiency view against opposite distributive views. I assume that sufficientarianism is the pattern that we should employ in distributive justice and that capabilities represent the most adequate metric. I will engage in discussions concerning the numerous objections that have been raised against the sufficiency view so far (for a brief inventory see Fourie

¹ For a recent value-satiability version of sufficientarianism see Nielsen (2019).

(2017) or Nielsen (2019)), *only* to the extent that one argument or another can actually help sufficientarianism escape them. Nonetheless, I take into account several powerful charges aimed at the sufficiency view in proposing a very specific account. This is why perhaps the most common criticisms that could be adduced are *internal* ones from within the sufficientarian camp. With this *caveat* in mind, I do believe that the particular account that I put forth is rigorous enough to be assertable by many authors who position themselves on sufficientarian grounds.

I find sufficientarianism to be a subject-centered, instead of a reciprocity-based, conception of justice, where “we focus on the needs and non-strategic capacities of the person and upon our collective ability to satisfy those needs and develop those capacities, not upon the bargaining power of the individual” (Buchanan, 1990, p. 233). Furthermore, I consider the most plausible version of the sufficiency view to be a multilevel one, in which we have different thresholds, each with its own weighing rule. In fact, I believe that we could identify something similar to a Lakatosian hard core of sufficientarianism, comprising the intrinsic value of sufficiency and the notion of a threshold. How they are operationalized, whether this is all justice requires, is something that belongs to the auxiliary assumptions (Lakatos, 1989). Sufficientarianism thus represents a research program, with the different versions and applications proposed in the literature representing the positive heuristic (stating what paths of inquiry to pursue). The purpose of the negative heuristic (what paths of inquiry to avoid), on the other hand, is meant to insulate the hard core of this research program from direct criticism. Under this conceptualization, sufficientarianism replaced egalitarianism, by trying to show that it better accounts for several intuitions we might have that are usually employed to defend that distributive principle. For instance, Frankfurt argues that some people mistakenly consider themselves egalitarians, when in fact their intuitions reflect a sufficientarian reasoning (1987, p. 32). This is similar to Lakatos’ explanation of the process of evolving research programs – changes occurring whenever “a rival research program begins to explain the previous success of its rival and supersedes it by a further display of explanatory power” (1989, pp. 69-72).

Carina Fourie’s recent primer reflects the intuition of considering sufficientarian views as comprising a hard core of concepts, although in her version, there are more elements that “have to be specified by

a comprehensive account of sufficiency”: currency, positive and positioning claims, justifications for sufficiency principles, weighing rules, sufficiency thresholds, scope of sufficiency (2017, p. 17). Nonetheless, several of these are not unique to sufficientarianism, but are encapsulated by other views on distributive justice as well. Weighing rules, scope, currency, and justifications are all requirements for comprehensive accounts of any distributive justice view. Stripping these away from the sufficiency view leaves what could be considered the backbone of sufficientarianism: positive and positioning claims and the notion of sufficiency thresholds. I will not have much to say regarding the first.

The positive claims incorporate what Benbaji (2005) labels “the basic intuition,” *i.e.* the consideration that at least some priority should be given to benefiting people who are absolutely badly off. More usually, they take the form of what Casal (2007) calls “the positive thesis,” *i.e.* that “it is important that people live above a certain threshold, free from deprivation.” Positioning claims can either take a strong or a weak form. The best known strong positioning claim is the negative thesis of sufficiency, which denies the relevance of additional distributive requirements above the threshold (Casal, 2007, pp. 297-8). A weak positioning claim is Shields’ shift thesis, which holds that “once people have secured enough, there is a discontinuity in the rate of change of the marginal weight of our reasons to benefit them further” (Shields, 2012, p. 108). That is, our reasons to benefit someone change once that person has reached a certain threshold of well-being (Shields, 2012, p. 107). The shift thesis is compatible with any number of operationalizations, in that above the threshold we could apply desertist or even egalitarian distributive criteria (Shields, 2017, p. 87). For the purpose of the paper I shall remain agnostic regarding what the shift thesis requires above the threshold. As far as I am concerned, this is something that should be determined by public debates, when we can be more open to what real people would actually endorse, once we have ensured that justice has been realized.

The concept of the sufficiency threshold, on the other hand, requires more attention, especially considering that I believe that the arguments I put forward are compatible only with a specific account of how these thresholds should be set. More specifically, I endorse a multiple threshold sufficiency view, where both thresholds are measured in capabilities. Employing a distinction advanced by Robeyns

(2000), the lower threshold should be measured in basic capabilities, while the upper threshold in non-basic capabilities. Whereas basic capabilities refer to the freedom to do basic things (such as having access to adequate shelter, adequate food, adequate healthcare), non-basic capabilities are more complex, and include things that allow individuals to have a flourishing life. They include aspects related to fundamental interests of human agents that are “needed to shape, reshape and pursue valuable ends” and aspects regarding the fundamental interests of a social being (Axelsen & Nielsen, 2017, pp. 113-4). Individuals situated below the minimal threshold should be given absolute priority, and between thresholds, strong-responsibility catering prioritarian considerations apply. That is, large benefits to a large number of people would trump a small benefit to a small number of people, even though the latter could raise someone above the superior threshold. Above the maximal threshold, I propose that sufficientarians embrace the shift thesis, though, as mentioned beforehand, I am not concerned here with specifying the exact way our reasons to benefit people should change. However, in non-ideal circumstances, I tentatively advance the view that, no matter how advantages are distributed above the maximal threshold, we should be preoccupied once we get to the realm of resources that are irrelevant from the standpoint of justice, *i.e.* that are superfluous for any kind of flourishing life that might reasonably be conceived. That is, whereas sufficiency represents a core normative principle, in non-ideal circumstances we need to set upper limits to the amount of resources people can reasonably hold, and this is where limitarianism (seen as an institutional scheme that can help us achieve the goals of sufficientarian justice) will be useful. I will postpone this discussion until the last section of the paper.

Prolonging Life as a Requirement of Justice

One of most influential early sufficientarians, Roger Crisp, mentions that “eighty years of high-quality life on this planet is enough, and plausibly more than enough, for any being” (2003, p. 762). This statement contains two assumptions that I would like to reject: that eighty years of life on *this* planet (as opposed to another planet) are what matter and that *eighty* years of life suffice for the realization of justice. Crisp’s position is that the compassion of an impartial spectator is appropriate “up to the point at which someone has a level of welfare such that

she can live a life which is sufficiently good.” Supposedly, once someone reaches the age of 80, the impartial spectator can no longer be compassionate concerning that person’s well-being. It is not surprising that one of the most recurrent critiques raised against sufficientarianism – the arbitrariness objection – is still a powerful one. As Axelsen and Nielsen point out, “this seems to imply that, according to Crisp, all of humanity could be allowed to die at seventy something for the sake of any nontrivial benefit to somebody just below the threshold” (2017, p. 105). I will leave aside for the moment the fact that Crisp’s account of sufficientarianism is further aggravated by giving absolute priority to the worst-off situated below the threshold (a problem also signaled by Casal (2007, pp. 298-9)), and focus only on the arbitrariness of the lifespan mentioned by Crisp.

In a groundbreaking study, Caspari and Lee (2004, p. 10,899) show that “longevity came late in human evolution and was a fundamental demographic component tied to population expansions and related behavioral innovations associated with modern humans.” Finch (2010) shows that, although there are no statistical data showcasing mortality rates or causes of death, in Antiquity the expected lifespan at birth ranged from 20 to 36 years. The increase of human lifespan is not “a biological attribute of a modern human species, but a culturally driven demographic change” (Caspari & Lee, 2006, p. 512). Data collected by Roser (2019) gives a global average life expectancy of 32 years in 1800, 48 in 1950 and 70 in 2012. In addition, Hughes and Hekimi (2017) argue that in principle there is no plateau to the maximum human lifespan. Irrespective of the accuracy of such predictions, based as they are on a limited dataset (the analyses are based on individuals living over 100 years, so the sample is rather small), what cannot be rejected is the fact that “human lifespan has been increasing steadily over the course of human evolution, leaping ahead by 30 years in the past century in developed countries” (Morris, 2005, p. 1,285). There is a wide array of explanations for this evolution, but most of them converge on the shared finding that longer lives are possible. Either by using hormones or protein-based drugs or by using embryonic stem cells to replace deteriorating organs and tissues, scientists have conceptualized and begun testing different solutions for increasing the average lifespan of the human being (Chapman, 2004, pp. 479–486).

What implications does this research have for sufficientarianism? I argue that the sufficientarian threshold should be conceived dynamically, at least when we adopt a capabilities metric. The concept, similar to Gilabert's notion of dynamic duties (2009), states that the threshold should be responsive to societal and technological changes. Furthermore, it should not take those for granted, but improve upon them when this would increase the justice of a society. Returning to Crisp's "80 years of good life," this would suffice under a static threshold. Nonetheless, static thresholds are subject to the arbitrariness objection (Arneson, 2002; 2006). For Gilabert and Lawford Smith (2012), dynamic duties are those that "do not focus merely on what can be done in given circumstances, but also on how to change circumstances so that new things can be done." Similarly, a dynamic conceptualization of the threshold considers that the length (and quality) of life are not set in stone, but subject to improvement. Given that the current average lifespan is not fixed by nature (Partridge & Hall, 2007), we should not take it as a given, especially if increasing it could lead more effectively to the realization of some key capabilities.

The question that should be answered at this point, however, is *whose* capabilities we are taking into account. Most probably, the technology that can allow us to live longer lives will not be available to our generation, and not even to the next one. The beneficiaries will be those from future generations, so the new problem that arises is one of prioritizing future generations versus current generations. However, future generations are in an asymmetric relation with us (cannot impact the past, and thus hold no responsibility towards us) and are *caeteris paribus* approaching an infinite number. Thus, even if we applied a 'discount rate' when it comes to benefits reserved for future generations (*i.e.* benefiting them would matter less than benefiting contemporaries), their sheer numbers and the strong priority for those between thresholds mentioned above means that we ought to take measures ensuring that they are raised above the maximal threshold (as long as no one falls under the minimal threshold in the present generation). The fact that the account proposed here tries to ensure that no one will fall under the minimal threshold means that it will fare better than some other versions of sufficientarianism or prioritarianism, which are number-sensitive and do not include such a requirement (Meyer & Roser, 2009, p. 235).

Add to this the fact that increasing average lifespan has sometimes been criticized as being "a frivolous investment compared to imperatives of providing basic medical services to everyone in our own country and ameliorating basic health needs elsewhere in the world" (Chapman, 2004, p. 488). Other critics point out that "there is no duty for individuals or society to fund research into developing life-extension interventions," although "individuals should be free to do so if they wish" (De Grey, 2005; Partridge & Hall, 2007, p. 891). As Farrelly shows, "proponents of the project of delaying, or even eliminating aging, must show that support for it is a requirement, not a violation, of the demands of justice" (2007, p. 514). I theorize that a lot of the work required for this can be done by adopting a capabilitarian metric. More precisely, I intend to show how increasing lifespan is necessary for the achievement of some non-basic capabilities which are necessary for defining an individual's life as flourishing. Starting from a set of basic, less controversial premises, I want to argue for a more contentious conclusion, that investing in projects that aim at prolonging average lifespan becomes a duty of achieving sufficientarian justice.

Although there are many papers discussing the relations between the capability approach and old age (Tsuchiya, 2000; Lloyd-Sherlock, 2002; Anand, 2005; Nussbaum, 2006; Coeckelbergh, 2012), there is a surprising dearth of analyses concerning the capabilities that can be primarily or even uniquely attained as people grow older. Lloyd-Sherlock mentions, for instance, that the capability approach fits well with the "active ageing school of thought, which argues that the focus should be on policies which optimize older people's capacity to contribute to their own quality of life and that of others" (2002, p. 1,665). Furthermore, a great part of the literature focuses on conceiving of old age as a disability. Nussbaum, for instance, includes elderly people among the categories which require caregiving, such as children or disabled people (2006, p. 61). This could be a consequence of the fact that bioethics, or discussions touching upon bioethics, are ill equipped to discuss the concept of increasing lifespans (Trotter, 2004). For instance, if increasing lifespan were possible with the help of stem cells, this would not lead to a "prolonged senescence," but would increase the period of time in which the human body could perform its functions. Furthermore, stem cell treatments are known for their curative and therapeutic properties (Shweta, Elmore & Gewirtz, 2016;

Watt & Hogan, 2000): With their application, reaching old age might not have to entail a period of suffering, as it does in many cases today. As such, reaching old age could be important for the realization of a number of key capabilities.

In Ram-Tiklin's (2017) sufficiency of basic human functional capabilities account, justice requires that health institutions guarantee that individuals get enough capabilities that they can live dignified lives and implement their life plans. Axelsen and Nielsen (2017) consider that for someone to lead a flourishing life she needs, *inter alia*, to fare adequately well in regard to aspects concerning the fundamental interests of a human agent that are "needed to shape, reshape, and pursue valuable ends." The stipulation that I make is that many of these valuable ends and aforementioned life plans require making sure that individuals reach old age.

Furthermore, I posit that one central non-basic capability that could be reached only by prolonging lifespan is the ability to see your children's lifelong goals be reached. This hypothesis rests on the premise that an individual's flourishing is affected by the opportunity to be a parent (Brighouse & Swift, 2006); otherwise, it would not matter from the standpoint of justice. In turn, this involves the existence of a "right to be a parent" (Gheaus, 2012). It is towards this justification for increasing average lifespan that I turn to in the next section.

Defending a Non-basic Capability to See your Children's Goals be Reached

According to Brighouse and Swift, the parent-child relationship is one that parents have an interest in, as it "enables them to exercise and develop capacities the development and exercise of which are, for many, crucial to their living fully flourishing lives" (2006, p. 95). Some of these capacities refer to someone's development as a person, others to the satisfaction derived from overseeing and ensuring her child's "cognitive, emotional, physical and moral development" (2006, pp. 94-5). For Brighouse and Swift, being a parent "makes a crucial contribution to the flourishing of the adult" (2006, p. 100). It is important to mention that in Brighouse and Swift's account what has normative force is not the special relation between a biological parent and a child, but the practice of parenting (2014, p. 27). Overall (2014) is also adamant that, from the standpoint of the benefits it brings to parents, procreation is not superior to adoption. She mentions that "pro-

creation is more costly to the planet in terms of resource depletion and waste generation," and, as such, it would be less defensible than adoption (Overall, 2014, p. 106). Furthermore, it must also be mentioned that this opportunity to flourish from the goods of parenting are contingent upon one's fulfillment of duties towards the child (for a comprehensive list of these duties see Brake (2016)).

According to Gheaus (2016), one of the responsibilities society has towards future generations is to ensure that "children are given environmental conditions in which they will be able to parent adequately. These include sufficient environmental resources for their prospective children to have lives not merely worth living, but adequate" (2016, p. 498). Furthermore, she mentions that, given the fundamental interest in rearing children that adults have, this also imposes further negative and positive obligations. Among the negative obligations is the one to not interfere with the right to be a parent, whereas the positive obligations refer to measures that can help prospective parents or people who have already become parents (such as endorsing, or even sponsoring, assisted reproduction or entitling parents to active support) (Gheaus, 2016, p. 487).

It is at this point that I posit a non-basic capability referring to reaching lifelong goals – one of these lifelong goals being to see your children's lifelong goals be reached. Ensuring that parents live long enough to see their children's lifelong goals realized would thus be important from the standpoint of justice, as otherwise they would not be able to be raised above the threshold. Thus, we have at least *pro tanto* reasons to benefit them in a way that increases the probability that they will be able to live long enough for this. The structure of the argument could be summed up in the form of three reasonable premises and a controversial conclusion emerging from these:

- P1: An individual's flourishing is *pro tanto* dependent on the opportunity to be a parent.
- P2: An individual's flourishing involves reaching lifelong goals, one of these being to see her children's lifelong goals be reached.
- P3: We have positive and negative obligations towards prospective parents.

∴ We ought to increase the lifespan
of human beings.

The viability of the argument hinges on showing that P2 is factually correct. Can we correctly posit such a capability? Empirical studies seem to support this claim. At first, we must renounce a more rational choice approach, according to which the parents' flourishing might depend instrumentally on the realization of their children's life goals. According to the findings of Silverstein et al. (2002), "parents interested in optimizing their chances of receiving old-age support would do well to invest in their children well before they reach old age" (p. S12). Children with better life prospects would be able to provide more efficiently for their aging parents. However, increasing average lifespan would (purportedly) bring several important changes to this "implicit social contract that ensures long-term reciprocity," (Silverstein et al., 2002, p. S12) by delaying the age at which considerations of caring for one's parents would kick in (so this would simply be a form of procrastination, as the main problem would not be circumvented). On the other hand, Nilsen, Brannen and Vogt (2015) mention that "parenthood today is ever more demanding as parents are increasingly required to take greater responsibility and are made morally accountable for their children's welfare and progress (or lack of it). Hitlin and Kirkpatrick (2015) develop a model of "future orientation," more specifically, "expectations about how life will turn out," in order to understand individual agency (p. 1,461). Although parental expectations concerning their children's future are largely dependent on their social and economic contexts, Irwin and Elley (2013) show that all parents are to some degree preoccupied with the future occupations of their children.

Furthermore, such a capability can be also defended on normative grounds. First of all, given the importance of parent-child relations, it seems odd to consider that the significance of a child's well-being withers away once the child grows older. Although the quality of the relation and its asymmetrical nature might change, parents' well-being continues to be influenced to a high degree by their children's well-being. An account of parental flourishing which does not include a capability to see the lifelong goals of one's children realized seems to miss a vital component, given how much focus there is in the sufficientarian literature on life plans (Ram-Tiklin, 2017; Axelsen & Nielsen, 2017).

Assuming that one concedes that this capability is an important one, and accepts the other two (less contentious) premises, the conclusion we reach is

that we ought to increase the lifespan of human beings, by investing in scientific research or allocating funds for projects meant to achieve such a result. Nonetheless, this leaves us with two horns of a dilemma: On the one hand, people have a right to a flourishing life, which requires extending their lifespan; on the other, prolonging life can lead to an overcrowded world, which seems to interfere with our capacity to benefit even the people who are under the minimal threshold (a similar argument is put forward by Partridge and Hall (2007)). It would be normatively undesirable to advance a theory which will leave future people under duress. It would be the equivalent of a time bomb that has no negative effects up to a point, but which will explode sometime in the far future (Mulgan, 2016). The solution that I envisage lies in funding projects that aim at conquering space and making the outer planets of our solar system inhabitable for future generations.

Terraforming Planets

In this section I want to defend the moral duty to invest in projects that aim at terraforming other planets by showing how it is necessary for carrying out the duty to prolong lifespan without violating the rights of any agents. In doing so, I shall employ what is known as a linkage argument, developed by Henry Shue (1996) in order to justify a controversial claim "by showing that it is necessary to the effective implementation of another that is already uncontroversial and accepted" (Nickel, 2009, p. 441).

Terraforming refers to the process of creating "uncontained biospheres on other planets, which emulate the properties of Earth [...] a life-support system as beneficent, reliable, unobtrusive and long lasting as our planet's history shows is possible and to allow life not just to settle in space, but also to persist" (Fogg, 1995, p. 42). The technical aspects of terraforming are not important for the purposes of this paper. Fogg's research (1993; 1995) explores the science behind the process of terraforming Mars or Venus, which would be extremely long-term, costly, and – most likely – impossible with technologies that exist at the moment. For instance, in the case of Venus, where CO₂ exists in much larger quantities than are required for sustaining life, the process of "either chemically reducing or sequestering CO₂ within the body of Venus as carbonate minerals would take some thousands of years" (Fogg, 1995, p. 58). Terraforming would require modifying the temperature of the planet, by either

“manufacturing a greenhouse effect to heat it up or by placing large sheets of molecular thickness mirroring in orbit around the planet to cool it down”; afterwards, genetically modified plants and animals would have to be introduced (Sparrow, 1999, p. 228). The length of these processes would most likely change should intermediate space objectives be accomplished. For instance, O’Neill (1975) thought that it would be possible to construct manufacturing facilities in the vicinity of the Lagrange libration points of the earth-moon system, which would increase the energy output needed for terraforming other planets.

Ethical aspects and implications of terraforming and space colonization have been advanced somewhat sparsely so far. McKay (1982) defended the terraformation of Mars for two reasons: 1) it would teach us how to “reclaim vast areas of Earth which are currently uninhabitable”; 2) it would represent a safety net should Earth be faced with a cataclysm. Anker (2005) argued that the colonization of space could be propitious for understanding and better protecting life on Earth. Schwartz (2013) mentions that terraforming is a moral requirement, to the extent that it can lead to better protection of Earth’s environment, by “providing valuable lessons for the intelligent management of Earth’s biosphere” (p. 27). Sparrow (1999) criticizes the potential terraforming of Mars due to the potential consequence of “drastically altering the character of a whole planet, a unique environment which includes complex inorganic systems and possesses many features of striking natural beauty” (p. 229). Stoner (2017) embraces a similar argument, claiming that by “fundamentally altering the Martian environment forever, colonizing Mars is morally wrong” (p. 350).

On the other hand, no discussion exists – to my knowledge – about terraforming planets as seen through a distributive justice lens. Framing it as a question of distributive justice allows me to eschew the discussions regarding the intrinsic value of colonizing space and to make the more limited claim that this is a project that should be tackled by human beings only to the extent that it would help raise a number of individuals over the threshold, now or in the future. Taking into account the possibility that a cataclysm might negatively impact the well-being of future generations, having a safety net in the form of a different inhabitable planet would be vital for ensuring that we minimize the number of individuals who fall under the minimal threshold. To the extent that such a scenario is a plausible one, investing a

part of the public budget in projects that aim at colonizing space seems like a small sacrifice, and it would continue to be a small sacrifice – even if we discounted the benefits received by future people and reduced the probability of such an event to close to (but not) zero – since the number of potentially affected people would theoretically be infinite.

Furthermore, given the fact that it seems problematic to impose limits on the number of children parents can have (for simplicity, let’s assume that all the children would have sufficiently good lives), and given that improving average lifespan seems to be a precondition for attaining the non-basic capability of reaching the lifelong goal of seeing your children’s lifelong goals be realized, the potential increase of the population would be unsustainable. Since prolonging life and terraforming planets would both be lengthy projects, we can stipulate that small budgetary allocations for such projects would eventually lead to their (almost) simultaneous realization. It is undeniably true that increasing the population of Earth would be unsustainable in the absence of new resources added to the equation. Nonetheless, such resources would not be infinite (there are a limited number of planets that can foreseeably be made inhabitable). Thus, even if we managed to inhabit not one, but two or more planets, we would still be subject to what Rawls (1971), following Hume, calls the “circumstances of justice,” and thus, we would still face a distributive justice problem (as infinite resources would not represent a distributive justice problem, but rather an allocation one.)²

In the following section I will defend the arguments made so far from the feasibility objection and from what I shall call the *violation of present people’s rights objection*, which states that allocating public money for such projects is unfair given the present state of affairs. To that end, I defend limitarian policies, which aim at setting an upper limit to how much money individuals are allowed to keep (Robeyns 2017; Volacu & Dumitru 2019).

² Rawls (1971, p. 88) distinguishes between distributive justice (which “cannot be judged in isolation from the system of which it is the outcome or from what individuals have done in good faith in the light of established expectations” and allocative justice, which applies “when a given collection of goods is to be divided among definite individuals with known desires and needs.”

Limitarianism and Moral Accessibility

According to Volacu (2017), much confusion in political philosophy can be traced back to conflating the notions of feasibility and fact-sensitivity. Whereas feasibility refers to the outputs of a theory and touches upon the possibility of implementing a theoretical account, fact-sensitivity refers to the inputs of a theory, when one must choose whether to introduce into the theory “descriptive assumptions, which make some form of empirical claims about the state of the world” (Volacu, 2017, p. 3). In this section I am concerned with feasibility in a narrow sense – the extent to which we can actually gather the resources necessary for funding the twin projects of increasing average lifespan and terraforming other planets, and whether or not this can be achieved without violating existing people’s rights. Or, as Allen Buchanan (2004, pp. 61-2) puts it, we should be concerned that our transition from non-ideal state of affairs to ideal state of affairs satisfies the criterion of moral accessibility, *i.e.* that no unreasonable moral costs are placed on an individual.

Limitarianism enters into the equation when we start talking about the institutional implementation of a core normative principle, such as the version of sufficientarianism that I defended in the previous section (Gilbert & Lawford-Smith, 2012). According to Gilbert and Lawford Smith, a state of affairs “may be feasible even if no agents have the current ability to bring it about.” Although a certain project might be infeasible now, it need not be so in the future. Moreover, we should place ourselves in the situation in which we could bring about that desired state of affairs. A similar perspective is embraced by Heroz (2012, p. 283), who also considers that there might be certain “dynamic duties to acquire knowledge about how to correct certain injustices.” Furthermore, as Gheaus (2013, p. 448) mentions, “a willingness to apply the label unjust to regrettable situations that we cannot fix is going to enhance the action-guiding potential of a conception of justice, by providing an aspirational ideal.” Given that the projects of terraforming and increasing average lifespan are not impossible to accomplish – only terribly long-term commitments – this worry is somehow alleviated.

The claim that I make is that limitarianism is an institutional scheme (or, more accurately, denotes a family of institutional schemes) that can ensure that no agent rights are violated in the process of raising resources for our supposedly worthwhile goals, and that can ensure that enough resources are raised for

said projects. Limitarianism was first proposed by Robeyns (2017)³ as a way to mitigate a shortcoming of most distributive justice theories – the fact that they are largely ignorant of the duty-bearers of redistributive processes. Limitarianism seeks to identify the level at which agents have “more resources than are needed to fully flourish in life” (Robeyns, 2017, p. 1) and argues that it is morally problematic to own more than that amount. As Volacu and Dumitru (2019) argue, limitarianism is a partial view of distributive justice since “it does not necessarily address what justice requires of those below the riches line, though it has something to say about what justice requires for the worse off.” As such, limitarianism is not a fundamental view of justice, “but rather a particular view on how to prioritize duty-bearers” (Volacu & Dumitru, 2019, p. 263). In the language employed in this paper, limitarianism represents an institutional scheme that aims at instantiating the fundamental view of distributive justice, sufficientarianism.

In line with Robeyns’ and Volacu and Dumitru’s interpretations, limitarianism should be seen not necessarily as a full-fledged view of distributive justice, but rather as an instrument that helps us achieve the ends of sufficientarianism. In her initial proposal, Robeyns mentions two possible grounds for limitarianism. Given its instrumental character, limitarianism does not argue that being rich is intrinsically wrong, only that having surplus money (*i.e.* money that is not needed for having a flourishing life) is problematic, given the fact that surplus money could be employed in politics and disrupt the value of political equality, or, alternatively, that it could be used to alleviate unmet urgent needs. These two values give rise to two distinct forms of limitarianism, labeled by Volacu and Dumitru weak and strong limitarianism.

Weak limitarianism is based on the democratic argument, which holds that usual institutional constraints are unable to stop the negative phenomena associated with money in politics – buying votes, gatekeeping, influencing opinion. Such phenomena disrupt political equality, and strong limitarianism therefore requires a taxation rate of 100 percent in order to safeguard this value. By contrast, weak limitarianism acknowledges that “such a fiscal policy would disincentivize the wealthiest individuals in society from being maximally productive and that there would be less income to redistribute for the

³ See also Robeyns (2019) for a defense of limitarianism in the more concrete case of climate change.

meeting of urgent needs” (Volacu & Dumitru, 2019, p. 255). As Robeyns has recently mentioned (2019, p. 12), there is “a potential tension among the different reasons for limitarianism.” Her conclusion is that “the incentives objection should prompt us to adapt

limitarianism as applied to fiscal policies in line with optimal taxation design, to the extent that we weight the value of meeting the unmet urgent needs higher than the effects of surplus money on the undermining of political equality” (2019, p. 12).

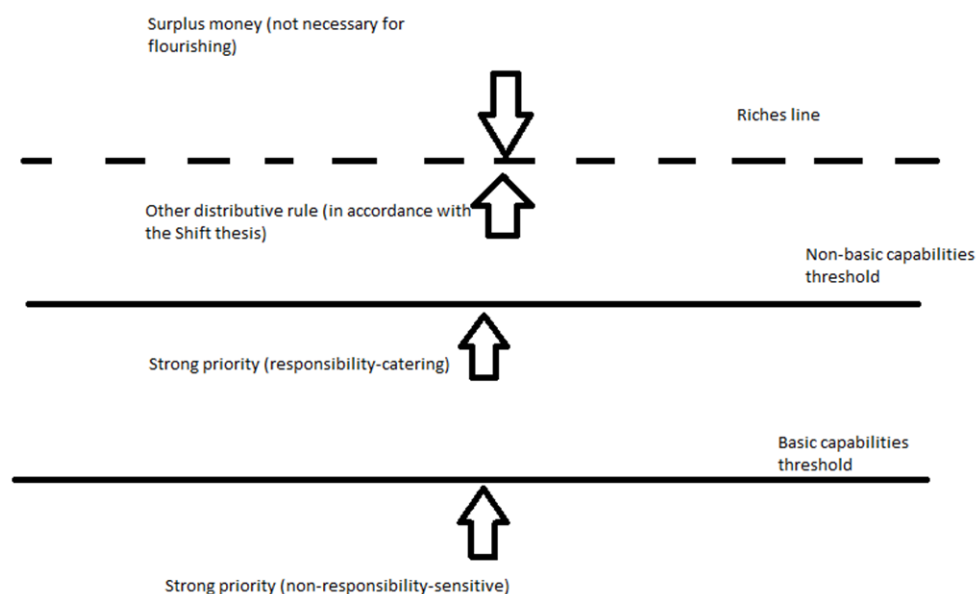


Figure 1: The distributive configuration

I am largely agnostic in the present paper with regard to choosing between weak or strong limitarianism. In fact, I want to propose a third reason for accepting limitarianism: achieving the projects of extending average lifespan and terraforming other planets. Choosing limitarian policies would represent the main way in which we increase the probability of achieving the aforementioned objectives, by gaining access to more public funds for sponsoring research that aims to extend lifespan or to terraform other planets. Simultaneously, given that limitarian policies would impose a high rate of taxes (the specific rate is not a question that should be settled right now, as we would need more empirical research in order to determine this matter), only surplus money would be used for these projects, and thus, no moral violation of an agent’s right would occur. (Surplus money is, *ex definitione*, that money which is not required for allowing an agent to have a flourishing life.) Furthermore, what constitutes surplus money for one person can represent the resources needed to achieve a basic or a non-basic capability for another person. Given the arguments made in the previous section for benefiting members of future genera-

tions, taxing the wealthiest individuals in a society in order to sponsor projects that are only going to help future people is not a violation of existing agents’ rights. Whereas we should still have a different, more lenient tax scheme in place in order to help bring existing individuals either to the basic capabilities threshold or to the non-basic capabilities threshold, the taxation regime imposed on the very rich ought to help members of future generations.

As such, we end up with a distributive arrangement in which we have three main thresholds. Two of these are sufficiency thresholds, which are of a different nature than the third one – they represent thresholds of justice, and are not contingent upon the satisfaction of an empirical condition. The third threshold is the limitarian threshold – the “riches line” as Robeyns calls it (2017; 2019) – which is contingent upon the realization of the reasons advanced here for limitarianism: increasing average lifespan and terraforming other planets. Although the implementation of limitarianism is dependent on the degree to which it satisfies these reasons, I posit that achieving infinite voluntary lifespan is an insatiable reason (*i.e.* it is always possible to satisfy it to

a higher extent), and as such, limitarianism would be a mainstay of the sufficientarian conceptual and action-guiding apparatus.

References

- Anand P (2005) Capabilities and health. *Journal of Medical Ethics* 31 (5): 299–303.
- Anker P (2005) The ecological colonization of space. *Environmental History* 10 (2): 239–268.
- Arneson R (2002) Why justice requires transfers to offset income and wealth inequalities. *Social Philosophy and Policy* 19 (1): 172–200.
- Arneson R (2006) Distributive justice and basic capability equality. In *Capabilities equality: basic issues and problems*, ed. Alexander Kaufman, 17–44. New York: Routledge.
- Axelsen D, Nielsen L (2017) Essentially enough. Elements of a plausible account of sufficientarianism. In *What is Enough? Sufficiency, Justice, and Health*, eds Carina Fourie and Annette Rid, 101–118. New York: Oxford University Press.
- Benbaji Y (2005) The doctrine of sufficiency: a defence. *Utilitas* 17 (3): 310–332.
- Brake E (2016) Parenthood and procreation. *Stanford Encyclopedia of Philosophy*
- Brighouse H, Swift A (2006) Parents' rights and the value of the family. *Ethics* 117 (1): 80–108.
- Brighouse H, Swift A (2014) The goods of parenting. In *Family-making. Contemporary ethical challenges*, eds Françoise Baylis and Carolyn McLeod, 11–29. New York: Oxford University Press.
- Buchanan A (1990) Justice as reciprocity versus subject-centered justice. *Philosophy & Public Affairs* 19 (3): 227–252.
- Buchanan A (2004) *Justice, legitimacy and self-determination. Moral foundations for international law*. New York: Oxford University Press
- Casal P (2007) Why sufficiency is not enough. *Ethics* 117 (2): 296–326.
- Caspari R, Lee S-H (2004) Older age becomes common late in human evolution. *Proceedings of the National Academy of Sciences of the United States of America* 101 (30): 10895–10900.
- Caspari R, Lee S-H (2006) Is human longevity a consequence of cultural change or modern biology?. *American Journal of Physical Anthropology* 129 (4): 512–517.
- Chapman A (2004) Ethical implications of prolonged lives. *Theology Today* 60:479–496
- Coeckelbergh M (2012) How I learned to love the robot: capabilities, information technologies, and elderly care. In *The Capability Approach, Technology and Design*, eds Ilse Oosterlaken and Jeroen van den Hoven, Dordrecht: Springer.
- De Grey A (2005) Life extension, human rights, and the rational refinement of repugnance. *Journal of Medical Ethics* 31: 659–663.
- Farrelly C (2007) Sufficiency, justice and the pursuit of health extension. *Rejuvenation Research* 10 (4): 513–520.
- Finch C (2010) Evolution of the human lifespan and diseases of aging: roles of infection, inflammation and nutrition. *Proceedings of the National Academy of Sciences of the United States of America* 107 (1): 1719–1724.
- Fogg M (1993) Terraforming: a review for environmentalists. *The Environmentalist* 13 (1): 7–17.
- Fogg M (1995) *Terraforming: engineering planetary environments*. Society of Automotive Engineers
- Fourie C (2017) The sufficiency view: a primer. In *What is Enough? Sufficiency, Justice, and Health*, eds Carina Fourie and Annette Rid, 11–30. New York: Oxford University Press.
- Gheaus A (2013) The feasibility constraint on the concept of jsuitce. *The Philosophical Quarterly* 63 (252): 445–464.
- Gheaus A (2016) The right to parent and duties concerning future generations. *The Journal of Political Philosophy* 24 (4): 487–508.
- Gilbert P (2009) The feasibility of basic socioeconomic human rights: a conceptual exploration. *The Philosophical Quarterly* 59 (237): 659–681.
- Gilbert P, Lawford-Smith H (2012) Political feasibility: a conceptual exploration. *Political Studies* 1–17.
- Herzog L (2012) Ideal and non-ideal theory and the problem of knowledge. *Journal of Applied Philosophy* 29(4).
- Hitlin S, Kirkpatrick M (2015) Reconceptualizing agency within the life course: the power of looking ahead. *American Journal of Sociology* 120 (5): 1429–1472.
- Hughes B G, Hekimi S (2017) Many possible maximum lifespan trajectories. *Nature* 546 (7660): E8–E9.
- Irwin S, Elley S (2013) Parents' hopes and expectations for their children's future occupations. *The Sociological Review* 61 (1): 111–130.

- Lakatos I (1989) *The methodology of scientific research programmes. Philosophical Papers I*, eds John Worrall and Gregory Currie, New York: Cambridge University Press.
- Lloyd-Sherlock P (2002) Nussbaum, capabilities and older people. *Journal of International Development* 14 (8): 1163–1173.
- McKay C (1982) On Terraforming Mars. *Extrapolation* 23 (4): 309–314.
- Meyer L (2016) Intergenerational justice. *The Stanford Encyclopedia of Philosophy*, available online at <http://plato.stanford.edu/archives/sum2016/entries/justice-intergenerational>.
- Meyer L, Gosseries A (2009) *Intergenerational justice*. New York: Oxford University Press
- Meyer L, Roser D (2009) Enough for the future. In *Intergenerational justice*, eds Lukas Meyer and Axel Gosseries, 219–249. New York: Oxford University Press.
- Morris B J (2005) A forkhead in the road to longevity: the molecular basis of lifespan becomes clearer. *Journal of Hypertension* 23 (7): 1285–1309.
- Mulgan T (2016) Intergenerational ethics. *The International Encyclopedia of Ethics*, ed. Hugh LaFollette, Wiley.
- Nickel J (2009) A defense of welfare rights as human rights. In *Contemporary Debates in Political Philosophy*, eds Thomas Christiano and John Christman, 437–457. Malden: Wiley-Blackwell.
- Nielsen L (2019) Sufficiency and satiable values. *Journal of Applied Philosophy*, online first, doi: 10.1111/japp.12364.
- Nilsen A, Brannen J, Vogt K (2015) Transitions to adulthood. An intergenerational lens. In *Handbuch Kindheits und Jugendsoziologie*, ed. A. Lange, Springer.
- Nussbaum M (2006) Capabilities as fundamental entitlements: Sen and social justice. In *Capabilities equality: basic issues and problems*, ed. Alexander Kaufman, 44–71. New York: Routledge.
- O'Neill G (1975) Space colonies and energy supply to the Earth. *Science* 190 (4218): 943–947.
- Overall C (2014) What is the value of procreation?. In *Family-making. Contemporary ethical challenges*, eds Francoise Baylis and Carolyn McLeod, 89–109. New York: Oxford University Press.
- Partridge B, Hall W (2007) The search for Methuselah: should we endeavor to increase the maximum lifespan?. *EMBO Reports* 8 (10): 888–891.
- Rawls J (1971) *A theory of justice*. Cambridge: The Belknap Press of Harvard University Press.
- Rawls J (1999) *The law of peoples*. Cambridge: Harvard University Press.
- Raz J (1986) *The morality of freedom*. New York: Oxford University Press.
- Robeyns I (2017) Having too much. In *NOMOS LVI: Wealth. Yearbook of the American Society for Political and Legal Philosophy*, eds J. Knight and M. Schwartzberg. New York: New York University Press.
- Robeyns I (2019) What, if anything, is wrong with extreme wealth?. *Journal of Human Development and Capabilities*, online first, doi: 10.1080/19452829.2019.1633734.
- Roser M (2019) Life expectancy, available online at: <https://ourworldindata.org/life-expectancy>.
- Schartwz J (2013) On the moral permissibility of terraforming. *Ethics and the Environment* 18 (2): 1–31.
- Shields L (2012) The prospects for sufficientarianism. *Utilitas* 24 (1): 101–117.
- Shields L (2017) Some questions (and answers) for sufficientarians. In *What is Enough? Sufficiency, Justice, and Health*, eds Carina Fourie and Annette Rid, 85–100. New York: Oxford University Press.
- Shue H (1996) *Basic rights*, 2nd edition. New Jersey: Princeton University Press
- Shweta C, Elmore L, Gewirtz D (2016) Is senescence reversible?. *Current Drug Targets* 17 (4): 460–467.
- Silverstein M, Conroy S, Wang H, Giarrusso R, Bengston V (2002) Reciprocity in parent-child relations over the adult life course. *Journal of Gerontology* 57b (1): S3–S13.
- Sparrow R (1999) The ethics of terraforming. *Environmental Ethics* 21: 227–245.
- Stoner I (2017) Humans should not colonize Mars. *Journal of the American Philosophical Association* 3 (3): 334–353.
- Trotter G (2004) Why bioethics is ill equipped to contribute to the debate about prolonging lifespans. *HEC Forum* 16 (3): 197–213.
- Tsuchiya A (2000) QALYs and ageism: philosophical theories and age weighting. *Economic Evaluation* 9 (1): 57–68.
- Volacu A (2017) Bridging ideal and non-ideal theory. *Political Studies*, online first, doi: 10.1177/0032321717730297.
- Volacu A, Dumitru A (2019) Assessing non-intrinsic limitarianism. *Philosophia* 47 (1): 249–264.
- Watt F, Hogan B (2000) Out of Eden: stem cells and their niches. *Science* 287 (5457): 1427–1430.

THE ROLE OF TRUTH IN EXPLANATORY UNDERSTANDING

Stefan Petkov

School of Philosophy, Beijing Normal University, Beijing, China
e-mails: yaggdrasil@yahoo.com

Abstract:

This paper discusses the polemical question of whether explanations that produce understanding must be true. It argues positively for the role of truth in reaching explanatory understanding, by presenting three lines of criticism of alternative accounts. The first is that by rejecting truth as a criterion for evaluating explanations, any non-factual account thereby effectively cuts ties with the central theories of explanations, which provide at least partial criteria for explanatory understanding. The second line of criticism is that some of the most well-known non-factual accounts implicitly operate over a notion of partial-truth, and as such, they do not provide a valid alternative. The final critical argument is that, in the place of truth evaluations, these accounts often offer a multiplicity of other criteria, and by changing a unitary criterion such as truth for a collection of other requirements, these non-factive theories introduce a level of ad hoc-ness, which diminishes their normative value.

Key words: scientific explanation, explanatory understanding, the factivism debates, quasi-factive accounts of explanatory understanding.

1. Introduction

In recent years the epistemic problem of analysing understanding has enjoyed a heightened interest. This interest has extended beyond epistemology into the philosophy of science (see Baumberger et al. (2017) for an overview). Two main questions have emerged within the philosophy of science: 1) What are the “vehicles” of understanding; and 2) What are the demarcating criteria of understanding? To un-

pack the first problem: If we take a scientific theory as an umbrella term that encompasses not only sentential generalizations, models and exemplars, but also problems, standards, skills, practices and tendencies (Savage, 1990), then which of these generate understanding?

Several ideas about these “vehicles” of understanding have been explored. Perhaps understanding can be due to an exemplar (Elgin, 2017); perhaps it can be reached by grasping a model revealing the dependency conditions for its targets (Dellsén, 2018); or perhaps understanding of the facts is reached when an explanation is available (Khalifa, 2012). Arguably, this final option has a higher value for understanding because in many cases other candidates for elucidating a puzzling fact are underlined by, or reducible to, an explanation (Khalifa, 2013).

I will take Khalifa’s defence of explanatory chauvinism¹ as a starting point and narrow my analysis only on explanatory understanding. Therefore, I will tackle the second normative problem of understanding – analysing the criteria by which we can distinguish between illuminating and non-illuminating explanations.

The literature on understanding has been divergent exactly on these criteria. Two major opposing camps have emerged. Factive accounts stay close to

¹ The claim that explanations have primacy over other possible “vehicles” of understanding, whether by the fact that forms of non-explanatory understanding are reducible to an explanation, or that whenever an explanation is provided in their place it will always extend the resulting understanding.

the three major theories of explanations² and assert that an explanation should be true (Grimm, 2010; Kvanvig, 2003; Greco, 2013; Khalifa, 2013). Non-factive accounts oppose this idea, and claim that explanations can contain false information and still succeed in illuminating the facts (De Regt, 2015; De Regt & Gijsbers, 2017; Doyle et al., 2019).

My position, which is similar to Khalifa's (2013), is that truth is essential for explanatory understanding. I will defend that idea by arguing that the truth evaluation of the explanans plays a grounding function for all three of the major models of explanations. As such, it is a unitary basic criterion upon which each theory of explanation builds further specific requirements for explanatory relevance. Taken together, the truth of the explanans and the specific requirements for each type of explanation exhaust the objective criteria by which we can distinguish between illuminating and non-illuminating explanations (**Section 2**).

By building on these accounts of explanations my argument against non-factivism will be mainly critical.

Firstly, I will argue that, by rejecting truth as a criterion for evaluating explanations, any non-veridical account effectively severs all links with the three major models of explanations. As these theories present structural requirements for good explanations, and soundness is essential to all of them, if non-factivists reject veridicality they effectively have to offer a novel theory of explanations or reevaluate the results from the existing literature on explanations (**Section 2**).

Following quasi-factive accounts (Greco, 2013; Kvanvig, 2009; Mizrahi, 2012), I will claim that the veridicality of explanations can also be based on weaker truth theories – approximate or partial truth. I will further support my position by briefly going over the general features of such weak truth theories, as they have been shown to meet the challenges raised by non-factivists (**Section 3**). In doing so, I will further develop two more arguments against non-factivism.

The second argument is that at least some of the most well-known non-veridical criteria of understanding reduce to a notion of partial or approximate truth, and, as such, do not offer a viable alternative to it. (**Section 4 to 4.2**).

The final argument is that due to the multiplicity of non-factive criteria, these accounts potentially introduce a level of ad hoc-ness in the analysis of explanatory understanding (**Section 4.3**).

2. The role of truth in the three central models of explanations

The literature on understanding has enriched the existing analyses of scientific explanations by showing that explanatory understanding depends on both the objective features of the explanation and on subjective epistemic factors. Some of these subjective factors are: the intelligibility of the explaining theory³ (de Regt, 2015), the group of related cognitive skills (Newman, 2014) which permit grasping and operating with the relevant fragment of the theory that generates explanatory inferences (Grimm, 2010); the ability to construct typical explanations for other similar facts based on a singular explanatory blueprint; and the ability to organize different supplementing explanations of the same phenomenon into a uniform explanatory nexus (Khalifa, 2017).

Whilst these requirements of epistemic agents are widely accepted, the objective features of explanations that should complement the above-mentioned epistemic skills and produce genuine understanding have sparked a much more heated debate. At the centre of this debate are the non-factivists' efforts to overthrow one of the central dogmas of the literature on explanations – namely that an explanation capable of providing genuine understanding needs to be veridical. Their idea can be schematized as:

“There exist some p and x such that:

- (1) p is false;
- (2) p provides understanding of x ; and
- (3) The understanding of x resulting from ei-

² I will assume that these are deductive nomological, causal and unificatory explanations. For simplicity, I will also accept explanatory pluralism, or, that under specific explanatory seeking questions all these models provide sufficient criteria for a good explanatory answer. Here, I will not present a specific discussion on distinctively mathematical explanations, by presuming that they are analysable as a species of deductive explanations. This is due to the fact that they are often presented as logical or mathematical proofs that are realized in some phenomena.

³ I list the intelligibility of the theory as a subjective epistemic requirement, because determining which theory is intelligible depends on prior knowledge and the abilities of epistemic agents. For instance, a sixteen-year-old might find contemporary statistical mechanics unintelligible, whilst a university student in physics should be able to grasp it.

ther not accepting p or accepting a more accurate proposition instead of p is not better than the understanding provided by accepting p " (Doyle et al., 2019, p. 346).

Narrowing down p and x to, respectively, explanations and facts requiring explanation results in the claim of non-factive explanations. This is that there exist some fact x and some explanation of that fact p , and that p can be false (or contain false premises) but still provide a better explanation than some other explanation p' which is more accurate (e.g., has only true premises).

Here I will focus on debunking the first premise of the argument for non-factive understanding. I claim that no false explanation can provide understanding, because that would make the explanation neither causal, nor deductive, nor based on genuine unification. To defend this claim, two specific questions must be addressed: What is a scientific explanation, and what is the role of truth evaluation of explanations?

The answer to the first question is a familiar one. An explanation can be seen as a discursive game. In it a question of the form "Why x ?" is given, where x stands for a puzzling empirical fact for which no understanding exists. In a scientific context an answer of the form "Because y " is offered, where y is some information derived by a theory which includes x as an element of its domain.⁴ Following Grimm (2010), if we presume the goal of explanation to be understanding, the answer will generate understanding of the question if the above-mentioned subjective factors are satisfied, but also if the information of the answer manages, so to speak, to "tie up" with the question in a correct way.

Besides the obvious requirement of the relevance of the answer, this "tying up" is analysed by the three major models of explanations as a form of consequence relation (or an inference), where x (or the explanandum) appears as a conclusion from y (the explanans). The consequence relations are, from the perspective of explanations, of two types: truth preserving and non-truth preserving, where (in the latter case) the truth of the explanans nevertheless

provides partial grounds for the truth of the consequent.

Under the first of these two types fall deductive-nomological explanations (Hempel, 1965) and classical forms of unification (Kitcher, 1989; Friedman, 1974). Since the consequence relation in these inferences is a necessary one, it also covers some more exotic types of explanations, such as those of highly specific events that have a unique causal vector and distinctively mathematical explanations (Lange, 2017). In these types of explanations, the consequence relation works by showing that the truth of the explanandum is necessitated by the truth of the explanans. Following this idea, if we assume that the explanandum is factually true, then the explanandum can be necessitated only by true propositions of the same class, e.g., the same truth theory should evaluate positively both the explanans and the explanandum. Hence, false premises, whilst in principle capable of rendering a valid inference, do not provide the correct type of information that can satisfy the truth preserving consequence relation in a way that can "tie up" the explanandum with the explanans.

Here one can argue that a valid inference can be true in a modal sense, and thus modal possibilities, whilst being strictly false in our world, nevertheless provide understanding. This is further supported by the idea that counterfactual inferences are routinely used in causal explanations, and also by the fact that some models might be thought of as representing their targets not as they are but as they can possibly be (Le Bihan, 2017). However, I shall argue that these forms of explanations can work only against a background of known factual explanation of causal interactions, or in reference to some version of partial factual truth (See section 4.2).

The second class of consequence relations are the non-truth preserving. This type of inference is typified by probabilistic causation (Pearl, 2000) and looser forms of unification (Petkov, 2015). Under this type of explanation, the truth of the explanans is, nonetheless, required on similar grounds. For instance, non-occurring causally relevant events typically do not change the probability of an occurring effect,⁵ and probabilistic explanations that rely on unification (such as those in evolutionary biology) are also grounded on the basis of coherence with a large body of empirical or factually true data (Petkov et al., 2016).

To summarize, all three major theories of ex-

⁴ This framing permits me to distinguish between understanding of the explaining theory and understanding of the facts using the explanatory apparatus of that theory. I will assume that x is a factual statement and not a theoretic statement which lacks understanding. Therefore, the problem of factivity of explanations does not concern the explanandum x (which is assumed to be factually true by definition), but only the statements of the explanans y .

⁵ For a more specific discussion on causal explanations in particular causal omissions see section 3.1.

planation have the basic requirement that the explanatory inference should be sound. To further appreciate the requirement for soundness we can take a step away from the logical structure of explanations and focus again on the psychological dimension of understanding. As some of the authors on both understanding (Grimm, 2010) and explanations (Bunge, 1973) have noted, the subjective cognitive abilities and skills related to understanding can be triggered (so to speak) by a merely valid deductive explanatory inference, a spurious causal narrative, or a non-genuine unifier. This will result in the same subjective epistemic “feel” of understanding, but the epistemic agent will reach no genuine understanding. Taking this into account permits us to see that the truth evaluation of the explanans does not merely play a logical function, but also provides a grounding criterion for all three major theories of explanation. Of course, beside truth, these models all require further specific criteria (those that filter out spurious and non-relevant causes, non-genuine unifiers, or accidental generalizations) in order to filter genuine from non-genuine explanations. However, all of these specific criteria are supposed to work not separately from, but jointly with, the truth of the explanans.

2.1. Criticisms of the applicability of truth - ontic explanations

The first and most immediate criticism could be that truth is not the right criterion to evaluate explanations simply because explanations are not the type of things that can be true or false. The ontic interpretation of explanations presents such an option. According to the ontic view, explanations cannot be true or false, because they are not texts. As Craver puts it: “[ontic] explanations are not texts; they are full-bodied things. They are facts, not representations. [...] There is no question of [ontic] explanations being ‘right’ or ‘wrong,’ or ‘good’ or ‘bad.’ They just are” (2007, p. 27).

There are two issues with such a view and a possible defence of non-factivism that could stem from it. Firstly, following the quote it is obvious that ontic explanations, if their existence is asserted, are plainly factive, simply because they are *facts* and not propositions. The second issue is related to the class of claims about ontic explanations themselves (for an excellent and more detailed review and a defence of the epistemic view see Wright and Dingmar (2018)).

There are two positions that a proponent of on-

tic explanations can hold. The first one is the claim that all explanations are disguised ontic explanations (Craver, 2007), and the second is that only some explanations are ontic (Salmon, 1989). Both are problematic. The first claim fails if there are distinctively mathematical explanations, or explanations in mathematics (Lange, 2017). In these cases, mathematical structures carry the explanatory load and abstract mathematical structures are typically not considered as ontic facts.⁶

The second one also remains disputable. Even if some events are explanatory for their effects, these events remain ambivalent and cannot explain anything until someone acquainted with them invokes them in order to produce an explanation (Collins, 1966). For instance, whenever causal chains are invoked in explanations the events, entities, and interactions to which they point are often re-described in a restricted or contextually limited way. As such, even if the resulting lexical explanation just points to specific features of the ontic facts (and the ontic facts carry the explanatory load), the lexical element in an ontic explanation should be minimally correspondently true.

2.2. Criticisms of the applicability of truth - theories as the vehicles of understanding

An extension of the same problem is the claim that theories offer understanding directly and we do not necessarily build explanations in order to achieve understanding (De Regt, 2017). This position could be at least partially correct, because as I have noted, a theory is often seen as an umbrella term. It is possible that other theoretic entities (such as graphics, experiments, some forms of experiments and operations with material models as exemplars, etc.) can yield understanding. In all of these situations the vehicles of understanding are not propositional, and therefore cannot be true or false. Nevertheless, and even without a specific argument for the primary importance of explanations, the presence of scientific explanations is a fact in theory development, didactic, and pragmatic contexts. Therefore, the problem of defining the conditions under which explanations do provide understanding is still a necessity.

⁶ An option for a proponent of ontic explanations could be to opt for Platonism and thus claim that explanations in mathematics are just helpers that point to a Platonic fact and its insight. However, I do not think that Craver or other proponents of the ontic view will endorse mathematical Platonism.

2.3. *Non-factive accounts, truth, and the existing theories of explanations*

Following the above arguments, it becomes apparent that analysing explanations and understanding from an epistemic perspective⁷ cannot be avoided. As a corollary, the requirement for positive veridical evaluation of explanations cannot be easily discarded. Any account that discards veridicality effectively neglects a large portion of the literature on explanations. As we have seen, all the criteria for explanations that the core theories of explanations offer operate under the assumptions that explanations are propositional and that some form of soundness of the explanatory inference is required. If this unitary veridical criterion is discarded, these theories become unavailable to non-factivists. It is incumbent, then, on non-factivists to reinterpret causal, deductive, and unificatory explanations or to provide a novel account of explanation. We shall see further, that possible non-factual interpretations of deductive and causal explanations create significant challenges.

3. **Explanations containing idealizations or derived from falsified theories as an example of non-factive explanations. Partial truth as a solution.**

To the above-mentioned ideas in favour of the essential function of truth in evaluating explanations, non-factivists can further respond with an argument familiar from the literature on explanations (Salmon, 1989) to the effect that understanding and truth might in fact pull in different directions. This can happen if the objective structural requirements for explanations raise a high bar that cannot be met by the subjective epistemic skills related to understanding.

This idea has been in circulation in the literature on understanding as early as the 1980s. Railton (1981) has suggested that, in a complete factually true explanation, the lawful model and the causal model will collapse into one singular explanatory text.⁸ The complexity of such an explanatory text, even if available in segments, will be useful only for Maxwell's demon-like creature – but not for limited epistemic agents like us. Arguably, then, the subjective

criterion of intelligibility plays an essential role for understanding and, in a sense, can trump veridicality. Causal omissions, flagging minute qualitative differences, and setting up simplified idealized models are all clever tools that we use to overcome our cognitive limitations. Nevertheless, non-factivists insist (Doyle et al., 2019) that theoretic and model-specific idealizations and causal omissions all result in strictly speaking false but explanatorily relevant claims.

This non-factive argument gains further traction with the observation that some explanations derived from falsified theories are still in circulation (De Regt & Gijsbers, 2017), and that the pessimistic meta-induction does not justify the belief that our current theories are closer to the truth than any past science that is now rejected (Laudan, 1981). There are two possible factive lines of defence: a reinterpretation strategy and weakening the factive condition.

The first strategy involves a reinterpretation of explanations. This type of argument exists both in the literature on scientific modelling (Rohwer & Rice, 2015) and within the debates on understanding (Strevens, 2017; Sullivan & Khalifa, 2019). According to their proponents, the explanatory power of explanations derived from idealized models does not depend on a literal interpretation of the explanans. Two options exist. Firstly, idealized claims do not figure in actual explanations and, as such, idealizations are, so to speak, explanatorily inert. The second option is that whenever idealizations are included in explanatory inferences, these inferences function against the background of true claims. In the second case, explanations work by simultaneously emphasising the importance of the represented factors and the irrelevance of the omitted factors (which are collapsed under the idealized assumptions of the model). This in turn brings sufficient explanatory simplicity that can accommodate our epistemic limitations.

Even if effective, the reinterpretation strategy is only a partial solution to the problems emphasized by the non-factivists. After all, a reinterpretation is possible only if there are available true claims on the basis of which the idealizations can be flagged. This leaves out the deeper worries related to pessimistic meta-induction, falsified but still used science, and our limited cognitive ability to actually formulate these true “flaggers” of idealizations.

Given these problems the second strategy – acceding to a weaker theory of truth – seems to be a

⁷ Understanding at least in some cases is a type of knowledge, a knowledge of the right explanation (Khalifa; Psillos).

⁸ The criteria for completeness Railton uses is a closure under relations of causal dependence, reduction, and supervenience.

much more viable option. This leads to the position of quasi-factivism, or the claim that an explanation does not necessarily have to be true *tout court*, but can be true in some weaker sense. The question, however, that so far has remained largely neglected in the literature is this: What does the adoption of these weaker theories of truth entail for explanatory understanding?

There are several accounts of truth which can meet the challenges of non-factivists. Accounts of approximate truth (Brown, 2002; Petkov, 2019), for instance, can cover omitted factors in idealized models without the explicit presence of true statements that can flag the idealization. Moreover, this account of approximate truth fits nicely with explanatory unification (Petkov, 2019).

Partial truth, developed originally by Da Costa (Mikenberg et al., 1986; Da Costa et al., 1998), is another intriguing avenue of research, because it has already been explored in relation to idealized modeling, falsified science, and considerations related to pessimistic meta-induction (Da Costa & French, 2003). Due to space constraints, I shall not pursue an explicit analysis of the merits and drawbacks of these theories. Nevertheless, a few points on adopting a weaker theory of truth for analysing explanatory understanding deserve specific attention.

A common thread to these theories is that they all define truth contextually. Truth evaluations of approximate or partial truth are all based on partial or limited numbers of observations. These observations are used to frame a factual domain and bind subsequent theoretic statements to it. Within this domain internal inferential consistency is sought between the descriptive statements about the basic observations and statements involving idealizations or theoretic terms (Mikenberg et al., 1986; Da Costa et al., 1998). In some more complex situations involving contradictions between theoretic entailment and observations, or contradictions between theoretic claims, inferential consistency is achieved through non-classic closure principles (Brown, 2002; Petkov, 2019). Since any sentence can be true within a specifically tailored context, trivialization in these theories is avoided by choosing appropriate closure and inferential principles: those which best organize, in a consistent way, a maximally large factual domain.

These contextual characteristics of weaker versions of truth have important consequences for an account of understanding. They entail that, whenever idealizations or falsified claims are involved in constructing explanations, these explanations are

evaluated only against the background of limited or partial observations. Strictly speaking, potential falsifiers (if available) are flagged outside this domain of evaluation. This grants only partial or approximate status to the resulting explanatory inferences. For example, they are considered explanatory only relative to their specific theoretical domain and not outside of it. In fact, extending the inferences which involve idealized claims beyond their domain can lead to serious theoretic or intra-theoretic contradictions (Priest, 2002; Petkov, 2019).

Employing Grimm's concept of grasping (the ability to extend a typical explanatory argument to other similar cases) means that weakly true explanations derived from such partially true theories result only in what he calls "conditional grasping" (Grimm, 2010).

An important worry here is that weak truth theories might not be what factivists seek. After all, their contextual nature and the focus on inferential closure principles as evaluative mechanisms are a far cry from the simple rules of correspondence upon which factivism is usually established. Moreover, some non-factivists (De Regt, 2015) have pointed out that adopting non-correspondence theories of truth, such as pragmatic truth,⁹ is a possible resolution to the difficulties of factivism.

A more elaborate response to this problem requires a special effort. But, briefly, we can assume that factivity can be defined in two ways: as a weaker claim that requires that explanations should be veridical and as a stronger claim which also asserts that truth evaluations should be made by a correspondence theory. If we assert the weaker claim, the availability of theories of partial and approximate truth clears the space for quasi-factive positions.

However, if factivity is taken as the stronger claim the problem becomes much thornier, as it transcends the epistemic analysis of explanations and understanding and shifts towards resolving deeper metaphysical commitments about scientific realism or instrumentalism (Khalifa, 2017). Nevertheless, even if the stronger factive stance is assumed, at least some of the aforementioned weaker truth theories have been interpreted as weakened forms of correspondence (Psillos, 2017; 2009; Brown, 2002).¹⁰

⁹ In fact, the initial drive behind the development of Da Costa's partial truth was to formalize Pierce's pragmatic truth (Mikenberg et al., 1986; Da Costa et al., 1998).

¹⁰ Academic honesty requires me to note that partial truth, which I personally believe is the most viable truth theory

4. Addressing specific non-factive criteria of explanatory understanding

Beside these positive arguments for the value of truth or partial truth for evaluating explanations, veridicality can also be defended by a critical analysis of the criteria for evaluating explanations offered by non-factivists. The central point is that, if non-factive accounts aim to reject veridicality, they must offer other ways of evaluating explanations. Given the scope of the present article, an exhaustive overview of all the non-factive criteria of understanding is impossible. Instead, my strategy in defending quasi-factivism will be to single out two non-factual assessment criteria which, on the surface, seem to be diametrically opposed to veridicality. I shall then argue that these criteria, on close scrutiny, collapse into partial or approximate truth.

4.1. De Regt's criteria of intelligibility and reliable scientific success

The first account is due to the collaborative work of de Regt and Gijsbers (2017). The authors' criterion for achieving understanding is that the scientific theory should be '*effective*,' where effectiveness is a summary of two complementary elements—*intelligibility* and *reliable scientific success*.

The notion of intelligibility is a subjective requirement for theories. It states, roughly, that a theory and its explanatory inferential apparatus should be in principle comprehensible by epistemic agents. Even if we emphasize the subjective psychological dimensions of understanding, intelligibility offers only a component of understanding, and some form of objective assessment of explanations and of background theories must be provided.

In de Regt and Gijsbers' terminology, this second objective criterion is '*reliable scientific success*.' It is defined as a type of propensity of the theory *to generate predictions, ground practical applications, and be a stepping stone for future, better science*. For reasons of brevity I will focus mostly on the notion of predictive success, as it is usually placed higher on the scale of theoretic vir-

tues and because predictions can be easily linked to practical applications.¹¹

Interestingly, de Regt and Gijsbers have not included explanations as elements of scientific success. One suspects that this is due to the strange consequence that would follow. For if one rejects veridicality as an evaluative criterion, one must accept that merely valid explanatory inferences will be consequence-bearing for factually true explananda. As we have seen, this simply goes against the grain of all theories of explanation. A viable option for a non-factivist is to point out that explanations and predictions are often presented by similar inferences. An explanation, then, can be evaluated only if a successful prediction can be based on it. Analysing the structure of predictions, however, makes this idea even more problematic for non-factivists.

Predictions can be represented as structured arguments in which the predicted phenomena appear as the consequent of premises that provide supporting information as to why the predicted outcome will turn out as foreseen. In this regard they are very similar to explanations, as the explanans should also offer information on which the explanandum depends. This idea is already present in Hempel's deductive-nomological model with its notion of nomic expectability (Hempel, 1965). Within the deductive-nomological model, the law-like regularities used to construct explanatory arguments can also be employed for the purpose of constructing predictions, as the laws of nature are regarded as being universally valid (the laws do not function for a specific temporal segment). If we take a closer look at the structure of predictions and explanations within examples conceivable within the deductive-nomological model, their only difference will be the addition of a temporal operator *X* (*neXt time*) for the conclusion in predictions. For instance:

Given the relative laws of nature *L* and the

applicable to the problem of understanding, has interpretations from both realists (Pillos, 2009), structural realists (Ladyman, 1998) and structuralists (Bueno, 2010). Such divergent interpretations of partial truth lead me to believe that partial truth can be accepted as a metaphysically neutral position. It is neutral in the sense that it offers a sufficient criterion by which we can judge between sound and merely valid explanatory inferences, but which, at the same time, is insufficient to strongly support a realist or an anti-realist position.

¹¹ However, a few short notes on heuristics and practical applications can be made. First, I do not think that adding the notion that a theory provides understanding if it leads to a better new theory is a viable criterion for understanding, as this will blur the demarcation between theories that are comprehensively enriching and those that are not. Any theory that is subsequently replaced could be thought of as providing understanding. For instance, the notion of demonic possession is the forerunner of modern theories of mental illness, but one would typically deny that such a theory provides any understanding. Practical applications, in their turn, have a significant overlap with predictive success, since knowledge of how a phenomenon will evolve is the basis for any practical applications.

auxiliary circumstances A at time t , at time $X(t) > t$, it will be the case that p .¹²

Notice that the soundness of the argument plus the extra nomological requirement (the general premise should not be accidental and the auxiliary conditions should not be irrelevant) limits the conclusion to a single possibility. Given a temporal twist, this constraint guarantees the predicted outcome.

If we pursue an analysis of causal explanations, as difference makers we arrive at an even stronger factive result. Causal dependencies are usually temporally asymmetrical, and the same causal narrative can be used for both explanations and predictions (so long as the causal conditions are repeatable). Woodward's account, for instance, is explicitly inclusive of the similarity between explanatory arguments and predictions (Woodward, 2003).

The important element in causal inferences is that they are evaluated based on assessing the factual correctness of the causal determinant for the consequent. If causal predictions and explanations have such a significant overlap, it is hardly imaginable that both can be judged without any notion of factual accuracy operating in the background. Moreover, causes and effects should have the very same ontological status (e.g., as facts), otherwise no causal interaction can ever occur (Dowe, 2000). Or, to put it bluntly, a false (non-occurring) causal event simply cannot necessitate a factual outcome.

Omissions and preventions are a notable exception. However, normally we do not describe a causal outcome as the absence of preventive factors but in terms of actually occurring interactions that bring about the effect. Nevertheless, Woodward has noted that omissions and preventions are sometimes considered, but only if they represent a "serious possibility." The notion of a serious possibility, in turn, is determined only against the background of already established causal interactions that bring about the effect. These known interactions are used to define realistic possible preventions and omissions (Woodward, 2003).¹³ Similarly, Dowe (2009) has argued that omissions and preventions can be meaningfully defined only with known, factually true,

e.g., occurring, causal interactions.

Finally, a non-factivist can protest that our explanations and predictions very often misrepresent causes, even potentially relevant ones (Doyle et al., 2019). However, such an argument can stand only against full-fledged correspondence and not against forms of partial or approximate truth. Without going deeper into the structural properties of these truth theories, some of their interpretations (Psillos, 2009) assume that weaker truth evaluations are established because they can (at least in part) trace some of the relevant causal structures that bring about the facts under investigation.

Similarly, the omission of causally relevant information by collapsing it under idealizations is only considered relative to an observational domain which does not include the resulting minimal differences between the effects. Factivists have insisted that these omissions are justified because they involve negligible effects or effects that cancel each other out (Strevens, 2017; Sullivan & Khalifa, 2019). Here the central point is not that idealizations assume deliberate falsities, which then are mirrored in explanations, but that these idealizations are inferentially and explanatorily relevant only for an appropriately framed observational domain. In other words, the idealizations are true but only within a partial theoretic structure (Da Costa & French, 2003).

Given these critical points it seems that De Regt's "reliable scientific success" implicitly operates over some form of weakly true evaluations of predictive and practical inferences.

4.2. *Le Bihan's modal understanding*

The second non-factive account of understanding which I am going to analyse is based on the idea that understanding is a modal notion. This account is credited to Le Bihan (2017). Her argument for non-factivism is that many theories and their associated models are known to misrepresent reality. If truth is employed to assess these models and theories, then they cannot be used to derive explanations, because the resulting inferences will contain false information. However, falsified theories and their models still have epistemic value, so they must provide some understanding of their targets. For her, this understanding is not based on factual truth, because it is brought about not by representing the phenomena as they are, but as they possibly could come to be.

Importantly for Le Bihan (2017), the phenome-

¹² Note that since the deductive-nomological model operates on systems with deterministic laws, a linear tensed logic can be used to formalise the predictive arguments functioning in a deterministic context (for an overview, see Goranko (2015)).

¹³ For more details on the idea that possibilities can provide understanding see the next section.

na are not facts. They are constructs based on partial empirically gathered data. This makes phenomena not completely isomorphic with empirical facts. Models describe dependency conditions sufficient to bring about these data structures but insufficient to describe actual world facts in all their abundance and completeness. Based on this idea, she defines understanding as a modal notion – the epistemic ability to navigate within the possibility space of dependencies that can potentially generate the phenomena under scrutiny. In her view, explanations are just descriptions of the dependence conditions which allow for the possibility of particular phenomena.

Similar to the intelligibility condition in De Regt's analysis, Le Bihan's modal understanding requires some grounding evaluative criteria that can classify the modal possibilities described by the model. With no evaluation grounded in the actual world (or in this case, in data extracted from actual observations and experiments) we cannot determine whether a given modal account is or is not representative of its targets. After all, each inference based on a possible dependence structure is trivially true relative to its possible world.

The only option for vindicating modal understanding is to assess possible dependence structures against the background of the available factual information. Le Bihan herself is not explicit on the assessment criterion, besides a rather vague remark suggesting that an idealised model should 'save' the phenomena in order to provide comprehension of its target. We can perhaps assume that the notion of 'saving the phenomena' echoes van Fraassen's (1980) idea of constructive empiricism and is meant to judge the empirical success of a given theory and its idealised models. This amounts to the fact that even though the theory represents phenomena as idealised data structures brought up from partial empirical information, in order for these data structures to be informative of the facts, they should be maximally consistent with the available corpus of empirical information. Such being the case, 'modal understanding' will also implicitly require an assessment of the inferences based on these models against the background of some partial factual correspondence.

Here, under close scrutiny, Le Bihan's view collapses into the notion of partial factivity. If we strengthen the evaluative criterion to require that our data structures are completely isomorphic with the facts, the account will collapse into classic correspondence. If we weaken the evaluative criterion to

require only contextually restrained partial isomorphism with the facts, we will arrive at the notion of partial truth (Da Costa & French, 2003). In fact, Da Costa's partial truth (da Costa et al., 1998) has simply been given a modal spin, where the partial theoretic structures upon which partial truth operates have been represented as worlds in Kripke's sense.

As such, if we use Le Bihan's account for evaluating explanations, an ideal explanation will describe the dependence conditions of the occurrence of the phenomena in the actual world, whilst any other explanation will bring only partial understanding.

Because the aim of any model is to be maximally consistent with a well-defined fragment of the observational data, even if we assume that understanding is a modal notion and that explanations simply describe how things could have possibly occurred, the understanding based on these explanations should be defined over some grading of the possible worlds relative to our actual world. In other words, the understanding that a modal explanation offers can be evaluated only in terms of how close the world (in which the explanation is sound) is to the actual world. Such an idea is not novel and is already well known for establishing the explanatory power of counterfactual causal explanations (Lewis, 1973; 2000). Therefore, any scientifically interesting assessment criterion of modal inferences requires some form of veridical evaluation that is factually grounded.

To summarize, if we judge understanding as a modal notion but reject any account of factual veridicality, we will reach the trivial conclusion that any possible world scenario provides understanding. Instead, if we judge the possible world narratives in terms of closeness to the actual world, we are implicitly endorsing an evaluative criterion based on partial factual truth.

The conclusion is that the criteria for non-factive understanding of de Regt, Gjisbers and Le Bihan, implicitly operate on a version of partial or approximate truth. Moreover, their criticism of veridicality as an evaluative criterion is warranted only if a full-blown correspondence notion is assumed, whilst their accounts are actually permissive of the application of partial or approximate truth.

4.3. The multiplicity of non-factive criteria and *ad hocness*

Besides the central disagreement about the role of truth, one of the key differences between factive

and non-factive accounts is that non-factive accounts are centralized only around a negative criticism of truth for evaluating understanding. They do not share a unified agreement on a core criterion for understanding that can replace it.

A possible reason for the multiplicity of non-factive accounts is that such analyses are often painted with a broad brush – as general accounts of understanding. Perhaps the strategy of non-factivists is not to focus narrowly on explanatory understanding, but instead to offer an account of theories that generate understanding and then to define explanations simply in terms of the explanatory practices that such theories employ. However, avoiding specific requirements of explanations cannot be a successful strategy.

Firstly, without any theory of explanations, it is possible to construct very bad explanations based on otherwise successful theories. Take for instance: Q: “Why do humans have two hands?” A: “Because the two hands enable the gene carriers of genes that code for two hands to replicate and persist.” In order to resolve this issue, a non-factivist must fall back on the criteria of informativeness and explanatory relevance, and these criteria essentially lead back to some of the central theories of explanations, all of which require the truth of the explaining premises.

Secondly, even if we can agree on a broad definition of explanations as the explanatory practices resulting from successful theories, this will, in turn, invite another set of problems. The available non-factive criteria include a plethora of epistemic goods and scientific virtues, and it is hardly imaginable that an explanation can satisfy all of them (Doyle et al., 2019).

This, however, invites the question of how to grade explanations, if we assume that the three models of explanation will simply reject those explanations that have false premises within the explanans. This becomes a problem since the multiplicity of available criteria may then lead to ad hoc-ness within the whole area of defining understanding and, more narrowly, in an account of explanatory understanding. After all, any explanation that is based on some scientific data can be intelligible, can lead to ease of calculation, a new better explanation, etc. So many different criteria of understanding simply blur the borders between genuine and non-genuine understanding, and also obscure the requirements that can distinguish good from bad explanations. As a consequence, this undermines the purpose of the whole project of the theories of explanations –

namely, defining the necessary and sufficient conditions of good scientific explanations.

5. Conclusion

The present text has aimed to provide arguments for the positive role of factive truth in evaluating explanations. The core idea of the text is that the central models of explanations offer the groundwork for evaluating explanations and the understanding that results from these. All of these models describe specific criteria for filtering genuine from non-genuine explanations, and all of these specific criteria operate only on the more basic requirement that the explanans be true. A further point is that whenever full blown correspondence cannot be applied, partial or approximate truth can be employed, and both of these truth theories preserve the essential features of the logic of explanations.

On the other hand, if truth is discarded by non-factive accounts, they have only three options: 1) define another criterion for evaluating explanations; 2) deny that explanations are valuable, or 3) deny that explanations are entities that could be true or false.

In the case of the first option, the multiplicity of criteria of non-factivists could be seen as introducing a level of ad hocness into the literature of explanations, or these non-factive criteria could even be perceived as implicitly operating over a notion of partial truth. The second option can be easily discarded, because explanations are scientifically valuable and their analysis cannot be overlooked. The third and final option also remains questionable. Ontic explanations, even if it is asserted that they are themselves facts, are such that factivity considered broadly also should apply to them. Therefore, the lexical fragment in ontic explanations, even in the most minimal cases, should at least point correctly to the fact doing the explanatory work. Even in this case, an explanation is best analysed as true or false in the correspondent sense.

A conclusion from all these cases is that a traditional account of explanations, supplemented with an appropriate notion of truth (partial or approximate), seems to offer the best choice in analysing explanatory understanding.

References

Baumberger C, Beisbart C, Brun G et al. (2017) What is understanding? An overview of the recent debates in epistemology and philosophy of science.

In *Explaining understanding: New perspectives from epistemology and the philosophy of science*, eds Grimm et al., 1–34. New York: Routledge.

Bueno O (2010) Structuralism and information. *Metaphilosophy* 41 (3):365–379

da Costa N C A, Bueno O, French S (1998) The logic of pragmatic truth. *The Journal of Philosophical Logic* 27: 603–620.

Bunge M (1973) *Philosophy of physics*. Dordrecht: Reidl.

Collins A (1966) Explanation and Causality. *Mind* 75(300): 482–500.

Craver C (2007) *Explaining the Brain*. Oxford: Oxford University Press.

da Costa N C A, French S (2003) *Science and partial truth*. Oxford: Oxford University Press.

De Regt H W (2015) Scientific understanding: Truth or dare? *Synthese* 192: 3781–3797.

De Regt H W (2017) *Understanding Scientific Understanding*. Oup Usa.

De Regt H W, Gijsbers V (2017) How false theories can yield genuine understanding. In *Explaining understanding: New perspectives from epistemology and the philosophy of science*, eds Grimm et al., 50–75. New York: Routledge.

Dellsén F (2018) Beyond explanation: understanding as dependency modelling. *British Journal for the Philosophy of Science*, doi: 10.1093/bjps/axy058.

Doyle Y, Egan S, Graham N, Khalifa K (2019) Non-factive Understanding: A Statement and Defense. *Journal for General Philosophy of Science / Zeitschrift für Allgemeine Wissenschaftstheorie* 50 (3):345–365.

Dowe P (2000) *Physical causation*. New York: Cambridge University Press,

Dowe P (2009) Absences, Possible Causation, and the Problem of Non-Locality. *The Monist* 92: 23–40.

Elgin C (2017) Exemplification in Understanding. In *Explaining understanding: New perspectives from epistemology and the philosophy of science*, eds Grimm et al., 76–91. New York: Routledge.

Friedman M (1974) Explanation and scientific understanding. *Journal of Philosophy* 71 5–19.

Goranko V, Galton A (2015) Temporal Logic, *The Stanford Encyclopedia of Philosophy*. (Winter 2015 Edition), ed. Edward N. Zalta, <https://plato.stanford.edu/archives/win2015/entries/logic-temporal/> Accessed 15 February 2020

Greco J (2013) Episteme: knowledge and understanding. In *Virtues and their vices*, eds Kevin

Timpe and Craig A. Boyd, 285–301. Oxford: Oxford University Press.

Grimm S R (2010) The goal of explanation. *Studies in the History and Philosophy of Science* 41, 337–344.

Hempel C (1965) Aspects of Scientific Explanation. In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, 331–196. New York: Free Press.

Khalifa K (2012) Inaugurating Understanding or Repackaging Explanation? *Philosophy of Science* 79:15–37.

Khalifa K (2013) The role of explanation in understanding. *British Journal for the Philosophy of Science* 64 (1):161–187

Khalifa K (2017) *Understanding, Explanation, and Scientific Knowledge*. Cambridge: Cambridge University Press.

Kitcher P (1989) Explanatory unification and the causal structure of the world. In *Scientific explanation*, eds P. Kitcher, W. Salmon, 410–505. Minneapolis, University of Minnesota Press.

Kvanvig J (2003) *The value of knowledge and the pursuit of understanding*. New York: Cambridge University Press.

Kvanvig J (2009) Responses to critics. In *Epistemic value*, eds Adrian Haddock, Allan Millar and Duncan Pritchard, 339–352. Oxford: Oxford University Press.

Ladyman J (1998) What Is Structural Realism?. *Studies. History and Philosophy of Science* 29: 409–424.

Lange M (2017) *Because without cause: non-causal explanations in science and mathematics*. New York: Oxford University Press.

Laudan L (1981) A confutation of convergent realism. *Philosophy of Science* 48: 19–49.

Le Bihan S (2017) Enlightening falsehoods: A modal view of scientific understanding. In *Explaining understanding: New perspectives from epistemology and the philosophy of science*, eds Grimm et al., 111–135. New York: Routledge.

Lewis D (1973) *Counterfactuals*. Cambridge MA: Harvard University Press.

Lewis D (2000) Causation as Influence abridged version. *Journal of Philosophy* 97: 182–197.

Mikenberg I, da Costa N C A, Chuaqui R (1986) Pragmatic truth and approximation to truth. *The Journal of Symbolic Logic* 51: 201–221.

Mizrahi M (2012) Idealizations and scientific understanding. *Philosophical studies* 160 (2): 237–252.

- Newman M (2014) EMU and inference: what the explanatory model of scientific understanding ignores. *European Journal for Philosophy of Science* 4 (1):55–74.
- Petkov S (2015) Explanatory unification and conceptualization. *Synthese* 192: 3695–3717.
- Petkov S, Wang Wei, Lei Yi (2016) Explanatory unification and natural selection explanations. *Biology & Philosophy* 31(5): 705–725
- Petkov S (2019) Studying Controversies: Unification, Contradiction, Integration. *Journal for General Philosophy of Science / Zeitschrift für Allgemeine Wissenschaftstheorie* 50 (1):103–128.
- Priest G (2002) Inconsistency and the empirical sciences. In *Inconsistency in science*, ed. J. Meheus, 119–128. Dordrecht: Kluwer Academic Publishers.
- Psillos S (2009) *Knowing the Structure of Nature*. London: Palgrave-MacMillan.
- Psillos S (2017) World-Involving Scientific Understanding. *Balkan Journal of Philosophy* 9 (I): 5–18
- Railton P (1981) Probability, Explanation, and Information. *Synthese* 48: 233–256
- Rohwer Y, Rice C (2016) How Are Models and Explanations Related?. *Erkenntnis* 81(5): 1127–1148.
- Salmon W (1989) *Four decades of scientific explanation*. Pittsburgh: University of Pittsburgh Press.
- Savage C W (1990) Preface. In *Scientific Theories. Minnesota Studies in the Philosophy of Science*. Volume 14, vii–ix. Minneapolis, University of Minnesota Press.
- Strevens M (2017) How idealizations provide understanding In *Explaining understanding: new essays in epistemology and philosophy of science*, eds Stephen Grimm, Christoph Baumberger and Sabine Ammon, 37–49. New York: Routledge.
- Sullivan E, Khalifa K (2019) Idealizations and Understanding: Much Ado About Nothing?. *Australasian Journal of Philosophy* 97 (4):673–689.
- Van Fraassen B (1980) *The scientific image*. Oxford: Oxford University Press.
- Woodward J (2003) *Making things happen*. Oxford: Oxford University Press.
- Wright, C, van Eck, D (2018) Ontic Explanation Is either Ontic or Explanatory, but Not Both. *Ergo: An Open Access Journal of Philosophy* 5:997–1029.

ON A REGULARITY OF THE INTELLECTUAL EVOLUTION OF MANKIND

Rovshan Sabir Hajiye

The Institute of Philosophy of Azerbaijan National Academy of Sciences,
Baku, Azerbaijan
e-mails: vinodelrovshan1@rambler.ru

Abstract:

The article explores yet another view of the history of mankind, and examines global problems related to historical processes, which are still far from receiving an unequivocal explanation. As an alternative to Marxism and other theories of social development that shed light on key historical events and global processes, I propose an account based on the hypothesis of the age periodization of the intellectual evolution of mankind. The main provisions of the hypothesis are set out in the content of the article. The methodological basis of the hypothesis is a comparative analysis of ontogenesis and phylogenesis. In other words, on the basis of known laws of intellectual development in ontogeny, I examine historical processes occurring in phylogeny, paying special attention to the substantiation of the main provisions of the hypothesis of the age periodization of the intellectual evolution of mankind.

Key words: ontogenesis, phylogenesis, historical process, age periodization of the development of Homo sapiens sapiens, intellectual evolution of humanity.

1. Introduction

Throughout history humankind has evolved, experimented, learned, and grown to reach a modern and more civilized stage in its intellectual and social development. The social evolution of human beings has allowed for the development of the concept of human freedom not only within various cultures but as a central tenet of international law and other norms that mediate relations between cultures. The formation of the concept of a 'unified

mankind' transcending cultural difference traces back to the age of European Enlightenment (J.G. Herder). Roughly from that time forward some of the greatest European thinkers have tried to discover the generalities shared by all human beings that indicate a uniformity in our social development, and thus an essential shared humanity. To this end both materialists (e.g. K. Marx) and idealists (e.g. G. Hegel) have offered theories of human development. The concept of social development as it occurs in certain theories, especially Marxist theory, which is prevalent in post-Soviet nations, does not consistently explain global events and their role within, and relation to, historical processes and human development. The following represent examples of these kinds of historical events and the unanswered questions they raise:

i) Why did Mesolithic or Proto-Neolithic humans (who for tens and even hundreds of millennia during the Paleolithic age survived through the practice of hunting and gathering) make the first steps toward the domestication of plants and animals, and how did they succeed in building agrarian societies in a relatively short period of time?¹ How did the tendency toward individual activity arise from a collectivist lifestyle? How did human beings invent complex tools, such as the bow and arrow, and generally develop a qualitatively new lifestyle? In G.

¹ Lamberg-Karlovsky C.V. V. Gordon Childe and concept of revolution // Journal of Ancient History № 4(207), 884; Goncharov N.P. Domestication of plants// Vavilov Journal of Genetics and Selection, No 4/2, 93.

Barker's words, "why did foragers become farmers?"²;

ii) Why did Neolithic hunter-gatherer communities suddenly abandon nomadism and adopt a settled lifestyle that precipitated developments in cattle-breeding, agriculture, pottery, and metallurgy?³;

iii) Why, in the period of the earliest civilizations, did rulers and high priests appear as leading cultural figures? Why were epics composed, written language founded, social hierarchies developed, and early scientific thinking cultivated?;

iv) Why did the first sages, prophets, and philosophers appear during the Axial age (in K. Jasper's words) – *i.e.*, why did people begin freely thinking and questioning the primitive mythological worldview of their culture, ultimately releasing humankind from taboos, rites, and magical belief?

These and other events of global significance flowed step by step in the historical development of human cultures for millennia, and their influence caused fundamental changes in lifestyle, leading to the development of contemporary societies. But a full and unambiguous explanation of the reason these events occurred hasn't been given yet. More specifically, Marxism, which retains intellectual authority in post-Soviet nations, cannot unequivocally answer the questions raised above. This becomes clear when we try (and fail) to explain the events that occurred in the period of humanity's earliest civilizations from a Marxist point of view.

II. An externalist approach to resolving these issues

According to Marxism, states were formed out of subsistence farming communities when agriculture based on irrigation was created at the end of the fourth century BCE, and as a result of this, communities obtained extra income that enabled the stratification of classes based on wealth. The development of class hierarchy and increased leisure time for members of the upper classes resulted in myriad social changes, including the appearance of the first political sovereigns and the invention of writing.

Prima facie such commentary on historical processes looks rational, and as such, the Marxist approach appears to successfully explain the formation of states. However, it fails to answer a crucial ques-

tion: Why didn't this process occur during previous ages? The traditional answer is as follows: Irrigation techniques did not exist before the fourth century BCE, so the surplus income it generated, which led to social stratification, was not previously available. However, archaeological research has cast doubt on this response.

According to specialists in the field, as far back as the sixth millennium BCE large cities (Hajilar, Chatal-Huyuk) existed in the Middle East and housed temples, indicating that social hierarchy and a social division of labor existed and that the existence of surplus income was undoubtedly possible.⁴ New evidence also supports this. Artifacts dating to the sixth millennium BCE were discovered by archaeologists on top of Tell-Zeydan hill in Syria, and appear to indicate that humans during this period had achieved the basic necessary conditions for the creation of the first states. Yet the first states appeared not in the sixth millennium but in the fourth. Indeed,

Research workers of the Institute of Oriental Studies of Chicago University in the USA and their Syrian colleagues have no doubt that this archaeological find is one of the prehistoric dwellings of Ubeyd (existed approximately in 5300-4000 BC). The artefacts prove that this city was a trade-industrial centre where copper metallurgy and pottery production were developed enough. The facts allow us to assert that the class society was already formed there. Thus, the stone seals proving the possession of private property by privileged people were found. So, the Ubaid culture was a period in the Middle East when farming based on irrigation and centralized temple agriculture was widely known. Namely during this period first tribal leaders appeared and social inequality was raised.⁵

It's interesting that despite all these facts, specialists present the cities mentioned above as 'protocities', and they do not qualify the civilizations which existed as long ago as the fourth millennium

² Barker G. The agricultural revolution in prehistory: Why did foragers become Farmers? Oxford etc.: Oxford univ. press.

Munchaev R.M. Cultures of preliterate Mesopotamia // Science and humanity, Moscow., 63

⁴ Mellaart J. Earliest civilizations of Near East, Moscow., 121–125; Gulyaev V.I. Mysteries of ancient cities: Near East and Mesoamerica, Moscow., 5; Ancient civilizations (under redaction of G. M. Bongard-Levin), Moscow., 28–30.

⁵ Stein, Gill J. Excavations at Tell Zeidan, Syria // The Oriental Institute. News & Notes., no. 203, 3–7 <http://oi.uchicago.edu/sites/oi.uchicago.edu/files/uploads/shared/docs/nn203.pdf>

BCE as a *civilization* or *state*. This is likely because written language did not appear in that period, and the first formal rulers had not arisen. Given other social developments that occurred (e.g. the attainment of surplus wealth), it remains unclear, according to Marxist theory, why these crucial markers of civilization had not also been achieved in this period.

Explanations of the development of human cultures and civilization rely not only on philosophical or sociological concepts; in some cases the reasons behind these events can be accounted for by ecological, *i.e.*, natural factors. Thus, most specialists explain widespread cultural-historical changes during the Mesolithic age mainly on the basis of these ecological factors. For example, experts believe that during the transitional period from Pleistocene to Holocene huge animals (mammoth, etc.) disappeared due to global warming, and consequently, people who relied on them as a primary source of food were forced to make fundamental changes in their lifestyle. Previously, entire communities hunted large animals together; forced to adapt to changing natural conditions, humans began focusing instead on hunting small game independently, and the growth of this individualistic practice may explain, in part, the impetus behind the development of the bow and arrow. But archaeological facts may fuel skeptical attitudes to this kind of explanation. For instance, artifacts show that during the Mesolithic age bow hunting spread not only throughout periglacial areas that were affected by climate change, but also warmer regions, including tropical and subtropical regions, where flora and fauna were not impacted. Archaeological findings in India⁶ and South Africa⁷ – Mesolithic arrow tips made from microlithic stone – appear to confirm this.

Explanations of historical events based on sociological or ecological theories fail to account for a more serious matter, namely, that major milestones in the development of modern civilization were achieved in a short period of time considering that the Paleolithic age covers hundreds of millennia. Specifically, the appearance of neanthropics humans (about 40,000 years ago), considered by experts to be *the revolution of the late Paleolithic age, oc-*

*curred during a very narrow window of time.*⁸ The Neolithic revolution was first fundamentally characterized by English archaeologist G. Childe, but the revolutionary events which happened during the Neolithic period amazed not only archaeologists and paleoanthropologists⁹ but also representatives of the natural sciences.¹⁰

An explanation of revolutionary historical events on the basis of the law of dialectics, which maintains that quantitative changes translate to qualitative ones, may offer a way out of the Marxist dilemma. However, the effort ultimately fails when confronted by a natural follow-up question: How can the synchronous step by step occurrence of historical processes in various parts of the world be explained? It is impossible to imagine that primitive people living in different areas of the world agreed on changing their lifestyle fundamentally, and as a result, achieved the realization of the law of dialectics.

Here it is crucial to bear in mind the relevant historical facts: 1. The appearance of neanthropic humans (this biological type is also known as *Homo sapiens sapiens* in modern scientific terminology) took place in different parts of the world approximately 39,000 years ago;¹¹ 2. In the Mesolithic age the identification of the initial effects of domestication and the invention of the bow and arrow appeared simultaneously in different parts of the world (in Mexico, Thailand, and the Middle East from the tenth to the ninth millennium BCE);¹² 3. The transition from hunting and gathering to agriculture (the seventh to the sixth millennium BCE) during the Neolithic age happened in different areas of the world concurrently, although irrespectively of each other;¹³ 4. The independent and synchronous ap-

⁸ Vishnyatskiy L.B. Upperpaleolithic revolution: geography, chronology, causes // «STRATUM plus», No1.

⁹ Teilhard de Chardin P. The phenomenon of man, Moscow., 164-165.

¹⁰ Prigogine I.R. Network society // Sociological researches, No1.

¹¹ Vishnyatskiy L.B. Cultural Dynamics in the middle of the Late Pleistocene and the transition to the Upper Paleolithic, 34; World history: in 6 vols, Vol.1, Moscow. 37.

¹² Zablozka Y. History of Near East in ancient times, Moscow., 20; Boriskovskiy P.J. Old Stone Age of South and Southeast Asia, Moscow., 159; Kabo V.R. Primitive preagricultural community, Moscow., 255.

¹³ History of primitive society. Epoch of primitive tribal community (Edited by academician Y.V. Bromley), Moscow., 263; Kabo V.R. Primitive preagricultural community, Moscow., 253-255; Chesnov I.V. Agricultural crops as an ethnogenetic source // Early ethnic history of the peoples of East Asia, Moscow., 115.

⁶ History of primitive society. Epoch of primitive tribal community (Edited by academician Y.V. Bromley), Moscow., 156.

⁷ History of primitive society. Epoch of primitive tribal community Moscow. 182.

pearance of the first states in different regions of the world (including Mesopotamia, Ancient Egypt, Ancient India, etc.) became possible in the fourth millennium BCE;¹⁴ 5. In the Axial Age (first millennium BCE), the first prophets and philosophers appeared in different parts of the world at the same time.¹⁵ The three-millennial difference between the global events mentioned above draws attention to the protracted period between the appearance of neoanthropic humans and the rise of the first states (which the law of dialectics fails to address). It is possible to increase the number of events under examination (*e.g.* the creation of the 'Paleolithic Venuses' in the Upper Paleolithic period, the simultaneous appearance of megalithic buildings in the period of the first civilizations, etc., although such events are perhaps less important than those examined in this article) that manifested themselves during this period, but the specific events that we examine ultimately do not impact the overarching explanation of the history of humanity. In all cases, the question logically arises: What were the reasons for the occurrence of global historical developments which appeared around the world simultaneously and in accordance with near-identical, step by step progressions?

III. The hypothesis of age periodization of the intellectual evolution of mankind

I've tried to give a comprehensive and scientifically based answer to this question for many years in my empirical investigations,¹⁶ speeches at international conferences,¹⁷ and in the lecture I delivered at the 23rd World Congress of Philosophers in 2013 in Athens.¹⁸ My response to the questions raised above rests on the hypothesis of age periodization of the

intellectual evolution of mankind,¹⁹ a hypothesis I arrived at through methodical analyses comparing phylogeny and ontogeny as they relate to human intelligence. The first, and one of the most well known analyses of human intellectual development that tracked the relationship between ontogenesis and phylogenesis was conducted by S. Hall, an American psychologist and founder of the discipline of pedagogy. Based on the ongoing experimental work and the biogenetic law of E. Haeckel, S. Hall concluded that the ontogenetic development of the individual child's psyche repeats all stages of the phylogenetic development of the psyche of humanity in general, a discovery referred to as recapitulation theory. Hall was a supporter of preformism and, on this basis, he believed that the consistency and content of developmental stages were set genetically and therefore the child couldn't evade or skip any stage of its development. Hall's theory was further developed by his students, in particular, C. Hutchinson and his followers (for example, E. Meumann).

Other pupils of Hall, such as E. Claparède, a prominent Swiss scientist, did not accept certain fundamental provisions of this theory. T. Martsinkovskaya, a modern researcher of children's psychology, writes: "Claparède believed that the well-known similarity between the phylogenetic and ontogenetic development of the psyche existed not because the child's psyche contained the stages of development of the species and the ancient instincts that he had to overcome (as the recapitulation theory suggests it), *but because there was a general logic of the development of the psyche in phylogeny and ontogeny.*" According to Claparède, this general logic of development determines the similarity of processes (but not their identity!). Therefore, there is no fatal predetermination in the development of the child, and external factors (including training) can accelerate its course and even partially reverse direction.²⁰ Identifying thinking with mental development, Claparède singled out four stages in the child's mental development where the criterion of division was the transition from one type of thinking to another.

Claparède's ideas and achievements were expanded upon by his famous student, psychologist J.

¹⁴ Oppenheim A. Ancient Mesopotamia, Moscow., 26; Tokarev S.A. A. Religion in the history of the peoples of the world, Moscow., 332; Jaspers K. The origin and goal of history, Moscow., 42-44.

¹⁵ Jaspers K. The origin and goal of history, Moscow. 33.

¹⁶ Hajiyev R.S. The dynamics of development the historical process or the intellectual evolutionary path towards globalization, Baku.

¹⁷ Hajiyev R.S. About one regularity of historical process // Theses of reports of the VII Russian Philosophical Congress. Ufa.

¹⁸ Hajiyev R.S. On substantiation of hypothesis of age periodization in intellectual evolution of humanity // XXIII World congress of philosophy: Philosophy as Inquiry and way of life, University of Athens, School of philosophy, Athens 04-10 august 2013, Abstracts.

¹⁹ Hajiyev R.S. Problem of the meaning of life, Baku. 118-138.

²⁰ Martsinkovskaya T.D History of child psychology: Textbook for students of pedagogical universities, Moscow. 91-93.

Piaget, who distinguished the following stages in the development of cognition: sensorimotor (sensually motional); preoperational intellect; concrete operational; formal operational.²¹ A comparative analysis of ontogenesis with phylogenesis was also conducted by L. Vygotsky, an outstanding researcher of the Soviet psychological school. He wrote: "Speaking about the cultural development of the child we mean a process that corresponds to the mental development that took place during the historical development of mankind."²²

I, for the first time in science, have used the inverse method, *i.e.* I have examined the process of the intellectual development of mankind on the basis of ontogeny, and more specifically, on the basis of Piaget's aforementioned theory.

Research suggests that Piaget's periods of cognitive development in ontogeny definitely correspond to the critical stages of the intellectual development of mankind. In particular, there is a correspondence that captures a critical turning point in human development: In individual children this stage of cognition is reached around roughly age seven, and for humanity generality, this stage can be identified with the Solutrean epoch (18,000 years ago), the historical period in which humankind realized the developmental achievements appropriate to contemporary seven-year-old children. During this period, the first drawings that reflect not singular objects but the relations between them (for example, some interaction between a woman and a beast) appear on the walls of caves. Further, at the end of the Madeleine (15,000 years ago, when mankind was "8" years old according to the developmental correspondence), engravings reflecting the perspective image (presumably, from that time on, the perception of a homogeneous and isotropic space was gradually formed) were created. With the advent of spatial thinking (the perception of a homogeneous, isotropic space, instead of the so-called "fixed-route" according to A. Leroi-Gourhan) in the Mesolithic age (12,000 years ago), primitive societies became increasingly more complex. Fishing techniques were developing, and people were eager to engage in individual hunting; they therefore invented the bow and arrow and other tools and techniques, including the microlithic technique.

We can offer a concrete reason for the drastic changes that occurred in the Mesolithic age by referring to the notion of correspondence between the cognitive development of individual children and the evolving faculties of the human race. Consider that during the Mesolithic age mankind reached the developmental age of "9" (if we multiply the number 9 by 3,000 years, we will get 27,000 years; we get 12,000 years by subtracting 27,000 years from 39,000 years); in other words, the cognitive milestones that children typically reach at age nine correspond to those reached by human beings during the tenth millennium BCE.

Following Piaget's thought on the process of ontogenesis, the perception of three-dimensional space and feeling of simultaneity manifest themselves in children aged nine.²³ According to phylogenesis, this indicates that at this point in human history Mesolithic humans – that is, humans with the cognitive abilities of average nine-year-old children – could begin to free themselves from the confines of "this place" (in terms of space) and "now" (in terms of time), and thus engage in more complex theoretical cognition that allowed them to achieve drastic changes in their lifestyle. For example, having newly acquired the ability to perceive three-dimensional space, Mesolithic humans were afforded the opportunity (in the words of A. Leroi-Gourhan) to develop and adjust to the practice of individual hunting.²⁴

Greater cognitive ability also gave rise to the cult worship of ancestors, as human beings were now able to consider points in time beyond the present moment and examine their connection to the past. It is not coincidental that collective cemeteries began to appear during the Mesolithic period, as such a practice would not make sense unless humans were capable of remembering the past. Crucially, and in connection with this, in order to understand when crops would be ready for harvest, it was necessary for humans to achieve the perception of three-dimensional space, the understanding of object permanence, and the feeling of simultaneity. Having reached these milestones, primitive people in the Mesolithic period took the first steps toward the move from nomadic gathering to a reproducing economy based on agriculture.

²¹ Piaget J. The chosen psychological works, Moscow. 177.

²² Vygotsky L.S. Psychology of human development. Moscow. 233.

²³ Piaget J. The chosen psychological works, Moscow., 199.

²⁴ History of primitive society. Epoch of primitive tribal community, Moscow. 510–511.

Three millennia had passed since the Mesolithic age, leading to the events of a historical period called the “Neolithic revolution.” Following the comparison between ontogenesis and phylogenesis, mankind reached the developmental age of ten during the Neolithic period, when human beings begin to demonstrate the first signs of higher order reflection.²⁵ Increasingly analytical thought processes formed an intellectual foundation for the revelation of individual consciousness and individual interests, gradually freeing individuals from rigid adherence to traditional collectivist thought. This process, in turn, led to social differentiation within the community: First, human beings developed and filled roles within the social units of family and tribe; then, they developed a more complex social hierarchy that acknowledged the authority of the leaders of tribes and council of elders. The complete establishment of the perception of three-dimensional space afforded the appearance of farming agriculture based on irrigation. However, these intellectual skills were not sufficient for the formation of the state and invention of written language. Mastery of the skills of higher order reflection was necessary for these achievements, (following the analogy of ontogenesis, mankind had to be “11” years old.)²⁶

This historical event became possible after three thousand years. As the result of the emergence of individual personality during the period of the first civilizations, sovereigns that were capable of organizing and leading their communities came to the fore and spearheaded the formation of the first states. The appearance of reflection, or abstract thinking, was also necessary for the emergence of written language, which came into being in the fourth millennium BCE, and the emergence of the sciences, which occurred during this historical period.

Despite these intellectual developments, knowledge and individuality remained limited. The constraints on individual personality were only dissolved when humanity reached the stage of intellectual development corresponding to a 12-year-old child, that is, when human beings became capable of formal logic. By this stage in ontogenesis, humans have acquired all of the cognitive faculties necessary to reflectively develop their individual thinking and personality.²⁷ And in phylogenesis, when mankind is

“12” years old (*i.e.*, first millennium BCE), K. Jaspers’s so-called Axial epoch (when “consciousness was conscious of consciousness”) manifests itself. Thus, at this point in human history, human beings became capable of free-thinking, and prophets, sages, and philosophers appeared. According to ontogeny this developmental age is crucial for the intellectual development of mankind, since the achievement of reflection marks the realization of formal thinking.

Before turning to a more detailed review of the key provisions of the hypothesis, let’s formulate the main results of the research here.

1. Mankind (neoanthropic humans) has been developing intellectually since the day of its emergence;

2. Intellectual development is of immanent character and occurs with certain consistent regularities;

3. This regularity is defined on the basis of the hypothesis of the age periodisation of the intellectual evolution of the human race;

4. The age periodisation is conducted on the basis of the comparative analysis of intellectual development which occurs in ontogenesis and phylogenesis;

5. Scientific investigations show that according to ontogenesis, there are some age periods in the process of the intellectual development of our biological species in phylogenesis;

6. The investigations conducted by the author afford ground for the assertion that every 3,000 years mankind ages “one year” in the process of its intellectual development;

7. Considering that *Homo sapiens* (*i.e.*, neoanthropus) as a single species was formed approximately 39,000 years ago, it follows that humanity is now “13” years old and is on the verge of reaching the developmental age of “14.”

IV. Hypothesis verification through scientific facts

What are the basic scientific sources and research that inform my primary judgments about the above provisions and theses? Some critical beliefs relate to the question concerning the origins of neoanthropus. As confirmed by fundamental research²⁸

²⁵ Psychology. Vocabulary, Moscow., 279.

²⁶ Piaget J. The chosen psychological works, Moscow., 202.

²⁷ Vygotsky L.S. Psychology of human development. Moscow., 779.

²⁸ History of primitive society. Epoch of primitive tribal community, Moscow., 99; Grigoryev G.P. The beginning of the Upper Paleolithic and the origin of *Homo sapiens*,

and texts,²⁹ neoanthropus originated approximately 40,000 years ago. In his doctoral dissertation,³⁰ (By humankind, or humanity, we mean a spiritual unity, the bearer of which is neoanthropus.) Vishnyatsky's work indicates that 39,000 years ago humanity appeared, and so too did culture, with its complexities and refinements. During the Upper Paleolithic period, independently of each other and synchronously (which is especially surprising!), human beings in different parts of the world began creating and participating in cultural institutions, practices, and arts that signal greater cognitive sophistication.

Interestingly, the first Paleolithic drawings appeared about 30,000 years ago, when the intellectual development of mankind corresponded to the age of "three." Detailed information on the first Paleolithic drawings is outlined in the fundamental works of Z.A. Abramova,³¹ A.D. Stolyar,³² Y.A. Sher, and others.³³ The theory of the dynamics of the development of Paleolithic art developed by Stolyar definitely corresponds to certain provisions of our hypothesis that are based on Piaget's theory. In particular, the changes that took place in Paleolithic art during the Solutrean epoch (approximately 20,000 years ago) are positively explained by the fact that at that time mankind was "7" years old. This is the age when, in ontogenesis, the intellectual development of an individual enters the phase of the manifestation of concrete operations, namely, when the child has mastered the ability to build relationships simultaneously between two objects and to draw from this the correct logical conclusions (that is, the child can cognitively engage in reversible operations, according to Piaget). As noted above, cave drawings from the Solutrean epoch depict the relationship between two objects, rather than merely representing discrete objects as such, indicating a more advanced stage

(equivalent to age seven) in humanity's phylogenesis.

Crucially, the development of human cognition and society during the early prehistoric era occurred across Eurasia, following parallel trajectories marked by identical events within communities that were disconnected from one another. This synchronism of events has allowed archaeologists to distinguish such consecutive development periods in the Upper Paleolithic era as the Aurignac, Solutre, and La Madeleine, based on the similarity (although not uniformity) of artifacts, including tools and artworks, discovered in numerous regions, and therefore, to regard these epochs as absolute stages within an authoritative historical chronology.

A notable example of this that is described by A.P. Okladnikov³⁴ is the parallelism of the emergence of artifacts referred to as the "Paleolithic Venus" in multiple regions during the Aurignacian period (20 – 22 millennia ago). Archaeologists G.P. Grigoryev³⁵ and A.P. Chernysh³⁶ also provide valuable examples of this phenomenon, as does A.A. Formozov who, upon comparing the paleolithic cave drawings of the Lascaux (France) and the Kapova Cave (Bashkortostan), concluded that the synchronism of human developmental events that took place across Eurasia derived from the (single) universal psyche of humanity,³⁷ which corresponds to the universal natural laws of the intellectual development of human beings. This is further supported by the research of Z. A. Abramova, who also notes the synchronicity of the origins of paleolithic images and iconography.³⁸

However, it should be noted that the *parallelity* of events in the sociocultural life of primitive communities inhabiting various and remote regions of the world (and in general, of the way of life, as a series of norms and social practices, of people of that time), most clearly and prominently appears in the Mesolithic age.

Leningrad.; Foley R One more unique species: Environmental aspects of the evolution of human., Moscow., 62.

²⁹ Alekseev V.P., Pershits A.I. History of primitive society: Textbook for universities Moscow., 40.

³⁰ Vishnyatskiy L.B. Cultural Dynamics in the middle of the Late Pleistocene and the transition to the Upper Paleolithic (Abstract of dissertation for the competition of degree of Doctor of Historical Sciences). S.-P.

³¹ Abramova Z.A. Image of human in Paleolithic art of Eurasia, Moscow.

³² Stolyar A.D. The origin of fine art. Moscow.

³³ Sher Y.A., Vishnyatsky L.B., Blednova N.S. The origin of symbolic behaviour. Moscow.; Blednova N. S., Vishnyatsky L. B., Goldshmidt E. S., Dmitrieva T. N., Sher Y. A. Primitive art: problem of origin (under the redaction of Y. A. Sher).

³⁴ Okladnikov A.P. Morning of art, Leningrad. 73.

³⁵ Grigoryev G.P. The beginning of the Upper Paleolithic and the origin of Homo Sapiens, Leningrad.; Foley R. One more unique species: Environmental aspects of the evolution of human, Moscow., 118–119.

³⁶ Chernysh A.P. About absolute age of the Paleolithic monuments of Pridnestrovie // Stratigraphy and periodization of the Paleolithic of Eastern and Central Europe, Moscow., 188.

³⁷ Formozov A.A. Monuments of primitive art in the territory of USSR, Moscow., 25

³⁸ Abramova Z.A. Image of human in Paleolithic art of Eurasia, Moscow., 154.

It is known that the rudiments of agriculture appeared approximately at the same time in different parts of the world, although the specific tools used features of the community's way of life, climatic conditions, and flora and fauna in these regions (including the Near East, Southeast Asia, and Mesoamerica) were significantly different. Given the breadth of difference between cultures, it may be surprising that the transition from hunting and gathering to subsistence agriculture occurred with near simultaneity. This raises the question: Which economic, social, environmental, or other external factors (as well as the complex interactions among, and impact of, these factors) explain the parallelity of the global events that occurred in the Mesolithic Age?

This question has been investigated by A.P. Chernysh,³⁹ P.I. Borisovskiy,⁴⁰ V.P. Kabo,⁴¹ and L.A. Fainberg.⁴² Between the tenth and the eighth millennium BCE, when important events happened synchronously and independently of each other in different regions of the world, there emerged petroglyphs with multi-figures (a phenomenon that had not yet occurred in the Upper Paleolithic era), such as mass hunting scenes. These drawings were schematic, representing not a static collection of figures but rather the actual complexities of the dynamics of the events depicted, thus marking a critical change in humanity's capacities of reflection and representation, which A.A. Formozov discusses in his work in detail.⁴³

These images reveal that during this period people switched from exclusively collectivist action, including collectivist hunting, or hunting that was carried out through the coordinated action of the entire community, to individual hunting, which was not practiced during the Upper Paleolithic era. This is further confirmed by the archaeological evidence from the Mesolithic period, including bows and arrows, bone hooks for fishing, and one-seater boats

hollowed out from a single piece of wood, all of which were introduced during this period.

Literally all reputable scientists are unanimous in the opinion that during the Mesolithic period the way of life for people around the world fundamentally changed. Still, it remains unclear why this is so. Why, for many thousands of years during the Pleistocene epoch (about 30,000 thousand years), did humans never realize the achievements noted above, and why, beginning around the start of the Holocene epoch, did the lifestyle of primitive communities change so dramatically?

My investigations have led me to believe that external factors cannot explain the essence of parallel events – rather, it is the internal spiritual, or rather, intellectual character of humanity that does.

Three thousand years after the beginning of the Mesolithic period, significant changes in people's way of life took place around the world (a phenomenon which English archaeologist G. Childe has termed the "Neolithic revolution"). During the Neolithic period primitive communities moved away from the appropriating economy and toward reproductive activity. Other distinctive features of Neolithic life are captured by the words of ethnographer V.A. Shnirelman: "communities have become much larger now ... material culture was enriched by many new elements which measurably was caused by the deepening of the processes of labour specialization and the formation of craft. Central became the concept of vertical kinship, the system of power was perfected, the main features of religion began to be determined by the cult of ancestors."⁴⁴ During this era intercommunal ties began expanding. Communities that increased intercommunal contact rapidly built more complex societies with great innovation and cultural and intellectual advancements, while primitive collectives that remained insular (or isolated) did not.

The transition to the producing economy, which took place in the Neolithic age, is characterized by the parallelism of events which transpired within communities that did not interact with one another.⁴⁵ V.A. Shnirelman specifically points out that qualitative changes in people's way of life did not depend

³⁹ Chernysh A.P. About absolute age of the Paleolithic monuments of Pridnestrovie // *Stratigraphy and periodization of the Paleolithic of Eastern and Central Europe*, Moscow., 150.

⁴⁰ Borisovskiy P.I. *Old Stone Age of South and Southeast Asia*, Moscow., 159.

⁴¹ Kabo V.R. *Primitive preagricultural community*, Moscow., 255.

⁴² *History of primitive society. Epoch of primitive tribal community* (Edited by academician Y.V. Bromley), Moscow., 132.

⁴³ Formozov A.A. *Monuments of primitive art in the territory of USSR*, Moscow., 53–62.

⁴⁴ *History of primitive society. Epoch of primitive tribal community* (Edited by academician Y.V. Bromley), Moscow., 236.

⁴⁵ *History of primitive society. Epoch of primitive tribal community*, 263.

on economic practices.⁴⁶ In other words the processes characterizing the Neolithic era unfolded simultaneously among both farmers and pastoralists and among hunters and fishermen. The parallelism of processes has been observed not only by archaeologists, ethnographers, and historians, but also representatives of the natural sciences. Indeed, Nobel Prize winner I. Prigogine wrote: "I am always amazed that the bifurcation of the Neolithic appeared everywhere about a period of about ten thousand years ago, but it was in different forms in the Middle East, China, or pre-Columbian America."⁴⁷

These significant global changes, once mysterious, can be solved on the basis of my hypothesis. Accordingly, in the Neolithic age, mankind was developmentally "10" years old. So, by recognizing the intellectual capabilities of 10-year-olds and applying these observations to human evolution, one may understand the cultural changes that took place in the Neolithic period. (For this, it is sufficient to refer to the thesis of A.A. Shemshurin, which is devoted to the development of reflective abilities of younger adolescents.⁴⁸)

Three thousand years after the beginning of the Neolithic revolution the first human civilizations appeared. With the emergence of the first states came the inauguration of the first persons of the country – the pharaohs, kings, and high priests. The cultivation of a more complex political and social life resulted in other cultural developments as well. Writing appeared, and science developed. Epics were composed and folk heroes were created. Notably, all the known qualitative lifestyle changes that appeared during the era of the first civilizations occurred synchronously, as documented by philosophers, scientists,⁴⁹ and modern researchers in multiple fields.⁵⁰

The parallelism of these synchronously occurring changes is best explained by the immanent spiritual, or more precisely, intellectual character of humanity itself, rather than any particular external factor, or combination of external factors.

During this period, mankind, in its intellectual development, turned "11" years old. This is the age in ontogenesis when the child achieves formal thinking, or in Piaget's words, the subject becomes capable of reasoning hypothetically-deductively, that is on the basis of common premises without any necessary connection to reality or to one's own convictions and experience. At this stage the subject can arrive at conclusions by giving in to the necessity of reasoning itself in view of one of its forms, rather than relying on experience in order to gain understanding.⁵¹ If you map these intellectual abilities onto the process of phylogenesis, you can understand why writing was born and science appeared, and also why the first human beings qua rational agents in different parts of the world appeared through parallel synchronous processes and independently of each other. However, the appearance of a fully-fledged personality in that era was a rare occurrence. The achievement of hypothetical-deductive reasoning was only a beginning stage in the process of humanity's development, the full manifestation of which occurred only in the first millennium BCE.

Three thousand years passed, during which there occurred new parallel events, leading K. Jaspers to conceptualize the so-called "Axial epoch" of humanity. Regarding the impetus for these changes, Jaspers states: All this happened through reflection. Consciousness was conscious of consciousness and thinking made its object thinking.⁵² He further notes: "All these changes in human life can be called spiritualisation."⁵³ While the deepening and maturation of reflection crucially underpin the changes that occurred during the Axial epoch, axial time itself remains somewhat of an enigma that K. Jaspers tries to explain in the following manner:

The simplest explanation of the phenomena of axial time can be as if given ultimately by bringing them to the general social conditions favourable for the spiritual creativity of small

⁴⁶ History of primitive society. Epoch of primitive tribal community, 238-244.

⁴⁷ Prigogine I.R. Network society // Sociological researches, No 1, 27.

⁴⁸ Shemshurin A.A. Development of reflective abilities of younger teenagers. The abstract of the dissertation on competition of a scientific degree of the candidate of psychological sciences, https://static.freereferats.ru/_avtoreferats/01002746993.pdf

⁴⁹ Oppenheim A. Ancient Mesopotamia, Moscow., 26; Tokarev S. Religion in the history of the peoples of the world, Moscow, 332; Jaspers K. The origin and goal of history. Moscow, 42-44.

⁵⁰ Jestokanov S.M. On the typology of ancient civilizations //Metamorphosis of history, No 11; 51; Kruglikova G. I. Genesis of ancient civilizations of Mesopotamia and

Egypt: general and special // Questions of general history. Vol. 10, 132.

⁵¹ Piaget J The chosen psychological works., Moscow, 202.

⁵² Jaspers K. The origin and goal of history. Moscow, 33.

⁵³ Jaspers K. The origin and goal of history, 34.

states and small towns, the time of political fragmentation and the universal struggle.... All these sociological considerations are very significant and can serve as a basis for methodological research, but in the final analysis, they only shed light on the factual data, but do not give explanations of their causes, for all these facts themselves are an integral part of the spiritual phenomenon of axial time.⁵⁴

As a result, Jaspers concludes: "No one can fully understand what happened here: how the axis of world history arose!"⁵⁵ However, if we begin from the hypothesis of the age periodisation of the intellectual evolution of mankind, we are able to explain what Jaspers could not.

As previously noted, in the first millennium BCE mankind was "12" years old – the precise age at which reflexive thinking has developed, and due to this, the individual first forms his relatively stable personality. This is the time when a teenager can form original opinions, act on the basis of his conscious decisions, and behave independently. Similarly, in phylogenesis, the first prophets, philosophers, and sages (i.e. individuals who can think critically and independently), appear, which previously was impossible given the developmental limitations of more primitive humans

With the passage of another 3,000 years modern humans have emerged, and humanity has come to experience the process of globalization (for a deeper understanding of my position on the essence of globalization see R.S. Gadzhiev⁵⁶). Fundamentally, globalization is characterized by novelty, but also by coherence. After all, humanity is now "13" years old, and is gradually emerging from its childhood and becoming conscious of itself, its unity. This process of integration is characterised not only by economic, informational, or other aspects. In addition, the process of spiritual *rapprochement* among people of different nationalities, races, and religions is developing. Globalization is an immanent phenomenon engendered by a desire for the spiritual unification of mankind. We are only at the beginning of this unification, but the process is natural and, apparently, irreversible.

V. Conclusion

As demonstrated by the examples cited above, the hypothesis of the age periodisation of the intellectual evolution of society can answer the question of why global events took place in the historical process at specific points in time, and moreover, why the unfurling of human history and cognition followed parallel trajectories within communities that had had no contact with one another. The hypothesis is compatible with archaeological and scientific evidence, and it offers several advantages over competing theories. It allows us to understand a number of complex matters related to the intellectual evolution of humanity – for instance, how the first states came into being; it provides a consistent (non-contradictory) explanation of historical events, such as the transition from nomadic hunting and gathering to subsistence farming; and it clarifies problems concerning the parallel, synchronous, independent achievement of cultural milestones that occurred across Eurasia. As such, it offers a compelling alternative to Marxism and to other conceptions of social development that have failed to provide comprehensive, historically accurate, evidence-based responses to the crucial questions addressed in this article, which have long puzzled historians, archeologists, sociologists, and philosophers.

References

- Abramova Z A (1966) *Izobrajenie cheloveka v paleoliticheskom iskusstve Evrazii [Image of human in Paleolithic art of Eurasia]*. Moscow: Nauka publ, 224 pp.
- Alekseev V P, Pershits A I (1990) *Istoriya pervobytnogo obshestva: Uchebnik dlya vuzov [History of primitive society: Textbook for universities]*. Moscow: Visshaya shkola publ., 351 pp.
- Antropologiya (2003) [Anthropology], Moscow: Vldos publ., 272 pp.
- Blednova N S, Vishnyatsky L B, Goldshmidt E S, Dmitrieva T N, Sher Y A (1998) *Pervobitnoe iskusstvo: problema proisxozhdeniya [Primitive art: problem of origin]* (under the redaction of Y. A. Sher). Kemerovo: Nikals publ., 211 pp.
- Boriskovskiy P J (1972) *Drevniy kamenniy vek Yujnoy i Yuquvostochnoy Azii [Old Stone Age of South and Southeast Asia]*. Moscow: Nauka Publ., 175 pp.
- Vishnyatskiy L B (2000) *Verhnepaleoliticheskaya revolyutsiya: geografiya, hronologiya,*

⁵⁴ Jaspers K. The origin and goal of history, Moscow, 47.

⁵⁵ Jaspers K. The origin and goal of history, 48.

⁵⁶ Hajiyeu R.S. On globalization and globalism // Philosophy and social sciences. Belarus State University. Minsk, c.43-47.

prichini [Upperpalaeolithic revolution: geography, chronology, causes] // «*STRATUM plus*» journal, No 1: 245–272.

Vishnyatskiy L B (2006) *Kulturnaya dinamika v sereдинe pozdneqo pleistotsena i perexod k verxnemu paleolitu* [Cultural Dynamics in the middle of the Late Pleistocene and the transition to the Upper Paleolithic]. Avtoreferat dissertatsii na soiskaniye uchyonoy stepeni doktora istoricheskikh nauk [Abstract of dissertation for the competition of degree of Doctor of Historical Sciences], Saint-Petersburg, 37 pp.

Vsemirnaya istoriya (2011) [World history]: in 6 vols, Vol.1, Moscow, Institute of Universal History RSA, 822 pp.

Vygotskiy L S (2004) *Psixologiya razvitiya cheloveka* [Psychology of human development]. Moscow: Smysl Publ., 1136 pp.

Hajiyeve R S (2014) O globalizatsii i globalizme [On globalization and globalism] // *Scientific journal Filosofiya i sotsialnye nauki* [Philosophy and social sciences], Belarus State University, Minsk, No 2: 43–47.

Hajiyeve R S (2015) Ob odnoy zakonomernosti istoricheskogo protsessa [About one regularity of historical process]. // *Filosofiya. Tolerantnost. Globalizatsiya. Vostok i Zapad – dialog mirovozzreniy: tezisy dokladov VII Rossiyskogo filosofskogo kongressa* [Philosophy. Tolerance. Globalization. East and West – a dialogue of worldviews: Theses of reports of the VII Russian Philosophical Congress]. In 3 volumes. Vol. II – Ufa, Editorial and Publishing Center of Bashkir State University, 300–301.

Hajiyeve R S (2007) *Problema smisla jizni* [Problem of the meaning of life]. Baku: Teknur Publ. 158 pp.

Goncharov N P (2013) Domestikatsiya rasteniy [Domestication of plants] // *Vavilovskiy Jurnal Genetiki i Seleksii* [Vavilov Journal of Genetics and Selection], 17(4/2): 884–900.

Grigoryev G P (1968) *Nachalo verkhnego paleolita I proisxojdenie Homo sapiens* [The beginning of the Upper Paleolithic and the origin of Homo sapiens]. Leningrad: Nauka publ., 175 p.

Gulyaev V I (2012) *Tayni drevnih gorodov: Blijniy Vostok i Mezoamerika* [Mysteries of ancient cities: Near East and Mesoamerica]. Moscow: AST-Press Kniga publ., 320 pp.

Drevnie tsivilizatsii (1989) [Ancient civilizations] (under redaction of G. M. Bongard-Levin), Moscow: Mysl publ., 479 pp.

Jestokanov S M (1997) K voprosu o tipologii drevnikh tsivilizatsiy [On the typology of ancient civilizations]. *Metomorfozy istorii* [“Metamorphosis of history” journal], No 11: 42–57.

Zablozka Y (1989) *Istoriya Blijnego Vostoka v drevnosti* [History of Near East in ancient times]. Moscow: Nauka publ., 415 pp.

Istoriya pervobytnogo obshestva. Epoha pervobytnoy rodovoy obshiny (1986) [History of primitive society. Epoch of primitive tribal community] (Edited by academician Y.V. Bromley). Moscow: Nauka Publ., 574 pp.

Kabo V R (1986) *Pervobytnaya dozemledelcheskaya obshina* [Primitive preagricultural community]. Moscow: Nauka Publ., 304 pp.

Kruglikova G I (2008) Genezis drevnikh tsivilizatsiy Mesopotamii i Egipta: obshee i osobennoe [Genesis of ancient civilizations of Mesopotamia and Egypt: general and special]. *Voprosy vseobshey istorii* [Questions of general history] 10: 131–175.

Lamberg-Karlovskiy C V (1993) Gordon Chayld i kontseptsiya revolyutsii [V. Gordon Childe and concept of revolution]. *Vestnik Drevney Istории* [Journal of Ancient History] 4(207): 90–105.

Martsinkovskaya T D (1998) *Istoriya detskoy psixologii: Uchebnik dlya stud. Pedagogicheskikh vuzov* [History of child psychology: Textbook for students of pedagogical universities]. Moscow: VLADOS Publ., 272 pp.

Mellaart J (1982) *Drevneyshie tsivilizatsii Blijnego Vostoka* [Earliest civilizations of Near East]. Moscow: Nauka publ., 149 pp.

Munchaev R M. (1991) Kultury dopismennoy Mesopotamii [Cultures of preliterate Mesopotamia]. In *Nauka i chelovechestvo* [Science and humanity], Moscow: Znaniye Publ., 400 pp.

Okladnikov A P (1967) *Utro iskusstva* [Morning of art]. Leningrad: Iskusstvo Publ., 136 pp.

Oppenheim A. (1990) *Drevnyaya Mesopotamiya* [Ancient Mesopotamia]. Moscow: Nauka Publ., 319 pp.

Piaget J (1969) *Izbranniye psixologicheskiye trudy* [The chosen psychological works]. Moscow: Prosvesheniye Publ., 659 pp.

Prigogine I R (2009) Setevoye obshestvo [Network society]. *Sotsiologicheskiye issledovaniya* [Sociological researches] No 1: 24–28.

Psixologiya. Slovar (1990) [Psychology. Vocabulary] (Under the general edition of A.V. Petrovsky and M.G. Yaroshevsky), Moscow: Politizdat Publ., 494 pp.

Stolyar A D (1985) *Proisxojenje izobrazitel'nogo iskusstva [The origin of fine art]*. Moscow: Iskustvo publ., 298 pp.

Teilhard de Chardin P (1987) *The phenomenon of man*. Moscow: Nauka Publ., 240 pp.

Tokarev S A (1976) *Reliqiya v istorii narodov mira [Religion in the history of the peoples of the world]*. Moscow: Politizdat Publ., 575 pp.

Formozov A A (1980) *Pamyatniki pervobitnogo iskusstva na territorii SSSR [Monuments of primitive art in the territory of USSR]*. Moscow: Nauka publ., 136 pp.

Foley R (1990) *Eshe odin nepovtorimyy vid. Ekologicheskiye aspekty evolyutsii cheloveka [One more unique species: Environmental aspects of the evolution of human]*. Moscow: Mir publ., 368 pp.

Chernysh A P (1974) *K voprosu o sootnoshenii granits paleolit-mezolit i pleystotsen-golotsen i osobennosti razvitiya materialnoy kultury i prirodnoy sredy Prikarpatya [On the question of the relationship between the boundaries of the Paleolithic-Mesolithic and Pleistocene-Holocene and features of the development of material culture and the natural environment of the Prikarpatie] // Pervobytny chelovek, ego materialnaya kultura i prirodnaya sreda v pleystotsene i golotsene [Primitive human, his material culture and natural environment in the Pleistocene and Holocene]*. Moscow: Publishing house of the Institute of Geography of the USSR Academy of Sciences, 315 pp.

Chernysh A P (1965) *Ob absolyutnom vozraste paleoliticheskix pamyatnikov Pridnestrovya [About absolute age of the Paleolithic monuments of Pridnestrovie] // Stratigrafiya i periodizatsiya paleolita Vostochnoy i Tsentralnoy Evropy [Stratigraphy and periodization of the Paleolithic of Eastern and Central Europe]*. Moscow: Nauka publ., 232 pp.

Chesnov I V (1977) *Zemledelcheskiye kultury kak etnogeneticheskiy istochnik [Agricultural crops*

as an ethnogenetic source] In *Rannyyaya etnicheskaya istoriya narodov Vostochnoy Azii [Early ethnic history of the peoples of East Asia]*, 109–136. Moscow: Nauka Publ.

Shemshurin A A (2005) *Razvitie reflektivnykh sposobnostey mladshix podrostkov. Avtoreferat dissertatsii na soiskanie uchenoy stepeni kandidata psikhologicheskikh nauk [Development of reflective abilities of younger teenagers. The abstract of the dissertation on competition of a scientific degree of the candidate of psychological sciences]*. https://new-dissert.ru/_avtoreferats/01002746993.pdf

Sher Y A, Vishnyatsky L B, Blednova N S (2004) *Proisxojenje znakovogo povedeniya [The origin of symbolic behaviour]*. Moscow: Nauchny mir, 280 pp.

Jaspers K (1994) *Smysl i naznachenkiye istorii [The origin and goal of history]*. Moscow: Republic Publ., 527 pp.

Barker G. (2009) *The agricultural revolution in prehistory: Why did foragers become Farmers?* Oxford etc.: Oxford univ. press, 598 pp.

Hacıyev R S (2015) *Tarixi prosesin inkişaf dinamikası və ya globallaşmaya aparıcı intellektual təkamül yolu [The dynamics of development the historical process or the intellectual evolutionary path towards globalization]*. Baku, Khazar University Publ., 352 pp. (in Azerbaijani)

Hajiyeu R S (2013) On substantiation of hypothesis of age periodization in intellectual evolution of humanity. XXIII World congress of philosophy: Philosophy as Inquiry and way of life, University of Athens, School of philosophy, Athens 04–10 August 2013, Abstracts, 818.

Stein G J (2009) Excavations at Tell Zeidan, Syria. The Oriental Institute. News & Notes. no. 203, 3–7. <http://oi.uchicago.edu/sites/oi.uchicago.edu/files/uploads/shared/docs/nn203.pdf>

HOW TO ASSIGN WEIGHTS TO VALUES? SOME CHALLENGES IN JUSTIFYING ARTIFICIAL EXPLICIT ETHICAL AGENTS

Silviya Serafimova

Institute of Philosophy and Sociology,
Bulgarian Academy of Sciences,
Sofia, Bulgaria
e-mails: silvija_serafimova@yahoo.com

Abstract:

In this paper, one of my primary objectives is to analyze why adopting particular machine-learning techniques and using a moral AI as an adviser is an insufficient condition for eradicating racist human attitudes. By outlining some difficulties in justifying what artificial “explicit ethical agents” in Moor’s sense should look like, I explore why, even if the development of machine-learning techniques can be accepted in epistemic terms, it does not follow that the techniques in question will have a positive impact in changing immoral human behavior.

Key words: machine ethics, machine-learning techniques, artificial explicit ethical agents, conscious and unconscious biases.

1. Introduction

In the era of the “big data,” we are witnessing the “transmogrification of the person as a coherent representational unity into a constellation of data points,” which is a result of “the symbolic alchemy of algorithmic computation” (Markham et al., 2018, pp. 1–2). However, the alchemy in question has very practical consequences for the transformation of moral and social practice, since it requires both re-

vising and enriching the ethical framework of what moral agency and moral responsibility should look like so that they maintain their normative validity within the “epistemic shift toward ‘algorithmic knowledge production’” (Metcalf et al., 2016, p. 6). In addition to emphasizing how the shift in question raises the issue of the tension between so-called good social research and big data research (Ibid.; Markham et al., 2018, p. 5), one should also pay special attention to the way in which algorithmic knowledge production is inevitably tied to the necessity of “producing” new knowledge about morality – having been influenced by the challenges brought about by the increase in algorithmic computation.

It is not an accident that decision-making based upon algorithms remains “a standout question in machine ethics” (Mittelstadt et al., 2016, p. 11), which in turn necessitates that one differentiate the requirements for moral agency when applied to humans and artificial intelligence agents respectively. In this context, special attention should be paid to the issue of whether or not elaborating upon machine-learning techniques, and then applying them for the purposes of building a strong AI scenario (in which human-level cognition is reached in a computer) (Savulescu and Maslen, 2015; Klincewicz, 2016), can enable algorithmic decision-making to achieve a level of moral autonomy similar to that of

human decision-making.¹ Since one of the crucial problems in the field of the ethics of algorithms concerns how unfair outcomes lead to discrimination (Mittelstadt et al., 2016, p. 5), my primary objective is to clarify why even when the development of machine-learning techniques may be non-contradictorily justified in epistemic terms, it does not follow that the techniques in question will have the positive effect of transforming immoral human behavior, as is demonstrated by the hypothetical scenario of John's racist attitudes (Klincewicz, 2016).

2. Exemplifying Some Machine-Learning Techniques from the Perspective of Machine Ethics

2.1. Savulescu and Maslen's Model of a Moral AI as an Adviser

Arguing for the opportunities introduced if an AI system is interpreted as a complex of subsystems, including moral environment monitor, moral prompter, moral adviser, and moral observer (Savulescu and Maslen, 2015, p. 84), Savulescu and Maslen claim that before offering advice in the first instance, the AI should ask the agent to "indicate which of a long list of morally significant values or principles he holds and is guided by" (Ibid., p. 88; Klincewicz, 2016, p. 174). The agent is also asked to assign a weight (between 0 and 1) to each value, including values such as autonomy, benevolence, non-malevolence, justice/fairness, legality, environmental protection, family/significant relationships, fulfilling duties/commitments/promises, and maximizing net utility.² Regarding the different scenarios of axiological priorities, "the AI would compute the extent to which the courses of action open to the agent would uphold or compromise these values (fully uphold value = 1; fully compromise value = -1), amplifying or diminishing based on the weight indicated by the agent" (Ibid.). The AI would then

use the ranking of weighted values to suggest the best course of action (Ibid.).³

2.2. Applying Savulescu and Maslen's Model to Klincewicz's Scenario of John's Racism

Extrapolating from Savulescu and Maslen's model of AI to one hypothetical scenario, namely, to that of a hypothetical person, a white American named John "who could report that he just saw a police officer verbally abusing a Chinese person" (Klincewicz, 2016, p. 177), M. Klincewicz examines how AI can function as a moral environment monitor, moral prompter, moral adviser, and moral observer in respect to the different aspects of the aforementioned scenario. In this context, he raises the following question: If John is "then hooked up with a moral AI with all of its different subsystems, as presented by Savulescu and Maslen, and faced with an identical situation," can we expect that he will change his behavior and react differently? (Ibid.).

According to Klincewicz, the moral environment monitor in the AI may tell John that the person is Chinese and that this is morally relevant. In turn, being a racist, John will use this information to make the decision that he should not report the police officer's abuse because "This is how racism works" (Ibid.). Regarding the function of AI as a moral prompter, it could tell John that the situation with the police officer involves a moral dilemma, which assumes the capacity to react "to an observed transgression of laws by someone else" (Ibid.). Thus, John may find it helpful to know that in the given case, the laws are broken by a police officer, but he will "also be compelled by his racist views to not do anything about it" (Ibid.). "Again, this is how racism works" (Ibid.). As Klincewicz relevantly argues, even if the AI functions as a moral adviser, it is

¹ For the differences between traditional (human) decision-making and decision-making based on algorithms, see Mittelstadt et al., "The Ethics of Algorithms".

² According to this model, an agent who strongly cares about not harming others, only a little about legality, and not at all about protecting the environment "might assign them weights of 1, 0.5 and 0, respectively" (Ibid.).

³ In turn, developing machine-learning techniques to suggest the best possible course of action is not an exception: it is preceded by systems such as those based upon the ILP method (inductive logic programming) (Lavrač and Dzeroski, "Inductive Logic Programming"). The latter concerns the role of inductively learning relations represented as first-order Horn clauses, which are supposed to provide a model of mediating abstract relationships between prima facie duties and an agreed-upon correct action when ethical dilemmas occur (Anderson, Anderson, and Armen, "An Approach to Computing Ethics.").

highly unlikely that John would act differently (Ibid., p. 178).⁴

3. Is the Project of Moral AI as an Adviser Accomplishable?

3.1. Some Moral Problems Regarding the Discrimination Preservation in Data Mining

Before specifying the moral challenges for a moral AI as an adviser, it is important that one outline why machine-learning techniques can be precisely applied from an epistemological point of view, but still make an unsatisfactory impact upon changing immoral human attitudes. One of the main reasons for the discrepancy between the criteria of epistemological verification and normative validity in morality, as projected into the field of algorithms investigations, rests upon “the ethos of reducing intelligence to a numeric value” (Markham et al., 2018, p. 4).⁵ In turn, this ethos raises the question of how one ascribes weights to values.

On a macro-methodological level, the origin of the deceptive assumption saying that epistemic accuracy is a necessary and sufficient condition for arguing for moral “accuracy” can be traced back to another, more general “deceptive premise” – to the premise that “knowledge derived from data analytics is true (or has strong truth value) because of the objective qualities of data” (Ibid.). Judging by the aforementioned specifications, one can argue that elaborating upon data mining investigations may bring some hints to light, namely, how one can ques-

tion the status of algorithms as being impartial, e.g. as being “without any “subjective” interference” (Ibid.). However, one should also take into account that relying upon such hints is just the tip of the iceberg.

According to Romei and Ruggieri, data mining and machine-learning models “extracted from historical data may discover traditional prejudices that are endemic in reality (i.e. taste-based discrimination), or patterns of lower performances, skills or capacities of protected-by-law groups (i.e. statistical discrimination)” (Romei and Ruggieri, 2014, p. 39). Mining algorithms themselves “may then assign the status of general rules to such practices, with the result that these rules may be deeply hidden within a data mining classifier” (Ibid.). What is of special interest in terms of discrimination prevention is Romei and Ruggieri’s theory of the four overlapping strategies for discrimination prevention in analytics. These strategies include 1) controlled distortion of the training data 2) integration of anti-discrimination criteria into the classifier algorithm 3) post-processing of classification models and 4) modification of predictions and decisions for the purposes of maintaining a fair proportion of effects between protected and unprotected groups (Ibid., p. 40; Mittelstadt et al., 2016, p. 8).

Referring to the aforementioned second criterion, while interpreting the possibility for building machine-learning techniques, as implied in Savulescu and Maslen’s AI model, I argue that the main reason for the difficulties posed is that the techniques in question can provide a certain epistemological clarity, but not a moral clarity by default. This is due to the fact that assigning a weight to a value is not a uniform process when an autonomous moral agent is supposed to weight only one value as highest once and for all. Another part of the problem is that moral agents’ plurality of choices regarding values is driven by complex factors which are neither fixed nor purely moral. These factors include biases that could be both conscious and unconscious, as well as being of a moral or immoral origin. Furthermore, their mutually inclusive configurations affect whether the ‘moral’ outfit of the AI is partially or fully dependent upon the values of its designers.

Regarding the particular problems with machine-learning techniques as displayed in the hypothetical case of John, I argue that the main issue is that he cannot identify only one value as being the highest, for reasons mentioned above. If John were

⁴ Some other attempts at building moral AI as advisers (such as those of MedEthEx and EthEl) show that even their creators (Anderson, Anderson, and Armen, “An approach to Computing Ethics.”) confess that they (especially, EthEl) only minimally meet the requirements of what Moor calls an explicit ethical agent – a machine which can calculate by itself the best action when faced by ethical dilemmas, relying upon a given ethical principle (Anderson and Anderson, “Machine Ethics”, 24; Moor, “Difficulty of Machine Ethics”).

⁵ Markham et al. argue that the ethic is indirectly engendered through “the absence of questioning the validity of variables” (Ibid.). I emphasize, however, that this is just the tip of the iceberg, since most problems derive from the fact that one’s choice is rarely determined by one single variable. In other words, even if one has critically questioned the validity of variables as such, one cannot accurately predict how their mutual interference affects one’s choice in both the long and short term. This issue brings us back to the debate about the complexity of different biases, as well as whether or not they should be overcome.

following the rules of the process of valuation as outlined by Savulescu and Maslen, he would rate only one value as being the highest – by ascribing to it the value of “one.” This, in turn, would make calculating the best possible scenario for John highly problematic, since it would present the AI model with one initially limited picture of John’s values.

For instance, in his general value system, John may rate both justice and legality highly, and double down on his act of non-interference by assuming (from the perspective of his racist ideology) that, in the case with the Chinese person, justice coincides with legality. Specifically, John can believe that Chinese people are not “valuable” as people, ergo it is “fair” (legal and just) for a Chinese person to be subjected to a discriminatory offense if the offense is committed by an authority figure. John may assign one point to both justice and legality. Thus, the computation of this morally problematic scenario will be evaluated as being morally appropriate according to his distorted vision of morality. Accordingly, moral computation would have zero impact upon John’s immoral attitudes, that is, it would fail to generate moral insights that could lead John to adopt morally good (anti-racist) attitudes toward Chinese people. This failure to achieve moral learning could also be the result of John’s conviction that verbal abuse is legal, firstly because it is exerted by a representative of the authorities, and secondly because he does not consider verbal abuse to constitute ‘real’ (physical) harm.

Let us assume, for hypothetical purposes, that John can assign one point to more than one moral value. As I will show below, this is just the tip of the iceberg of the process of moral computation, since highly weighted values may vary in their mutual relations under different circumstances (i.e. depending on the actual situation, certain values may ‘win out,’ or carry greater weight), while having an equal value in John’s general theoretical value system. Secondly, John may ascribe completely different “contents” to a given category which is ranked by the designers of the AI as belonging to the set of values. In other words, the problem may also derive from the fact that, by one and the same value, John and the aforementioned designers can understand different, sometimes even mutually exclusive ‘things,’ which in turn would undermine the successful application of machine-learning techniques to moral questions. A reason for such a methodological discrepancy is the complex role of biases – those

of both John and the designers of the given AI system – which are directly or indirectly influenced by what John and the designers understand by moral facts (data) and moral reasons.

3.2. The Role of Biases in Building Artificial Explicit Ethical Agents

Arguing for the distinction between conscious and unconscious biases is also of crucial importance for providing a distinction between the human decision-making process and that based on algorithms.⁶ Both processes result from and have an impact upon the specificities of so-called artificial implicit and explicit ethical agents in Moor’s sense (Moor, 2006, pp. 19-20). These clarifications are required, since the differences in the decision-making processes assume that one distinguishes between different types of responsibilities,⁷ taking into account the way in which they are related to the diverse requirements for holding one accountable as an autonomous (im)moral agent. Algorithms that make decisions and are considered blameworthy agents, as Floridi and Sanders suggest (Floridi and Sanders, 2004), cannot be regarded as blameworthy in the same sense as human agents – if for no other reason than, at least, due to the fact that humans, as moral agents, have more complex moral motivations; consequently, their moral behavior is apparently guided by both conscious and unconscious biases of different origins.

⁶ Algorithms “inevitably make biased decisions” (Mittelstadt et al., “The Ethics of Algorithms,” 7) because their design and functionality reflect the values of the designers “and intended uses, if only to the extent that a particular design is preferred as the best or most efficient option” (Ibid.). This in turn raises the issue of to what extent the inevitable errors, as manifested in the datasets processed by algorithms (Ibid.), can be described as ‘human’ errors and to what extent they can be examined as purely algorithmic errors. Since this topic is not an object of investigation in the current article, I would briefly argue that there are at least two possible scenarios which are relevant in moral terms. Either there are no specifically algorithmic errors, which shifts the focus back again to the origin of the biases behind human mistakes, or there are purely algorithmic errors. If so, however, the chances of building ‘moral’ machines which are supposed to be not only explicit ethical agents, but also agents with unquestionably moral (in its ‘content’) behavior, become slight.

⁷ For the different types of responsibilities addressed from the perspective of algorithmic decision-making or so-called traditional decision-making, see Floridi and Sanders, “On the Morality.”

3.2.1. *The Role of Unconscious Biases*

Pointing out the occurrence and the crucial role of unconscious biases, however, is only the first step in the analysis which aims to reveal why a moral AI cannot function as an adviser for changing immoral human behavior. Elaborating upon the previous discussion, we may argue that John can assign one point to non-malevolence, ranking it as the highest value in his general moral system, without necessarily considering verbal offense a matter of harm. That is why John may refuse to intervene on behalf of the offended Chinese person – not only because he has some general racist attitudes towards Chinese people, but also because he believes that verbal abuse is not ‘real’ abuse, as I have pointed out in Section 3.1. Judging by the given example, I argue that the moral motivation of a full ethical agent in Moor’s sense, who demonstrates immoral behavior, is guided by different biases. It is their complexity that complicates the evaluation of the machine-learning techniques when examined from the perspective of the moral AI as an adviser.

On a macro-methodological level, I demonstrate that adopting act utilitarianism,⁸ which is based upon recognizing what could be coined moral arithmetic (Anderson and Anderson, 2007, p. 18) and thus, is applicable to artificial explicit ethical agents, is an insufficient condition for elaborating upon successful machine-learning techniques in moral terms. Returning to the previous example, John’s neglecting to report abuse is certainly evaluated as being morally wrong, but this does not say enough about John as an autonomous moral agent or his complex (im)moral motivation.

3.2.2. *The Role of Conscious Biases*

The assumption that if a given bias is conscious, it can be changed when one acquires relevant knowledge about it, is a rather wishful prediction. Klincewicz examines two hypotheses: firstly, whether or not John will change his mind in light of rationally acceptable reasons provided by the AI if he is aware of the conflict between his racism and ethical theory, and secondly, whether or not John will be somehow tempted to trust the AI more than other human agents and thus change his racist attitudes (Klincewicz, 2016, p. 180). However, we

should keep in mind that even in an inter-human context, knowing that a given ideology is an ideology (in the best possible scenario) does not contribute to making its performative power less influential by default.

As Klincewicz’s hypotheses show, making agents conscious of the role of a given bias is a necessary condition for beginning to evaluate the reasons behind human actions, but it does not necessarily follow that such knowledge would lead to one’s emotional, moral, and social willingness to reject the ideology that serves as a trigger for one’s immoral behavior. In this context, one could point out that the methodological difficulties concerning moral and any other bias are due to the fact that bias is firstly a problem for and of humans, and only secondly (and in consequence of this) a problem for machine learning. Judging by the aforementioned investigations, I argue that the gist of these dilemmas in the field of machine ethics can be formulated as follows: What would make agent-computer trust a more successful means of achieving a moral transformation than agent-agent trust, and thus facilitate the process of eradicating racist and any other ideology?

3.2.3. *Agent-Computer Trust versus Agent-Agent Trust*

Before analyzing whether or not it is possible for a moral AI as an adviser to become an explicit ethical agent in Moor’s sense, we should also trace the role of the varieties of biases as they determine the behavior of the AI’s designers.

The challenges for designers in building machine-learning systems can be summarized as having two main aspects, at least. Firstly, there is no fool-proof way for the designers of moral AI as advisers to ensure that they avoid implementing their own biases, even if they are aware of them. In other words, just as humans who are advised by the AI may still perceive moral computation through the lens of their own biases, the AI’s designers are inevitably influenced by both conscious and unconscious biases. Secondly, if the designers cannot avoid the aforementioned biases, we can argue for only weak AI scenarios,⁹ and consequently, for AI to

⁸ According to act utilitarianism, a given action is evaluated as right or wrong due its consequences, namely, whether or not we argue for a good or bad consequence in moral terms.

⁹ According to the weak AI scenarios, weak AI does not aim at human-level cognition, but only to gather, update, and compute data to assist human agents in the process of moral-decision making (Savulescu and Maslen, “Moral

be implicit ethical agents in Moor's sense – that is, tools of moral computation which can only follow the particular ethical rules with which they have been programmed.

On a practical level, this would mean that one can consider artificial intelligence agents 'as if' they are full ethical agents, which in turn implies that a given moral AI as an adviser would be completely dependent upon the designer's way of assigning weights to values. If we apply this insight at a macro-methodological level, it indicates that it is problematic to argue for agent-computer trust, since this is just a matter of mediating agent-agent trust by AI systems. Thus, we face one more fundamental question: What would be the necessary and sufficient reasons to give priority to agent-computer trust over agent-agent trust, taking into account the fact that the chances of humans successfully recognizing each other as full ethical agents in Moor's sense (Moor, 2009)¹⁰ are significantly higher than those for artificial intelligent agents?

In turn, assigning weights to values depends upon whether the designers of the AI consider this scale to be objectively or subjectively determined. In the first case, one could argue that these values are objective, and therefore, unquestionable. Going back to the hypothetical scenario with John, this could mean that there is an increased risk that his racist ideology will get stronger; if the machine-learning techniques construe a given scale of values as objective, then John could simply become even more confident in the moral rightness of his scale due to his racist ideology and psychology.

By contrast, if the designers accept that moral values are subjectively determined, this could impede their ability to rank moral values, as they would be unable to assign an objective weight to values and to designate only one value as being objectively the highest. This in turn would imply that

the designers of the machine-learning techniques that should be adopted when building a moral AI as an adviser would face many difficulties in providing value scales that cover the broadest set of possibilities for non-contradictorily coupling weights with values.

Furthermore, adopting such an approach would entail accepting that every single moral AI which functions as an adviser would be predetermined to function as an implicit ethical agent in Moor's sense. This would result not from insufficient possibilities for developing machine-learning techniques, but due to the lack of one single, unquestionable outline of the normative validity of the chosen values. Regarding the scale of values itself, it would mean that every attempt at assigning weights to values faces the dilemma, "Whose morality? Which rationality?" Consequently, even if a given scale of values were somehow justified, it would not avoid the issue of how moral this moral AI as an adviser could be, nor would it specify whether or not it was sufficiently moral in normative terms.

4. Conclusion

Outlining the challenges which accompany the implementation of certain machine-learning techniques in building an AI as an explicit ethical agent requires taking into account the following specifications. One should clarify why even if we have precisely elaborated some of the aforementioned techniques in epistemic terms, their application to the project of designing a moral AI as an adviser does not guarantee that the latter obtains an ethical competence which is similar to that of humans as full ethical agents in Moor's sense.

By examining a hypothetical scenario in which one's racist values can be changed by a moral AI as an adviser, I demonstrate that although discrimination-aware data mining can be non-contradictorily designed and applied by properly weighting a set of values, this does not necessarily result in changing someone's discriminatory attitudes towards others. The particular reasons for reaching the conclusion that AI cannot turn into an explicit ethical agent in Moor's sense trace back to the level of how the epistemological verification of machine-learning techniques is only a necessary condition for arguing for the normative validity of artificial intelligence agents' decision-making process in moral terms.

Enhancement," 84; Klineciewicz, "Artificial Intelligence," 175).

¹⁰ As Moor relevantly argues, whether or not robots can become full ethical agents is "a wonderful and speculative topic, but the issue need not be settled for robot ethics to progress" (Moor, "Four Kinds of Ethical Robots"). That is why he suggests that *explicit ethical agents* become "a paradigm target example of robot ethics," since "such robots would be sophisticated enough to make robot ethics interesting philosophically and important practically, but not so sophisticated that they might never exist" (Ibid.).

I raise a hypothesis that the aforementioned difficulties occur due to the implementation of three mutually related factors. Firstly, due to the complexity of humans as autonomous moral agents (full ethical agents) whose moral behavior cannot be evaluated by adopting act utilitarianism alone, namely, by judging the moral agent by his or her actions alone. This complexity brings the second factor to light, namely, that of the distinction between a decision-making process conducted by humans and an algorithmic decision-making process. In turn, the reasons for the differences regarding the decision-making processes in question can be ascribed to two additional factors: to that of moral motivation, which is driven by both conscious and unconscious biases in a different manner when applied to humans and artificial intelligence agents respectively, as well as to the way in which so-called moral data is considered verifiable while arguing for moral facts whose epistemological validity should be representative of their normative validity. Last but not least, the third factor concerns the differences between humans as full ethical agents and artificial implicit and explicit ethical agents. These differences affect the lack of unquestionable proofs for arguing that agent-computer trust can better contribute to changing one's immoral attitudes.

As a crucial part of the problem concerning whether or not a moral AI as an advisor can function as an explicit ethical agent in Moor's sense, I would point out the fact that there is no unquestionable correspondence between being a good (in the sense of acting properly) explicit ethical agent in Moor's sense and acting morally as an explicit ethical agent. In other words, the debate on machine-learning techniques faces one more challenge if elaborated upon within strong AI scenarios. Even if such a model of a moral AI as an advisor is successfully built in the distant future, there are significant risks that it could reinforce, or encourage the recognition of, an immoral decision-making process, disguised as a moral one. In this case, we risk ending up with what Dellot and Wallace-Stephens call an AI bubble,¹¹ namely, the conclusion that machines are more intelligent than moral.

¹¹ According to Dellot and Wallace-Stephens, an AI bubble is a bubble with machines that are more artificial than intelligent (Dellot and Wallace-Stephens, "What is the Difference between AI & Robotics?").

References

- Anderson M, Anderson S L, Armen C (2006) An Approach to Computing Ethics. *IEEE Intelligent Systems* 21 (4): 56–63.
- Anderson M, Anderson S L, Armen C (2006) MedEthEx: A Prototype Medical Ethics Advisor. In: *Proceedings of the Eighteenth Conference on Innovative Applications of Artificial Intelligence*, 1759–1765. Menlo Park, CA: AAAI Press.
- Anderson M, Anderson S L (2007) Machine Ethics: Creating an Ethical Intelligent Agent., *AI Magazine* 28 (4): 15–26.
- Dellot B, Wallace-Stephens F (2019) What is the Difference between AI & Robotics? Last modified 23.09.2019. <https://medium.com/@thersa/what-is-the-difference-between-ai-robotics-d93715b4ba7f>.
- Floridi L, Sanders J W (2004) On the Morality of Artificial Agents. *Minds and Machines*. 14 (3): 349–379.
- Klincewicz M (2016) Artificial Intelligence as a Means to Moral Enhancement. *Studies in Logic, Grammar and Rhetoric* 48 (61): 171–185.
- Lavrač N, Dzeroski S (1997) *Inductive Logic Programming: Techniques and Applications*. Chichester, UK: Ellis Horwood.
- Markham A N, Tiidenberg K, Herman A (2018) Ethics as Methods: Doing Ethics in the Era of Big Data Research – Introduction. *Social Media+ Society*, July-September: 1–9.
- Metcalf J, Keller E F, Boyd D (2019) Perspectives on Big Data, Ethics, and Society. The Council for Big Data, Ethics, and Society. Last modified July 29, 2019. <http://bdes.datasociety.net/council-output/perspectives-on-big-data-ethics-and-society/>
- Mittelstadt B D, Allo P, Taddeo M, Wachter S, Floridi L (2016) The Ethics of Algorithms: Mapping the Debate. *Big Data and Society*, July-September: 1–21.
- Moor J H (2006) The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems* 21 (4): 18–21.
- Moor J H (2009) Four Kinds of Ethical Robots. *Philosophy Now. A Magazine of Ideas* 72, https://philosophynow.org/issues/72/Four_Kinds_of_Ethical_Robots.
- Romei A, Ruggieri S (2014) A Multidisciplinary Survey on Discrimination Analysis. *The Knowledge Engineering Review* 29 (5): 582–638.

- Savulescu J, Maslen H (2015) Moral Enhancement and Artificial Intelligence. Moral AI?. In *Beyond Artificial Intelligence. The Disappearing Human—Machine Divide*, eds Jozef Kelemen, Jan Romportl and Eva Zackova, 79–95. New York: Springer, 2015.

“PAPER WAVE,” “PAPER PRESSING” AND “PAPER GENOCIDE” AS APPLICABLE TO THE RUSSIAN EDUCATION SYSTEM*

Alexander M. Osipov

Department of Sociology,
A.-I.-Hertzen Russian State Pedagogical University,
St.-Petersburg, Russia
email: osipov.al58@gmail.com

Abstract:

The article considers the metaphors of “paper wave,” “paper pressing” and “paper genocide” as reflecting the social realities of the Russian education system, which are nonetheless poorly understood in sociolinguistics and mostly tabooed within respectable Russian academia and top-management. The relevancy and applicability of these metaphors are substantiated as their criteria, social contexts, and basic connotations are specified. “Paper genocide” is analyzed in journalistic and academic contexts as a term that reproduces the most significant aspects of genocide but with a social and non-criminal meaning. “Paper genocide” helps draw attention to the most acute social and managerial problem, a deadlock within the contemporary Russian education system.

Key words: education system, bureaucratic management, paper work, metaphors, paper wave, paper pressing, paper genocide.

Post-soviet realities of the Russian education system have prompted the emergence of certain linguistic innovations reflecting the social and professional ailments afflicting its personnel. At issue here are the terms “paper wave,” “paper pressing,” and “paper genocide,” which have recently entered

into the vernacular as well as into public and scholarly discourse. They are metaphors. Metaphor is a figure of speech in which a word or a phrase is applied to an object or action to which it is not literally applicable. Metaphors can be classified according to their degree of unexpectedness. Absolutely unexpected, unpredictable metaphors are called genuine. They belong to language-in-action. Metaphors which are commonly used in speech are called trite, or dead. They belong to the language as a system and are often fixed in dictionaries.

The metaphors mentioned above are genuine. They lie at the intersection of political-legal and social-science contexts and require clarification. The last of the three metaphors (paper genocide) demands the most attention, firstly, insofar as it initially had a narrower, concrete-historical meaning tied to specific legal practices of the past that run counter to the latest connotations. Secondly, even in its figurative meaning, it touches on the sphere of formal law and ethics and thus can sound insulting for education policymakers in contemporary Russia.

All of these metaphors deal with the vernacular circulation of the term “paperwork” (or more commonly in English, ‘bureaucracy’), which denotes either individual professions in their entirety (usually called office jobs) or the sizable and labor-intensive components of professional work in different sectors. All those jobs result in the production of various kinds of departmental documents (reports,

* Supported by a grant from the Russian Science Foundation (18-18-00047).

presentations, plans, statements, etc.), whether on paper or in electronic formats.

The emergence of these metaphors and their entry into professional and academic language in Russia was predicated in view of the new trend in education policy that has dominated since the 1990s (Osipov, 2017). According to data from regional surveys between 1991–2017, the average daily time spent on non-educational paperwork among high school teachers increased from 0.6 to 2.2 hours (82–103 different forms to fill out during a year, with the actual amount of documents produced by a teacher reaching up to 3400 pages in printed and electronic formats), among school administrators (5–7 people in each urban school) from 1.5 to 5.0 hours (thus, one urban school spends on documentary reporting an average of about 2000 man-hours monthly, and in the city of Novgorod with its 43 public schools the municipal education system supports about 600 administrative positions only for the sake of non-educational document turnover), among university teachers from 0.3 to 0.8 hours, and among positions at municipal education bodies from 2.2 to 5.6 hours daily. Reporting has become more important than actual productivity and vitally necessary for successful official accreditation and receiving budget funding, so educational establishments do their utmost to report, even in view of its irrelevance in terms of the mission and purpose of education (Осипов, Бояджиева, 2018: 37–38).

This trend is a consequence of the overarching bureaucratization of the economic and social policy of the Russian state. As applied to the sphere of education, the trend has been expressed in marketization and the introduction of a market principle aligned approach to public administration that is known internationally as *new educational management* (Dehli, 1996; Ball, 2001; Tolofari, 2005). This form of management positions education as a service, introduces the priority of consumer choice among institutions, imposes increasingly unwieldy reporting requirements, and ranks institutions by an array of quantitative indicators, with competition as the basis for the allocation of subsidies.

As international studies show, ostensible state budget savings notwithstanding, the marketization of education generally reduces the social efficacy of the public education system. It inevitably leads to the segmentation of the system, makes more unequal the chances of getting a quality education, introduces inappropriate benchmarks into the activities of educators and educational institutions, and gives rise to

a perverse management model at all levels of education (Seppänen, et al., 2015).

All three of the neologisms under consideration are used figuratively, and as a rule, are perceived as being negative, critical, or satirical. They aptly convey the social nature of painful phenomena in formal education, which is why their correct definition and scope of connotation is so important to academic discourses. Russia's academic 'pedagogy and educational management' officials have tended to ignore for years the corresponding issues, as evident from the content of *Education Policy* (published quarterly by the Federal Institute for Education Development, affiliated with Russia's Ministry of Education and Science¹) and *Educational Management: Theory and Practice* (monthly electronic journal of the Institute of Education Management of the Russian Academy of Education²).

The metaphors deserve study from the standpoint of different academic disciplines.

1. "Paper wave," in the context of journalism, dates back earlier than the other two metaphors. It emerged and caught on as a result of the transference and satirical extrapolation of *The Ninth Wave*, a famous painting by Ivan Aivazovsky (1817–1890), to the periodic swells in paperwork that cause psychological discomfort in the life of office workers. These swells are irregularly transmitted by people and groups, at times reaching the outer limits of the corporate loyalty and professional forbearance of individual workers.

An important feature of the connotations associated with *paper wave* lies in the assumption that this wave, just as its natural maritime prototype, is by its very nature an elemental, cyclical, transient, and relatively fleeting phenomenon. It will most certainly pass by and subside, leaving hope for the survival of the people finding themselves in its wake by virtue of circumstances beyond their control. That said, all "wanderers" sailing the paperwork sea realize that the temporary calm should be followed by thorough preparations for the next (seasonal) *paper wave* so as not to be overwhelmed by it and suffer a professional drowning.

Corporate personnel usually view paper wave, erroneously, as a spontaneous situation not directly associated with the interests of particular corporate groups. It is partly for this reason that the paper

¹ <http://edupolicy.ru/ru/>

² <https://www.kursobr.ru/ojs/ojs/index.php/Journal/issue/archive>

wave is accepted with a patient attitude as an *unintentional* disaster that strong people are expected to persevere through while maintaining their position and strength.

2. “Paper pressing,” unlike *paper wave*, which has Russian-language roots, contains an English-language borrowing (pressing – pressure, pressuring). A wider meaning is held by “informational pressing” as the general outline of a civilizational trend uniformly distributed across all people experiencing an overload of incoming information whose growing scale is related to the diversification of consumer rights, health-and-safety issues, and the ability of different entities to serve as sources of information, advertising, creative products, and images. *Paper pressing* is associated with enduring social practices typical of far from all professions and groups of people. Such pressing is most frequently experienced by workers in sectors (segments) of the economy that have embraced an all-encompassing operational regulation that assumes the documentary accompaniment of many if not all operations entailed by the production cycle.

An apt synonym for *paper pressing* might be “paper press,” although such a *press* has long since been associated with the corresponding piece of equipment (machine) used in the printing industry. Moreover, *equipment* is perceived as a passive object only temporarily switched on at will by the worker in the interests of the production cycle. *Pressing*, which sounds like one of today’s more popular sports terms, conveys a special tactic found in team games – the active, subjective nature of situations in which some of the players apply systemic “pressure on the opponent,” while the rest bear the brunt of the *paper pressing* and find themselves in the role of grudging “resisters.”

Paper pressing is a technique in the arsenal of management practices that presupposes the compulsory nature of a high volume of reporting by subordinates to superiors within the scope of the organizational (corporate) hierarchy. It’s a technique that gives certain advantages *to the one doing the pressing*: finding oneself under *paper pressing*, the one in defensive mode (the subordinate) has less time and effort at work to devote to the conceptualization of corporate management practices – to the formulation of a critique aimed at their superior (the one doing the pressing) – or to mount an organized resistance to their defective management practices.

3. Since 2016, the Russian vernacular has increasingly seen the figurative use of the phrase “paper genocide.”³ The key term here is “genocide,” which has been documented in international legislation since 1948 (Office 2018) and is currently under discussion for a review of the relevant criteria (Arutyunyan 2009; Bailey 2008); it therefore requires clarification in terms of the application in the phrase ‘*paper genocide*’ in scholarly contexts. Furthermore, there is an unwelcome likelihood of today’s understanding of *paper genocide* in education coinciding with another, narrower use of the term in mid-20th-century legal opinion journalism in the United States (State of Virginia), which was associated with instances of the improper recording of Native American children on the official list (census) of the country’s races.

Let’s briefly pause to consider the term “genocide.” Its first extensive definition appeared in the UN Convention of 1948, which expanded on a number of previous UN documents: Genocide is the intentional denial of the right to exist, the physical destruction of, or the creation of unbearable conditions for, particular demographic groups, whether in whole or in part. The Convention enumerated five main manifestations of genocide, inspired by the works of R. Lemkin on the Jewish Holocaust. Modern interpretations hold that such destruction could embroil not only ethnic and racial groups but others as well – that in the practice area of genocide, it would be wise to delineate its different levels (stages) (Stanton, 1996) in the context of intentions to commit genocide and concealment of the corresponding practices and situations (Bailey, 2008).

In a number of Russian academic journals and among some Russian scholars, the neologism “paper genocide” provokes antagonism. In connection with the use of “paper genocide” in its *figurative meaning* to describe a systemic problem in Russian education, we note three of its characteristics affirming the appropriateness of using the term “genocide,” which is so harsh in terms of its social and legal resonance – and its implications.

Destruction.

Paper genocide has become, according to regional studies, an obstacle to the quantitative and

³ Use of the term “paper genocide” has been recorded at international conferences (“The Value of Human Existence”, 2016, Barnaul; All-Russian Congress of Sociology, 2016, Ekaterinburg) and in publications (Osipov, 2018).

qualitative reproduction of the sector's workforce. In the motivational area, taking teaching as an example, the hardships of *paper genocide* occupy a ranking statistically on par with low pay among the top reasons for leaving the profession, and one particularly typical of young teachers. The numbers quitting the profession for that very reason now break even with the annual replenishment of the ranks with college graduates. Confronted with daunting paperwork at school-based internships, students often respond by deciding not to pursue a professional career in teaching after all. The workforce at rural schools has catastrophically aged, and the same process is now spilling over to urban schools in the provinces. The teaching shortage is becoming chronic. In a certain sense, the destruction of the teaching profession itself has begun: As school teachers and administrators admit, "we don't work with pupils anymore – we work with papers."

Creation of unbearable conditions for particular demographic groups.

As studies have shown, the practical (functional) irrelevance and sheer pointlessness of much if not most of the document shuffling that goes on in the education system, the physical or technical impossibility of its proper compilation, and the unavoidability of false reporting and "grabbing figures from thin air" are all fully recognized in the regions and at organizations. Workers – from kindergarten teachers to college professors – are being made to sign many documents (including the curricula of study disciplines) whose appropriateness and practicability they themselves doubt. It's this recognition that rests like a psychological burden on the sense of professional wellbeing and social relationships within labor collectives, that leads to corruption and the widespread burnout of individual workers.

Intentionality of action.

No one is likely to think that *paper genocide* is something that occurs spontaneously, such as in connection with the wavelike development of large social organizations. In Russia, it is associated with the accepted strategy of the marketization of education, giving rise to an aggressive *new educational management* that ignores both the status quo as well as the institutional functions of the sector (Osipov, 2017). The financial prosperity and high power status of this management rely on vast and growing

volumes of paper reporting and administrative information flows. These volumes form the basis for staffing numbers and management body budgets, determine resources for outsourcing structures that earn a hefty sum on the processing of information and on the maintenance of its scale, and ensure regulatory control over the sector's lower levels.

Paper genocide presupposes a conscious and coordinated action, expressed in the buildup by management bodies of the volumes of paper reporting that bind subordinate organizations. This buildup drives subordinate organizations and masses of workers into a situation in which:

- the rights and interests of subordinate regional and municipal education systems and masses of workers – whose resistance remains internal – are regularly violated,
- proper professional functioning is carried out at the outer limits of human capacity or rendered thoroughly impossible,
- the quality of educational management and its institutional functioning does not improve, and the legislatively heralded public state management in education is not implemented,
- instances of overwork, unpaid position work, and staff stress occur regularly but go undocumented,
- decision-making and the measures taken in education policy are perceived at the system's lower levels as being unsubstantiated and erroneous,
- in terms of a course correction, the system's lower levels only find moral support at the higher public-state instances (Russian Union of Education and Science Professionals, RF Presidential State Council), but not more broadly throughout the federal Ministry of Education.

In conclusion, therefore, application of the term "paper genocide" would seem justified in today's emerging realities of Russian education. Glimpses of it in the social communications system of linguists and education researchers would make it possible to enrich the linguistic tools available to the scholarly discourse and focus attention on the problems of the Russian education system that are currently driving it into a virtual dead-end and erecting barriers to the preservation and innovative development of the Russian science and education system's potential.

The neologisms *paper wave*, *paper pressing*, and *paper genocide* are deserving of further study in conjunction with an analysis of the corresponding social practices, including comparative analysis. It would be prudent not to restrict this research to the sphere of education, but rather to expand it to encompass other spheres of public life as well.

Conclusions

Under the situation of paper pressing, the information and managerial flows in the Russian education system go beyond the societal mission and responsibility of education, contain institutional dysfunction, and are extremely overloaded, unreliable, and ineffective both in terms of ensuring the quality of education and solving managerial tasks. They are deprived of any standards for labor wages/costs, or for elementary order and planning. They also do not aim at socially necessary and justified results. They are based on the peremptory right of a higher authority to demand any and even irrelevant reports, and on the distrust of the mass of skilled educators who comprise the young generation of Russians. These flows cause hidden resistance, but the problem itself has remained taboo for years and has not been reflected in the National Priority Project “Education.”

Teachers, a previously stable and respectable professional group in the USSR, under the pressure of socially unreasonable information flows, in their everyday practices and in the public consciousness, more and more retreat to the periphery of society, immersed in the “precariat.” The group seems to have lost the ability to fulfill its basic social function. The danger of this kind of deformation of teachers as a group produces new societal risks for Russia.

The theoretical grounds of paper pressing in the Russian education system are the strategy chosen by top management – which is supposedly effective, but not justified in practice – for the marketization of education, and its conversion to competitive principles and a “competency-based paradigm.” These grounds do not in fact ensure the achievement of the goals of the social state in the field of education, but rather increase the “noise” and inaccuracy – that is, the obvious and hidden risks for education, for the political, socio-economic, intellectual, and cultural development of the country, all the while ignoring the global competition. Paper pressing is a way to reduce the quality of education in Russia and disin-

tegrate the education system in Russia. The waves of the “paper abyss” are not receding; they deserve wider interdisciplinary study and call for expert analyses of educational policy.

References

- Arutyunyan M A (2009) The notion of genocide. *Aktualnie problemi rossiiskogo prava [Actual Problems of the Russian Law]* No 3: 403–414. (in Russian)
- (2019) *Information in educational management: theoretical problems*. Eds A. M. Osipov, P. A. Boyadjieva. – 2nd ed. improved and enlarged. S.-Petersburg: RSPU. (in Russian.) <http://www.novsu.ru/file/1566571> (accessed November 11, 2019).
- Osipov A M (2015) (ed.) *Global sociology of education: foreign experiences in solving social problems through education*. Book supplement to the Russian academy of sciences’ journal “Sociological Studies”. Veliky Novgorod: NovSU. (in Russian) https://elibrary.ru/download/elibrary_32667826_21965157.pdf
- Osipov A M (2017) “Trojan horses” of neoliberalism in education. *Sotsiologicheskie issledovania [Sociological Studies]*. No 8: 137–148. (in Russian)
- Osipov A M (2018) Teachers in the context of sociology: theoretical approaches and research priorities. *Sotsiologicheskie issledovania [Sociological Studies]*. No 3: 45–52. (in Russian.) https://elibrary.ru/download/elibrary_32782971_65059029.pdf
- Ball S J (2000) *Sociology of Education: major themes*. Volumes 1–4. L.: Routledge Falmer.
- Dehli K (1996) Travelling Tales: Education reform and parental “choice” in postmodern times. *Journal of Education Policy* 11 (1). 75–88.
- (2008) New Definition of Genocide. *Undergraduate Research Journal at UCCS* 1 (2): 2–8.
- Office 2018 – Office of the High Commissioner for Human Rights. “Convention on the Prevention and Punishment of the Crime of Genocide”. <https://www.ohchr.org/en/professionalinterest/pages/crimeofgenocide.aspx> (accessed 04 August 2018).
- Seppänen P, Kalalahti M, Rinne R, Simola H (toim.) (2015) *Lohkoutuva peruskoulu – perheiden kouluvalinnat, yhteiskuntaluokat ja koulutuspolitiikka*

ka. Kasvatusalan tutkimuksia 68. Jyväskylä: Suomen Kasvatustieteellinen Seura.

Stanton G H (1996) *The 8 Stages of Genocide*. Paper presented at the US Department of State.

Tolofari S (2005) The new public management and education. *Policy features in education* 3 (1): 75–89.

HEALING THE BLACK SEA'S PATRIMONY: HOSPITALS, HEALTH AND HERITAGE. AN ARGUMENT IN FAVOR OF CREATING A NETWORK OF HOSPITALS-MONUMENT BELONGING TO THE STATES ACROSS THE BLACK SEA*

Oana Șerban

University of Bucharest, Faculty of Philosophy
Bucharest, Romania
email: oana.serban@filosofie.unibuc.ro

Abstract:

The main aim of this paper is to examine the tangible forms of cultural heritage represented by European hospital buildings from states across the Black Sea that are still functional or have been closed, and that are subjected, due to the lack of sustainable financial means for conservation and restoration, to degradation, abandonment, and destruction. For the purpose of this analysis, I will tackle both elements of *the operational plan* of hospital buildings that have been evaluated and registered as national monuments, from the perspective of their clinical functionality, and the elements of *architecture* and *aesthetic forms* behind such structures that embrace medical canons and particularities. Therefore, hospitals will be treated as entities of *tangible cultural heritage* that develop, through their complementary medical and cultural history, *forms of intangible cultural heritage*.

This wide range of buildings can be reduced to two operational categories: *hospital buildings* designed from the beginning to fulfil a clinical functionality, and *cultural buildings* – from ecumenical establishments, castles, or villas, such as hermitages and churches, to military

structures, such as garrisons – which have been adapted for historical, social, or political reasons to *clinical conversion*. I will analyse not only the national constraints, prejudices, and values that contributed to a certain medical and cultural imaginary of state hospitals as monuments, but also the similar strategies and cultural policies that different states across the Black Sea have adopted in preserving the memory and structure of these buildings. The main question I address is: To what extent is it possible to create a network Black Sea region state hospitals as European cultural monuments, and what advantages might this bring to the attempt to perform a more reflective and inclusive notion of European identity? The current research is designed to be a starting point for the development of transectorial public policies, which could lead to an improvement in standards for quality of life, the infrastructures of medical units, and the preservation of tangible forms of cultural heritage, such as the public state hospitals classified as monuments.

Key words: hospitals, monuments, Black Sea, health as a cultural resource, tangible / intangible cultural heritage.

Health as a cultural resource of the Black Sea

Bordered by six states, the Black Sea is currently considered one of the most important maritime areas; yet it has been underexplored from a cultural point of view, in terms of the generous junction be-

* Publishing current paper was financed through Entrepreneurial Education and Professional Counseling for Social and Human Sciences PhD and Postdoctoral Researchers to ensure knowledge transfer Project, co-financed from European Social Fund through Human Capital Programme (ATrIUM, POCU380/6/13/123343). The main idea of this article is inspired by a previous lecture supported by research project PATCULT#RO ("Platforma pluridisciplinara complexa de cercetare integrativa si sistematica a identitatilor si patrimoniului cultural tangibil si nontangibil din Romania") - UEFISCDI [PN III-P1-1.2-PCCDA-2017-0686, 52PCCDI/2018].

tween the cultural heritage that port cities have developed throughout history and the context of economic-territorial relationships that foster good practices in the management of a shared cultural heritage. We place the Black Sea at the core of the European agenda for raising awareness on and safeguarding monuments and sites throughout the Black Sea region; the urgent task of promoting the cultural diversity of this area is mainly focused on cultural tourism, through the development of so-called “cultural routes” represented by clusters of hospitality traditions, values, and creative expressions in various cultural and artistic sectors in which the six states perform. However, increasing the regional attractiveness of the states across the Black Sea has never been prospected as a feasible project that might be achieved by cultural and medical tools; whenever experts attempt to identify new and sustainable cultural routes in the Black Sea by relating transnational elements across the Black Sea countries, the role of hospitals established as monumental buildings of cultural heritage, as well as the capacity of related intangible forms of cultural heritage resulting from a certain historical and cultural understanding of health and quality of life standards across the Black Sea, are ignored or regarded as a peripheral concern.

The intention to identify and promote a certain transnational cultural paradigm of health through the material and symbolic capital of the Black Sea is not included on the list of priorities that different European institutions agree on as part of their strategic coordination in emphasizing culture as a reliable tool for the social and economic development of this area. Through these considerations, I am attempting to promote the inclusion of culture in a very specific sector (toward the goal of encouraging inter-european cooperation and policies) other than education, tourism, technology, industry, or maritime commerce, namely health. Nowadays, representatives of different working groups belonging to the European institutions and commissions responsible for culturally regenerating the Black Sea insist on developing networks of museums, cultural sites, or other cultural agencies such as festivals, in order to strengthen the partnership between states that border

this maritime area, which needs a powerful and renewed branding.

My research indicates that there is no transparent effort to address hospitals that are established in historical buildings as part of the cultural heritage of the Black Sea, nor any intention of insisting on their potential to be reliable candidates for inclusion in the UNESCO World Heritage List and Representative List of Intangible Cultural Heritage. The Black Sea coast is recognized for its balneary potential, but health is not discussed as a cultural repository of resources and practices that correspond, from their material order, to a set of comprehensive values, inherited rituals, and medical knowledge that have evolved through an intercultural dynamic. In fact, health should be treated as a community-based value: It depends on certain standards of living and welfare that endure from generation to generation; it can be tracked as a cultural good on a comparative basis from one community to another; it involves certain skills of self- and community-care; and its definition is relative to a social cohesion capable of bringing together urban and rural traditions prescribed both by clinical/scientific and popular/alternative/nonscientific aspects of medicine. Given all of this, how can the hospital-monuments across cities that belong to the Black Sea states (not limited to those located in port cities) be valued and treated as forms of both tangible and intangible heritage, and how might this valuation contribute to the overarching project of expanding and promoting the cultural capital of the Black Sea?

My argument stresses that the Black Sea’s patrimony is intensively and commonly explored as an underwater maritime heritage. This includes prehistoric archaeological sites, natural protected areas, caves,¹ etc. But related intangible forms of cultural heritage are currently neglected. The coastal resources specific to the states across the Black Sea have historically raised and shaped health as a notion related to care and therapy. As an example, in Romania, Eforie Nord evolved as a health resort

¹ See Marilena Papageorgiou, “Stakes and Challenges for Underwater Cultural Heritage in the Era of Blue Growth and the Role of Spatial Planning: Implications and Prospects in Greece,” in *Heritage*, 2019, 2, 1060–1069, doi: 10.3390/heritage2020069, p. 1060.

before earning its reputation as a place of tourism and entertainment: King Carol I facilitated the opening of a clinic and a resort that included sapropelic mud from Techirghiol Lake as a curative element.² But this approach is singular, and is widely discussed in the field of knowledge devoted to the value of health across the Black Sea.

My hypothesis is that natural material resources have highly influenced certain medical practices and specializations in territories where they are available; therefore, techniques of medical recovery and care for certain diseases have been inspired by the health resources of the Black Sea. Such techniques and procedures have found their place nonetheless in clinics and hospitals that transformed them into symbolic and cultural capital. Reconstructing this history retrospectively obliges us to revisit certain hospitals from states across the Black Sea as shareholders of not only tangible, but also intangible forms of cultural heritage specific to this area.

Looking exclusively at the balneary institutions housed by small hospitals and clinics is a trap: From this, we can understand at most certain historical traditions of folk medicine (alternative medicine) practiced in communities along the coast of the Black Sea. Meanwhile, cultural capital in the form of historical buildings of national patrimony remain exquisite and unexplored. Addressing proper concern and attention to such institutions, which embody both tangible and intangible forms of heritage following cultural patterns specific to the Black Sea, means healing this area of its one of its most malignant diseases: a political, economic, and educational underestimation of health as a fertile, cultural resource.

Unusual forms of tangible and intangible cultural heritage in the patrimony of the Black Sea: Hospital-Monuments, Hospital-Museums

The main objective of my paper is to concentrate, for the first time, on hospital-monuments from states across the Black Sea region, recognizing them as elements of tangible and intangible cultural heritage. (I also tackle the possibility of extending this

critical perspective to other areas on the Danube-Black Sea Canal.). This perspective invites us to understand the way in which the architectural particularities of each medical space bring functionality, and the structure as a whole – from the main buildings to the annexes – imparts a certain medical hermeneutic relevant to clinical protocol, that is, an interpretation of the phenomena of healing, disease, and the tension between life and death, which tempers the cultural memory of the building before being reified, through daily practices, as a medical institution.³ Where buildings have become inoperative, the ways in which they can be reconfigured – for example, as medical museums, with outdoor spaces (hospital yards) transformed into open gardens and parks – will be studied. Depending on the medical profile of the building, some will prove more promising than others in compensating for the shortcomings of urban centers or peripheries and improving the arrangement of places for leisure, rest, or relaxation in the natural landscape. In Romania for example, the history of hospital-monuments is quite under-discussed, despite the generous number of such forms of cultural heritage. Relevant examples can be tracked across the entire territory of our state, and in order to limit the scope of this article, I will critique the situation of Romanian hospital-monuments exclusively.

The Carol Davila Clinical Hospital of Nephrology is one of the most impressive local buildings, and the only institution in Romania constructed in neoclassical style, modeled on Ospedale di S. Lazzaro e Medicanti in Venice.⁴ Colțea Hospital, established 300 years ago, was one of the first European medical units with a social-medical mission similar to the Charité hospital in Berlin and Guy's Hospital in London. The architectural curiosities of Colțea Hospital are as remarkable as the strictly medical ones. The latter include the following: Clinical prac-

³ See Michel Foucault, *Madness and Civilization*, Taylor and Francis Ltd, 2001 and Xavier Greffe, "Heritage conservation as a driving force for development", in *Heritage and Beyond*, Council of Europe Publishing, 2001, p. 103.

⁴ Bogdan Horea Chicoș, „Spitalele românești în secolul al XVIII-lea” (en. “Romanian Hospitals in the 18th century”), in *Medicina românească în evoluție, viziune enciclopedică*. Volumul I, Vasile Goldiș University Press, 2008, p. 223.

² <http://romaniatourism.com/spa-romania-eforie-nord.html>, last accessed October 20, 2019.

tice was carried out exclusively by foreign doctors until 1840; the first documented surgery at the institution was an amputation; under the aegis of the hospital the first “minor surgery manual” appeared.⁵ In 1844, the innovative medical practice carried out by leading personalities in the field and policy carried out by hospital managers, who in time became government Ministers of Health, placed the institution on the map of the most important local medical units, and today includes a renowned bone marrow transplant department. It is one of the so-called successful museum-hospitals, functional and effective, a celebrated feature of a culturally and medically “healthy Romania.”

A worrying class of monuments, however, is represented by the headquarters of former hospitals that have ceased their functional activity, including I.C.Brătianu Hospital in Buzău. The building, a group B historical monument, was considered for renovation or demolition by the Local Council of the Municipality, but until 2011 no concrete plans were proposed. This medical unit was the third hospital established in Muntenia after the Colțea and Pantelimon Hospitals, a historical fact reflected in the design and particularities of the building itself. From its founding in 1792 until the middle of the nineteenth century, the building gradually evolved; the first rooms for minor surgery and first pharmacy appeared, and in the twentieth century the building was selected to become the County Hospital, due to its departments of ENT, Ophthalmology, and Dermato-Venereology. Through the 2014-2020 Regional Operational Program, and with European funding, the building will be restored for public use (due to be completed in 2021) and renovated and reopened as a museum/annex of the existing County Museum.⁶ Short broadcast interviews have revealed details of the medical protocol regarding the ophthalmology department, including testimonials of former doctors and hospital employees suggesting that greater documentation of the monument is needed, not only from a cultural perspective but also from a medical perspective, if the long-closed, non-

functional, and degraded building is to become an important part of the monograph of the city.

Local authorities have managed to attract funds in order to restore the historical heritage building of the former hospital. This example sheds light on the synchronization of the cultural and medical history behind heritage buildings as well as the mechanisms for identifying problems and solutions aimed at reconfiguring these spaces for their public reopening; moreover, it demonstrates the benefits of similar restoration projects, including buildings currently slated for demolition or permanent abandonment.

The most recent example in this regard is the oldest maternity ward in the country (located in the city of Fălțiceni), founded 120 years ago and demolished after the space was allocated for the construction of a supermarket, since local authorities and institutions neglected to request that the building be classified as a historical monument when the hospital administration relocated in 2008.⁷ In such cases, I propose the reconstruction of a memorial archive of the monument. This would include interviewing former medical staff of the hospital and architects or urbanists who know the file of the retrocession. Doing so might hopefully succeed in pointing out the institutional, human, and material errors that led to the loss of the heritage building, in an attempt to prevent future losses.

Health may be tracked as a form of symbolic capital with tangible and intangible expressions in the hospital-monuments of port cities, which thus possess the power to revive the cultural, economic, and educational agenda of the Danube-Black Sea Canal region. The Lady Elisabeth Hospital of Psychiatry, currently recognized as a national historical monument, was constructed at the end of the nineteenth century based on the orders of Aristide Seriphoti, a doctor and a powerful reformer of the Romanian medical system. A three-meter-tall bronze bust of Seriphoti created by Oscar Spathe, lavish gardens, and architecture combining classical and

⁵ See <http://www.coltea.ro/>, last accessed October 20, 2019.

⁶ See <http://rba.spatiuconstruit.ro/arhiva/2019/proiect/166/view>, last accessed October 20th, 2019.

⁷ See PressHub, <https://presshub.ro/proiecte-editoriale/orasul-uitat/2018/05/22/falticeni-cea-mai-veche-maternitate-din-tara-nu-poate-fi-salvata-de-la-demolare-deoarece-directia-pentru-cultura-suceava-nu-are-personal-si-buget-pentru-a-intocmi-dosarul-de-clasare/>, last accessed October 20, 2019.

gothic elements represent the most important aesthetic elements of this cultural assemblage.

The Military Hospital of Galați⁸ is another important historical monument: In 1831, doctor Czihak, a physician and leader of the Moldovan army, decided to move 15 beds for the sick members of his legion into the house of a local merchandiser, from which the current hospital emerged. Unfortunately, the memorial house of Seriphoti is not yet included on the list of historical national monuments, despite its aesthetic and cultural significance, nor is the tomb of the Seriphoti family.⁹ The two-story vault of the funerary construction imitates a basilica, including a basement and a terrace with gothic stairs and a fortification strengthened by ramparts. The terrace, carved in stone, repeats artistic elements belonging to the neoclassical style, with the mortuary chamber narrowed by two apses. The tombs are made of blue marble. This artistic construction, which may be easily recognized as a valuable historical monument, has been highly neglected; it belongs to the family of Serfioți, whose relatives recently visited the crypt without any administrative engagements. As it constitutes private property, the state cannot proceed to include the construction on its list of historical monuments. The memorial house shares the same fate: The exterior is well-known for its balcony supported by statues of winged angels, similar to caryatides, arranged around niches with ancient statues. In this place the doctor Serfioți laid to rest his decorations received after the War of 1877: the Order St. Stanislav in the rank of commander awarded by the Tsar, the Order of the Crown of Prussia, the Order Medgidia awarded by the Ottoman Empire, and the order the Crown of Romania in the rank of commander.

Such elements capture certain patterns of tangible cultural heritage that can be fully assigned to the material capital specific to the Black Sea. Nonetheless, these structures rely on intangible forms of

cultural heritage that can be equally projected transnationally across the Black Sea. For example, the medical quality of cures based on the internal use of seawater, sea salts, aerosols, baths, and maritime mud, applied in a large spectrum of therapies for rheumatism, arthrosis, dermatology, neurology, etc., lies at the core of thalassotherapy and thermalism.¹⁰ All these material uses of the resources of the Black Sea establish standards of antistress, dietetics, reshaping, beauty, and well-being, and all are applied in specific procedures, rituals, and methods of cure that are used in clinics, balneary and spa resorts, or sanatoriums. They also contribute to a certain understanding of what a maritime community should be and should provide in terms of physical and mental care and support. For the purposes of the current research, I will ignore the regional particularities of such elements of intangible cultural heritage; I have mentioned them here in order to provide a general understanding of the correspondences between tangible forms of cultural heritage (such as those encapsulated by the hospital-monuments) and intangible – and more versatile and vulnerable – forms of cultural heritage.

Instead of conclusions: the advantages of a network of hospital-monuments across the Black Sea, from A to Z

Regarding the possibility of creating a network of hospital-monuments belonging to states across the Black Sea, some important historical buildings should be considered for inclusion. Among them are the Russian hospital building from the coast of Sevastopol, named after Pirgov on Pavlovsky cape;¹¹ the city hospital of Krasnodar, built together with a complex of buildings, including a poorhouse and a church¹² in the beginning of the nineteenth century; the former hospital of Kavala established during the Balkan wars, which currently functions as an all-

⁸ See <https://directiamedicala.mapn.ro/pages/view/137> last accessed October 20, 2019.

⁹ See “The inheritance of doctor Aristide Serfioți: SOS for the monument-tomb, nowadays in waste”, in *Archaeoheritage*, the Romanian archeological watchdog, 2013: <http://archaeoheritage.ro/2013/04/mostenirea-doctorului-aristide-serfioți-sos-pentru-cavoul-monument-azi-in-paragina/>, last accessed October 20, 2019.

¹⁰ Charlier & Chaineux, “Thalassotherapy”, in *Journal of Coastal Research*, Vol. 25, No. 4, 2009, pp. 839-40.

¹¹ See <https://www.alamy.com/stock-photo-admiral-building-of-the-hospital-of-the-black-sea-fleet-named-after-137079267.html>, last accessed October 20, 2019.

¹² See <https://goodplaces.travel/en/pretty/krasnodar/culture/kompleks-zdaniy-voyskovoy-bogadelni-i-gorodskoy-bolnitsy/19828>, last accessed October 20, 2019.

male high school;¹³ the old hospital of Izmail, currently registered as a historical monument; the Museum of Health and Sultan Beyazit II Mosque (located in Edirne), built in the fifteenth century and known for treatment of mental disorders.¹⁴

There are many possible advantages to establishing this network of historical hospitals across the Black Sea. I consider that such a project would immediately:

- a) Introduce a cultural, educational, and economic platform of hospital-monuments divided between hospital-museums (that no longer operate and fulfill their clinical functions, but receive the necessary financial support for conservation, restoration, and conversion into medical museums, preserving the cultural and historical memory of the buildings) and hospital-monuments (that continue to perform as medical centers of the highest medical and academic rank, and whose architectural splendor is preserved as a fixture of historical and cultural heritage).
- b) Procure public space for recreation, transforming the gardens of hospital-museums into public parks.
- c) Create an independent source of funding emerging from the costs of visiting tours of the hospital-museums.
- d) Allow the legal representatives of hospital-museums direct access to a new category of funds for conservation and restoration, e.g. funds allotted for cultural institutions. (Note that hospital-monuments were previously financed by the Ministry of Health exclusively.)
- e) Develop, at a national level, strategic partnerships between the Ministries of Health and Culture.
- f) Increase the number of inscriptions of different historical buildings belonging to hospital-monuments or to the memorial houses of authoritative figures from the domain of medicine on the tentative list of UNESCO World Heritage Sites.
- g) Create a source (site) of reflection on the elements of intangible cultural heritage behind the notion of health as it has been understood throughout history at the crossroads of the Black Sea.
- h) Configure a sense of hospitality routes around and across the Black Sea, using health as a strategic component, not only in terms of clinical and medical practices, but also in terms of cultural symbols and values.
- i) Mix European and non-European perspectives on health as they are conceived and designed by the states across the Black Sea, developing, through this, a more inclusive and reflective notion of the maritime states that compose the Black Sea's basin.
- j) Force the institutions of education and culture to align their priorities through strategic partnerships in order to support the creation of synergic public policies on health, care, standards of living, and welfare, inspired by the resources and the problems revealed by the Black Sea basin.
- k) Raise awareness and encourage educational forums to recognize and promote health as a cultural value.
- l) Increase the authority of the Ministry of Education in working alongside the Ministry of Culture and the Ministry of Health from each country across the Black Sea in order to create the necessary legal conditions and financial instruments to open and maintain hospital-museums (converting the historical buildings of clinically non-functional hospitals into museums of medicine).
- m) Increase the annual revenues obtained from tourism by correlating the cultural capital of the hospital-monument with the curative potential of spa resorts from the port cities of the Black Sea coast.
- n) Create a powerful connection between the tangible forms of cultural heritage that compose the patrimony of the Black Sea and the intangible elements of cultural heritage, such as medical balneary practices, rituals related to alternative medical techniques of cure and therapy, values and understandings of life, wellness, and health.

¹³ See <http://www.limenproject.net/kavala-cultural-route/>, last accessed October 20, 2019.

¹⁴ See <https://turkisharchaeonews.net/museum/health-museum-edirne>, last accessed October 20, 2019.

- o) Design an instrument to overcome the intercultural barriers and political conflicts between the states across the Black Sea by concentrating their efforts on protecting the maritime basin as a common resource.
- p) Offer a proper context for creating new cultural visiting-tours of the Black Sea, which would be of great interest for a new target group (medical representatives).
- q) Set a new list of priorities for public health policies proposed by the states across the Black Sea.
- r) Contribute toward attracting new categories of European funds for the hospital-monuments, which function by using the latest technologies in their clinical domains of intervention, despite their cultural history and their material conditions as national buildings of cultural heritage.
- s) Improve the reputation of old hospital-monuments as relevant and sustainable medical centers within their clinical domain of intervention.
- t) Attract new medical personnel.
- u) Create new job positions in the cultural field: curators, restorers, conservators, historians of medicine.
- v) Generate a proper context for an educational lobby in favour of including health as a cultural resource in the curricula of different classes of cultural studies.
- w) Save the buildings of clinically non-functional hospitals from destruction or conversion into consumption spaces (malls, hypermarkets), preserving their cultural memory.
- x) Urge, in some cases, the closing of certain non-functional hospital-monuments, in order to convert them into museums.
- y) Create new spaces for the archive of the hospital-monuments (that should ultimately be digitized).
- z) Create a cultural platform for additional events, such as medical congresses, cultural conferences, festivals.

References

Books and Articles

Charlier R H, Chaineux M-C P (2009) The Healing Sea: A Sustainable Coastal Ocean Resource: Thalassotherapy. *Journal of Coastal Research*, 25 (4), 838–859.

Chicoș B M (2008) Spitalele românești în secolul al XVIII-lea (Romanian Hospitals in the 18th century). *Medicina românească în evoluție, viziune enciclopedică*. Vol. I, Vasile Goldiș University Press (in Romanian).

Foucault M (2001) *Madness and Civilization*, London-New York: Taylor and Francis Ltd.

Grefe X (2001). Heritage conservation as a driving force for development. In *Heritage and beyond*, Council of Europe Publishing.

Papageorgiou M (2019) Stakes and Challenges for Underwater Cultural Heritage in the Era of Blue Growth and the Role of Spatial Planning: Implications and Prospects in Greece. *Heritage 2*: 1060–1069, doi: 10.3390/heritage2020069.

Online sources:

<http://romaniatourism.com/spa-romania-eforie-nord.html>, last time accessed at October 20, 2019.

<http://www.coltea.ro/>, last accessed October 20, 2019.

<https://presshub.ro/proiecte-editoriale/orasul-uitat/2018/05/22/falticeni-cea-mai-veche-maternitate-din-tara-nu-poate-fi-salvata-de-la-demolare-deoarece-directia-pentru-cultura-suceava-nu-are-personal-si-buget-pentru-a-intocmi-dosarul-de-clasare/>, last time accessed at October 20, 2019.

<https://directiamedicala.mapn.ro/pages/view/137> last accessed October 20, 2019.

<http://archaeoheritage.ro/2013/04/mostenirea-doctorului-aristide-serfioti-sos-pentru-cavoul-monument-azi-in-paragina/>, last accessed October 20, 2019.

<https://www.alamy.com/stock-photo-admiral-building-of-the-hospital-of-the-black-sea-fleet-named-after-137079267.html>, last accessed October 20, 2019.

<https://goodplaces.travel/en/pretty/krasnodar/culture/kompleks-zdaniy-voyskovoy-bogadelni-i->

gorodskoy-bolnitsy/19828, last accessed October 20, 2019.

<http://www.limenproject.net/kavala-cultural->

route/, last accessed October 20, 2019.

<https://turkisharchaeonews.net/museum/health-museum-edirne>, last accessed October 20, 2019.

MUSLIMS IN EUROPE: IDENTITY CRISIS BETWEEN “EUROPEAN” AND “ISLAMIC” VALUES

Gyulnara Gadzhimuradova

Moscow State Institute of International Relations (MGIMO),
Federal Scientific Research Center of the Institute of Social and Political Studies
of the Russian Academy of Science
e-mails: gadzhimuradova7@gmail.com

People and Nations are trying to answer the most important question they can face: who are we?¹

Abstract:

A world in constant motion, in a state of migration turbulence, presents humanity with new challenges and risks. Globalization is a blessing or a tragedy for humanity, occasioning the problem of how to preserve one's identity, remaining “one's own among strangers” while, at the same time, not becoming “a stranger among one's own.” Integration processes in the world today are met with resistance by multidirectional processes that encourage a critical engagement with all spheres of life in modern society in order to counteract forces of depersonalization and the disappearance of one's identity – one's self – as expressed in the preservation of one's ethnic group, culture, religion, and so on. This is especially evident in attempts at preserving identity within Muslim communities in European countries.

Given the growing Muslim population in Europe, it has become obvious that “European” and “Islamic” values are opposed in the context of preserving one's own identity, which is increasingly manifested in a religious

context. Europe today has become a hostage of its values, which are despised by many of the immigrants who have poured into its borders. These are tolerance, political correctness, multiculturalism, democracy, and freedom of speech, among others, which are perceived as weakness and indecision. Eastern mentality, habits, and traditions are sometimes very different from European ones. The author examines the transformation of Muslim identity and the compatibility of “European” and “Islamic” values. The article also presents the opinions of various researchers on this issue, and provides possible scenarios for the trajectory of events, given continued intercultural contact through immigration and given the stakes and state of this collision of values.

Key words: Globalization, integration, migration crisis, identity, “Islamic” and “European” values.

1. Introduction

The era of globalization has blurred borders not only in terms of the simple movement of peoples and individual citizens, but also in the sense that self-awareness of one's own identity has become hazy. Paradoxically, the growth of objective integration trends (globalism) in the world, which lead to interconnection and interaction in all spheres of modern society, has also been accompanied by a parallel, no less stable process of resistance; individ-

¹ Huntington S. The Clash of Civilizations and the Transformation of the World Order // New Post-Industrial Wave in the West: Anthology / Ed. B.JI. Inozemtseva. M., 1999. P. 532.

ual ethnic and cultural communities in different regions and countries have opposed the myriad manifestations of globalization in the economy (anti-protectionist global free trade initiatives) and spiritual sphere (erasure of cultural traditions and identities). The fear of losing one's culture, identity, and uniqueness under globalist conditions, which is shared by many people and communities, is also manifested at the state level, with states proclaiming defense of their national identity and their national and state interests in an increasingly globalized world. As such, humanity itself faces a dilemma. On the one hand, self-awareness and a sense of cultural identity have increased. On the other hand, mutual exchange between people, nations, and civilizations has significantly expanded.

The growth of national self-awareness and the focus on self-identification can be seen as a defensive reaction to the 'standardization' of social life. Cultural diversity and national identity are under threat today, as globalist trends increasingly impose a single form of life and system of values as the only true and valuable model to aspire to and to emulate. Hence, the natural reaction of nations is to defend themselves and their uniqueness. The growing anti-globalist tendencies are linked to the fact that people do not want to be representatives of a common, faceless world, but rather bearers of a specific ethnocultural, national community.

The problem of identity is multifaceted. It includes the semantic and metaphysical questions: "Who are we?," "What is a person?," "What is the value and meaning of a person's existence?" In other words, the process of globalization (and its impact on cultural identity) makes relevant the question of the deepest foundations of genuine human existence.

In Europe, the boundaries between communities formed over centuries of living together have traditionally been porous, with communities retaining their identity while sharing certain values common to the West. By contrast, this tradition does not exist between the West and the East. Between Europe and the rest of the world, there has been more limited intercultural interaction and exchange, particularly regarding high-stakes and deeply personal matters of religious belief. Of course, modern life has brought about many changes, but rarely in terms of the cultural codes and deeply-held traditions and beliefs of different civilizations. Thus, two dominant forms of global civilizational processes – globalization and the commitment to traditional identities –

remain incompatible, particularly in the context of East-West relations.

In addition to a Muslim religious identity, Muslims have many other identities: local, national, regional, tribal, linguistic, ethnic. Islamic identity takes on meaning in connection with these concomitant factors, which may condition and determine each other. Therefore, in developing an understanding of identity, an important role is played by the juxtaposition of "self" and "other." Depending on the political, socioeconomic, or other aspects of one's situation, the individual can bring to the fore different context-dependent identity traits – these may be salient even without contrast to an "other." As such, it can be said with confidence that there is no true central identity or single aspect of identity that defines the self without remainder; there exists only cultural strategies of self-definition and meaning-making that can be used when faced with a variety of situations.

The individual develops his political and social viewpoint within the framework of his internal consciousness or identity, and projects it onto and through political and economic institutions. For Muslims, this means that national identity is often expressed through Islamic symbols and values.

2. Islam in contemporary Europe

Worldwide, the Muslim population is growing around two times faster than the non-Muslim population. The average annual growth rate of the Muslim population is 1.5 percent compared to 0.7 percent of the non-Muslim population. According to preliminary forecasts, by 2030, the Muslim population in Europe will exceed 58 million people, and its share of Europe's population will be 8 percent.² Today, Islam is the largest religion in Europe after Christianity in terms of its representatives. Muslim communities in Europe differ significantly from each other in terms of country of origin, language, religious preferences, customs, traditions, and socioeconomic status, among other categories.

Modern Europe faces a great challenge: Muslim migrants are either very poorly integrated or not integrated at all into their country of residence. This is facilitated by a fairly low level of educational attainment. For example, the number of children who drop out of school is significantly higher among Muslims than among other groups of the population (including other immigrants). The low performance

² The Future of the Global Muslim Population. Projections for 2010–2030. – Pew Research Center. 2011.

of Muslims in education can be partly explained by widespread socioeconomic limitations for Muslim students and by a lack of social mobility. Due to existing social relations in which Muslims are branded with a certain (often negative) image from birth, young Muslims are convinced that they will not be able to succeed even if they have a good education. In addition, the school system in Europe raises many issues for Muslims, including school dress codes, lack of religious education, school curricula with a secular character, and coeducational instruction in schools. It should be noted that the growth of modern forms of communication and media have not contributed to more rapid integration since many migrants maintain much closer contact and ties with their homeland than with the host country. They may invest more time in keeping up to date with events in their homeland than they do in getting to know the country that is supposed to be their “second home.”

Today in Europe it has become possible to speak of the emergence of so-called “European Islam.” European Muslims consider themselves citizens of European countries; many were born in Europe and have absorbed “European values.” They may be Europeans who have converted to Islam but have not abandoned their European way of life. One of the problems caused by the growth of the Muslim population in Europe is related to questions of identity and self-identification. This problem concerns both Europeans in general and European Muslims in particular. When Muslims assume a European identity, they may find themselves torn between two lifeworlds: On the one hand, they belong to a certain ethnic group, represent a certain country, and are considered part of the Muslim umma, and on the other hand, they feel that they belong to a particular European country of which they are citizens.

Increased migration forces people of different cultures and different faiths to live side by side and actively interact with each other. In regard to this, the question of the compatibility of Islam with democratic norms and the acceptance of “European values” by Muslims has long been overdue in Europe.

Immigration problems in Europe itself are also twofold: The most energetic and talented leave, while the birth rate of those who remain is very low. Moreover, immigrants from low income countries do not have enough education to fill these vacancies, and their reproduction rates are quite high.

Islam in Europe can be viewed as a heterogeneous and diverse phenomenon. On the one hand, there

are Muslims living in Europe who peacefully practice traditional Islam. On the other hand, there are adherents of radical Islam and its different forms. Muslims today are scattered throughout Western Europe and live mainly in large industrial cities. According to rough estimates, by 2010 the number of Muslims in Europe exceeded 12 million people, and in 2015 their number exceeded 20 million people. The Muslim population is concentrated to a large extent in the three main EU countries- Germany, France, and the UK. According to the World Bank, in 2009, there were 4,026,000 Muslims in Germany, 3,554,000 in France, and 1,647,000 in the UK.³ It is difficult to determine the exact number of the Muslim population since most Western European countries, such as France, avoid the question of religious affiliation when conducting a census.

The migration crisis of 2015-2016 led to a rapid increase in uncontrolled migration and accelerated the formation of closed, isolated communities (ghettos). The French and German practice of turning “Muslim” areas into de facto autonomous enclaves only reinforces their isolation, makes it difficult for them to access traditional state institutions, and further contributes to the marginalization of Muslims. Examples of self-isolated Muslim communities are, for example, Molenbeek and Skarbek in Belgium, Rinkeby in Stockholm, Sweden, or the Tingbjerg district in the Danish capital, Copenhagen. Once in Europe, these people become marginalized, existing in a limited space, and because of this, the interpenetration of their own cultures takes place within the Muslim community itself. Therefore, in Muslim communities, the differences between people from different countries are somewhat blurred, and members of these communities become more monolithic. This process is quite smooth and painless, because all communities are united by a common religion – Islam, based on traditional values, focused primarily on Dar al-Islam, i.e. the space regulated by the laws of Islam (Sharia).

A significant number of Muslims in Europe deliberately refuse to accept the Western way of life, and the morals and values of Europeans. In modern Europe, the religious system of values has been replaced by a secular one, and the priority of values has changed from religious to secular ones. European identity is no longer based on Christian values, but on so-called “European values.” The concept of

³ Interactive Data Table: World Muslim Population by Country. URL: <http://www.pewforum.org/2009/10/07/mapping-the-global-muslim-population23/>

“European values” for many Europeans has now become a fetish or, as Yu. Fedorova has rightly noted, “a kind of theoretical construct that allows Europeans to declare their identity and isolate themselves from representatives of another culture that produces a different set of value imperatives” (Fedorova, 2014). In Europe, it was initially assumed that Muslim migrants would share the Western European system of values – democracy, cultural and religious freedom, individual desire for well-being and self-realization – and that differences would be erased in no more than two or three generations. This has not happened. On the contrary, there is a growing trend towards cultural and territorial self-segregation among Muslims who have entered the country, their children, and grandchildren.

The increasing number of mosques established every year plays an important role in migrants’ advocacy of “Islamic values.” For example, in Hamburg in 2013, the building of the Lutheran Church and the adjacent 44-meter tower with a land plot were sold to the al-Nur Islamic center, which intended to open a mosque in this space. In recent years, Methodist, Lutheran, and Catholic churches have been transformed into mosques in Berlin, Dortmund, and several other cities. In Germany, more than 400 Catholic and 100 Protestant churches have been closed since 2000. Since 1990, according to the Evangelical Church of Germany, at least 277 Protestant churches have been sold or demolished. There are hundreds of mosques in the country, and more than 120⁴ are under construction. This suggests that religion remains the basis of the identity of Muslims in the West.

3. European Muslims and Muslims in Europe: the problem of “values”

The majority of Muslims living in Europe resist assimilation because they are strongly attached to “Islamic values” and are convinced of the advantages of their mores and customs. At the same time, they enjoy many aspects of Europeans’ lives: a high level of consumption of goods and services, honesty, clean and safe cities, thirst for knowledge, politeness. But some of the European customs are absolutely unacceptable to them; they cannot, in particular, accept the “liberal values” that are so dear to Europeans. In the eyes of some Muslims, these values look like permissiveness and licentiousness. This applies primarily to family and marriage rela-

tions, in particular same-sex marriage, certain rights and freedoms, and feminism. The majority of European Muslims are increasingly intolerant of certain “European values” such as gender equality, freedom of religion, freedom of speech, ensuring the rights of sexual minorities, and so on. Islam for them is not just a religion with its pillars of faith; Islam is a way of life that includes ethical norms, social behavior, customs, and traditions that have remained unshakable for centuries. Therefore, anything that, in their opinion, does not comply with the norms of Islam is harmful and alien.

Religion is exactly what allows one to not dissolve in another civilizational paradigm and to preserve one’s ethnic and other identity. According to Olivier Roy (2004) “Muslims are experiencing the deterritorialization of Islam. When they turn to religion, they are more attracted to other ways, including neo-fundamentalism. Neo-fundamentalism has found ground among the rootless Muslim youth, particularly among the second and third generations of migrants in the West.” At the same time, according to G. Heinzon, (2003), a professor at the University of Bremen (Germany), in the growing Islamic population of Europe, the problem of self-identification of young Muslims is increasing.

In his book *What it Means to be a European Muslim* (1999), Tariq Ramadan, a Professor at the University of Fribourg (Switzerland), rejects the division of the world into Western and Muslim, as well as the confinement of Muslims in voluntary ghettos. He further insists on bringing the followers of Islam living in Europe closer to European culture. Calling themselves Muslims living in Europe, people deliberately reject the reality surrounding them as alien; but calling themselves European Muslims, on the contrary, emphasizes that they belong to both Islamic and Western civilization. For Muslims, Islam consists of general principles of existence. Every Muslim accepts the five pillars of the faith,⁵ regardless of what tradition they follow. According to T. Ramadan, it is necessary to distinguish between the principles of Islam and its traditions, which may be different considering the history and culture of the people, as well as the laws of the country of residence. Islam is one, but there are many diverse cultures in which it is practiced. There is a place for unity and diversity in Islam. We can’t help but agree

⁴ <http://www.islam.ru/news/2015-09-04/34624>

⁵ Some religious scholars also speak of the sixth pillar of faith. Most often talking about sixth pillar of Islam they mean Jihad which is from a theological point of view “the struggle against one’s own passions.”

that when we talk about European, Canadian, American, and Western Islam, “European” or “Western(ized)” refers to culture, while the principles set out in the Qur’an remain religious. Arab culture is not the culture of Islam, because religion and culture are not coterminous entities.

According to another researcher, B. Tibi (2001, p. 208), Euro-Islam is an attempt at the creation of a liberal form of Islam acceptable to both migrant Muslims and European societies, consistent with European ideas of secularism and individual citizenship. In other words, Euro-Islam is an Islam that is culturally adapted to secular European societies, just as, for example, Islam in Africa is adapted to local African cultures. Accordingly, “Euro-Islam should be compatible with liberal democracy, individual human rights and the requirements of civil society ... Euro-Islam should be directed both against assimilation and against ghettoization.”

Today, European leaders are discussing important problems, such as the preservation of “European values” through the process of Muslim assimilation and their adoption of these European values. But in fact, under the guise of protecting European values, there is a process of promoting neoliberal values that are not always understood – even by Europeans – and that have replaced traditional Christian values. These largely secular values of post-industrial society are associated with the legalization of same-sex marriage in most European countries, children’s social protection (which may have applications that some consider excessive), ideas of tolerance (which often discriminate against the indigenous population), active and public promotion through various platforms (competitions, festivals, etc.) of transgender people, or people with non-traditional sexual orientation (for example, the participation of Conchita Wurst in the Eurovision song contest), and others. These conflict with conservative beliefs, like those held by traditional Muslim and Christian communities. Therefore, it should not be taken for granted that this new citizen with conservative values will necessarily tolerate local laws and norms of behavior, such as feminism and same-sex marriage. And it should be kept in mind that the promotion of neoliberal values (under the guise of European values) is deeply incompatible with “Islamic values,” which are based on traditional, conservative values and attitudes.

Does this mean that identity transformation, and thus cultural integration, takes place only in terms of accepting or rejecting “values”? I believe that the

solution to this issue lies in the sphere of a broader discourse. When we talk about those *European Muslims* who are already integrated into the country of their residence, they often do not face the question of “values” or identity very sharply because they have several sets of values rooted in their various coexisting identities. For example, a person may be at once a Muslim, a Shi’ite, a Syrian, a European, a Parisian, and so on. And in each part of his identity there is no doubt, neither for him, nor for others. If he perceives these identities critically, then he is consciously aware of, and understands, the values he holds, which may be “European and “Islamic.”

But when it comes to *Muslims in Europe*, the situation is slightly different. These are people who contrast themselves and their way of life with people of a different culture and identity, creating a comfortable environment around themselves that reproduces a small sense of their homeland with its laws and regulations. It is here that identity takes shape or changes, depending on a personal understanding of the “self” and “other.” As such, the marginalization and identity crises of young Muslims in Europe who feel alienated from local society create favorable conditions for recruiting young immigrants to extremist and terrorist organizations. The problem of the radicalization of young Muslim Europeans has become obvious, and has become particularly acute against the background of the conflict in Syria and Iraq, where hundreds of young men and women with European citizenship have left Europe and flocked to territories controlled by jihadist groups. The radicalization of European Muslims, and their desire to return to their origins rooted in a distorted understanding of Islam is also at the core of the problem of finding one’s identity.

Against this background, O. Rowe’s postulate about the transformation of modern Islam into post-Islam is of great interest. In this transformation, the relationship between religion and politics is altered, and the superiority of the political over the religious is given in the name of religion itself. If religion used to define politics, now politicians use religion to achieve certain goals in the name of religion. It can be said that identity in Islam is no longer determined by religion, but by political goals. And probably, the juxtaposition of the values of different cultures is merely a means of political manipulation of various influential groups. Such interplay between religion and politics gives rise to the identity crisis within those who subscribe to “Islamic values,”

which can be interpreted differently depending on the political environment.

Conclusion

The modern globalizing world is faced with the need to find new approaches to the relationship between religious and cultural communities. In the last decade, in particular, in connection with the migration crisis of 2015–2016, the issue of self-identification of European Muslims has become particularly important. The problem of European Muslims holding onto their religious and legal traditions has become one of the key issues in the political discourse of many European countries.

We can observe a counterreaction to the factors of globalization both in the Muslim communities of Europe and among the autochthonous population. And this brings to the fore the problem of identity, which is, in my opinion, the code of self-preservation of ethnic groups, nations, and entire states. But the indicators of identity preservation are not only related to national identity; they may also be commitments related to religion, culture, traditions, or something else. Failing to account for these and other complexities, the policy of multiculturalism has not had a positive effect, and in most countries has proven to be a failure.

The persistent policy of integration has raised the problem of self-identification of Muslims. If the first generation of immigrants pursued the goal of survival outside of their country of origin, and therefore had to integrate, then for the second and third generations of young Muslims, the problem of identity is important. Throughout this process, “European values” have come into conflict with “Islamic values”, which has given birth to an identity crisis.

But regardless of that, all people living in Europe will have to continue to live next door to each other and seek strategies for inter-civilizational and intercultural dialogue. And to live in such proximity with each other, it is important for all parties to understand, tolerate, and have mutual respect, as well

as the ability to defend the “self” without thereby denying the “other.”

Today, no one questions the fact that Islam has become an integral part of Europe. Therefore, it is necessary that we find an opportunity to reconcile “European” and “Islamic” values. Perhaps Euro-Islam can be the compromise that will help young Muslims successfully integrate and become full citizens without losing their identity. But this scenario is possible only with certain efforts on both sides, “new citizens” and native Europeans. Robert J. Pauly’s (2004) words serve as confirmation of this, for he argues that “the integration of Islamic communities in the context of a fully unified Western Europe requires joint efforts on the part of both Muslims and their predominantly Christian neighbors in the EU member States” (2004).

References

- Heinsohn G (2003) *Söhne und Weltmacht. Terror im Aufstieg und Fall der Nationen*. Bern: Orell Füssli Verlag AG.
- Pauly R J (2004) *Islam in Europe: Integration or marginalization?* Hants (England): Ashgate.
- Ramadan T (1999) *To Be a European Muslim*. Leicester: The Islamic Foundation, 273 pp.
- Roy O (2004) *Globalised Islam: the search for a new Ummah*. New York: Columbia University Press.
- Tibi B (2001) *Islam between culture and politics*. New York: Palgrave.
- Huntington S (1999) *The Clash of Civilizations and the Transformation of the World Order // New Post-Industrial Wave in the West: Anthology*. Ed. B.JI. Inozemtseva. M., 532 pp.
- Fedorova Yu E (2014) Islam in the perception of the modern European community: stereotypes and reality. *Philosophical thought* No 7: 99–125.

Book reviews

A COHERENT TEMPORALIST METAPHYSICS

[Book review of George Allan. *Whitehead's Radically Temporalist Metaphysics: Recovering the Seriousness of Time*. Lexington Books, UK, 2020 – 200 pages]

Vesselin Petrov

Institute of Philosophy and Sociology, Bulgarian Academy of Sciences
e-mail: petrov.vesselin@gmail.com

Process philosophy is one of the greatest philosophical approaches in our days. Its main feature is the metaphysical understanding of the nature of things emphasizing their radical temporality. It is substantiated by Alfred North Whitehead – the founder of contemporary process philosophy – in his philosophical books. However, some philosophers are of the opinion that Whitehead's theory lacks coherency with respect to the nature of temporality. The present year (2020) is marked with one new remarkable book that aims at the elimination of this incoherency. It is the book by George Allan – professor emeritus at Dickinson College, Pennsylvania, USA – *Whitehead's Radically Temporalist Metaphysics: Recovering the Seriousness of Time*. Lexington Books, UK, 2020 – 200 pages. My task in the present review is to present that book.

Let me first explain what Whitehead's metaphysical incoherency consists in. In his magnum opus *Process and Reality* (1929) Whitehead introduces a notion of God as the reason why there is stability and progressive change in the world, and he considers all possibilities to be eternal. George Allan affirms in the book presently under review that by eliminating both of these notions we will be able to recover the coherency of the temporalist foundation of Whitehead's metaphysics. The aim of Allan's book is, first, to argue that introducing possibilities as eternal and God as the wellspring of stability and progressive change makes Whitehead's temporalist metaphysics incoherent: "there is no way to make the radically temporalist aspects of Whitehead's metaphysics compatible with his notion of God as it is developed in *Process and Reality*" (p. ix); second, "to show how what remains in *Process and Reality* after eliminating the notions of eternal objects and God can readily be reshaped into a coherent radical-

ly temporalist metaphysics" (p. vii); and third, to argue "how this reshaping can be furthered by what is found in [the last Whitehead's books] *Adventures of Ideas* and *Modes of Thought*" (ibid).

The author realizes the aforementioned aims in seven chapters of his book.

The first chapter "The God that failed" is devoted to Whitehead's explanation of four metaphysical functions provided by the notion of God in *Process and Reality*, namely, three reasons associated with the primordial nature of God and one reason associated with the consequent nature of God. The author argues that the discussed functions fail to resolve the systemic difficulty they were designed to address. In the Categorical Scheme of Whitehead's metaphysics in *Process and Reality* he identifies two basic kinds of existence: actual entities and eternal objects. (Whitehead refers to possibilities as "eternal objects.") "Eternal objects are *eternal* because they can be objectified not just once but endlessly" (p. 12). Whitehead's difficulty derives from the claim that their eternality entails they are essentially self-contained. Whitehead needs "an ultimate actuality to provide ontological location for the infinity of possibilities required for a metaphysical account of the creative advance ... [he] thinks that God as primordial suffices for that task" (p. 16). However, as Allan George has pointed out, "God is able to give the eternal objects a reason by relating them in all the various ways possible. ... And not only relating all possibilities in all possible ways, but doing so primordially! Which is surely impossible to do in a world characterized as a continually creative advance into novelty" (ibid). So, Whitehead proposes an actual entity as "a *deus ex machina*."

Whitehead shifts to God what one might presume belongs to Creativity. Allan thinks that "the

reason Whitehead shifts from Creativity to God is that the introduction of novelty means the introduction of hitherto unactualized possibilities" (p. 18). "According to Whitehead God envisions all possibilities, all eternal objects, in their individual essences and in their relational essences to all other eternal objects" (p. 20). "Whitehead insists there are no exceptions to the ontological principle, then introduces a notion of God that is obviously an exception. Whitehead gets into difficulty because of his presumption that possibilities must be essentially eternal and self-contained, having no essential relatedness to either the actual world or one another" (p. 25). "Whitehead's notion of God is not a part of the Categorical Scheme ... Rather, it is the first of the Derivative Notions. ... Whitehead ... added the non-primary notion of God in order to assure the coherence of the primary notions. ... The result is quite opposite of what was intended..." (p. 26). "If coherence is to be restored to Whitehead's system, it must be purged of the unfortunate addition of God responsible for the primordial tasks assigned in *Process and Reality*" (p. 27).

The second chapter "The power of the present" is a continuation of the author's thoughts from the first chapter. However, "whereas chapter 1 undertook the negative task of showing that the notion of God that Whitehead includes in *Process and Reality* renders his metaphysics incoherent and inadequate, chapter 2 has undertaken the positive task of showing that the Categorical Scheme in *Process and Reality*, which makes no mention of God, suffices to account for an adaptively stable cosmos, the world as we find it" (p. 51). The chapter is about the power of the present, the capacity of actual entities to achieve through their concrescing a spatio-temporal order of enduring significance. George Allan thinks that "Whitehead has the same problem of ontological incoherence as Descartes, and solves the problem in the same unsatisfactory manner: by the addition of a divine agent" (p. 29). According to Allan, the philosophy of organism can be transformed into a metaphysics that is compatible with experience and internally coherent, and the task of the chapter is to explore this resource, focusing on the Categorical Obligations which are organized by the author into three "transcendental" concepts that he terms Aristotelian, Leibnizian, and Hobbesian.

The keystone in *Process and Reality* is "Creativity"; hence it is also a recovery of the radically temporalist foundations of the philosophy of organism. The author emphasizes the vector character of Creativity as one of the main necessary conditions governing Creativity. He calls this feature of the

nature of things an 'Aristotelian transcendental' "because of the way it focuses on the dynamics of natural processes" (p. 33). On the other hand, "the multiple actual instances of the Creativity principle are vectors toward the best of the outcomes possible for each of them" (p. 38-39), and the author calls this necessary feature of the nature of things "a Leibnizian transcendental" (p. 39). Finally, "a new kind of transcendental condition makes its appearance. It applies its constraints to the concrescent process of each of the actual entities that are members of an enduring object, but it does so for the sake not of each entity but of the interrelation among them, for the sake of the nexus. The condition this transcendental imposes is that one of the aims of the new actual entity emerging within a nexus be that it replicates the character its predecessors have in common" (p. 44). The author calls this third transcendental condition 'Hobbesian' because it governs a process for ensuring the continued survival of a kind of nexus.

The third chapter "The power of the past" is devoted to the argument that the notion of God is in a sense a redundant one and that the resources for novelty in an actual entity's initial aim are provided by the inherited past. Allan rejects Whitehead's claim in *Process and Reality* that without God "the way of the world would lead necessarily toward the degradation of enduring achievement, toward the irreversible loss of whatever truth, beauty, and goodness there might happen to be" (p. 56). "Whitehead has insisted that possibilities must be essentially timeless and therefore without relation to any particular temporal reality" (p. 61). Instead of that Allan proposes that we understand possibilities as inherent features of temporal processes. "Possibilities are ordered intrinsically with respect to the actual world they qualify. Their relatedness is internal to their temporal context, nor something imposed on them by a nontemporal orderer" (ibid). And also: "Possibilities, like actualities, come to be and perish ... The possibilities to be found in any actual entity's past are of three sorts: previously actualized possibilities, possibilities previously entertained but never actualized, and possibilities never before entertained" (ibid). The important question that should be answered here is "If nothing comes from nothing, and if there is no God to provide a nascent actual entity with its initial aim, how can that aim ever arise?" (p. 64). The author's answer is that "physical purposes are a sufficient resource for explaining how a nascent subject can form its own subjective aim and devise novel approaches to its actualization without the need for a *deus ex machina*" (ibid).

To summarize, we can say that the present chapter concentrates on the importance of order over will. It argues “for the primacy of entities able to discover what should be amid the overweening pressure of what is, and then able to act on its behalf” (p. 76). The author criticizes Whitehead’s notion of God “as the agency by which Creativity becomes particularized, the timelessly constant source of the initial aims by which actualization can occur” (p. 77). His critique of Whitehead’s notion of God is “a metaphysical insistence on the ability of temporal momentary events to originate and sustain their cosmos, and therefore to obviate the need for some grander kind of entity to do it for them” (ibid).

Chapter four “The shadow of truth” focuses on Whitehead’s alternative metaphysical account in *Harvard Lectures I* that includes four notions that resonate very well with what George Allan is arguing. Let me summarize them briefly: The first of these notions is Whitehead’s insistence “on the dynamic interdependence of a fundamentally atomic actuality, and the failure of science and metaphysics to articulate adequately its complexity” (p. 83). The second notion “has to do with the open-ended character of the natural order, and the need for theories about it that echo that openness” (ibid). “A third notion is Whitehead’s proposal that the coming to be and perishing of things requires a Ground that does neither: The Eternal. As eternal it cannot act and thus cannot change anything, neither itself nor other entities. Yet it can be changed because it is the Past, which is changed constantly by the perishing of present accomplishment” (ibid). The fourth notion can be named ‘the shadow of time’ and it is an elaboration of the argument from the previous chapter 3 “that the Past is sufficient resource for explaining the Creative Advance” (p. 84).

What is meant by the phrase “shadow of truth” that Whitehead himself uses briefly? According to George Allan “the shadow of truth” is “a myth about a metaphor”; “A shadow is an absence. Truth is a primary feature of any fact, a feature without which it is not fact, without which it has no intrinsic value” (p. 100). The author emphasizes that “the Past is what makes novel possibilities relevant, not God. And it makes them relevant not by any purposive intent, not by a function of some transient final cause, but by the character of what it is. The shadow cast by perishing achievements conveys the possibility of fresh alternatives, often by conveying the impossibility of most of them” (p. 102). Whitehead thus replaces The Eternal with God in *Process and Reality*. Allan reminds us that “the Eternal is divided into two components: an Eternal aspect that houses

all possibilities and an Everlasting aspect that houses all actualities. The Eternal aspect, rather than the Past, is the agency for shaping pure possibilities into initial-aim options; the Everlasting aspect is the agency for translating the achievements of the past into resources for devising better aims and effecting better ends” (p. 107). George Allan explains in the present chapter why he thinks Whitehead’s suggestion about God is metaphysically incoherent. He underlines that “to understand our cosmos metaphysically as through and through contingent ... is to ... take time seriously” (p. 107).

Chapter five “Erotic power” analyzes the main metaphysical notions in the first of Whitehead’s books published subsequently to *Process and Reality*, namely, *Adventures of Ideas*. There is here only a passing reference to the idea of God developed in *Process and Reality*. Instead, the author investigates new ideas serving as a resource for developing his radically temporalist interpretation of Whitehead’s metaphysics. In *Adventures of Ideas* Whitehead articulates seven basic notions connecting science and philosophy. They are: The Ideas, The Physical elements, The Psyche, The Eros, The General Harmony, The Mathematical Relations, The Receptacle. George Allan explores the fundamental notions of any adequate metaphysics, focusing first on Psyche and Eros. Allan criticizes Whitehead’s inclination to deify these two notions. He says: “On four occasions in *Adventures of Ideas* ... Whitehead seems to identify Eros with the God of *Process and Reality*” (p. 115). And then he concludes: “Whitehead ends up with a set of ten different terms, which mutually ‘correct each other’: ‘Together’, ‘creativity’, ‘concrecence’, ‘prehension’, ‘feeling’, ‘subjective aim’, ‘data’, ‘actuality’, ‘becoming’, and ‘process’ ... *Notably missing is the word ‘God’ or any of its synonyms*” (p. 118, italics mine, V.P.).

Next, George Allan considers four of the five notions Whitehead says are fundamental to any civilized mode of living. The first of them is Truth. He considers it the agency which grounds a stabilized social order by relating Appearance to Reality. Then he turns to Beauty and Art as “providing a way to foster and then exploit weaknesses in any stable system, opening pathways to possibilities not otherwise possible” (p. 109). Next is Adventure, which “adds a moral dimension to the quest for novel possibilities” (ibid).

Chapter six “Metaphoric power” explains the radically temporalist ontology of Whitehead’s last book *Modes of Thought*. Here Whitehead “abandons the claim that possibilities are eternal objects, which permits abandoning the idea of a God” (p. 137). In

the book “we are given a sketch of that world as thoroughly contingent, a world in which actualities are all finite components of finite dynamic systems” (ibid). George Allan displays in this chapter “the radically temporalist ontology of that book, focusing on its key ideas about how existence and meaning – fact and value – are intertwined forms of finitude” (ibid). The author has developed his understanding of *Modes of Thought* in four parts through four gerunds: mattering, interpreting, rationalizing, and civilizing. According to him, “Whitehead thinks that inquiry and action are functionally integrative finite processes, and that we are able to understand things and accomplish things more effectively if we take our inquiries and our courses of action to be the deployment of functionally effective specific aim” (p. 143). The author’s second claim in the chapter is that Whitehead “thinks functionally integrated processes are interpretive standpoints, that we ourselves and what we know are limited – finite and contingent – accomplishments” (p. 149). The author’s next claim is that Whitehead “understands reason not as a fixed faculty of the mind but as a useful way of being conscious. He sees it as a method for devising fresh interpretive standpoints, ones able to resolve theoretical and practical problems that have resulted from the clash of divergent purposes and incommensurable truths and values they entail” (p. 156). Finally, the author’s fourth claim in the chapter is that Whitehead “thinks both poets and philosophers have a crucial role to play in sustaining the viability of civilized modes of thought and practice by helping to prevent the success of currently effective ideas and institutions from blocking the emergence of successors with differing perspectives, ones that may be more adaptively suited to altered conditions” (p. 162).

The last, **seventh chapter** “The solemnity of the world” is devoted to the discussion of finalities: “the ultimate meaning of the temporal process” and “the endless emergence and perishing of finite things” (p. 163). The chapter considers two notions of ultimate meaning: The first is the notion of Peace in Whitehead’s book *Adventures of Ideas*. Earlier, in Chapter 5 of the book, George Allan considers four essential qualities (Truth, Beauty, Art, and Adventure), but misses the fifth essential quality – the sense of Peace. This missing essential quality is considered here – for good or bad – in the last chapter of Allan’s book. The inclusion of this discussion in the final chapter of the book, rather than Chapter 5, is a potential disadvantage, but ultimately Allan addresses all of the essential qualities. The second notion of ultimate meaning is that of Totality, considered in

Whitehead’s book *Modes of Thought*, which is fundamental to understanding Whitehead’s argumentation in *Modes of Thought*. “Totality of things is not a thing; it does not organize everything into some great absolute all-encompassing unity. It is not God or some other transcendent agency” (p. 170). Both these notions of Peace and Totality are explicated “as being about the sense finite things have of belonging to an ever-emerging, ever creating, ever perishing ultimate community composed of predecessors, contemporaries, and possible successors” (p. 163). The chapter continues by considering the antithesis of this view: that “this ultimate community is an everlasting ontological reality ... the consequent nature of God” (ibid). The author ends the chapter with a discussion of Tragic Beauty as a way to summarize the metaphysical hypothesis developed in the whole book. According to Allan, Tragic Beauty is crucial to an understanding of Peace. He says, “The pursuit of ideals is not only a high-risk adventure in which we are all the time suffering setbacks and needing to refashion our plans and rethink our purposes. We must also recognize that it is an effort doomed to failure” (p. 185). Tragic Beauty is “a fact of the world ... located not at the periphery but at the core of the cosmos, at the heart of the very nature of things” (ibid). Allan points that “The tragedy of Tragic Beauty is that we failed to get the job done. ... The beauty of Tragic Beauty is that it is also Truthful, that it was a reasonable goal and so remains, despite the failure, a viable possibility” (p. 186).

The chapter concludes – and it can be accepted as a conclusion of Allan’s book as a whole – with the remarkable statement that “The keystone of Whitehead’s philosophy of organism ... is the denial that anything can be imperishable, serving as the perfect uncreated timeless foundation for what is created or as the perfect noniterated everlasting unity of what is created. ... All existent entities are creatures, all creatures are created, and everything created, everything that comes to be, also perishes” (p. 181). In addition, the “counter to those who insist on the metaphysical necessity of a power able to rescue the things of this world from their seemingly endless perishing is that the ultimate meaning of life is to be found solely in the finite power of finite entities to create things of value” (p. 184).

Finally, let me firmly stress again that George Allan’s book is a remarkable investigation on temporalist metaphysics that should be read and studied by every philosopher who is interested in the recent developments of metaphysics.

Rudolph Hermann Lotze: Mikrokosmos: Ideen zur Naturgeschichte und Geschichte der Menschheit. Versuch einer Anthropologie. 3 vols., edited by Nikolay Milkov, Hamburg: Felix Meiner, 2017. ISBN 978-3-7873-3180-2.

Michele Vagnetti

University of Florence, University of Paderborn
e-mail: michelevagnetti90@gmail.com

The first edition of this work by Rudolph Hermann Lotze was published in Leipzig by Salomon Hirzel in three volumes under the title *Mikrokosmos* (1856/58/64). The text was soon translated into Russian (1866) by Evgenij Korsch, into English (1884–85) by Elizabeth Hamilton and E. E. Constance Jones, and also into Italian (1911–16) by Francesco Bonatelli and Gaetano Capone Braga,¹ who actually translated the first and second volumes of *Mikrokosmos*; the Italian translation is undoubtedly well rendered, although the register is laborious and obsolete. For this reason a new (abridged) Italian translation of *Mikrokosmos* by Luigi Marino and Gianstefano Stella (1988) was necessary.²

The genesis of the book is “well documented,” although it is long and complex.³ In a letter dated February 18, 1837 to his friend Ernst Friedrich Apelt—who was of great importance in Lotze’s philosophical education and through whom Lotze gained access to the thought of Apelt’s master, Jakob Friedrich Fries—Lotze outlined his plans to write essays on human nature and its place in nature.⁴

The letters between Lotze and Salomon Hirzel, his publisher and close friend, allow us to understand in greater depth the complexity of the genesis of the work. In a letter (March 8, 1853) to Hirzel,

Lotze outlines the fundamental pillars of his work: the *mechanics* of human bodily life; the *mechanics* of mental life and the general laws of its interaction with the bodily system; the characterization of the relations (*die Beziehungen*) between man as a psycho-physical organism and a) the system of the animal kingdom, b) external nature in general, and c) other human beings (these are the *conditions* of the development of specifically *human* life); the laws of history; the philosophy of history (the *practical*, *aesthetic*, and *theoretical* struggle with nature and the world); the state; the social order; religious consciousness; the exposition and criticism of different historical epochs.⁵

There is a clear difference between the 1853 project and the final version of *Mikrokosmos*. Whereas the 1853 project ended with an extensive analysis of the philosophy of history, in *Mikrokosmos* the metaphysical vision of the world plays a central role. The ninth and last book of the work is dedicated to the being of things as being-in-relation. Lotze’s metaphysics is based on the concept of the interconnection of things (*der Zusammenhang der Dinge*) and, therefore, opposed to Johann Friedrich Herbart’s metaphysics. According to Herbart, there is a radical difference between “real essences” (explored by metaphysics) and the subjective way in which essences are perceived and known. Herbart embraced the Kantian distinction between noumenon and phenomenon. Real essences are simple and have no relation with other essences. The relations between essences are due to the contingent point of view of the knowing subject on reality. According to Herbart, relations are mere psychological constructs and have no metaphysical value; human knowledge construction consists of judgments, which are associations of ideas or representations. Lotze emphatically rejected this doctrine, arguing that, metaphysi-

¹ F. Bonatelli mentioned his Italian translation of *Mikrokosmos* in his letter to Lotze, dated May 4th, 1861, (R. H. Lotze, *Briefe und Dokumente*, R. Pester and E. W. Orth (ed.), (Würzburg: Königshausen & Neumann, 2003), 376.

² R. H. Lotze, *Microcosmo. Idee sulla storia naturale e sulla storia dell’umanità. Saggio di antropologia*, L. Marino and G. Stella (ed.), (Torino: Unione Tipografico-Editrice Torinese, 1988).

³ R. H. Lotze, *Mikrokosmos*, Nikolay Milkov (ed.), (Hamburg: Felix Meiner Verlag, 2017), XXV.

⁴ R. H. Lotze, *Briefe und Dokumente*, 83–84; see: F. C. Beiser, *Late German Idealism*, (Oxford: Oxford University Press, 2014), 251.

⁵ R. H. Lotze, *Briefe und Dokumente*, 229–30.

cally speaking, reality is relational. In *Mikrokosmos* (III, IX) Lotze wrote about the interrelation of things and opposed the Herbartian concept of the position (*Setzung*) of being. According to Lotze, relation is rather the ontological form of substances, since the whole substance consists of single elements (atoms). Lotze's metaphysics is grounded on three fundamental concepts: atoms, substances, and relations. Substances, consisting of atoms and relations, are 'states of affairs' (*Sachverhalt*), which are furthermore expressed as logical forms through judgments. Judgments (logical forms) connect ideas in the same way as substances (ontological forms) interrelate in reality. This is possible because, unlike Herbart, Lotze no longer considers judgment a simple association of ideas or representations; for him, judgment connects things or contents. This means that the logical and the epistemological forms of judgments are secondary and depend on the ontological forms, which are primary and independent.

Herbart is mentioned only once in the ninth book of *Mikrokosmos*, which focuses on the doctrine of the ideality of space.⁶ Lotze rejected associationism, according to which judgments are connections of ideas or representations. In his opinion, the associationist theory is not able to justify rationally the validity of judgments. Judgment is a connection of things or contents according to the criteria of the retrospective activity of thought. Quite consistently, Lotze considered space not as an innate idea but as a product of the retrospective activity of thought. This does not mean that the idea of space is less universal and necessary (a priori). The validity of this idea is assured by the fact that it is produced by the activity of thought (*die Tätigkeit des Denkens*).⁷ The activity of thought is what constitutes coherence (*die Zusammengehörigkeit*) and validity (*die Gültigkeit*).

An important idea that Lotze presented in *Mikrokosmos* is that the mind is not a mirror which reproduces an exact copy of the outside world in us; it is a continuous critical activity that receives external stimuli from the outer world and re-elaborates them according to its own criteria. Precisely for this reason, Lotze adopted Kant's conception of the ideality and priority of space, although, at the same time, he denied innatism, rather opting for the doctrine of mental production.⁸

Even if the outer world were really extended, we could not know it because, in order to be subjectively known, it would have to be translated into our inner language, i.e. into interactions of sensations or

inner states of our own nature (*Wechselwirkung von Eindrücken oder inneren Zuständen unseres Wesens*). Ideas or representations do not organize themselves according to a spatial pattern; it is their *content* that assumes a spatial meaning in the mind which elaborates them.⁹

Lotze rejected Kant's attempt to base a spatial a priori on innatism and, at the same time, refused Herbart's attempt to base the spatial a priori on the interaction of representations. According to Herbart, it is the mechanism of the representations that is provided with spatial relations. In other words, Lotze rejected Kantian innatism and Herbart's psychologism, opting for a third perspective, which considers space as the *content* or the mental *value constructed* by the activity of mind.¹⁰

In *Mikrokosmos* Lotze focuses then on the recognition of the serious inner opposition between human spiritual needs (*die Bedürfnisse des Gemütes*) and the results of modern human science (*die Ergebnisse menschlicher Wissenschaft*).¹¹ He wrote:

But all the same it is in such mediation alone that the true source of the life of science is to be found; not indeed in admitting now a fragment of the one view and now a fragment of the other, but in showing how *absolutely universal is the extent* and at the same time how *completely subordinate the significance, of the mission which mechanism has to fulfil in the structure of the world*.¹²

In order to solve the problem of the possible explication of the structure of the world, human knowledge can follow two different paths, both necessary to satisfy the deepest human needs: science and metaphysics. They consist of two different methods of investigation: mechanism and teleology. As said above, in *Mikrokosmos*, mechanism has a

⁹ *Ibid.*, 485–88.

¹⁰ Interestingly enough, Lotze's idea that space (and time) are mental constructions corresponds to Stephen Hawking's "holographic principle" of understanding the universe (*The Universe in a Nutshell*, Bantam Books, 2001; Hawking and Mlodinow, *The Grand Design*, Bantam Press, 2010, 44, and Max Tegmark, *Our Mathematical Universe*, Knopf, 2014, 130). Furthermore, the current cosmology mentions 12 dimensions of which human beings only perceive three or four. It is possible to imagine "intelligences" which perceive only two dimensions and others which, starting from the same "reality", can perceive 12 dimensions.

¹¹ *Ibid.*, first vol., III*.

¹² *Ibid.*, XIII*; English trans., *Microcosmos: An Essay Concerning Man and his Relation to the World*, E. Hamilton and E. E. Constance Jones (trans.), (Edinburgh: Clark, 1885), 2 vols., I, xvi.

⁶ R. H. Lotze, *Mikrokosmos*, third vol., 494.

⁷ *Ibid.*, first vol., 257–59.

⁸ *Ibid.*, third vol., 494.

universal extension: all the processes and events that happen are subject to this universality and realized through a *mutual interaction* (*Wechselwirkung*). However, mechanism is not sufficient to *elucidate* (*erläutern*) the meaning of the true structure of the world. Nevertheless, in “Selbstanzeige des ersten Bandes des Mikrokosmos”, mechanism is presented as necessary for human knowledge, since the scientific *description* of the world, according to the mechanical method, constantly *stimulates* (*anregen*) questions about the meaning of the human being and human life.¹³ Mechanical science provides a description of the outer world, but does not explain its meaning.

By applying the philosophical principle of the Teleomechanism, Lotze employed his specific dialectical method as well as the method of perspectivism. This allowed Lotze to show that several philosophical contradictions were only apparent (*scheinbar*) and due to the limitations inherent in different points of view (*Ansichten*) (this is a *terminus technicus* of Lotze’s philosophy) from which we consider reality.

For example, Lotze considered the philosophical systems of his time, such as Herbart’s realism and Hegel’s idealism, as simply different points of view on reality with different finalities. For this reason, they did not contradict each other. Reality is complex and articulate; philosophical systems are branches of this complex. One cannot grasp the totality of reality from one specific perspective.¹⁴ Lotze tried to reach a mediation between realism and idealism through the *critical selection* of single elements of these doctrines. If, on the one hand he rejected realistic metaphysics according to which the structure of the world was “real”, on the other he adopted the realistic method and embraced its focus on concrete scientific research. The realistic method deals with the causal description of the development of events, starting from some given forces and conditions of essences that are related to other essences.

Other thinkers and scientists who contributed to and influenced the scientific and realistic approach that Lotze employed as the starting point of his philosophical thought were Jakob Friedrich Fries, Ernst Friedrich Apelt, Ernst Heinrich Weber, and Gustav Theodor Fechner. Lotze embraced the idealistic metaphysics, according to which the structure of the world was “ideal”, but he rejected the formalistic outcome of this doctrine, which considered the whole of reality as the development of the pure form

of the idea. In Lotze’s philosophy the focus on the *concreteness of experience* is fundamental. In this sense, Lotze reached a *critical idealism* which was a sort of attempt to reconstruct idealism on a realistic basis.

Besides his attempt to mediate between these two main philosophical systems of the time, Lotze carried on a constant fight against the revival of radically materialistic positions. It is important to remember that Lotze published *Mikrokosmos* precisely at the time of the *Materialismusstreit*.¹⁵

In 1856, the year of publication for the first volume of *Mikrokosmos*, Immanuel Hermann Fichte published his *Anthropologie. Die Lehre von der menschlichen Seele. Neubegründet auf naturwissenschaftlichem Wege für Naturforscher Seelenärzte und wissenschaftlich Gebildete überhaupt* (Leipzig: F. A. Brockhaus). In this work Fichte (junior) formulated his theory of the mind-body problem in response to Lotze’s theory, of which he was quite critical. In his *Anthropologie*, Fichte used Lotze’s *Medizinische Psychologie oder Physiologie der Seele* (1852) as a constant element of comparison. He is frequently quoted by Fichte (§ 110, §§ 121–23, §§ 130–32, §§ 185–87). The omnipresence of the mind in the body is the central idea of *Anthropologie*. Based on this theory, Fichte criticized Lotze’s psychology. It is not accidental that, in a letter to Salomon Hirzel dated June 1, 1856, Lotze defined Fichte’s *Anthropologie* as a “polemic almost exclu-

¹⁵ The controversy between materialists and anti-materialists is a key element to understanding mid-19th century German philosophy. The critical works by K. Fischer (*Die Unwahrheit des Sensualismus*, 1853), J. Frauenstädt (*Der Materialismus*, 1856), and others were published in the same years as the critical works of many important scientists of the time: Liebig, Andreas Wagner, and Rudolf Wagner. An important controversy developed between Rudolf Wagner and Karl Vogt in Göttingen at the 1854 conference of the “Gesellschaft Deutscher Naturforscher und Ärzte”. The following year there was another controversy between Lotze and Czolbe. In his 1855 *Neue Darstellung des Sensualismus* Czolbe stated that Lotze’s article “Lebenskraft” had inspired the idea of the uselessness of the trans-sensual explanation principle. The same year, Lotze reviewed Czolbe’s work in “Göttinger gelehrten Anzeigen” (pp. 1521–38) and defined Czolbe’s materialistic position as “bad metaphysics”. Lotze’s critique of Czolbe’s mechanistic explanation of the phenomena of consciousness, devoid of all directly controversial aspects—Czolbe is never directly quoted—became part of *Mikrokosmos*. For more on the materialistic controversy, see: F. Gregory, *Scientific Materialism in Nineteenth Century Germany*, Dordrecht-Boston: D. Reidel Publishing Company, 1977.

¹³ *Ibid.*, 462.

¹⁴ See: R. H. Lotze, *System der Philosophie*, G. Misch (ed.), Leipzig: Meiner, 1912, xciv.

sively against me.”¹⁶ One can detect in *Mikrokosmos*, especially in its first volume, the trace of Lotze’s reading of *Anthropologie*. However, there are no clear reactions to Fichte’s work: in *Mikrokosmos* there is no explicit reference to their personal controversy.¹⁷ In “Selbstanzeige des ersten Bandes des Mikrokosmos” published in the same year, Lotze again addressed the question of Fichte’s *Anthropologie*.¹⁸ A direct answer to this controversy is contained in *Streitschriften*, which we can consider, following Lotze, a response to *Mikrokosmos*. Fichte considered Lotze’s psychology a development of Herbartian theory of the mind as substance.¹⁹ To be true to the facts, we must underline that this criticism of Lotze was unearned, since he was not a rational psychologist; hence, there is no reason to analyze this issue in greater depth. Furthermore, Fichte’s criticism also concerned § 10 (“Von dem Sitze der Seele”) of *Medizinische Psychologie* and the occasionalist theory of the interaction between mind and body.²⁰ Besides the rather technical and sophisticated theoretical differences between the two philosophers, they both opposed the materialist direction of German philosophy. Whereas materialists tried to reduce the mind to the body, Lotze, in his *Mikrokosmos*, presented the mind as a specific activity and dynamic which is irreducible and not mechanically explainable. Here it is important to bear in mind that these were not the considerations of a philosopher but of a scientist²¹ who was undoubtedly the most brilliant physiologist of his time.²²

In his work Lotze presented a new possible way to interpret the results of this scientific research, which indicated the existence of a specific mental sphere that could not be reduced to the simple physical and physiological level. In other words, Lotze’s philosophy is compatible with the results of scientific investigation, but, at the same time, affirms that philosophy itself is irreducible to scientific data. Precisely for this reason Lotze’s ideas had a great impact and spread until the First World War not only

in continental Europe and Russia, but also in Great Britain and the United States of America. This impact was essentially due to his vision of the man and the world outlined in *Mikrokosmos*, a work which, because of its stylistic completeness, became “a classical German literary monument”, as we read in Friedrich Überweg’s *Grundriss*.²³

The seventh edition of *Mikrokosmos: Ideen zur Naturgeschichte und Geschichte der Menschheit. Versuch einer Anthropologie* was published in 2017—year of the 200th anniversary of Rudolph Hermann Lotze’s birth (1817–1881)—by Felix Meiner Verlag, Hamburg and edited by Prof. Dr. Nikolay Milkov (University of Paderborn). The seventh edition follows the indications of the sixth edition (1923) directed by Raymund Schmidt and published by Felix Meiner Verlag, Leipzig: the stylistic choice of Antiqua typeface instead of Blackletter (Gothic script) and the modernization of the original title from *Mikrokosmos* replaced by *Mikrokosmos*. However, Schmidt’s 1923 introduction is not included in the seventh edition. Nikolay Milkov’s introduction is wide-ranging, exhaustive, and accurate. The index of concepts is also very detailed and, alongside the index of names, represents a useful instrument for navigating the text. In the appendix to the first volume of the text we find the “Selbstanzeige des ersten Bandes des Mikrokosmos”; in the appendix to the second volume there are the “Selbstanzeige des zweiten Bandes des Mikrokosmos” and the “Einschub” about Darwin’s theory of evolution. The second “Selbstanzeige” are not included in the second print volume of the work, but are available online (<https://meiner.de/mikrokosmos-13884.html>). Thanks to Nikolay Milkov and Meiner Verlag for this edition that contributes to the development of the research on such a great philosopher and psychologist who has been undeservedly forgotten for almost a century.²⁴

¹⁶ R. H. Lotze, *Briefe und Dokumente*, 282. Lotze wrote: “Vor einigen Tagen hat mir Fichte seine neue Anthropologie geschickt, die fast nur eine Polemik gegen mich ist...” (translation mine).

¹⁷ *Ibid.*

¹⁸ R. H. Lotze, *Mikrokosmos*, first vol., 463–64.

¹⁹ I. H. Fichte, *Anthropologie. Die Lehre von der menschlichen Seele. Neubegründet auf naturwissenschaftlichem Wege für Naturforscher Seelenärzte und wissenschaftlich Gebildete überhaupt*, (Leipzig: F. A. Brockhaus, 1856), 306.

²⁰ *Ibid.*, 307 ff.

²¹ *Ibid.*, 309–10.

²² *Ibid.*, 276.

²³ To be more explicit, Friedrich Überweg wrote: “Die allgemeinste und umfassendste (auch Früheres wiederholende) Darstellung seiner Weltansicht hat Lotze im *Mikrokosmos* (1856–64) gegeben. Derselbe ist durch den formvollendeten Charakter seines Stils zu einem klassischen deutschen Literaturdenkmal geworden...” (F. Überweg, *Grundriss der Geschichte der Philosophie* (4. Teil: *Die deutsche Philosophie des XIX Jahrhunderts und der Gegenwart*), 12th edition reworked by T. K. Oesterreich, Berlin: Mittler u. Sohn, 1923, 303), (translation mine—M. V.).

²⁴ I would like to thank Nikolay Milkov, my *Doktorvater*, for his helpful comments and critical notes. I also thank him for the important teachings on the critical analysis of texts and on writing in the history of philosophy.