

BALKAN JOURNAL OF PHILOSOPHY

Vol. 13, Issue 2, 2021

The issue is sponsored by the Bulgarian National Science Fund, competition “Bulgarian Scientific Periodicals – 2020”, project No: KII-06-HII2/11 from 02 Dec. 2020

TABLE OF CONTENTS

Articles

Invited paper

Platonism and Mathematical Explanations:

Some Critical Remarks on Explanatory Proofs and Debunking Arguments (II) / 113

Fabrice Pataut

Subject in Classical Logic and Speculative Logic / 123

Dragos Popescu

Dostoevsky's Christ and Nietzsche's Jesus as “Conceptual Characters” / 133

Tamara S. Kuzubova

Revisiting Hull's Evolutionary Model of Science / 145

Mohammad Mahdi Hatef

Political Ontology of Alain Badiou and Sylvain Lazarus: Turning to the Subjective Principles of Politics / 153

Irina Zhurbina

Integration of Learning and Neuroscience Theories with AI-based Technologies in Intelligent Learning System in Accordance of Whiteheadian Tradition and Contemporary Process Theory / 161

Rossitza Kaltenborn

Learning Emotion Concepts: Further Thoughts on Emotional Granularity / 175

Marina Bakalova

Learning with ANIMA / 181
Rosen Lutskanov

On Some Aspects of the European Knowledge Economy / 193
Petar Iliev

Struggling to Achieve High Creative Standards in Romanian Advertising Industry/ 205
Mădălina Moraru

Book review

Pepka Boyadjieva, Petya Ilieva Trichkova (2021). Adult Education as an Empowerment. Re-imagining Lifelong Learning through Capability Approach, Recognition Theory and Common Goods Perspective. Pallgrave Mcmillan, 354 pages / 213
Nina Dimitrova

Articles

Invited paper

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

Fabrice Pataut

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

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 non numerical 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 derivations resorting to conservative mathematics and proofs involved in non numerical best explanations also require abstract objects, concepts, and principles under the usual reading of “abstract,” one might complain that such accounts turn out to be as metaphysically loaded as their platonistic counterparts. One might then urge that ontological parsimony is also required of these nominalistic accounts. It might, however, prove more fruitful to leave this particular worry to the 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 claims. Our knowledge that applied mathematics is conservative is established at a cost, and so is our knowledge that nominalistic proofs play a genuine theoretical role in best explanations. I will suggest that the knowledge one must acquire to show that nominalistic deductions and explanations do indeed play their respective theoretical role involves some question-

begging assumptions regarding the nature and validity of proofs.

As for debunking, even if the face value content of either non numerical claims, or conservative mathematical claims, or platonistic mathematical claims didn't figure in our causal explanation of why we hold the mathematical beliefs that we do, construed or understood as beliefs about such contents, or as beliefs held in either of these three ways, we could still be justified in holding them, so that the distinction between nominalistic deductions or non numerical explanations on the one hand and platonistic ones on the other turns out to be spurious with respect to the relevant propositional attitude, i.e., with respect to belief.

Key words: Abstract Objects; Clarke-Doane (Justin); Colyvan (Mark); Conservativity; Debunking Arguments; Explanatory Proof; Field (Hartry); Indispensability Argument; Inference to the Best Explanation; Mathematical Explanation; Mathematical Realism; Mathematical Truth; Nominalism; Nominalization; Numbers; Ontological Economy; Platonism; Proof (Classical vs. non Classical or Intuitionistic); Quantification (Objectual vs. Substitutional); Realism; Rizza (Davide); Safety; Sensitivity.

* The first part of the article, covering sections 1 to 3, was published in *The Balkan Journal of Philosophy*, vol. 12, Issue 2, 2020, pp. 63-74. Note that the short abstract and the list of key words and names common to both parts has been modified.

4. Knowledge (II)

Suppose that in order to give a nominalistically respectable explanatory proof you need to know more than you need to know in order to give a mathematically loaded explanatory proof. The extra knowledge isn't mathematical but requires a fair amount of extra-mathematical or extra-numerical assumptions, theories, and techniques to be applied.

How should we assess the merits and demerits of the non mathematical cognitive cost? We may want to measure three things: (i) the sum total of theory one must appeal to, (ii) the length of proofs or formal derivations obtained given (i), and (iii) the range of application of the selected methods. It may be that, in certain cases, the theory appealed to is more complex, or that proofs and derivations happen to be longer and more cumbersome, or that the range of application of the methods is restricted compared to a bona fide mathematical derivation or numerical explanatory proof. In other words, it may turn out that in specific cases we give up our right to epistemic economy ((i) and (ii)), to elegance or beauty ((ii)), to strength ((iii)), or to some combination thereof.

It's important to find out how (i)–(iii) interact in order to assess the merits and demerits of nominalistic proofs. Should these proofs be preferred in spite of these drawbacks? If so, there must be a trade off between costs and rewards that speaks in favor of the nominalistic choice even though the cost happens to be high. The trade off isn't just a question of the psychological or mental collective effort we humans must muster in order to obtain a bona fide nominalistic proof or a nominalistically respectable derivation. We must explain *why* the effort has the features that it has rather than others it could have but just doesn't have. The assessment of epistemic cost is an objective matter, and the problem in this respect is to determine the means by which one will measure the costs of the trade off. It isn't merely a question of observing and computing how much time has been spent, how much collective stamina has been built up to secure the results, and so on and so forth. It is a question of why it should be so, and in this or that particular way.

Once we know why the nominalistic proof or derivation must be obtained in some particular way, we might perhaps appeal to principles that would justify a conclusion to the effect that the trade off is

indeed worthwhile *in spite of the cognitive cost*; but that is a different matter altogether — a matter that must be assessed on independent grounds, i.e., on grounds that must be independent of any consideration that would belittle the cost from the outset, or presuppose that its very consideration is downright irrelevant. Although we don't want to rule out the possibility that some philosophical principle or argument overrides the epistemological drawback or concern, we certainly mustn't put the cart before the horse.

Let us look once again at Field and Rizza's proposals. However different they may be, they share a common feature: In order to avoid a commitment to numbers in the derivation (Field) and in the explanation (Rizza), one must appeal to non numerical axioms whose models have suitable numerical representations. It is essential that the ordinary numerical treatment of the theory one wants to dispense with — i.e., the treatment that *prima facie* commits one to numbers given a face value reading of the theory — may be recovered from a non numerical system of axioms, so that the platonistic reading of the mathematical theory taken at face value is undermined.

In Field's case, we're allowed to use mathematical existence assertions provided that the theory they belong to is conservative (*Field 1980: 14*). This is established by showing that, for number theory and ZF, consistency implies conservativeness given Robinson's joint consistency theorem (*Robinson 1956*). The idea that mathematics applies conservatively to nominalistic versions of physical theories such as Newtownian gravitational theory (*Field op. cit.: 19*) requires modellability in ZF. What one gets then, in this case, is a platonistic proof that a physical theory "has all the nominalistically statable consequences that the usual platonistic formulation of the [...] theory [...] has" (*Field 1980: 5*). What makes it platonistic, presumably, is the recourse to ZF (with the axiom of choice and with or without urelements).

One may, according to Field, use platonistic methods of proof in arguing in favor of nominalism: Since the argument shows that abstract entities are not needed, either for ordinary inferences or for physical science, the Quine-Putnam indispensability argument thereby loses its clout (*Field op. cit.: 6*).

In Rizza's case, one establishes that mathematics plays an explanatory role "in a non-ontological

way” (Rizza 2011: 113) by conceptualizing the structure of an empirical phenomenon so that “the explanation may be presented as a derivation from certain formal conditions” (Rizza: *loc. cit.*). Obviously, in this case, the epistemological worry has shifted from a focus on non causal abstracta, or on the reliability of beliefs in non causal mathematical objects, to a focus on formal or structural properties of some bona fide non-numerical explanation.

What I want to suggest here is that, even if we grant that platonistic methods are nominalistically respectable (Field), or that an emphasis on a commitment to concepts is preferable to an emphasis on a commitment to objects (Rizza), it remains that the axioms one relies on must possess certain features. Whether they be Suppes’s axioms (Rizza) or Zermelo-Fraenkel’s (Field), they must allow the recovery of what’s usually appealed to in a mathematical practice that is unconcerned with philosophical worries about abstracta (whether they be concepts or objects in the ordinary (non-Fregean) sense).

What are the features that allow and indeed secure the recovery, and how do these axioms inherit them? The features are structural in the following sense. In Suppes’s case, a relation that holds between consecutive points in a relational structure whose domain is both finite and non-empty, and on which a quaternary relation is defined, determines congruent intervals. So if one supposes that the life-cycles of *Magicicadas* do indeed satisfy Suppes’s axioms, one may show that the numerical properties one wishes to dispense with are mere counterparts of purely empirical properties of time intervals that an empirically minded nominalist may be satisfied with. The reconstruction of a mathematical or numerical reasoning within a non-numerical, i.e., geometrical theory is possible only provided axioms of measurement theory correctly characterize finite equal-interval structures (see Rizza *op. cit.*: 107-108). They do so by fiat, and one has of course, no right to complain that this is how they enable us to secure the much-wanted recovery. Axioms are devised for a purpose, and one may certainly use them in any way one pleases as long as the use is consistent.

Now the fact that one has to establish the dispensability of numerical properties by showing that they have the particular ontological status of counterparts might be a source of worry. The counterpart

claim that justifies the dispensability claim that would allow us to shoo platonism away is established given the acceptance of Suppes’s axioms. These counterparts are not just “mere” or innocuous, ontologically light counterparts. They inherit a light ontological status *because* Suppes’s axioms have models that may be described or investigated set-theoretically as relational structures. The problem, of course, isn’t with Suppes’s axioms *per se*, but with the way in which they are used, especially Axiom 4.

As for Field, the conservativity principle guarantees that a nominalistically-stated assertion A^* resulting from restricting each quantifier in A with the formula “not $M(x)$ ”¹ isn’t a consequence of a mathematical theory unless it is logically true. The claim that the mathematical theories one resorts to do indeed satisfy that principle implies that our mathematical claims are true in all possible worlds, i.e., that they are a priori true (Field *op. cit.*: 13). Field considers the objection that standard, i.e., classical mathematics might after all turn out *not* to be conservative because it might “conceivably” be inconsistent. The weak reply to this objection is that a proof of the inconsistency of classical mathematics that would motivate a revision of mathematical theories and of mathematical practice is highly unlikely. Of course, Field offers more than a quasi-inductive reason for the belief that classical applied mathematics is consistent.

The key point here is that the conservativity of number theory and ZF is an (almost direct) consequence of their consistency, i.e., of Robinson’s joint consistency theorem. It hardly matters whether such theories are true. What matters is that a mathematics that applies conservatively to a nominalized physical theory is modellable in ZF, and that the conservativity claim we need for the vindication of nominalism follows from the consistency of ZF. If conservativity is secured in this way, i.e., if it can be shown that the gap between the consistency of classical mathematics and its conservativity is that narrow, then both consistency and conservativity are indeed sufficient for the purpose of application.

It must be noted that no matter how narrow the gap might be, and although some form of elegance and beauty might be preserved in the cases covered

¹ The predicate “ $M(x)$ ” means intuitively “ x is a mathematical entity.” See the first part of this article, sect 2, p. 68.

in Field's account (in the arithmetical, the geometrical, and the Newtonian gravitational theory cases), the consistency of set theory that allows us to use mathematics conservatively is something we must *assume*. We must assume that the usual axiomatizations of set theory are indeed consistent (and we know from Gödel and Church respectively that if ZF were to be *proved* consistent, it would be incomplete and undecidable). So the avoidance of abstract counterparts that *prima facie* refer to numbers and the obtaining of "concrete" conclusions using the homomorphisms with those counterparts is secured at a cost. "Descending back to the concrete" requires us to find a consistent mathematical structure that allows us to prove the existence of structure-preserving mappings from nominalistically respectable concreta to the abstract structure.

Nominalistically respectable inferences and derivations are deemed correct and truth-preserving only provided that we may obtain the very same results *without* ascent into the abstract, so that the secured homomorphisms are just convenient devices. The epistemic cost of *not* ascending into the abstract must be evaluated in terms of what it must take to prove a "representation theorem." The proof must establish that for any model of the appropriate axiom system, there is a function that satisfies the appropriate "homomorphism conditions," mapping whatever nominalistically respectable objects one needs in a nominalistically respectable derivation onto whatever platonistically respectable objects one needs in a downright platonistic derivation. Although it would certainly be unjust to object that the vindication of nominalism thus conceived rests on the assumption that ZF is consistent, or on Gödel and Cohen's relative consistency proofs (because these may be turned into relative conservativeness proofs), it seems fair to note that the epistemic cost of showing that classical mathematics isn't true but merely useful in deducing "concrete conclusions" from carefully crafted nominalistic or nominalized theories is high indeed, both in terms of the derivations themselves and of their theoretical background.

5. Debunking

What is debunking? To debunk is to expose what was formerly held to be both true and warrant-

ed as false and of vastly exaggerated value. When you debunk a belief or a theory, you don't merely expose the pretension of falseness, you somewhat ridicule the belief or the "theory" you take not to be a genuine one by undermining it in a strong way. In the particular case we're looking at, the debunking strategy purports to undermine both the mathematical beliefs and the claim that there are grounds in their favor by exposing their common causal origin. The causal origin disclosed by the explanation of how we come to have such beliefs is supposed to play an epistemic role with respect to both aspects of the undermining. The fact that the beliefs-*cum*-justifications exhibit a given causal pattern (or perhaps the fact plain and simple that they are *caused*) is supposed to provide sufficient grounds for rejecting both the beliefs and the justifications in their support as equally unwarranted, that is, as not constituting knowledge.

In particularly cutting versions of the strategy, the argument purports to denounce both as preposterous, perhaps even as irrational, i.e., as clearly not controlled or governed by reason, or by considerations that appeal to objective grounds and rationales. There's a lesson to be learned here, and the epistemic lesson is that one's confidence in the beliefs and their justifications must decrease as much as possible and possibly reach 0 on the scale from 1 (certainty that the beliefs are true and their justifications genuine) to 0 (certainty that the beliefs are false and their justifications spurious). The intermediate degrees of confidence exhibit one's particular sensitivity to the (alleged) force of the debunking argument.

What causes us to have the mathematical beliefs that we have? A number of simplistic answers may come to mind. One factor is their outstanding success. We successfully apply mathematics in a wide array of cases and that gives us reason to believe that the mathematical beliefs we happen to have are both true and warranted. We might argue counterfactually on this particular point: How could false mathematical results be successfully applied? Try to apply $2+2=5$ and see what happens.

Another factor is provided by proofs and computations. We obtain the mathematical results we believe to be true and correct by the appropriate mathematical means, and this is how we come to secure our mathematical beliefs. A similar kind of

argument might apply: How could we secure mistaken mathematical beliefs with proofs and computations that have stood the test of time? Try to add 2 and 2 with an addition that misbehaves—that behaves like, say, Kripke’s quaddition. Unless, of course, proofs and computations are themselves defective and not to be trusted, i.e., unless we are mistaken about what we take to be the appropriate mathematical means to establish our mathematical results.

To address debunking properly, one must therefore distinguish (i) the mathematical beliefs that we have from (ii) the means by which we come to justify the mathematical beliefs that we have, and both from (iii) the philosophical arguments we devise to show that a particular account of our having (i) and of possessing (ii) is the true and correct one.

The debunking strategy targets (i)-(iii) for different reasons, with distinct albeit closely related arguments. One may see how the strategy is supposed to work by expressing the problem in terms of the relation between the content of our beliefs, their causal origin, and our second order views about such contents.

Suppose the content of our mathematical beliefs, say the content of the belief that $2+2=4$, plays a role in the explanation that we have of our having that belief. The advocate of the debunking strategy argues that since the best explanation involves causal patterns and evolutionary advantages, the belief is undermined. This is sometimes contended by way of an analogy with moral or evaluative beliefs.

The analogy typically appeals to counterfactuals and works as follows. Just as the evolutionary origin of our evaluative beliefs is supposed to undermine the confidence we have both in the beliefs as such and in whatever we appeal to in order to justify them (maximum utility, religious prescriptions, the pure consideration of a moral law to the exclusion of any consideration that would *de facto* count as external, etc.), the causal origin of our mathematical beliefs is supposed to undermine the confidence we have both in the beliefs as such and in the particular means we appeal to when justifying our mathematical results. Why? Advocates of the Evolutionary Debunking Argument for moral beliefs argue in the following way: Had the contents of our moral beliefs been false (or perhaps neither true nor false) because of the absence of moral facts or of whatever would

make our evaluative beliefs warrentedly true, we still would have the very same beliefs with the very same contents. We have them and judge them to be worthy of our approval not because of their truth or because of any genuine justification we possess in their favor, but because it has benefited our ancestors to have them and because, throughout the causal evolutionary process, we have nolens volens and unbeknownst to us inherited these beliefs for the very same reason—the term “reason” referring in this instance to something purely mechanical or factual, i.e., to something that excludes any contribution on our part in terms of an appeal to grounds, purpose, or justification. These beliefs, whether true, false, or neither, whether justified or not, or beyond justification, have kept providing us with the very same benefits.

So, the argument goes, our evaluative beliefs have no truck with truth. The very fact that their content appears in the causal explanation of why we have them suffices to free these beliefs from the clutches of the semantic notion of truth. Does this evolutionary speculation (or established evolutionary fact), which appeals to benefits as opposed to truth, establish what the advocate of the Evolutionary Debunking Argument would like it to establish, namely that just because natural selection makes us believe certain moral contents independently of their truth or their correspondence to moral facts, these contents are spurious or of no intrinsic value? Does it thereby follow that they are false, or neither true nor false? Are they, just for that reason, not in the market for truth, or not in the market for either truth or falsity?

Hardly so. At this point, a confusion lurks in the argument concerning the distinction between standpoints one may wish to defend regarding (i), (ii), and (iii).

Think in particular of a different but closely related consideration. One could hold a moral view, the view that, say, killing one’s offspring is wrong, or hold a mathematical belief, say the belief that $2+2=4$, and rest content with evolutionary speculations (or established evolutionary facts). One could have these beliefs and deny that there are either moral facts or mathematical facts that make them true. Should one then move to the level of (ii) to be presented with a causal explanation of the evolutionary advantage of securing the well-being of one’s

offspring, or of the evolutionary advantage of the mathematical practice that allows us to prove or conclude that $2+2=4$ and apply the result of that proof or of that computation, the very same kind of consideration would apply.² One could be convinced by the causal explanation and still hold onto one's preferred means of justifying the moral view and to the particular computations or proofs that allow one to establish the arithmetical result.

In other words, one could accept the evolutionary speculation (or established evolutionary theory) and deny that the methods and justifications we appeal to thereby fall into disrepute. The explanation by way of evolutionary advantages doesn't undermine the appeal to say, religious prescriptions or moral law (in the case of moral beliefs), or to arithmetical computations or Peano Arithmetic (in the case of arithmetical beliefs).

As far as morality is concerned, this is of course closely related to the question of relativism in ethics. J. L. Mackie once remarked that one could be thoroughly conventional in one's strict moral views and believe that these are just attitudes others may very well not share (see *Mackie 1971: 16*). One could also hold a given moral view (either strict or liberal in the sense of open-mindedness and freedom from prejudice or bigotry) and remain agnostic with respect to the evolutionary speculation that would explain why one has that particular view.

In the evaluative case, it might well be *psychologically unlikely* that one would hold a first order view without being strongly opinionated about the second order one, ie., about the reasons one has to hold the first order view. If you think that killing is wrong whatever the consequences might be, or that slavery is unjust, or that abortion is murder, you usually also believe that you have compelling reasons to view the matters at hand in that particular way, reasons that, more often than not, you are willing to disclose and defend, and that you take to be *non negotiable*. A strongly held first order view often demands, if only psychologically, or perhaps for sociological or historical reasons, an equally strong second order view with respect to the grounds of the first order view.

Now, think of the converse possibility. Mackie argues that one may very well "reject all established morality while believing it to be an objective truth that it was evil or corrupt" (Mackie: *loc. cit.*). In other words, one may hold a first order view and a second order view that, as it were, don't match psychologically or sociologically (an awkward combination of which there is perhaps no genuine historical example). It looks like, in that case, one shuns the current moral laws or prescriptions of one's community for no particular reason at all, or because one fancies oneself a skeptic about them, but nevertheless agrees that these laws and prescriptions are objectively bad — although agreement at the second order level in fact plays no role in one's first order rejection.

Since the motivation for the rejection isn't based on an argument to the effect that a particular law of prescription is wicked, what is the content of the first order attitude *with respect to the second order one*?

What is indeed unlikely, at least psychologically in terms of sensibility to the very purpose of morality and as far as choosing certain values rather than others is concerned, is to opt for such a rejection while not caring whether some particular course of action is taken, or some moral judgement voiced for objective reasons rather than for subjective motivations, or because it happens to be customary or fashionable to do so, or on the contrary because one fancies the fashion should be different. In this particular case, the first order rejecter throwing away second order considerations is one dyed-in-the-wool cynic. Diogenes might approve. One rejects all established social norms and behavior not because one believes they are evil or corrupt, but because one happens to dislike them for no particular reason at all. One might have done otherwise and have accepted them after all. Anything goes. Inconsistency lurks in the background.

In the mathematical case, where beliefs do not have the controversial evaluative and subjective component moral beliefs conspicuously have, the situation is of course different. Anyone rejecting a given kind of mathematical practice, e.g., a particular way of proving theorems, may do so for very different kinds of reasons. One may choose, say, to prove theorems of geometry by geometrical means rather than by algebraic ones because of a belief in

² Just modify (ii) so that it reads "the means by which we come to justify the moral beliefs that we have" to cover the moral case.

the purity of methods. But the rejection of certain kinds of proofs might also be motivated by considerations pertaining to the *illegitimacy* of some of the means by which we come to justify the mathematical beliefs that we have. One may, e.g., decide to reject classical mathematical practice because of a belief that constructive proofs are the only legitimate ones. In this case, the disagreement about the means targets a specific feature of classical proofs, namely the fact that they do not necessarily specify constructions that, at least in principle, we should be able to carry out. The disagreement, which yields a strong repudiation of a given kind of mathematical practice, is thus grounded on a conception of intuitionistic proofs according to which the content of the conclusions of these proofs (i.e., the very content of our mathematical beliefs) must embed the contention that *non*-constructive proofs may *not* count as genuine justifications of the very same conclusions.

What is of philosophical interest here, though, is that psychological considerations are nevertheless unhelpful and that *they are so for the same reason in both cases*, in spite of the fact that psychological considerations and emotive attitudes aren't to be found anywhere in the vicinity of mathematical practice, while the idea of a moral psychology and of a moral behavior brought about by Humean passions makes sense.

As for (iii), why should whatever philosophical argument one may devise to the effect that, say, a platonistic account, or a nominalist one is in order for both (i) and (ii) be threatened by considerations pertaining to debunking?

Since, in the present context, this involves the view that our mathematical beliefs are reliable insofar as they are both safe and sensitive (see *Clarke-Doane 2016*), let me end this section with a brief comment on truth, safety, and sensitivity. One may consider two epistemological challenges pertaining to mathematical belief formation. One is to show that our mathematical beliefs are *safe*, i.e., that we could not (at least not so easily) have false mathematical beliefs using the methods we actually use to form these beliefs. Another is to show that our mathematical beliefs are *sensitive*, i.e., that, had the content of our mathematical beliefs been false, we would have abjured them, using the methods we actually use to form these beliefs. The challenge to be met with respect to the explanation of why we

hold the mathematical beliefs that we do, is to show that these beliefs are counterfactually persistent, i.e., that if mathematical truths had been different, or if there had been no mathematical truths at all, our mathematical beliefs would have been correspondingly different.

The construal of the reliability challenge that requires the content of our mathematical beliefs to be both safe and sensitive implies that mathematical truths are both contingent and defeasible. What matters here, the discarding of the necessity of mathematical truths notwithstanding, is that, as far as debunking goes, if it turned out to be impossible, given our cognitive capacities, or because this would involve a contradiction, to show that the content of our mathematical beliefs figure, via some causal explanation or otherwise, into the best explanation of our having them, this would still give us no reason whatsoever to give them up. It would not unless we were also ready to give up the requirement that they must be both safe and sensitive.

6. Conclusions

It has been argued all along that considerations of parity must play a part in the debate about the role and nature of mathematical explanations vs. the role and nature of non mathematical ones, in particular when they must be counted as being the best, so that competing platonistic and nominalistic devices used in accounts of the usefulness or, on the contrary, of the dispensability of mathematics, may not be treated differently, either epistemically or ontologically. The point was that it must be assessed in both cases which kind of account is more (or, on the contrary, less) ontologically or epistemologically loaded.

By Quine's criterion of ontological commitment, *any* account has commitments. Since both mathematical explanations and proofs, and nominalistic ones, wear such commitments on their sleeves, the fact that we may dispense with number talk, either by way of paraphrase or by way of a conservativeness proof, hardly neutralizes the commitment. We just can't dispense with ontology.

It is sometimes assumed that platonism is a second-best option one accepts as it were by default, a position that would be dismissed just in case a nominalistic alternative to it could be devised. This pre-

supposes, of course, that ontological parsimony should be preferred. It isn't clear why it should. In any event, it might be clearer why epistemic parsimony should be preferred. Pragmatic motivations might be called to the rescue: There is no reason to make our life more difficult with certain kinds of derivations or best explanations when an easier course is available. Of course, one might then offer an open question argument: Why should we grant a privilege to epistemological economy as such, or to anything that might imply it has advantages? Perhaps, then, epistemic economy is good *sui generis*.

In any event, one has no right to complain that there is no privileged neutral point of view from which to judge the advantages and disadvantages of either type of account or explanation. There is no *petitio principii* at stake in the contention that this is indeed possible. We can't argue that we're always somehow begging the question, either in favor of platonism or in favor of the opposite view. An assessment of the respective merits of these views doesn't necessarily take for granted that platonism offers the only possible account of mathematical truth, or that anti-platonism offers the only possible account of mathematical knowledge.

But since — given what has been said in section 5 — we're justified, after all, in holding our mathematical beliefs, the problem must lie elsewhere. We've been concerned with worries that presuppose that classical mathematics and classical reasoning are unproblematic as far as either mathematical knowledge or some nominalistically respectable knowledge that would replace it in best explanations of natural phenomena are concerned.

But suppose that proofs which establish the existence of something by providing the effective means of finding it, as is the case of constructive proofs, provide an insight that classical proofs conspicuously fail to provide. In a nutshell, a constructive proof of a closed statement of the form ' $(\exists x) Fx$ ' with the bound variable ranging over the natural numbers, either proves a specific instance $A(n)$ or provides, at least in principle, an effective means for finding one. Now, a classical logician or mathematician might prefer constructive proofs for particular reasons. But the intuitionistic logician or mathematician believes he has no choice: Proofs *must* be constructive.

Of course, we're now looking at a kind of worry that is different from the one concerned with the availability of nominalistic derivations and best explanations. We're dealing with disputes that happen to be of a quite different nature. There is, on the one hand, the question whether platonism may be vindicated by an inference to the best explanation of mathematical success or, on the contrary, whether anti-platonism may be vindicated by the possibility of explaining the applicability of mathematics and of providing explanations of natural phenomena by nominalistic means. There is, on the other hand, the question whether classical proofs and classical reasoning should be preferred to constructive ones. This constitutes a second dispute, clearly distinct from the first, and one might complain that the dispute regarding the classical vs. the non classical character of proofs has no bearing on the one we've been concerned with so far.

Let me briefly suggest that it isn't so and that the two disputes aren't as unrelated as it may seem at first blush.

Whatever the theoretical role of a classical explanatory proof may turn out to be, it must indeed be different from that of an intuitionistic explanatory proof, or from that of a proof or derivation which resorts to conservative mathematics only. The role must be played differently, i.e., it must yield a *different* explanation of a natural phenomena.

This suggests that we want to distinguish two kinds of explanations : the explanation of why classical mathematics is conservative from the explanation of why intuitionistic mathematics is conservative. Likewise, we may want to distinguish the explanation of why a geometrical classical explanation of some natural phenomenon does the job, perhaps even a better job than the numerical one, from the explanation of why a geometrical intuitionistic explanation of the same natural phenomenon has the same advantages.

So what are the question-begging assumptions regarding the nature and validity of proofs that the claim to the effect that nominalistic proofs do play their theoretical role is guilty of?

If it were shown that mathematics isn't conservative, or has no genuine platonistic claims embedded in it, or need not be intuitionistic, the result would have deep implications regarding what we must argue for in order to establish that some suita-

ble theoretical role is played. The scrutiny might show that nominalistically respectable proofs, or classically respectable proofs, or conservatively respectable proofs, happen not to be on a par with respect to playing a theoretical role.

This suggests that, because both the validity and the nature of proofs and derivations might differ in each case, the cost of undermining platonism is different from the cost of undermining nominalism or, for want of a better phrase, from that of undermining conservatism.

The differences will be brought about by features that might surface at different levels: at the level of applied number theory and applied geometry, but also at the level of logic. To show that there is indeed a difference and, moreover, that it matters a great deal as far as our problem is concerned, requires further inquiry. What I'm suggesting here is that the difference must matter for reasons that have to do with the most basic features of, respectively, classical proofs and intuitionistic proofs, namely their logic.

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. *Synthese* 193: 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*, 2nd ed. English translation by C. D. Parsons, P. Benacerraf and H. Putnam eds, 258–271. Cambridge, Mass, Cambridge UP.
- 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, Oxford Logic Guides.
- Field H (2016), *Science without Numbers – A Defence of Nominalism*, 1st edition: Princeton UP, Princeton, New Jersey, 1980 ; 2nd edition 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.
- 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.
- Mackie J L (1971) *Ethics – Inventing Right and Wrong*. London: Penguin Books.
- Molinini D, Pataut F, Sereni A, eds.(2016) Special Issue on Indispensability and Explanation, *Synthese* 193(2): 317–558.
- Pataut F (1999) Vérité morale et justification morale. In *Le réalisme moral*, ed. R. Ogien, 457–502. Paris: PUF.
- Pataut F (2016) Comments on ‘Parsimony and inference to the best mathematical explanation’. *Synthese* 193: 351–363.
- Putnam H ([1967] 1979) The Thesis that Mathematics is Logic. in *Mathematics, Matter and Method: Philosophical Papers*, vol. 1, 2nd ed., 12–42. Cambridge, Mass: Cambridge UP.
- Putnam H ([1971] 1979) Philosophy of Logic.

In *Mathematics, Matter and Method: Philosophical Papers*, vol. 1, 2nd ed., 323–357. Cambridge, Mass: Cambridge UP.

Quine W V O ([1968] 1969) Existence and Quantification. In *Ontological Relativity and Other Essays*, 91–113. New York: Columbia UP.

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.

Robinson A (1956) A Result on Consistency and Its Application to the Theory of Definition. *Koninklijke Nederlandse Akademie van Wetenschappen*, Proceedings, Series A, 59: 47–58.

Shapiro S (1997) *Philosophy of Mathematics: Structure and Ontology*. Oxford: Oxford UP.

Mark Steiner M (1978) Mathematics, Explanation, and Scientific Knowledge. *Nous*, 12: 17–28.

Tallant J (2013) Optimus Prime: Paraphrasing Prime Number Talk. *Synthese* 190: 2065–2083.

SUBJECT IN CLASSICAL LOGIC AND SPECULATIVE LOGIC

Dragoș Popescu

Romanian Academy of Sciences, Bucharest, Romania
e-mail: dragos.gp@gmail.com

Abstract:

The paper presents the classical theory of the subject in the predication judgment, and then the Hegelian doctrine on the subject, with the intention of conducting a comparative analysis.

The results of the analysis sustain the viewpoint according to which between the classical subject and the subject of speculative judgment there are some relations that entitle one to consider speculative judgment as a development of classical judgment, for the cases in which the subject is taken as a process.

Key words: metaphysical subject; logical subject; predication judgment; speculative judgment; middle.

New Logic (Mathematical logic) tends to leave the outdated notions of subject and predicate outside its area of interest, systematically replacing them with the modern notions of function and argument.¹ One might say, without the fear of making terrible mistakes, that step by step, the subject was almost removed from the list of notions used by contemporary logicians. Nowadays, the great majority of them understands it as part of the prehistory of logic, a far-off and largely unfamiliar period when logical form had not yet been differentiated from linguistic form or other types of 'form', more or less suited to being taken into account.

¹ The stages of transitioning from 'classic' logic to 'new' logic are well known to researchers. Leibniz, Boole, Frege, Whitehead and Russell are the representative figures of the gradual transformation of Aristotle's long-standing science. The publishing of *Principia Mathematica* (1913) establishes the full instauration of the new logic.

The reason for this exclusion will not be the subject of this article; we think that there is no way to find the answer from the *inside* logic, as this is a matter that metaphysics traditionally contends with. But incidentally, since the fate of metaphysics has also proven unfavourable, we cannot question metaphysics about the nature of subject.

Equally, the aim is not to outline an assessment of 'old' and 'new' logic. Such evaluation can't be done without constant reference to each of their 'accomplishments', which implies a continuous reference to the configuration of sciences and their development parallel to the stages of logic's evolution. Logic is *organon*, *instrumentum*, which means a tool of the sciences, and the most interesting thing in the study of any instrument is how it fulfils the purpose of its creation. An instrument that fails to do its job has no value, no matter how sophisticated it is. So, examining the issue from this angle – although interesting – will lead research into an overly broad field in which we should set aside the desire to reach a conclusion.

Our purpose is much more modest. Below we will outline the classical theory of subject, then its speculative interpretation, and afterwards we will compare the two. In the case of classical logic, we are dealing with issues that have been known, analysed and solved in multiple ways for ages now. Great logicians and philosophers of the past have often rendered their impression that nothing unusual can happen here, and so they reserved for classical logic only a propaedeutic function. Once in a while, new studies will contest and deny this long-lasting impression (although these remain few and far between), but they never truly set aside and eliminate it.

In the case of speculative logic things are completely different: There is significant confusion, sometimes fueled by the intermingling of strange foreign interests within the field – a phenomenon especially characteristic of the last two centuries, when traditional domains and methods of philosophy have gone through a continuous and intense transformation due to involvement with extra-philosophical activities – which has led to a denial of the validity of speculative logic's place within philosophy.

Our aim is not didactic, although we focus in our paper on those facts which corroborate our thesis rather than those which may foster disagreement. Furthermore, in this context we are not moving toward a historical perspective on the subject, but rather toward a systematic one. The historical references below are merely a rough guide.

It has often been claimed that there is an insurmountable gap between classical logic and speculative logic. Our purpose is to examine whether this point of view is correct or whether, in fact, there is a much closer connection between classical logic and speculative logic.

I. Subject in classical logic

I.1. Fundamental sources

No matter how we approach it (and not only with regard to our paper), we owe the fundamental sources for classical logic to Aristotle, the undeniable founder of this field of study. The representative Aristotelian paper on this topic, strictly from a logical perspective and as part of what would later be called 'judgement of predication', is the book *On Interpretation*. Sure enough, Aristotle recalls at the beginning of this work some results from his investigations on this topic that were already present in *On the Soul*. However, they focus on a different domain than that of logic – precisely the domain of psychology – so I will not discuss them here.

The Greek word that Aristotle uses for subject is ὑποκείμενον. The meaning of the word is not connected solely with logic, but also with metaphysics. This ambivalence is in no way proof of Aristotle's lack of concern with terminological accuracy; on the contrary, this is an essential requirement for the validity of any knowledge. There are attributes for the metaphysical subject (substance) and predicates for the logical subject. Without a correlation between these two domains, knowledge is not possible (see *infra*, the case of the relation called contradiction). Science owes its existence to the correlation estab-

lished between the metaphysical subject and the logical subject. Seen from a logical point of view, the subject is a *noun*. The noun has a few properties, specified by Aristotle:

Ὄνομα μὲν οὖν ἐστὶ φωνὴ σημαντικὴ κατὰ συνθήκην ἄνευ χρόνου, ἧς μηδὲν μέρος ἐστὶ σημαντικὸν κεχωρισμένον [*De Int.*, 2, 16a 19-21],

translated into English:

By a noun we mean a sound significant by convention, which has no reference to time, and of which no part is significant apart from the rest [*Aristotle, On Interpretation*, translated by E. M. Edghill],

And (just for a comparison) the Romanian version:

Numele este o glăsuire semnificativă pe bază de convenție, fără indicație de timp, în care nici o parte izolată nu este semnificativă [Ammonios, Stephanus, *Comentarii la tratatul Despre Interpretare al lui Aristotel, însoțite de textul comentat*, Traducere, cuvânt înainte, note și comentariu de Constantin Noica, Editura Academiei RSR, București, 1971, p. 9].

If any logical subject is a noun, not every noun is a logical subject. In order to be a logical subject, the noun has to be part of an λόγος ἀποφαντικός, which can be of two types: κατάφασις (affirmative) and ἀπόφασις (negative) (*De Int.*, 5, 17a, 8-9). The subjects of other types of λόγοι are not logical because those λόγοι are neither true nor false, so the relation between metaphysical subject and logical subject cannot be gleaned from them. This is precisely the case of interrogatives, imperatives and desideratives.

Undefined nouns present a special situation (for example: non-human, non-alive); strictly speaking, they are not nouns, but they can be a logical subject (*De Int.*, 2, 16a 30). Λόγος ἀποφαντικός represents a unit from a logical point of view. The parts of this unit, the noun and the verb, taken separately are also units because they indicate something (objects, actions) that has meaning; however, they are not *logical* units (λόγος) but *linguistic* ones (φάσις):

τὸ μὲν οὖν ὄνομα καὶ τὸ ῥῆμα φάσις ἔστω μόνον [*De Int.*, 5, 17a 17-18],

translated into English:

consent to call a noun or a verb an expression only [Aristotle, *On Interpretation*, translated by E. M. Edghill],

and also translated into Romanian:

Așadar numele sau verbul urmează să fie socotite o simplă vorbă [Ammonios, Stephanus, *Comentarii la tratatul Despre Interpretare al lui Aristotel, însoțite de textul comentat*, Traducere, cuvânt înainte, note și comentariu de Constantin Noica, Editura Academiei RSR, București, 1971, p. 9].

Therefore, the subject cannot be studied separately from the whole it is a part of.

Κατάφασις (the affirmative) is the enunciation of something about something (ἀπόφασις τινὸς κατὰ τινὸς – *De Int.*, 6, 17a 25); ἀπόφασις (the negative) is the separation of something from something else (ἀπόφασις τινὸς ἀπὸ τινὸς – *De Int.*, 6, 17a 26). This ‘separation’, as we can see, is also ἀπόφασις, which means the unity of λόγος ἀποφαντικός is also preserved in the case of negation. It does not tear apart or destroy the unity between ὄνομα and ῥῆμα, decreasing each component of λόγος to a simple φάσις. The logical subject and the logical predicate remain connected in a negation as well.

Each affirmation and each negation has its own opposite. The opposition of affirmation and negation is called contradiction (ἀντίφασις – *De Int.*, 6, 17 a 33). Affirmation and negation exclude each other, meaning they cannot be homonyms (which would mean, according to *Categ.*, 1, 1a 1 that they have a common noun and a different λόγος τῆς οὐσίας, so that the ὑποκείμενον targeted by affirmation and negation respectively would be a different one in each situation, and therefore, we wouldn’t associate it with mutual exclusion, but only with different λόγοι ἀποφαντικοί). For Aristotle, the contradiction is not, therefore, a simple operation at the level of language. The grammatically (syntactically) contradictory sentences are not instantly contradictory from a logical point of view; for this, they must consider the same ὑποκείμενον.

The noun is universal when enunciating about many and singular when not enunciating about many. The affirmation and negation regarding the whole domain of the universal noun set such sentences into an opposition rapport called contrariety. The affirmation and negation regarding just a part of the universal noun’s domain indicates a relation of opposition at the level of the concerned part of the universal noun domain (subcontrariety).

1.2. The definitive doctrine

The modern studies of formal logic call the Aristotelian λόγος ἀποφαντικός by the name of *proposition*. In the ‘proposition’ we can distinguish between [see J. Dopp, *Leçons de logique formelle*, Première partie, Logique ancienne, *La logique des jugements prédicatifs*, Éditions de L’Institut Supérieur de philosophie, Louvain, 1949, p. 18-19]:

a) ‘predication’ = thought content structured in such a way it can be either true or false; and

b) ‘predication’ = thought content we can regard as being true or false, which means it abides a predication act.

There is, therefore, a double meaning to the term predication (Aristotle would have said that the notion is a homonym). The second meaning, b), does not represent a point of interest for logic, because it concerns the thinking act of the person who thinks, that is, the act by which he or she adheres (acknowledges, subscribes) or not to the content of the proposition, considering it true or not [J. Dopp, *op. cit.*, p. 19].

The first meaning of proposition, a), which is a matter of study in formal logic, needs an additional specification: The structured content of thinking is stated orally, in writing or through gestures – and so, the proposition is, in fact, ‘predication’.

Propositions gather into two large families [J. Dopp, *op. cit.*, p. 29]:

a. *Propositions with a definite/singular subject* (definite/singular propositions); and

b. *Propositions with a general subject* (general propositions).

The two types of propositions are based on the Aristotelian distinction between universal nouns and particular nouns (see *supra*). Both have in common the relation between base/support (single or multiple) and the quality/property which is given to it. The base/support is called *subject*, the quality/property is called *predicate* and the relation is called *predication* [J. Dopp, *op. cit.*, p. 29].

In the study of proposition, the focus is not on the nature of the subject, nor on that of the predicate (meaning on their ontological status), but on their relation (predication). Nonetheless, subjects and predicates cannot be included in the same category. The bases/supports that can also assume the role of a predicate must never be mistaken with those that cannot perform this function (primitive bases/supports) [J. Dopp, *op. cit.*, p. 30].

Taking the proposition to be a) ‘predication’ – meaning a structure in which a stated content of

thought arises, capable of being true or false – we can single out three constituent components: 1. subject; 2. predicate; 3. copula (the connection between subject and predicate).

The founders of new logic (and not just them) called into question the rigid nature of ‘*predication*’, as it cannot render satisfying approaches to any type of ‘*predication*’ (content of thought upon which one can perform a predication act). This is the case for predications like: ‘It is raining’ or ‘Plato is older than Aristotle’, but also for other acts of predication specific to mathematics. Classical logic attempted, more or less convincingly, to transpose these predications into the classical form of ‘*predication*’.

In classical logic, the subject (*sub-jacere*) formally defines the base or the support (*sub-ponere*) upon which the predicate develops (determines), support which is also commonly named ‘object’. This does not mean that classical logic demands (requires) that the support/base (the object) need be *something*, e.g., a certain thing or a sum of things noticeable on an ontological level, but only *something the predicate asserts itself about*, such that the predication can be true or false [J. Dopp, *op. cit.*, p. 75]. The matter of the subject’s ontological status is not a topic for logic, but for metaphysics. Natural language doesn’t always allow the possibility of easily identifying the subject (logic), but ‘*predication*’ requires that it should be one and the same throughout the predication’s analysis [J. Dopp, *op. cit.*, p. 74].

Subject in definite (or: singular) subject propositions

Propositions with a *definite/singular* subject explicitly indicate the support (object) using specific expressions like *this*, *that*. They convert into predications the direct internal or external experiences of the person who performs the predication, they don’t require a previous transposing of the internal/external experience into the concept, and they also imply the gestures that signal the accomplishment of the predication act. Therefore, they indicate without any doubt the second meaning, **b)** ‘*predication*’, of the proposition.

Within propositions with a *definite/singular* subject we can distinguish different employments of expressions that stand for subject (*this*, *that*), keeping in mind that they all have in common the use of a certain gesture: to indicate. These are as follows: 1. the unique object, certain support/base indicated through predication (example: ‘*This is red*’); 2. the collective object, varied things brought together by

expressions like *this* or *that* (example: ‘*This, the other and that are red*’); 3. a certain object within a collective (specific expression: *each*, taken along with a gesture. Example: ‘*Each is red*’); 4. collective within a collective (example: ‘*Each one of these is red*’); so on and so forth.

Subject in general subject propositions

The subject of propositions with a general subject can either be particular or universal.

The universal subject takes into account each support/base for which the assigning stated by the term-subject (noun) is valid [J. Dopp, *op. cit.*, p. 86]. A particular subject does not take into account each support/base for which the assigning stated by the term-subject (noun) is valid [J. Dopp, *op. cit.*, p. 88]. Unlike Aristotle’s definitions of universal noun and singular noun, we can notice a distinction here between *noun* and *subject*. In this case, the noun stands for *grammatical* subject only, whereas *logical* subject can be stated by special words like: ‘some’, ‘all’. There is also a distinction that arises between *general* and *universal*: a universal subject is nothing but one of the types of general subjects, the other one being the particular subject [J. Dopp, *op. cit.*, p. 87].

II. Subject in speculative logic

Speculative logic is not unequivocally accepted as logic. There is a thesis which maintains that, in this case, we are rather dealing with metaphysics disguised by classical-logic terminology. Indeed, this thesis has some reason to it. There is also a thesis according to which speculative logic represents an ‘overtaking’ of metaphysics – that is to say, a replacement of this traditional field of study with a faultless method: dialectics. This view has few followers today.

While Aristotle is without doubt considered the father of logic, Hegel’s considerable influence comes from laying the bases of speculative logic. Previous attempts at constructing this logic have been identified in Plato, Aristotle and Kant by Hegel himself.

II.1. Fundamental sources

Hegel’s first attempts to add a speculative description to the subject are found in *Phenomenology of Spirit*, which is not a book about logic, but it does pave the way for logic. Hence, we can read the following in the *Preface* of the book:

Es kommt nach meiner Einsicht, welche sich nur durch die Darstellung des Systems selbst rechtfertigen muss, alles darauf an, das Wahre nicht als *Substanz*, sondern eben so sehr als *Subjekt* aufzufassen und auszudrücken' [*Phänomenologie des Geistes*, p. XX of first edition, 1807],

in English translation:

In my view, which can be justified only by the exposition of the system itself, everything turns on grasping and expressing the True, not only as *Substance*, but equally as *Subject* [*Phenomenology of Spirit*, translated by A.V. Miller with Analysis of the Text and Foreword by J.N. Findlay, Oxford University Press, 1977, p. 9-10],

and in Romanian translation:

Rezultă, în concepția mea, care va trebui să se justifice numai prin expunerea sistemului însuși, că adevărul trebuie să fie conceput și exprimat nu ca substanță, ci totodată ca subiect.

Let's first stop and consider the last two verbs in the sentence: *auffassen* (translated into English: grasping; and into Romanian: conceput) and *ausdrücken* (in English: expressing; in Romanian: exprimat). The first one clearly calls forth the reasoning activity (grasping, conception) and the second one, just as clear, confirms the result of the reasoning activity (by expression) which has both the meaning of 'predication' **b**) and 'predication' **a**).

The key term *Substanz* represents the Hegelian correspondent for the Aristotelian ὑποκείμενον, clearly attested to by its interconnection with Truth. Finally, on the other key term, *Subjekt*: as it has already been stated, we can also find in Aristotle's work the idea of placing the subject into relation with substance, because only through *expressing* substance – as the subject of λόγος ἀποφαντικός – can truth or falseness be applicable. But where is the novelty introduced by Hegel?

We can find the answer to this question in the following excerpt:

Die lebendige Substanz ist ferner das Sein, welches in Wahrheit Subjekt, oder was dasselbe heisst, welches in Wahrheit wirklich ist, nur insofern sie die Bewegung des Sichselbstsetzens, oder die Vermittlung des Sichanderswerdens mit sich selbst ist

[*Phänomenologie des Geistes*, ed. cit., p. XXI].

In English:

Further, the living Substance is being which is in truth *Subject*, or, what is the same, is in truth actual only in so far as it is the movement of positing itself, or is the mediation of its self-othering with itself [*Phenomenology of Spirit*, ed. cit., p. 10],

and translated into Romanian:

Substanța vie este mai departe ființa care este în adevăr *subiectul* sau, ceea ce înseamnă același lucru, este în adevăr realul doar întrucât este mișcarea autopunerii sau medierea dintre autodevenirea-ca-altul și sine însuși.

The quote immediately draws attention to the movement (*Bewegung*) of substance – something Aristotle does not mention (the noun, which expresses substance, does not have any kind of time indicator and the subject does not undergo any movement within a sentence – see *supra*). Furthermore, *die Bewegung des Sichselbstsetzens* (movement of self-positing) and *die Bewegung des Sichanderswerdens* (movement of self-development into something different) are for Hegel requirements of the substance's truth, since substance is the subject. In other words, in Hegelian acceptance, these *Bewegung* types, or at least one of them, are of major necessity for the subject. A substance lacking in one of these movements is not a subject, reality, truth; it is somehow similar to the Aristotelian undefined noun. Without offering further details regarding the two forms of movement, we can already draw the conclusion that, between the subject of classical logic and that of speculative logic, there are significant distinctions.

The historical and philosophical research can easily provide information regarding the two prior types of movements mentioned by Hegel. However, our aim is to discover the logical validity of their use, and not the historical and philosophical one. Taken from this perspective, we can use of yet another short excerpt from the *Phenomenology of Spirit*, but then we would also need to focus on the *Science of Logic*. This is where we can truly find the key to this issue. In the fragment considered below, Hegel describes the longstanding Aristotelian meaning of the subject and this allows him to make a comparison with his own acceptance of the subject's meaning:

Das Subjekt ist als fester Punkt angenommen, an den als ihren Halt die Prädikate geheftet sind, durch eine Bewegung, die dem von ihm Wissenden angehört und die auch nicht dafür angesehen wird, dem Punkte selbst anzugehören; durch sie aber wäre allein der Inhalt als Subjekt dargestellt [*Phänomenologie des Geistes*, ed. cit., p. XXVII];

English translation:

The Subject is assumed as a fixed point to which, as their support, the predicates are affixed by a movement belonging to the knower of this Subject, and which is not regarded as belonging to the fixed point itself; yet it is only through this movement that the content could be represented as Subject [*Phenomenology of Spirit*, ed. cit., p. 13].

and Romanian translation:

Subiectul este luat ca punct fix de care predicatele sunt ataşate ca de o staţie, printr-o mişcare ce aparţine celui care are cunoştinţă de el şi care, de aceea, nici nu sunt văzute ca aparţinând punctului însuşi; dar numai prin ea s-ar expune conţinutul ca subiect.

In other words, the classical proposition is presented as a structure around one of its three components: the subject. The movement (*Bewegung*) from which predication arises, is drawn from the *exterior* of predication and is also the process that enables predication to happen, starting with a subject to which a predicate is attached; thereby it differs from both the ‘movement of self-positing’ and the ‘movement of self-development into something different’ previously mentioned and is not a consequence of that unspecified *Wissender* activity, but of the subject itself.

These three types of movement (of self-positing, of self-development into something different and of external attaching of predicates) are not studied in the science of logic but by the phenomenology of spirit (or: the science of experiencing consciousness). But the *subject* itself, where all these movements take place, belongs to the field of logic. It is in fact an entire domain of logic applied to *subjective logic*.

It is necessary to mention at this point an extremely important observation: If the issue of the subject in classical logic is looked at from a simple context created by λόγος ἀποφαντικός, speculative logic approaches a different one, where λόγος ἀποφαντικός is just an aspect of this issue, its

groundwork. The other aspect, the one related to form, is pointed out in the following quote by Hegel:

Der Begriff, insofern er zu einer solchen *Existenz* gediehen ist, welche selbst frey ist, ist nichts anderes als *Ich*, oder das reine Selbstbewußtsein. Ich *habe* wohl Begriffe, das heißt, bestimmte Begriffe; aber Ich ist der reine Begriff selbst, der als Begriff zum *Daseyn* gekommen ist [*Die subjektive Logik oder die Lehre vom Begriff*, p. 11 of the first edition],

in English translation:

The concept, when it has progressed to a concrete existence which is itself free, is none other than the “I” or pure self-consciousness. True, I *have* concepts, that is, determinate concepts; but the “I” is the pure concept itself, the concept that has come into *determinate existence* [*The Science of Logic*, translated and edited by George di Giovanni, Cambridge University Press, 2010, p. 514],

and the Romanian version:

Conceptul, întrucât a ajuns la o asemenea *existenţă*, care este liberă ea însăşi, nu este altceva decât *eul* sau conştiinţa-de-sine pură. Eu *am* concepte, bineînţeles, adică concepte determinate; dar eul este însuşi conceptul pur, ajuns *fiinţă-în-fapt* ca concept.

Therefore, in speculative logic it is relevant *who* carries out the declarative sentence, λόγος ἀποφαντικός. But this *who* is not just a random speaker, it is a well-defined type of self (‘I’, a concept which is the subject in the predication) with a structure built by its own experience and by its specific concepts.

Subjective logic is the logic of concept as subject. Being that concept is the self, the subject in speculative logic is not *just* a noun or a formula that nothing can actually correspond to, but a noun with *something* that mandatorily complies with it, and that something is expressed by the notion of existence (concrete existence). This notion doesn’t have the trivial meaning of a certain being or an array of certain beings (e.g., Kallias, Socrates, people, each human), because this would transform the logical undertaking into a psychological one, and existence into a matter that could only be resolved through classical predications with an objective subject. Existence means that in any λόγος ἀποφαντικός the subject is the one asserting its predicates. The Aris-

totalian ἀπόφανσις leaves unspecified the way the enunciation of something about something is done; in speculative logic, the enunciation of something about something is done by the subject.

So, there is no distinction between **a)** ‘*predication*’ and **b)** ‘*predication*’ inside speculative logic. Instead, we can notice a different distinction: ‘**proposition**’ (*Urtheil*) and ‘*statement*’ (*Satz*). Speculative proposition assumes the meaning of both **a)** and **b)**, meanwhile ‘*statement*’ accounts for any ἀπόφανσις, where the relation between subject and self (meaning a structure of language that meets the requirements of **a)** but not those of **b)**), remains unmentioned.

Statements are still under the expertise of classical logic, but the study of **propositions** belongs to speculative logic.

The opposition of propositions gets a totally different meaning within speculative logic than it had in classical logic, because in this logic, the subject is the self, and the opposition between propositions is settled according to it. The subject of opposite propositions is the same, only the place of the subject inside the movement (*Bewegung*) differs. The status of subject in speculative logic is determined precisely by its development and progress; therefore, it is no longer in question, like in classical logic, whether the subject of opposite propositions could be the same or not, and that could only be fixed through an explicit requirement that it doesn’t undergo any alteration.

The most important aspect of a speculative contradiction is not the mutual exclusion of oppositions, something that was exploited by classical logic in its own manner through varied forms of argumentation, but rather taking notice of the transformations undergone by the subject. The purpose of speculative logic is not to advance an argument or proof, but to incorporate the subject’s stages of development. In the most rigorous sense, it is a systematic logic.

The contrary opposition is no longer taken into account by speculative logic, because ‘some’ and ‘all’ can no longer apply to the subject as self (or: Ich, the ‘I’, pure self-consciousness as Hegel has it described); the self is not more or less inclusive than another self in speculative propositions, because these propositions don’t relate to varied selves, but to one and the same. In fact, the speculative subject is always a wholeness/totality. Replacing the subject inside the sequence of speculative propositions does not lead to opposition, but to sophistry. Hence, speculative propositions can only oppose in a contradictory way, depending on the predicates the subject

gives itself, that is, within the bounds of its own movement of self-positing or the movement of self-development into something different.

There are two ways of regarding the movement of self-positing and the movement of self-development into something different: as a movement in full development or as a complete (or: finished) one. In the first instance, the subject hasn’t gained all of its conceptual determinations, which takes place in the second one. The latter can act as an agent [term] inside the speculative syllogism (*Schluß*).

The distinction between a proposition with a definite/singular subject and a proposition with a general subject encountered in classical logic disappears entirely, because all speculative propositions can be regarded as definite (their subject being the self) and as general at the same time (the subject always has particular conceptual determinations or universal ones given by the predicate). So, instead of a proposition with a definite/singular subject and a proposition with a general subject, speculative logic uses:

a’. propositions with an incompletely determined subject; and

b’. propositions with a completely determined subject (or Media).

Subject in speculative propositions with an incompletely determined subject

The subject of speculative proposition evolves by gaining conceptual determinations. This is its main distinction from the classical subject, which remains fixed and does not undergo any transformation. Each new conceptual determination adds up to the determinations that the subject of a speculative proposition already holds. The conceptual determinations are: universality, particularity, singularity. This means that the subject as self is consecutively determined as: universal, particular and singular. These are the predicates the subject can give itself in a speculative proposition; but we should mention that the consecutive structures the predicates account for within the subject are more important than the actual predicates. The subject is incompletely determined when it doesn’t hold all the conceptual determinations, but only a part of them. Of course, the subject holds all possible conceptual determinations at its own level of development, but, inside the general movement of development, it only holds those acquired in stages that are previous to the stage that it is currently at.

Conceptual determinations are not randomly assigned to a subject. The first conceptual determination of a subject is universality – the only case when the predicate can determine a subject which entirely lacks any determination.

Subject in speculative propositions with a completely determined subject or Medium

The completely determined subject appears in speculative propositions in tautological form. According to the three conceptual determinations, there are three types of predications here: ‘The Universal is universal’, ‘The Particular is particular’ and ‘The Individual is individual’. Actually, the resulting speculative propositions are not tautological in the classical meaning. For instance, in the predication ‘The Universal is universal’ (or, in Hegelian terms: problematic predication), the subject is assigned the predicate ‘universal’ which holds both determinations of ‘individual’ and ‘particular’, gained in previous movements when the subject was the universal. The same applies to the particular and individual subjects, which also hold these conceptual determinations.

These three mentioned propositions are still different in structure. The sequence of conceptual determinations is different for each one of them, so they can’t be confounded, but they do complement each other. Since each of them still holds all the conceptual determinations, they can act as the medium of a speculative syllogism, something that doesn’t happen with any speculative proposition with an incompletely determined subject. By contrast, since the latter lacks some conceptual determinations, it can’t act as a medium for speculative syllogisms. Instead, it has a different aim: to accurately indicate the stage of the speculative movement where the subject is currently engaged.

The keystone of the whole Hegelian reconfiguration of classical logic resides, as by now the reader can see, in the matter of the double movement (*Bewegung*) laid down in *Phenomenology of Spirit*. In the absence of this movement, we can no longer talk about any speculative logic. The immovable subject of classical logic would not allow it. But if we had to name the main source of the flawed representations of speculative logic, we would be compelled to name that precise movement.

In the historical context of Hegelian logic, we admit that the possibility of accurately defining the subject’s movement was quite remote. Various sciences were already coming up with their own suggestions: The physicalist movement (translation)

used to be the most familiar form of movement; chemical and biological transformations were far less understood, but it was known that they were related to another type of movement than that of physical translation. Otherwise, particularly from the latter perspective (chemical and biological transformation), we could have expected an intuitive support for the new speculative logic – something that didn’t occur.

The alternative came from appealing to the traditional conceptual resources of metaphysics, from which Hegel never wanted to fully come adrift. His effort of locating his own contribution within a historical sequence had an unintended outcome: It prevented the acceptance of the novelty inherent to his new point of view.

It is quite clear nowadays that both options had their own shortcomings. In fact, the movement taken into account by speculative logic is not physical, chemical, biological or ‘metaphysical’. If we had to find an equivalent, it would be transitioning from premises to a conclusion through reasoning (what we today call *inference*, understood as a type of *process*). This is a movement which is not physical, chemical, psychological or ‘metaphysical’, but yet still well-known by the scientists and metaphysicians, inasmuch as those who practice it continuously make use of it.

Placed in the context of more recent scientific developments, speculative logic could have been reassessed. However, it had already been introduced in the context of a historical undertaking (although this was not a conscious intention), and it was not able to free itself for almost a century and a half. Maybe this is the reason why scientific research like the experiments of genetic epistemology (just to give one example) failed to account for the fact that the following point of view was already a gain for the Hegelian approach to the subject:

la connaissance ne saurait être conçue comme prédéterminée, ni dans les structures internes du sujet, puisqu’elles résultant d’une construction effective et continue, ni dans les caractères préexistants de l’objet, puisqu’ils ne sont connus que grâce à la médiation nécessaire de ces structures et que celles-ci les enrichissent en les encadrant (ne serait-ce qu’en les situant dans l’ensemble des possibles) [J. Piaget, *L’Épistémologie génétique*, deuxième édition, Presses Universitaires de France, 1972, p. 5],

in Romanian variant:

cunoașterea nu poate fi concepută ca predefinită nici în structurile interne ale subiectului, deoarece ele rezultă dintr-o construcție efectivă și continuă, nici în caracterile preexistente ale obiectului, fiindcă nu sunt cunoscute decât datorită medierii necesare a acestor structuri și pe care ele le îmbogățesc, încadrându-le (chiar dacă le situează în ansamblul posibilelor) (J. Piaget, *Epistemologia genetică*, traducere: I. Pecher, Editura Dacia, Cluj, 1973, p. 5).

What captures and describes speculative logic in great detail is precisely the effective and continuous intrinsic transformation structure of the subject (seen from a logical perspective, not a psychological one, which doesn't radically change the data, especially if we acknowledge the particular relation between speculative logic and the science of experiencing consciousness, i.e., the phenomenology of spirit). Indeed, objectivity is merely the mediation of this becoming structure. But for the representatives of genetic epistemology, the Hegelian contribution came under the notion of 'dialectic movement', which, as they believe to be the case with all dialectical thinking:

insistent sur l'idée de nouveautés et en cherchant le secret en ces «dépassements» qui transcenderaient sans cesse le jeu des thèses et des antithèses [J. Piaget, *L'Épistémologie génétique*, ed.cit., p. 6]

which is rendered in Romanian as:

insistă asupra ideii de noutăți și caută secretul acestora în «depășiri» care ar transcende neîncetat jocul tezelor și antitezelor [J. Piaget, *Epistemologia genetică*, ed. cit., p. 5].

However, it is easy to notice that for speculative logic the 'jeu des thèses et des antithèses' is not the source of 'nouveautés', but only a by-product of the subject's movement of self-positing and that of self-development into something different. In their absence, 'le jeu des thèses et des antithèses' is nothing but hollow dialectics, often sophistic.

II.2. Is there a conclusive theory?

The fate of speculative logic was in no way similar to that of classical logic. In all fairness, not even their temporal and cultural intervals were alike. Over the course of two thousand years, Aristotelian logic became the quintessential tool of science in the Hellenistic and Roman world, in the Byzantine, Arab and Latin Middle Ages, and even in the Renais-

sance. The Modern age had a more sceptical view of it, but it also didn't completely dismiss it. Speculative logic underwent a severe misrepresentation right after its emergence, which almost completely compromised it.

The misrepresentation of speculative logic is a result of the logic's own movement originating from different types of movement studied by various sciences (where each of them occurs at different levels of development, of course). The concern with organizing the types of movement into quasi-systematic notions, apparently philosophical, became more and more widespread, as scientific development speeded up beginning around the second half of the nineteenth century. A whole new universal method of science was relentlessly sought through a process of classifying the types of movements, as researchers envisioned the replacing of traditional metaphysics, which had been incessantly blamed for the long-lasting setback of scientific knowledge. Metaphysics became synonymous with pseudo-science, as its meaning became more and more unclear. The scientist, attentive to his field of research and its particular movements, replaced the logician and accused him of being bound beyond remedy to metaphysical preconceptions.

But from a logical perspective (both classical and speculative), any effort to determine the privileged type of movement that could institute knowledge is based on a preconception, or implicit assumption, that emerges in any quasi-systematic endeavours, namely, we somehow have access to a sort of individual knowledge, *a priori* from the knowing subject. Therefore, precisely these quasi-systematic undertakings utilize the metaphysics we criticize classical and speculative logic for relying on. The criticism applied to classical and speculative logic slowly vanishes – along with the knowing subject – as we get closer to modern times.

Mathematics is the science where *processes* have a privileged status; furthermore, the study of processes in mathematics is done independently of any other subject. There is no surprise that, after a few failed attempts, the task of playing the role of universal method was taken over by mathematics. The increasingly rare attempts of continuing the speculative endeavour were therefore looked down upon.

III. The classical and speculative subject

The discussion above is not meant to wear out the entire topic of the classical and/or speculative subject. Its purpose was to clear the way for comparative analysis. So let us leave aside the previous examination regarding the historical complexity of the development of logic and focus on strictly logical issues concerning the subject. From everything previously mentioned, we can draw the following conclusions:

1. The subject of classical logic and the subject of speculative logic can only be studied as long as they are understood as parts of a syntactic structure. Taken independently of the syntactic structure they are a part of, the classical subject and the speculative subject are just nouns, not subjects.

2. The syntactic structures to which the classical subject and the speculative subject belong are: classical predication and speculative predication.

3. Classical predication focuses on the fixed, stable aspect of the subject. Classifying the classical subject intends to obtain an accurate demarcation tool for the subject as noun in regard to the corresponding object, so that through predication we can achieve a correct description of all the relations between subject and object.

4. Speculative logic considers that between the subject as noun and its corresponding object there is a permanent coincidence. The purpose of predication is to express this coincidence at different stages of the subject's development.

5. Classical predications – placed in opposition from a syntactic point of view – express subjects that are different from a real perspective.

6. Speculative predications – placed in opposition from a syntactic point of view – express changes of the real subject.

7. Viewed from the perspective of subject, classical predications and speculative predications refer to distinct processes:

a. Classical predication is the expression of a process *about* the subject. The operator (the one performing the predication) is different from the subject;

b. Speculative predication is the expression of a process *of* subject. The operator (the one adding conceptual determinations) is the subject itself.

8. Both classical logic and speculative logic are built around the notion of subject. In both cases the subject comes first in relation to the process. Classical and speculative syntactic structures cannot be analysed independently of the type of subject each use.

9. New logic (mathematical logic) on the other hand gives priority to the process (operation). The subject, understood as the argument for a certain function, gives the syntactic structures the possibility of being analysed independently of the value of the argument of function.

10. The subject of classical predication is the limit case of the speculative subject.

References

Aristotelis (1949) *Categoriae et liber De Interpretatione* recognovit brevisque adnotatione critica instruxit L. Minio-Paluelo, Oxonii e typographeo clarendoniano, MCMXLIX

Aristotle (1908) *On Interpretation* [Organon II], translated by E. M. Edghill. In *The Work of Aristotle translated into English under the editorship of. J.A. Smith, W.D. Ross*, Oxford.

Aristotel (1971) *Despre Interpretare*, in: Ammonios, Stephanus, *Comentarii la tratatul Despre Interpretare al lui Aristotel, însoțite de textul comentat*, Traducere, cuvânt înainte, note și comentariu de Constantin Noica, București: Editura Academiei RSR.

Hegel G .W .F (1999) *Phänomenologie des Geistes*, In G.W.F. Hegel, *Hauptwerke in sechs Bänden*, Band 2, Hamburg: Felix Meiner Verlag.

Hegel G .W .F (1999) *Die subjektive Logik oder die Lehre vom Begriff*. In G.W.F. Hegel, *Werke in sechs Bänden*, Band 4, Hamburg: Felix Meiner Verlag.

Hegel G .W .F (1977) *Phenomenology of Spirit*, translated by A.V. Miller with Analysis of the Text and Foreword by J.N. Findlay, 9–10. Oxford: Oxford University Press.

Hegel G .W .F (2010) *The Science of Logic*, translated and edited by George di Giovanni, Cambridge: Cambridge University Press.

Dopp J (1949) *Leçons de logique formelle*, Première partie, Logique ancienne, *La logique des jugements prédicatifs*, Éditions de L'Institut Supérieur de philosophie, Louvain.

Piaget J (1972) *L'Épistémologie génétique*, deuxième édition, Presses Universitaires de France.

Piaget J (1973) *Epistemologia genetică*, traducere: I. Pecher, Cluj: Editura Dacia.

DOSTOEVSKY'S CHRIST AND NIETZSCHE'S JESUS AS "CONCEPTUAL CHARACTERS"

Tamara S. Kuzubova

Federal State Autonomous Educational Institution of Higher Education,
Yekaterinburg, Russia
e-mail: tamara.kuzubova@mail.ru

Abstract:

In the present article, the author analyses the interpretation of the phenomenon of Christ by Dostoevsky and Nietzsche. The author uses comparative and hermeneutic methods of historical and philosophical research. Dostoevsky's Christ and Nietzsche's Jesus are interpreted as "conceptual characters" (G. Deleuze), occupying an important place in the philosophical constructions of both thinkers. Stating the epoch-making event of the "death of God" in European culture, they discover the origins of nihilism in Christianity itself and attempt (each in his own way) to recreate the original, pristine Christianity. Reconstruction of the original image of Christ makes it possible to comprehend not only the historical destiny of Christianity and the European portion of humanity, but also the prospects for overcoming the crisis of European and Russian (in the case of Dostoevsky) self-consciousness. It is argued that both interpretations, although far from orthodox Christianity, play the role of a central link in the development of the philosophic thinking of the Russian writer and German philosopher from the critical deposition of European humanism and metaphysics to new projects of human existence in the world. The conceptual images of Dostoevsky's Christ and Nietzsche's Jesus personally embody the spiritual attitudes and models of life that are timeless in nature, and at the same time serve as an expression of the "fundamental metaphysical positions" (M. Heidegger) of existential thinkers. The assertion of the absolute genuineness and beauty of the moral ideal of Christ allows Dostoevsky to return transcendence to the godless world – to substantiate the neo-Christian version of metaphysics, the religious-existential ontology. The "Glad Tidings" of Jesus, his life and death, appear in

Nietzsche's works as a practical elimination of transcendence, the Platonic dualism of the "true" and "visible" worlds. The spiritual attitude of Jesus reveals a direct affinity to Nietzsche's anti-metaphysical "philosophy of becoming".

Key words: Christ, Jesus, philosophical construction, conceptual character, transcendence, philosophy of becoming.

Introduction

Numerous studies devoted to the issue of congeniality, the coincidence of the main lines of thought of F.M. Dostoevsky and F. Nietzsche, according to K. Miller, have long "turned into a small genre of literature about Nietzsche" (Miller, 1973, 202). A similar "small genre" exists in the vast field of philosophical literature about Dostoevsky. The work of the Russian writer and German philosopher, who practically simultaneously stated the "death of God" and the ascent of nihilism, marked a turning point in European and Russian self-awareness, a "spiritual threshold" (K. Jaspers) of the new era. One of the key issues in the literature devoted to comparing the philosophical worlds of Dostoevsky and Nietzsche is the understanding of Christianity and its historical destiny. The figure of Christ is inevitably in the focus of this understanding. In interpreting the Christianity of Dostoevsky, researchers take two main positions: Some create the image of an Orthodox writer or, at least, a deeply religious

Christian, others emphasize the unorthodox nature of his philosophical and religious thinking. These positions, which were formed back in the Russian philosophy of the “Silver Age”, are also present in modern publications. N.O. Lossky believed that “all of his (Dostoevsky's – T.K.) worldview has an Orthodox Christian character” (Lossky, 1994, 80). This position is still held by some researchers today. Thus, K. G. Isupov asserts that in the concept of Beauty, which, according to Dostoevsky, “will save the world”, the divine Providence comes true (Isupov, 2011). V.K. Kantor believes that all Dostoevsky's philosophy is determined by faith that allowed him to become the creator of Russian religious philosophy (Kantor, 2011). A number of authors see the closeness and even identity of Dostoevsky's religious and ethical project to the project of Western Christianity (despite the well-known critical attitude of the writer to Catholicism) (Márquez, 2017; Admirand, 2021). A more moderate position is taken by researchers who, noting the unorthodox nature of Dostoevsky's religious experience and reflection, argue that his search for genuine Christianity ultimately leads to the harmonization of anthropological and theological thinking, the recognition of Revelation and the transcendent God (Sanz, Nzewi, 2017), to a deep renewal of the Orthodox tradition (Medvedev, 2018). In Dostoevsky's image of Christ, his openness to the transcendental life is emphasized (Ghini, 2017); only the sensation of the existence of God in the world and in the soul of man turns out to be a guarantee of salvation and a proof of the existence of God (Kibalnik, 2020). These authors point to the relevance of Dostoevsky's spiritual quest for modern theology.

The second position in its most radical version is presented in the works of I.I. Evlampiev, who believes that following the general anthropological orientation of non-classical philosophy, Dostoevsky comes in his spiritual quest to the idea of a “this-world” religion: There is no place for the transcendent in his world, and “Christ is understood as a perfect man, and God is understood as the inner principle of the human personality, which determines movement of the personality to perfection” (Evlampiev, 2013, 121). Another version of “this-world” religion is presented, according to the author, in the philosophy of Nietzsche, who, like Dostoevsky, removes the border between man and God from the being of Christ. In other words, according to Evlampiev, the desire of Dostoevsky and Nietzsche to present the phenomenon of Christ in its genuineness leads to two versions of religious immanentism.

In recent publications on Nietzsche's views on Christ and Christianity, two versions can also be found. According to one, Nietzsche's attempt to restore the true psychological type of Jesus, and thereby, the original Christianity, makes it possible to eliminate the rigorism, the coercive nature of the morality of dogmatized Christianity. Nietzsche's reconstruction of early Christianity opens up the possibility of a religion “maximizing” life, the possibility of a freer and more humane ethics (Bittencourt, 2011; Feiler, 2017). In another version, Nietzsche appears as an anti-Christian and a pagan; all his inner spiritual work is reduced to renouncing Christ and driving him out of himself (Bonetskaya, 2013). The neo-Christianity of the philosophers of the Russian Renaissance, inspired by Nietzsche, is unambiguously interpreted as neo-paganism.

It seems that a correct assessment of Dostoevsky's Christianity and Nietzsche's “anti-Christianity” is impossible without specifying the place of the phenomenon of Christ in their philosophical constructions, or without revealing the essential differences between these fundamentally non-classical and strikingly similar conceptual systems. Despite the antisystemic pathos of Nietzsche's philosophy, due to the efforts of many researchers, a certain conceptual structure has been identified, which concentrates in itself the internal interconnection of the main motives of his thought. In Nietzsche's draft notes, the following main issues can be outlined: “European nihilism”, “the criticism of higher values”, and “the principle of setting new values” (Nietzsche, Vol. 12, 292). According to M. Heidegger, Nietzsche's philosophical construction consists of “five main issues: “nihilism,” “the reassessment of all previous values,” the “will to power,” “the eternal recurrence of the same,” and the “Übermensch,” which “show Nietzsche's metaphysics every time in one, but always defining the whole, aspect” (Heidegger, 1993, 67). “A work of art”, wrote A. Camus, “is also a construction, and everyone knows how monotonous great creators are” (Camus, 1990, 77). Dostoevsky's novels are “completely art and completely philosophy, united unmixedly and inseparably” (Pomerants, 1990, 144). At the basis of the writer's creativity, there is a figurative-conceptual construction, the dominants of which are the favorite motives of Dostoevsky's existential-philosophical reflections. D. Merezhkovsky, L. Shestov, Vyach. Ivanov, N. Berdyaev, N. Lossky, M. Bakhtin, and other Russian and Western thinkers have worked on the identification of such dominants. Based on these investigations, the following version of Dostoevsky's

philosophical construction can be proposed: nihilism and "the rebirth of all beliefs" (F. M. Dostoevsky), the "instinctively creative nature of life" (Viach. Ivanov), the "man-god", "the ideal of Christ", God and the immortality of the soul, and the beauty of the "earth."

In both constructions, the movement of thought can be clearly traced from the statement of the loss of former spiritual values to their radical reassessment and, further, to new projects of the world and human existence, which are of a non-classical nature. However, both in the case of Dostoevsky and in the case of Nietzsche, these constructions are not simply logical sequences of concepts; in fact, behind each of these concepts there is a certain spiritual attitude, lived experience, and special understanding of the world and the purpose of man. Therefore, in comparing the philosophical constructions of Dostoevsky and Nietzsche, Deleuze's idea of conceptual characters that are hidden or explicitly present in the thinking of any philosopher and play an important role in the creation of his original concepts can be very fruitfully used (Deleuze & Guattari, 1998, 80-109). The present article will show that the role of the central link in the named constructions is played by the conceptual characters of Christ and Jesus.

Conceptual characters are the characters of thinking, contributing to the definition and the birth of a concept – the "supreme defenders" (G. Deleuze) of the philosopher's thought. Deleuze emphasizes that the conceptual character "has nothing to do with an abstract personification, symbol or allegory, since he lives and insists" (Deleuze & Guattari, 1998, 84). Probably, the very idea of conceptual characters arose in Deleuze largely under the influence of Nietzsche, who was the first to raise the fundamental questions, "Who speaks?" and "Who interprets?" which replaced the question of classical metaphysics, "What is existence?" According to M.M. Bakhtin, "Dostoevsky, speaking paradoxically, thought not with thoughts, but with points of view, minds, and voices" (Bakhtin, 2002, 106). The hero never acts as the personification of the idea; the idea always contains the integral spiritual attitude of a unique personality. The philosophical worlds of Dostoevsky and Nietzsche are constituted by means of conceptual characters in the experimental testing of their "perspectives" and spiritual attitudes. Parallel conceptual characters are not only "man-god" and "Übermensch", but also the "direct man" and "kind man", the "Russian wanderer" and "pilgrim", the "Grand inquisitor" and the "ascetic priest", "Christ" and "Jesus". A similar approach can be

found in some recent publications (Paschoal, 2016; Bittencourt, 2011). Thus, A.E. Paschoal uses Nietzsche's concept of the "psychological type", which acquires a certain meaning and stability, depending on the argued construction. The author points to the possibility of interpreting as Nietzsche's psychological types not only the figures of Jesus, but also, for example, Goethe, Kant and Wagner (Paschoal, 2016, 637–638).

In the present article, Dostoevsky's Christ and Nietzsche's Jesus will be considered as key conceptual characters, closely related not only to reflection on the essence of Christianity and nihilism, but also to the interpretation of Being and the meaning of human existence, as well as to attempts to formulate an alternative to both nihilism and historical Christianity as its source.

Methods

The author uses comparative and hermeneutic methods of historical and philosophical research. Dostoevsky's Christ and Nietzsche's Jesus are compared as "conceptual characters" occupying an important place in the philosophical constructions of both thinkers. Considerable attention is paid to the interpretation of the original texts related to the topic of research.

Results and discussion

Dostoevsky's Christ: "The Ideal of Beauty"

In Dostoevsky's spiritual experience, the thirst for faith coexisted, by his own admission, with the torments of unbelief and doubt. However, throughout the spiritual evolution he experienced, his "symbol of faith" (Dostoevsky, vol. 28: 1, 176) remained unchanged, namely the conviction of the unattainable beauty of Christ as a moral ideal. "Ecstatic love for the Holy Face of Christ" (N. A. Berdyaev) is closely connected with the writer's conviction in the meaningfulness of the world and human existence. This conviction is one of the most important elements of Dostoevsky's philosophical construction. "The writer had a completely a priori, not amenable to any reflection, deeply intimate rejection of the idea of the *absurdity* of human existence, and, consequently, confidence in some kind of high and purposeful mission of mankind on Earth" (Erofeev, 1990, 14). Christ, according to Dostoevsky, is the only possibility of generating meaning in a world experiencing the "eclipse of God" and the loss of all the former spiritual foundations of life. The writer

puts this idea into the mouth of Kirillov, the protagonist of *Demons*: "This man was the highest on the whole earth, he was what the earth existed for. The whole planet, with everything on it, without this person, is one madness. There was no such person either before or after *Him*, or ever, even before the miracle. This is the miracle that there has never been and never will be such a person" (Dostoevsky, vol. 10, 471). These words concentrate the main points of the interpretation of the image of Christ by Dostoevsky himself: the uniqueness of the phenomenon, the embodiment in his personality of the meaning of the world and human existence ("the ideal of man in the flesh"), a "miracle", equivalent to beauty and rational incomprehensibility. In response, F. A. Stepun formulated "Kirillov's antithesis", which, undoubtedly, was the antithesis of Dostoevsky himself in this way: "... either Jesus, whose name Kirillov does not pronounce, is the only meaning of the universe and its existence, or the entire universe is a lie and madness" (Stepun, 1991, 369). It is noteworthy that the theomachic logic of the nihilist characters Versilov, Kirillov, and Stavrogin not only does not deny the absolute genuineness of the ideal of Christ, but is paradoxically combined with the recognition of his holiness and infinite beauty. The death of Christ as "the highest man on earth" aggravates the feeling of the absurdity of human existence and serves as an argument in favor of rejection of "this world of God".

Genuine Christianity, inextricably linked with the person of Christ, seemed to Dostoevsky the only alternative to nihilism. The "Symbol of Faith" of the writer definitely indicates that he undoubtedly accepted the thesis of Kirillov's dilemma, rejecting the antithesis. However, in order to maintain this position, it was not enough to recognize Christ as the highest person or moral ideal. "The point is in the urgent question: is it possible to believe, being civilized, that is, a European, that is, to believe unconditionally in the divinity of the Son of God Jesus Christ (for all faith consists only in this)?" (Dostoevsky, vol. 11, 178). All the writer's doubts are focused on this issue, and its solution determines his philosophical construction. The acuteness of this issue is characteristic of the consciousness of a Christian who experiences, if not the "death of God," then the state of being forsaken by God, the situation of "God's eclipse" (Buber, 1999). Therefore, reflecting on the question of the divinity of Christ, the writer came to its unorthodox justification. The task of recreating a genuine image, distorted, according to Dostoevsky, by Catholicism, was one of the most

important stimuli for Dostoevsky's philosophical reflections, his artistic and publicistic work. The theme of Christ was persistently repeated in notebooks, letters, and journalistic works of the writer, and was one of the recurrent themes in his novels. The justification of the divinity of Christ, which subsequently underwent some changes, was formed by the writer mainly by 1864. It is contained in the famous note made by Dostoevsky on the day of his wife's death and in the outline for the article "Socialism and Christianity".

In these texts, as in a number of others, two lines of argumentation can be distinguished, ultimately coinciding with each other. The starting point of the first line of thought is the affirmation of the beauty of the ideal of Christ: "Christ in Himself and in His Word carried the ideal of Beauty" (Dostoevsky, v. 29, 85). The same recurrent theme can be found in notebooks and letters: Christ is the "ideal of man in the flesh", "the eternal ideal", "the great and final result of the development of mankind", "the ultimate ideal on earth", "the greatest positive ideal of man". In all these definitions, a certain discrepancy between the language in which Dostoevsky speaks about Christ and the very subject of his statement is striking. "Ideal", "the ultimate result", "purpose", etc., are terms referring rather to the vocabulary of German classical philosophy and Romantic aesthetics than to Christian theology. And in this language a completely secular thought is expressed: Christ is a man who, in an incomprehensible way, has risen to moral heights inaccessible to an ordinary person, and thereby revealed a model of life that represents the ultimate guideline of all human aspirations. Actually, the "wonderful and miraculous beauty" of the ideal of Christ lies in its "moral unattainability".

In "Socialism and Christianity", Dostoevsky formulated the "law" of this ideal. Christ united in his personality, life, and death, the things which, as it seems to Dostoevsky, cannot be united: the "law of Self," the law of personality, and the law of love and humanism. Personal freedom, both in the conviction of many of Dostoevsky's heroes and in his own, is incompatible with love for one's neighbor. "It is impossible to love a person as oneself, according to the commandment of Christ. The law of personality on earth is a limitation to it. One's Self is a limit. Christ alone could do it, but Christ was the eternal ideal, which man is striving to attain and, according to the law of nature, must strive to attain" (Dostoevsky, vol. 20, 172). Close contact with Another is always a threat to the original, unique Self.

"The ecstatic sense of personality" (N. A. Berdyaev) is characteristic of Dostoevsky as a "Russian European" into whose flesh and blood the Cartesian cogito has penetrated. Being an internal self-evidence, it coexists in the spiritual world of Dostoevsky with the equally absolute presence of Christ as the "ideal of Beauty". The idea of the uniqueness and intrinsic value of each Self serves in the philosophical and artistic world of Dostoevsky as the basis for a themachic rebellion, a protest against the structure of the universe, which threatens the individual with transformation into a thing among things, suffering, and death. A person embarks on the path of self-assertion, but on this path, chaos awaits him, generated by willful freedom. Describing the modern era as a "time of transition" marked by the "disintegration of the masses into personalities" and "the loss of the living idea of God" (Dostoevsky, vol. 20, 192), Dostoevsky regards the ideal of Christ as the only hope for man: "If no purpose were shown to man in this state of his, it seems to me that all mankind would have gone mad. Christ was shown as such purpose. (NB. Not a single atheist who disputed the divine origin of Christ denied that HE is the ideal of humanity...)" (Dostoevsky, vol. 20, 192). Christ as the possibility of gaining meaning and salvation is clearly opposed here to the perspective of madness and the death of mankind.

In the ideal of Christ, Dostoevsky hopes to find an outcome to the torment of duality, which is well known to the "heightened consciousness" of the nihilist: the awareness of the inconsistency of the former ideals coexists with the "need for a higher thought" (Dostoevsky, vol. 24, 334) and a higher life. The ideal of Christ also contains the possibility of overcoming nihilism as a transitional state in the history of mankind. Dostoevsky, on the other hand, is looking for an opportunity to unite individualism with the fraternal unity of people; for this, the "impossible on earth" miracle of Christ's love is necessary. This miracle is interpreted in the following way: "To attain the full power of consciousness and development, to fully realize one's self and to give it all by one's own will to everyone" (Dostoevsky, vol. 20, 192); "... To be the lord and master of oneself, one's own self, and to sacrifice this Self, give it to everyone" (Dostoevsky, vol. 20, 193). And then the peak of personal freedom, "the highest self-will" "is at the same time the highest renunciation of one's own will" (Dostoevsky, vol. 20, 192). The all-mighty Self with its "unrestricted and free will" (Dostoevsky, vol. 5, 115), which the hero of *Notes from Underground* so fiercely defended, is pre-

served, and at the same time its destructive energy is diverted into a safe channel by pointing to an unattainable model of transformation of the highest self-will into the highest renunciation of will. Here, a certain return of the personality, which has reached the "extreme limits of development," to the lost wholeness is meant. Consequently, at first, an abyss had to be formed between the Self, self-consciousness, and the world "offending" it, devoid of meaning, so that there was a need to overcome this abyss. It is unlikely that such a problem could have been faced by the historical Christ, since for him such a distance simply did not exist. The one-time personal embodiment of the "ideal of Beauty" and its rational incomprehensibility induced Dostoevsky to interpret the phenomenon of Christ as a miracle that evokes aesthetic delight. This interpretation opens up the possibility of proving his divine nature.

Understanding Christ as "the ideal of man in the flesh," Dostoevsky seemingly focuses on the dogma of Christianity, but his interpretation of divine Incarnation is not orthodox. Here the meaning is shifted: In fact, we are talking about the perfect embodiment of the human essence, about the implementation "in deed and in the flesh" of the model of humanity, ideal, dream, but not about the divine Incarnation in the Christian sense. That was why V.V. Zenkovsky looked on Dostoevsky's philosophical position as "Christian naturalism". One gets the impression that "...the mystery of the God-manhood of Christ, which Dostoevsky felt well in general, is lost in this context, and we can see the purely human face of Christ for he becomes the 'ideal' of man! The taste of naturalism is reflected in the very fact that perfection, which is conceived in Christ as a person, as it were, refers to an ordinary person, as his hidden, primordial, sacred quality" (Zenzovsky, 1991, 114). In his reflections on Christ, Dostoevsky went from the "ideal of man in the flesh" to the affirmation of His transcendental nature; "starting from Christ the man, step by step he cognized his divinity" (Lauth, 1996, 331), but he did not overcome 'Christian naturalism.'

The second line of substantiation of the divinity of Christ and the idea of the immortality of the soul is based on the motive of introducing finite human existence to world infinity, interpreted as transcendental divine being. This line of justification is developed by Dostoevsky in parallel with the theme of the beauty of the ideal of Christ, the merging in His personality of the law of Self and the law of humanism. More precisely, the antithesis of the law of per-

sonality and the law of love in Dostoevsky coincides with the antithesis of becoming and eternity, and the ideal of Christ embodies the synthesis of these opposites. According to Dostoevsky, the striving for self-affirmation (the law of personality) and the eternal openness and incompleteness of human existence associated with it are inherent in human nature. The attainment of the ideal of Christ as the fusion of the Self and everyone would therefore mean the end of individual existence and, in terms of humanity, the end of history. "Thus, on earth, man strives for an ideal that is opposite to his nature" (Dostoevsky, vol. 20, 175). Dostoevsky does not agree to regard this end as death and nothingness; it is nothing but the moment of transition into Eternity. "But to achieve such a great goal, in my opinion, is completely meaningless if, upon reaching the goal, everything fades out and disappears, that is, if the person does not have life even after the goal is achieved. Consequently, there is a future paradisaal life" (Dostoevsky, v. 20, 17). The interpretation of the phenomenon of Christ as "the final ideal of man," in other words, as a person who has fulfilled his destiny, allows Dostoevsky to assert the immortality of the soul, which embodies the ideal of Christ in eternity.

The summary of the writer's reflections on the theme of the "ideal of man in the flesh" is contained in the draft materials for *Demons*. Here the writer emphasizes that faith in Christ as a moral ideal alone is not enough to save a person in a situation of nihilistic despair. The divinity of Christ is derived from the "need for adoration", or from the irrational belief that Christ "is the ultimate ideal of man, the entire incarnate Word, God incarnate" (Dostoevsky, vol. 11, 187). "Note that human nature certainly requires adoration. Morality and faith are one thing, morality follows from faith, while the need for adoration is an inalienable property of human nature. This property is high, not low – the recognition of the infinite, the desire to spill over into the world infinity, the knowledge that one is coming from it. And for adoration, "God is needed" (Dostoevsky, vol. 11, 187–188). Transcendence is deduced, therefore, from the immanent consciousness of absolute certainty, self-evidence of the connection of finite human reality with the "infinity of the world", which the writer identifies with the infinity of God. Faith in Christ as "God incarnate" (tantamount to faith in the "final ideal of man") serves as living proof of the possibility of "contact with ... other mysterious worlds" (Dostoevsky, vol. 14, 290–291). This faith, experienced as the highest degree of delight, makes a breakthrough from the finiteness and egoism of man

into the infinity of divine being and love possible. This interpretation of "adoration" is essentially at odds with the Orthodox idea of deification. V.M. Lurie notes that Dostoevsky's reasoning is based on two meanings of the word "adoration": literal, which was preserved even in Dostoevsky's time, i.e., "deification", and figurative, or psychological, i.e. "extreme degree of love". "But both meanings, with the awareness of their difference, are identified: adoration is interpreted as a psychological and even natural state – a person's relationship to Christ, whom he believes in as in God. From such 'adoration' does not follow and cannot follow the deification of the person himself – on the contrary, the person remains as he was, with his psychology" (Lurie, 1996, 305). Lurie emphasizes the non-Orthodox character of such a soteriological perspective, although he sees in this interpretation a clear movement of Dostoevsky towards Orthodoxy.

In Dostoevsky's argumentation, one can discern the contours of a version of neo-Christian metaphysics, which would later be implemented by religious-existential philosophy. From the depths of existence, the consciousness of a person who experiences his finitude and God-forsakenness perceives and affirms transcendence and God; the metaphysical "leap" (Camus, 1990, 41–43) from a finite and absurd existence to a higher meaning and immortality of the soul is carried out in an irrational act of faith. For example, P. Vaskovic writes about a similar metaphysical leap in the philosophy of Kierkegaard and Dostoevsky, calling it "existential-religious transformation" (Vaskovic, 2019, 81). The latter represents the transition from an "internally fragmented and selfish personality to a truly Christian, humble and loving personality" (Vaskovic, 2019, 81), which takes place at the fateful moment of "meeting with God". However, unlike the interpretation of the "metaphysical leap" described above, this author does not assign the phenomenon of Christ a central role in existential transformation. On the contrary, the latter is possible due to the fact that God meets a person (in fact, thanks to divine grace), and only after this spiritual transformation is it possible to follow Christ as a model of love and humility. It seems that the philosophical solution found by Dostoevsky only seems to be a return to traditional Christianity, to the values that were forever lost by his nihilist heroes. The negative experience of a person's self-affirmation and "self-will," depicted many times in Dostoevsky's novels, requires a breakthrough to the transcendental foundations of human existence. Dostoevsky's philosophical con-

struction presupposes the identity of the existence of God and the immortality of the soul. The ideal of Christ as a "denial of the earth" indicates the possibility of a different, transcendental reality and a different form of human existence. Dostoevsky insists: "denial is necessary", and "denial of the earth is necessary" (Dostoevsky, vol. 24, 112), otherwise human existence is meaningless; that is, as Kirillov says in *Demons*: "God is necessary, and therefore must exist" (Dostoevsky, vol. 10, 469).^{1□}

Nietzsche's Jesus: "The one who felt that he was without sin"

For Nietzsche, as well as for Dostoevsky (but in a different way and for different reasons), the image of Christ is "painfully close and therefore atheistically 'swept away'" (Svasyan, 1990, vol. 2, 792). None of Nietzsche's works, from *Human, All Too Human* to *The Antichrist* and the draft notes of the late 1880s, is complete without addressing the mystery of the personality, life, and death of the "Son of Man". On the one hand, the image of genuine, original Christianity, embodied in the exemplary practice of the life of Jesus, plays an important role in the critical reappraisal of Christianity. On the other hand, like in Dostoevsky's works, the phenomenon of Jesus reveals the closest connection with the key concepts of Nietzsche's own philosophical construction: life as becoming, the "eternal return," and the *Übermensch*. The key to understanding the psychological type of the Savior can be given only by psychology, understood in the Nietzschean sense as a kind of interpretation of existence in general and of human existence in particular. The philosopher's interpretation of the "word" and life of Jesus reveals an unmistakable similarity with his own philosophical position and a radical divergence with the ascetic ideal of Christianity.

According to Nietzsche, Jesus discovers the "psychological reality" of salvation (Nietzsche, 1990, vol. 2, 659) by experiencing the Kingdom of God as a "state of the heart" (Nietzsche, 1990, vol. 2, 660). He is an "anti-realist," – a "great symbolist" (Nietzsche, 1990, vol. 2, 658, 659), no word of whom can be taken literally:

What is 'evangelism'? – True life, eternal life has been found – it is not only promised, but it is here, it is *in you*: like a life in love, love without concession and exclusion, without distance. Everyone is a child of God – Jesus has no claim to anything for himself alone, – as a child of God, everyone is equal to everyone ... How can one make a *hero* out of Jesus! (Nietzsche, 1990, vol. 2, 655).

In Nietzsche's interpretation of the "Glad Tidings" of Jesus, two conceptual moments are combined: firstly, the Kingdom of God – true life is only a psychological reality; "the experience of the heart: it is everywhere, it is nowhere..." (Nietzsche, 1990, vol. 2, 660), and secondly, this reality is eternal life here and now. The combination of these two points plays an important role in Nietzsche's philosophical construction and in his personal spiritual experience.

Jesus does not need to seek the "true world" as opposed to this world; this distinction simply does not exist for him. Jesus lives in truth, "in the bosom of God", and this life, this reality is the only one there is. "If I understand anything in this great symbolist, it is that he only accepted *internal* realities as realities, as 'truths' – that the rest (natural, temporal, spatial, historical) he understood only as a symbol, as an excuse for a parable" (Nietzsche, 1990, vol. 2, 659). It is obvious that in Nietzsche's portrayal Jesus is a spontaneous anti-metaphysician, he instinctively eliminates Platonism and the metaphysics of Christianity, as if following the demand of the philosopher himself to abolish the antithesis of the "true" and "apparent" worlds as a fiction. He knows nothing about the transcendent God, about "being" that goes beyond the "experience of 'life,' which is the only one available to him" (Nietzsche, 1990, vol. 2, 658). The deepest instinct prompts Jesus to understand the "external" reality only as a language of signs, allowing Him to talk about the "internal", that is, the only important thing. In the world of Jesus, there are no opposites and, above all, there is no distance between man and God, between the "inner" and "outer" worlds. However, Nietzsche himself declares the concept of opposites to be nothing more than a prejudice of the philosophers. Jesus appears as a "free mind": "after all, for him everything solid, stable is nothing (word, formula, church, law, dogmas)" and "everything solid kills..."; he believes only in life and in all living things: and this does not 'exist' – it becomes..." (Nietzsche, vol. 13, 154). And again, according to Nietzsche himself, the nature of the becoming world defies formulation as

^{1□} It should be noted that in the project of Dostoevsky's neo-Christian metaphysics there are two lines of thought that are fundamentally irreducible to each other: the reflection of the religious-existential type described above and the Christian version of aesthetic pantheism, the famous doctrine of the salvation of the world by beauty (see: Kuzubova, 2017, pp. 27–38).

“*cognition and becoming* are mutually exclusive” (Nietzsche, 2005, 294). Furthermore, the characterization of Jesus as a “free mind” testifies to a lot: after all, under the guise of this “conceptual character” in the works of Nietzsche, in most cases, he is hiding himself. To be a “free mind” means not only to think differently, but also to be different; driven by the will to self-determination, he “with pain constantly breaks the net around him” (Nietzsche, 1990, vol. 1, 426). Jesus, as Nietzsche understands him, is incapable of any kind of denial, and yet he unwittingly came into conflict with tradition, with the “good and just”. “Glad Tidings” meant “the *denial* of everything that was a priest and a theologian” (Nietzsche, 1990, vol. 2, 654), although the “evangelist” himself was hardly aware of this antagonism, hence his characterization as a “holy anarchist” (Nietzsche, 1990, vol. 2, 654). “Not to believe, but to do, and above all *not* to do much, another *existence...*” (Nietzsche, 1990, vol. 2, 663). The “Glad Tidings” are projected into life practice, the essence of which is expressed in the “deepest word of the Gospel” (Nietzsche, 1990, vol. 2, 655): “Do not defend yourself, do not be angry, do not hold others accountable...But also do not resist the evil, love it...” (Nietzsche, 1990, vol. 2, 660). In his understanding of the “deepest word of the Gospel”, Nietzsche is close, as is well-known, to Leo Tolstoy. Genuine life, eternal life, is a life in love that knows no distances and exceptions, extending even to the acceptance of evil, to love for the enemy. Only such love can abolish all opposites, overcoming the distance between man and God. Jesus, according to Nietzsche, leaves as a legacy to mankind the evangelical practice, his own life and death, as “the existence, the fulfillment, *the reality* of this Kingdom” (Nietzsche, 1990, vol. 2, 665). The “Glad Tidings” are a unique experience of a person living “in the bosom of God” and opening the way to this life for everybody – that is, everyone who wants to follow it. “Neither ‘repentance’ nor ‘prayer for forgiveness’ is the essence of the path to God: nothing but the gospel practice leads to God; it is ‘God!’” (Nietzsche, 1990, vol. 2, 659). This means that Jesus, by virtue of his earthly existence, his exemplary life and death, eliminated all transcendence, including the notion of “God” as something beyond the boundaries of the only material and psychological reality.

Following the genealogical method, Nietzsche discovers at the basis of the word and practices of Jesus “two *physiological* realities” (Nietzsche, 1990, vol. 2, 656): an instinctive hatred of reality and an instinctive aversion to enmity. Therefore, the love of

Jesus and “the deepest word of the gospel” calling “not to resist evil” are simply a natural manifestation of the “incapacity to resist,” or of “Love, as the only one, as *the last* opportunity for life” (Nietzsche, 1990, vol. 2, 656). Freedom, the call “to be different,” is a manifestation of the same instinct – that is, an instinctive understanding of “how one should live in order to feel in ‘heaven,’ in order to feel ‘eternal’” (Nietzsche, 1990, vol. 2, 659). “The psychological reality of ‘salvation,’” wholly actual and immanent, is generated by the interpretative force of the instinct of *décadence*, the fear of any touch from outside. It takes no effort for Jesus to live this way. At least, this is the impression Nietzsche achieves, calling such a belief “a return to childhood in the mental realm” (Nietzsche, 1990, 657), and drawing the image of “the most interesting of *décadents*” (Nietzsche, 1990, 656). The possibility of interpreting Jesus as “the most interesting of *décadents*” is reinforced by Nietzsche’s appeal to Dostoevsky, and specifically, to the image of Prince Myshkin in the novel *The Idiot*. In Myshkin, Nietzsche recognized the psychological traits of “his” Jesus: almost child-like innocence, the absence of boundaries between himself and the other, “this withdrawal from ordinary life relations, this constant irrelevance of his personality and his behavior” (Bakhtin, vol. 6, 196). Jesus, this “strange saint,” is distinguished by “the fascinating charm of such a mixture of the sublime, the sick and the childish” (Nietzsche, 1990, vol. 2, 657), which only Dostoevsky could feel and convey. Reading the fragments of *The Antichrist*, dedicated to the psychological type of the Savior, gives the impression of a dissonance between the exalted pattern of life and death of Jesus, created by Nietzsche himself, and the desire to “lower” it, pointing out the roots of Jesus’ extraordinary inner freedom and love in the “higher development of hedonism on quite an unsound basis” (Nietzsche, 1990, vol. 2, 656). Nietzsche seems to ignore the inner resilience of the “strange saint”, explaining his life practice as a consequence of the inability to resist. And yet, in the interpretation of the death of Jesus, this trait cannot but make itself obvious: “this way of death” is a symbol of freedom, of “superiority over any sense of *ressentiment*” (Nietzsche, 1990, vol. 2, 664–665). In the rough notes, this psychological trait is revealed more clearly: “I don’t want to do anything that would disturb my inner peace, and if I have to suffer from this, nothing will save peace for me better than suffering...” (Nietzsche, vol. 13, 151).

This motivation for accepting suffering by Jesus does not agree with the phenomenon of a “return

to childhood in the mental field" or with "hedonism on quite an unsound basis". However, Nietzsche deliberately avoids any moral assessment of the worldview of Jesus. In contrast to Dostoevsky, who looks on the love of Christ as the apotheosis of the moral development of the personality, Nietzsche deprives the "Glad Tidings" of Jesus of a moral connotation; they are in the fullest sense of the expression "beyond good and evil". "Jesus destroyed even the very concept of 'guilt,' he completely denied the gap between God and man, he presented with his life this unity of God and man as his 'gospel'..." (Nietzsche, 1990, vol. 2, 665). By instinctively eliminating transcendence, Jesus affirms the "innocence" of becoming and abolishes sin, guilt, and morality. In the world of this unique person, "the one who felt that he was without sin" (Nietzsche, 1990, vol. 1, 598), there is no opposition of good and evil, just as there are no opposites at all. In regard to this, Aphorism 164 of *Beyond Good and Evil* directly states: "Jesus told his Jews: 'Law was for slaves; love God, how I love him, the son of God! Morale does not matter for the sons of God!'" (Nietzsche, 1990, vol. 2, 303). With the exception of *Thus Spoke Zarathustra*, no work before *The Antichrist* hints at a future shocking characterization of Jesus as an idiot. "I know only one psychologist who lived in a world where Christianity is possible, where Christ can arise every moment...This is Dostoevsky. He guessed Christ, and instinctively he first of all avoided imagining this type with the vulgarity inherent in Renan...But is it possible to make a mistake worse than making a genius out of Christ, who was an idiot? Than to falsely make a hero from Christ, who is the opposite of heroic?" (Nietzsche, vol. 13, 373). According to K.A. Svasyan's commentary, "the reference to Dostoevsky unambiguously clarifies the semantics of the word: 'idiot' is equal to 'saint' here, although this reference does not completely remove the blasphemous shade of Nietzsche's characteristics" (Svasyan, 1990, vol. 2, 802).

Jesus as a conceptual character of Nietzsche embodies one of his favorite ideas: the idea of life as becoming, in which transcendence is absent and fundamentally impossible, or the idea of life as a moment in the "eternal return". The second conceptual moment of the "Glad Tidings" of Jesus, connected with the first one (denial of any reality of the "Kingdom of God" other than the psychological one) is the attainment of bliss, or the "Kingdom of God" here and now. As in the case of Dostoevsky, in Nietzsche's works the communion with the eternity and completeness of the whole is an inalienable

characteristic of human existence. For Dostoevsky, the experience of this connection, the "need for adoration", indicates the existence of a transcendental reality that carries in itself the denial of the earthly human existence. For Nietzsche, the very experience of this connection is the state of divinity and is "God." In the philosophical world of Nietzsche, Jesus implements through daily practice the imperative "to exist in becoming." According to Dostoevsky, "the eternity of the moment" (Bakhtin, vol. 5, 80–81) is tantamount to affirming the transcendental meaning of life and the immortality of the soul. The "Kingdom of God" of Nietzsche's Jesus is a symbol of "eternal life here" (Dostoevsky, vol. 10, 188), in which Kirillov believes in the novel *Demons*. The "Glad Tidings" and practice of Jesus affirm only one reality, namely the "innocence" of becoming, eliminating the transcendent God and morality as the world of what is desirable, what should be. As a result, as noted by Jaspers, "the entire world reality melted away like a mirage" (Jaspers, 1994, 19). However, according to Nietzsche, reality is a "mirage", an interpretation "interpreting existence", and Jesus instinctively lives in harmony with Nietzsche's phenomenalism. This is why the "Glad Tidings" and the sublime symbolism of Jesus are completely alien to Christian metaphysics and theology. "Christianity at any moment is still possible....It is not tied to any of the shameless dogmas that flaunt its name: it does not need the doctrine of a personal God, sin, immortality, redemption, or faith, it needs absolutely no metaphysics, let alone asceticism, let alone any kind of Christian 'natural science'..." (Nietzsche, vol. 13, 151). The metaphysics and theology of Christianity embody the opposite of the meaning of the "Glad Tidings", giving an arbitrary interpretation of the life and word of Jesus, and representing a completely different interpretation of existence. This interpretation grows out of a pessimistic view of the world: "God created man happy, idle, innocent and immortal: our real life is a false, fallen, sinful existence, a kind of punitive existence.... Suffering, struggle, labor, and death as objections and question marks to life are regarded as something unnatural, something that should not continue; against what people seek remedies – and they find them!" (Nietzsche, vol. 13, 93).

For Nietzsche's Jesus, suffering is not an argument against life, it does not condemn it or call it into question. Jesus says "yes" to life, affirming it in his own way as the psychological reality of bliss in which hatred, suffering, and death dissolve like a mirage. The only inner reality of Jesus is opposed by

the dualistic world of Christianity, where a sinful, false existence appears only as a bridge to the other world. Christianity distorts the meaning of the “Glad Tidings” by the fact that it infects “the innocence of becoming with the concept of a ‘moral world order,’ ‘punishment,’ and ‘guilt.’ Christianity is the metaphysics of an executioner...” (Nietzsche, 1990, vol. 2, 583).

If Dostoevsky considers the immortality of the soul to be the only guarantee of the meaning of human existence, Nietzsche is convinced that “the great lie about personal immortality” (Nietzsche, 1990, vol. 2, 667) deprives life of all meaning. According to Nietzsche, the nihilistic essence of Christianity is nowhere found more clearly than in the doctrine of the immortality of the soul. It is in this that the philosopher sees a particularly gross distortion of the “Glad Tidings” of Jesus. The resurrection dogma deprives of meaning the gospel concept of “bliss”; thus, the only reality that is known to Jesus is withdrawn in favor of the afterlife. But the suffering and death of Jesus on the cross had, according to Nietzsche, nothing to do with the reality of bliss, because they could not violate the peace of the soul once it had been acquired. That is why the concepts of “death” and “death hour” belong only to the seeming world, the world of symbols. The death of the Savior as the completion of his exemplary life practice was the reality of bliss and the present existence of eternity. “It was this kind of death that was the ‘Kingdom of God’” (Nietzsche, 1990, vol. 2, 665).

Dostoevsky as a Christian sees in the idea of the immortality of the soul the only guarantee of morality, while the “anti-Christian” Nietzsche looks on it as “a shameless elevation of any kind of egoism into the infinite, into the shameless” (Nietzsche, 1990, vol. 2, 667). In the world of Jesus, there is no death or transcendental purpose, which means that the focus on the end loses its significance as a fundamental characteristic of human existence. “The ‘Celestial Kingdom’ is a state of the heart, and not something that is ‘above the earth’ or comes ‘after death.’ The Gospel generally *lacks* the concept of natural death: death is not a bridge, not a transition, it does not exist, because it belongs to a completely different, only seeming world, a world that has only symbolic meaning” (Nietzsche, 1990, vol. 2, 660). The immortality of the soul cannot be the measure of meaning in such a life. The recognition of the “innocence of becoming,” which is transformed in Nietzsche’s philosophy into the idea of eternal return, opens up the possibility of “affirming the immortality

of this life without any props of ‘other life’” (Bely, 1996, 84). Every moment of “life here” turns out to be a moment of generating meaning, touching Eternity. That is why, if a person succeeds in acquiring meaning, as Jesus did through a unique way of life, “in such a life there is no task” and “everything has already been achieved” (Nietzsche, vol. 13, 152). According to Jaspers, “the problem of the experienced bliss of the *presence of eternity in the present*, which Jesus solved with his life practice, is also Nietzsche’s own problem” (Jaspers, 1994, 78).

Interpreting the phenomenon of Christ, Dostoevsky and Nietzsche develop in their metaphysical worlds “one of the fundamental possibilities of human existence in general” (Jaspers, 1994, 19). In the case of Nietzsche, this possibility was realized in the most perfect way by the character of Jesus. The concept of the “Son of man” is an “eternal reality”, a “psychological symbol freed from the concept of time” (Nietzsche, 1990, vol. 2, 659–660). Nietzsche himself, in his own life, tried to implement this fundamental possibility of human existence in general. Jaspers notes: “There are many such almost literal analogies when Nietzsche speaks of Jesus and himself in the same terms” (Jaspers 1994, 19). The most important condition for the possibility of such a way of existence is “superiority over any sense of *ressentiment*,” understood as a refusal to condemn life. This is also captured by the notion of Nietzsche’s *amor fati*, the mandate “do not resist the evil”, and Kirillov’s conviction that “all is well”. The love of Jesus as a path to bliss, “the last opportunity of life”, at the same time inevitably attracts him to death, manifesting itself as the will to Nothing. Jesus, who managed to gain eternity in a moment and yet was possessed by longing for death, foreshadows, in the world of Nietzsche, the possibility of a different psychological type, for which “*being itself is blessed enough*” to also justify the immeasurable bulk of suffering” (Nietzsche, vol. 13, 246). The embodiment of this type and attitude is Dionysus.

Conclusion

In the epoch of “the death of God,” a person finds himself before a world devoid of purpose, meaning, truth, before the abyss of Nothing. Dostoevsky and Nietzsche saw the image of this abyss in all its horror and strove with all their might to discover the possibility of salvation, developing new projects of human existence in the world in which the interpretation of the phenomenon of Christ played an important part. Dostoevsky’s Christ and

Nietzsche's Jesus appear as a unique personal embodiment of "one of the fundamental possibilities of human existence," and these conceptual characters do not fit into the framework of Christian theology. Both interpretations, although far from Orthodox Christianity, play the role of a central link in the development of the philosophic thinking of the Russian writer and German philosopher from the critical deposition of European humanism and metaphysics to new projects of human existence in the world. For both thinkers, recreating the genuine image of Christ is a way of understanding the history of Christianity, and thus the history and fate of the European part of humanity. Dostoevsky and Nietzsche pass from personal experience of the phenomenon of Christ to conclusions about the nature of human existence "in the midst of being as a whole" (M. Heidegger). In the philosophical worlds of both thinkers, the phenomenon of Christ is a symbol of the "eternity of the moment", that is, the highest meaningfulness of the final human life, the communion with eternity and the completeness of the whole. Dostoevsky arrived at the affirmation of the "eternity of the moment" by means of an unorthodox substantiation of the immortality of the soul, an important role which was played by the movement from the experience of absolute certainty in the ideal of Christ to the recognition of his divine nature. Christ plays a colossal role in the philosophical world of Dostoevsky – both the role of theodicy and "anthropodicy", as Berdyaev puts it. Belief in the meaningfulness of human existence, associated with the interpretation of the phenomenon of Christ as a unique embodiment of this meaning, serves as the basis for both theomachist rebellion and the acceptance of the transcendental divine existence, identical to the immortality of the soul. The image of Christ allows us to preserve freedom as an essential feature of human existence, to "provide" it with Being in eternity and at the same time to indicate the measure of freedom which has a transcendental nature. The metaphysical leap from the tragic meaninglessness of existence to the assertion of its transcendental meaning, from the self-assertion of the Self to the "complete synthesis" of the Self with *everyone*, from becoming to eternity is mediated by the absolute genuineness of experiencing the ideal of Christ. Christ as a conceptual character is the central link in the philosophical construction of the religious-existential type, which is found at the heart of Dostoevsky's artistic world. Nietzsche's interpretation of the phenomenon of Christ is determined by the key provisions of his radical philosophy of becoming. Jesus, depicted by

Nietzsche, with his "Glad Tidings" and remarkable life and death practically eliminates the transcendent God and the supersensible "true" world, affirming the only psychological reality: the phenomenal world. He unwittingly destroys metaphysics and follows Nietzsche's imperative "to exist in becoming". The new way of life proposed by Jesus shows an affinity for the "formula of human greatness" of the philosopher ("*amor fati*"). The psychological reality of bliss experienced by Jesus has nothing to do with morality and does not need the idea of the immortality of the soul as a measure of the sense of life. Eternity, rejected as the idea of the immortality of the soul, returns in the form of the idea of eternal return, becoming the source and criterion of the meaningfulness of human life. Jesus as a conceptual character embodies in Nietzsche's philosophical construction the possibility of generating meaning in the aimless stream of becoming devoid of transcendence, practically realizing the experience of the presence of eternity in the moment. Feeling his closeness to the deepest word of Jesus, Nietzsche, nevertheless, characterizes Jesus as "the most interesting of *décadents*," escaping from the hated reality "by fleeing into the 'incomprehensible,' into the 'inexplicable'" (Nietzsche, 1990, vol. 2, 655). In an effort to overcome Christianity through "super-Christianity" (K. Jaspers), he sets in opposition the "psychological type of the Savior" with another religious type – Dionysus, the type of "completely successful character delightfully pouring out in its abundance", and expressing "apology and deification of life" (Nietzsche, vol.13, 246).

References

- Bakhtin M M (2002) Problems of Dostoevsky's poetics. In *Collected works*, 6–300. Moscow: Russian dictionaries, Languages of Slavic culture.
- Bakhtin M M (1997) Additions and changes to "Rabelais". In *Collected works*, 80–129. Moscow: Russian dictionaries.
- Bely A (1996) Friedrich Nietzsche. In *Friedrich Nietzsche and Russian Religious Philosophy: Translations, Research, Essays by the "Silver Age" philosophers*, in 2 volumes, Vol. 2. Minsk.
- Bonetskaya N K (2013) Russian Nietzsche. *Issues of philosophy* No.7: 133–143.
- Buber M (1999) Eclipse of God. In *Two images of faith*, 440–541. Moscow: AST Publishing House.

- Deleuze G, Guattari F (1998) What is philosophy? Moscow: Institute of Experimental Sociology, St. Petersburg: Aletheia Publishers.
- Dostoevsky F M (1972–1990) Complete works: in 30 vols. Leningrad: Nauka Publishers.
- Evlampiev I I (2013) The “real-world” religiosity of F. Dostoevsky and F. Nietzsche (On the issue of the religious content of non-classical philosophy) *Issues of philosophy* No.7: 121–132.
- Erofeev V V (1990) Faith and humanism of Dostoevsky. In the labyrinth of damned questions, 11–37. Moscow: The Soviet Writer publishing house.
- Zenkovsky V V (1991) The problem of beauty in Dostoevsky's worldview. *The art of cinema* No.11: 108–116.
- Camus A (1990) The myth of Sisyphus. In *The Rebel. Philosophy Politics Art*, 23–92. Moscow: Politizdat Publishing House.
- Kantor V K (2011) Confession and theodicy in the works of Dostoevsky (reception of Aurelius Augustine). *Issues of philosophy*, No. 4: 96–103.
- Kuzubova T S (2017) The path to salvation: two versions of F.M. Dostoevsky. *Context and reflection: philosophy about the world and man* 6(5A): 27–38.
- Lauth R (1996) *Dostoevsky's philosophy in a systematic presentation*. Moscow: Respublika Publishing House.
- Lossky N O (1994) Dostoevsky and his Christian worldview. In *God and world evil*, 6–249. Moscow: Respublika Publishing House.
- Lurie V M (1996) Dogmatics of the “religion of love”. Dogmatic representations of the late Dostoevsky. *Christianity and Russian literature*: 2nd book of collected articles, Ed. V. A. Kotelnikov, 290–309. St.Petersburg: Nauka Publishers.
- Nietzsche F (1990) *Works*: in 2 volumes. Moscow: Mysl Publishing House, Vol. 2.
- Nietzsche F. (2005–2014) *Complete Works*: in 13 volumes. Moscow: Cultural revolution.
- Nietzsche F (2005) *The Will to Power. The experience of re-evaluation of all values*. Moscow: Cultural revolution.
- Pomerants G S (1990) *Openness to the abyss: Meetings with Dostoevsky*. Moscow: ‘The Soviet Writer’ publishing house.
- Svasyan K A (1990) Comments. Nietzsche F. *Works*: in 2 volumes, vol.: 769–827; vol. 2: 770–809. Moscow: Mysl Publishing House.
- Stepun F A (1991) “The Demons” and the Bolshevik Revolution. In *Russian Diaspora in the Year of the Millennium of the Baptism of Rus*. Comp. by M. Nazarova, 365–376. Moscow: Stolitsa Publishing House.
- Heidegger M (1993) European nihilism. In: *Time and being: Articles and speeches*, 63–176. Moscow: Respublika Publishing House.
- Jaspers K (1994) *Nietzsche and Christianity*. Moscow: Moscow Philosophic–I Foundation; Medium Publishing House.
- Admirand P (2021) Theist–Atheist Encounters in Les Misérables, The Brothers Karamazov, and The Plague. *Religions*, 12:12. <https://doi.org/10.3390/rel12010012>
- Bittencourt R N (2011) Nietzsche e a idiotia divina de Jesus. *Kriterion* 123: 105–119.
- Feiler A F (2017) A ética cristã a partir de uma leitura de Hegel e Nietzsche. *Revista de Filosofia* 16(2): 25–35.
- Ghini G (2017) Imago Dei anthropology in Dostoevsky's The Brothers Karamazov. *Church, communication, and culture* 2(3): 206–219.
- Isupov K G (2011) Dostoevsky's Transcendental Esthetic. *Russian Studies in Philosophy* 50(3): 68–87.
- Kibalnik S A (2020) “If there is a God, then anything is permitted” (Dostoevsky's meta-theme in Jacques Lacan's psychoanalytic interpretation). *Studies in East European Thought* 72: 227–239.
- Márquez M D (2017) El Cristianismo de F. M. Dostoevski y su correspondencia con la experiencia del amor en Ignacio de Loyola. *Pensamiento* 73(276): 673–691.
- Medvedev A A (2018) The Elder Zosima as a Renovation of Orthodox Tradition (K.N. Leontiev and V.V. Rozanov's Polemic about the Novel by F. M. Dostoevsky “The Brothers Karamazov”). *Journal of Siberian Federal University. Humanities & Social Sciences* 3 (11): 385–395.
- Miller C A (1973) Nietzsche's Discovery of Dostoevsky. *Nietzsche-Studien: Intern. Jb. für die Nietzsche-Forschung* 2: 203–207.
- Paschoal A E (2016) A Figura de Jesus nos escritos de Nietzsche. *Rev. Filos., Aurora*, 28(44): 633–652.
- Sanz S G, Nzewi C (2017) Dostoevsky and the religious experience. An analysis of The Possessed. *Church, Communication and Culture* 2(3): 292–299.
- Vaskovic P (2020) A path to authenticity: Kierkegaard and Dostoevsky on existential transformation. *International Journal for Philosophy of Religion* 87: 81–108.

REVISITING HULL'S EVOLUTIONARY MODEL OF SCIENCE

Mohammad Mahdi Hatef

Iranian Institute of Philosophy, Science Studies Department
e-mails: Hatef.mehdi@gmail.com, m.hatef@irip.ir

Abstract:

Evolutionary models for scientific change are generally based on an analogy between scientific changes and biological evolution. Some dissimilarity cases, however, challenge this analogy. An issue discussed in this essay is that despite natural evolution, which is currently considered to be non-globally progressive, science is a phenomenon that we understand as globally progressive. David Hull's solution to this disanalogy is to trace the difference back to their environments, in which processes of natural selection and conceptual selection occur. I will provide two arguments against this solution, showing that Hull's formulation of natural selection prohibits him from removing the environment from the selection process. Then I point to a related tension in his theory, between realism and externalism in science, and give some suggestions to solve these tensions.

Key words: evolutionary model of science, scientific progress, adaptationism, realism, externalism.

Acknowledgements: I am grateful to Hassan Miandari and Hossein Sheykh Rezaee for their generous support of this study. I am also thankful to anonymous referees for their constructive comments. This research was funded by the Iranian National Science Foundation (INSF).

1. Introduction

Employing natural selection to account for scientific change is almost as old as the natural selection theory itself. Two years before Darwin published *The Origin of Species* (1859), Herbert Spen-

cer declared that the connection between scientific change and organic evolution is not merely an analogy, but a manifestation of a greater all-compassing general law of the universe (Spencer, 1857). This early recourse might be partly motivated by a will to integrate nature, society and knowledge in a unified image in the nineteenth century. However, this analogy has also been employed ever since to solve another problem, i.e., to articulate progress in science. Despite an old consensus on the possibility of progress in science, this issue has long been very difficult to articulate. While science's general image is clearly progressive, introducing a particular criterion in terms of which science has actually made progress is obscure. But, when a problem is hard to solve directly, we can try indirect ways, among which include using analogies and borrowing models from other fields to represent the problematic phenomena – and here is where the evolutionary analogy comes out. It is likely because of the utility of this strategy that Kuhn invoked this analogy in his 1969 postscript to *The Structure of Scientific Revolutions* in order to reject relativist interpretations of his work. More interestingly, Popper, who disapproved of the evolutionary theory for several decades, also used it in his later works as an effective model to represent scientific progress (Popper, 1972, 1981). Like any other analogy, everything begins with observing similarities between nature and science, including the entities, processes, and mechanisms working in both of them. Then the evolutionary model is applied to represent the target phenomena, namely, scientific progress.

There is a clear relationship between the analogy's explanatory power and the similarities found

between nature and science. The more similar the definitions and functions of the counterparts within these two domains, the more successfully the analogy works. So, there is a two-stage program, positive and negative, before us. Positively, we must build up similarities by finding proper counterparts in science for entities and processes operating in the natural world. Negatively, we must remove dissimilarities observed between these two. Various evolutionary models of science have taken different strategies to meet these ends. Among these models, David Hull's work has a special place for at least two reasons: He articulated a model in detail and addressed some problems, available but not considerably noticed in former efforts, deliberately. He also enriched his model with a detailed history of biology in his *science as a process*, to show its conformity with the actual history of science in advance.

2. Hull's evolutionary model of scientific change

Hull's idea, in his evolutionary model of scientific change, which can be regarded as 'an extension of Dawkins' suggestion of memic evolution' [Wilkins, 1998, p.482], is that there is a general notion of the selection process which instantiates at several levels; acting on organisms in the biological environment it amounts to natural selection, and acting on ideas in a scientific environment it amounts to conceptual selection. Based on this general notion, he proposed an analogy between the mechanisms working in biological and scientific worlds. Accordingly, the mechanisms underlying biological evolution and scientific development are the same and work with similar rules in their domains. To maximize this analogy's explanatory power, Hull attempted to introduce proper counterparts in science for the essential constituents in the natural selection process, such as replicator, interactor, lineage, adaptation, and fitness. However, like any other analogy, there will always remain some dissimilar elements within the two processes, which one may regard as adverse disanalogies, causing the analogy to fail. Interestingly, Hull anticipated critics by addressing these troublesome elements and attempted to dissolve them to save the analogy. Regarding this aim, his strategy was to show that either there is no real dissimilarity, or rather, that a recognized dissimilarity does not lead to the analogy's failure.

Among these dissimilarities are the counterparts suggested in science for biological notions of

replicator and interactor (Sterelny, 1994: 50; Rosenberg, 1992: 220; Oldroyd, 1990: 478), the fact that Conceptual selection in science seems to be more Lamarckian than Darwinian-Wisemannian (Hull, 1988: 452; Sterelny, 1994: 52), and that variation-making changes in biology are blind and directionless, while variation-making changes in science involve intentionality and goal-directedness (Hull, 1988: 455; Cohen, 1973). The next dissimilarity that poses a particular challenge is related to progress: Despite biological evolutionary processes, which are thought of as not globally progressive, science is often supposed to be a globally progressive enterprise (Ruse, 1986, 1989, 1995; Grantham, 1994). This issue, which is known as *the problem of global progress in the selection process*, is something we will focus on in this essay.

This dissimilarity is an immediate consequence of Hull's two separate claims: on the one hand, and following most contemporary biologists, he denied global progress in evolution and advocated only its local progress. On the other hand, he understood science as a progressive enterprise that aims to uncover 'stable regularities' in nature. As a realist, Hull believed that, beyond variable aspects of nature, there are some stable regularities, spatiotemporally unconstrained, understanding of which is science's goal. This dissimilarity brings about a problem: on the one hand, the analogy's functioning depends on removing the dissimilarity; on the other hand, the dissimilarity is an inevitable consequence of his different assumptions about the nature of evolution and science. This results in a dilemma: Hull must either solve it by altering his ideas about either side of the analogy, or dissolve it by demonstrating that there is no real disanalogy here. Hull went the second way, saving his original views, and attempted to solve the problem by relocating the origin of dissimilarity. According to him, the dissimilarity was not due to the selection processes themselves. Instead, while the selection process is identical in both cases, what leads to dissimilarity is that relevant environments, natural and conceptual, are different. As he stated:

If this difference were due to differences in the mechanisms involved in biological versus conceptual change, then it would count against any attempt to produce an analysis of selection processes, that is equally applicable to both sorts of change, but the genuine differences between rational selection and natural selection that I have discussed thus far do

not. Instead, the source of this difference lies in differences in the relevant "environment" of these two processes [Hull, 1988: 466].

According to Hull, while there is a general notion of the selection process, the difference between the two kinds of progress is related to differences in the environments in which the two processes operate. The crucial point in this suggestion seems that it requires a definite distinction between the selection process and the environment. In other words, the selection process must be abstract enough, so it can remain identical while working in two different environments. Therefore, his solution seems to involve two steps: Separating the environment from the selection process and lying the difference in the environment. I doubt the possibility of the first step and will argue against that in the next section, demonstrating that the selection process cannot legitimately be separated from the environment.

3. Hull's solution revisited

There seem to be two difficulties in separating the selection process and the environment, when these terms are understood in the same way as Hull articulated in his evolutionary account:

1. Hull formulates natural selection using two concepts as its basic entities, replicator and interactor, which are defined as follows (Hull, 1988: 409):

- A replicator is an entity that passes on its structure largely intact in successive replicators
- An interactor is an entity that interacts as a cohesive whole with its environment in such a way that this interaction causes replication to be differential

Accordingly, the selection process is defined as a process in which differential extinction and proliferation of interactors cause differential perpetuation of the relevant replicators. Then, replicator and interactor seem to be defined functionally, since they are defined in terms of the role these entities play through natural selection (Brandon, 1998: 179). As Cain and Darden said, this focus on functional entities, rather than structural ones (e.g., in Lewontin's formulation), which involves an ontological displacement from 'what things are' to 'what they do' in the selection process, is an advantage for Hull's model (Cain & Darden, 1988: 168). This functional character is what enables him to identify these entities with any biological or mental entity, from insulin to ideas existing in mind (Rosenberg, 1992: 220).

It also underlines the concept of interaction between interactors, and since the interaction is an operation proposed to run in an environment, such a definition makes the environment conceptually tied with the interactor. Now the argument goes: Since a difference in the environment can influence the interaction, and this, in turn, can influence the interactor, and since this permeates the selection process itself, any difference in the environment can influence the selection process.

Moreover, Hull's fourth element regarding the selection process, i.e., the lineage, is formed through interaction with the environment. Insofar as selection is made according to lineages, and these entities are formed by interaction with the environment, the selection cannot be detached from the environment. Hence, again, the environment cannot be legitimately eliminated from the selection process.

2. Another notion which underlines the environment in the selection process is *adaptation*. The relation involved in this notion contains two relata: interactor and environment. Due to the adaptation's dependence on the context, both its definition and characterization rely on the environment. In fact, the adaptation depends on two classes: the interactor's characters and formed environmental elements. In Dawkins' terms, from whom Hull borrowed his formulation of the selection process, adaptation is obtained when these two classes match each other to some extent (Dawkins, 1988: 21). However, it is unlikely that all constituents of the environment have an equal part in the adaptation. For example, although the color of the keyboards' keys, by which this text is typed, is part of the environment, it does not considerably affect the typist's life. But the energy radiating from the monitor is much more significant. We can give a functional definition of the environment, as well; everything leading to a quantitative or qualitative adaptive effect is a part of the environment. By this definition, the environment turns out to be an ecological environment, and elements related to adaptation will be more distinguished (Cain & Darden, 1988: 167). For example, in the case of moths living in industrially polluted regions, the darkening of trees leads to an essential adaptation: over time, moths have developed darker coloration that allows them to camouflage themselves. Also note that their predators are part of their environment, and the fact that they are visually cued, rather than cued by, for example, odor, is a crucial factor that allows for the moths' color to be adaptive (ibid).

Hull's failure to exclude the environment from the selection process is rooted in two ideas involved in his approach to evolutionary theory: an externalist approach to organisms and an adaptationist approach to evolution. In an externalist approach to organisms, in Dawkins' terms, there is always "information from the environment inside the organism" (Dawkins, *ibid*). In contrast, an internalist approach explains an organic system's properties without appealing to characteristics of the external environment. On the other hand, in an adaptationist approach, fitness depends mostly on how close an organism is to an optimal state, while an optimal state cannot be characterized independently of the environment. Therefore, for his solution to be successful, Hull must leave these two ideas and replace his evolutionary theory with a theory of neutrality, in which changes in organisms are explained merely by random drift. Note that leaving one of them intact not only does not help but also makes his theory incoherent. For example, an internalist approach to organisms is not compatible with an adaptationist approach to evolution (Godfrey-Smith, 1996: 32).

Eliminating the environment from the selection process is even more difficult if we consider the phenomenon of co-evolution as a form of organic evolution. Co-evolution, which is used to explain co-adaptation, is a form of evolution through which the organism and the environment interact to maximize overall adaptation. This process ultimately leads to the evolution of the environment itself following the evolution of the organism (Griffiths, 2000: 305; Ridley, 2004: 613-16), in which organisms form their environment by altering the dimensions of the environment which shape their, or other organisms', niche (Odling-Smee, 1966: 641-48; Godfrey-Smith, 1996: 141). This complex form of adaptation further complicates the separation of the environment from processes of adaptation.

Consequently, it can be argued that the difference between the two environments inevitably leads to differences in the two processes. Suppose one environment enjoys a stable situation as the goal of the process, and the other lacks such a goal. In that case, since the environment is present within the process, this difference affects the processes themselves as well.

Now we are in a crucial situation; on the one hand, as Hull acknowledged, this dissimilarity cannot be tolerated. On the other hand, Hull seems unsuccessful in solving it. My suggestion in this essay is to reconsider the other horn of the dilemma, Hull's

views on progress in evolution and science, to escape the dilemma. To lend support to Hull's view on progress in evolution, I will show that Hull's understanding of scientific progress may be corrected to resolve the disanalogy. Let us reconsider progress in these two enterprises in more detail.

4. Challenging the dilemma of progress

Historically, the notion of evolution was related to the eighteenth-century idea of progress, such that the arguments for evolution in nature enjoyed the support of a progressive conception of society and history, aiming to give a unified perception of the world by subsuming nature under the law of progress. Then, from Lamarck and Erasmus Darwin to E. O. Wilson, the idea of progress enjoyed prominent advocates who saw the *scala naturae* in an anthropomorphic way and took a whiggish viewpoint in natural historiography, according to Ruse (1995). Darwin's view on this problem is, however, controversial. Fisher (1986) and Futuyma (2005) believe that what distinguished Darwin from his contemporaries was that he denied the necessary association between evolution and progress. Ruse (1995) views Darwin's attitude as anthropomorphic and progressive, and Hull (1988) found an ambivalence in his works. According to Hull, in his technical writings, Darwin didn't understand evolution necessarily as progressive, while, in his public works, he pointed to evolution as something progressive.

However, due to difficulties regarding the notions in terms of which progress in evolution can be objectively formulated, its advocates have considerably diminished in recent decades. According to Futuyma, 'it is difficult, if not impossible, to specify a universal criterion by which to measure improvement that is not laden with our human-centered value' [Futuyma, 2005: 519]. The global progress thesis, in general, requires three essential conditions (Ruse, 1995:110):

- The variance of the frequent units of evolution
- Availability of a stable target outside the evolution process (such that the relation and distance of varying steps to the target can be specified)
- An overall directed movement toward that target (that is, that both trend and tendency, in Rosenberg and McShea's terms, are directed toward one target (Rosenberg & McShea, 2008: 147-51))

As candidates in terms of which progress can be formulated, Hull pointed to several items like the

number of organisms, overall biomass, number of species, complexity and adaptability of organisms, variety of structural plans exhibited, amount of energy transduced, efficiency of energy conversion, and number of adaptive zones and/or niches occupied. However, he found none of them a proper indicator for global progress (Hull, 1988: 460-1). Indeed, biological evolution cannot satisfy the second and third conditions. Moreover, the theory of natural selection *per se* does not imply global progress. What is implied, at most, is the local progress of species' movements in the spatiotemporal periods, constrained by their environment. It is progressive because it contains an assumption concerning an optimal state of adaptation for each organism or population, and it is also confirmed against its rival, i.e., the theory of random drift, which implies that the organism or population moves gradually toward that optimal state (Sober, 2008: 193). But frequent changes in an environment preclude the stability necessary for a general optimal state, and make the optimal state necessarily a local state, which consequently localizes progress in evolution. Therefore, local progress in evolution, a theory that Hull (*ibid.*: 466) admittedly supported, would be difficult to refute.

Some of Hull's predecessors who employed the evolutionary metaphor for science, e.g., Popper and Kuhn, nonetheless, regarded biological evolution as progressive (Popper, 1996; Kuhn, 1970: 206, 2000: 160), and it was precisely on this point that Hull disagreed with them (Hull, 1988: 464). He agreed with Popper, however, on another issue, i.e., the theoretical global progress in science. To Hull, the stable goal of science is awareness of those aspects of the world that remain constant, something which he called 'spatiotemporally unrestricted regularities in nature' [*ibid.*: 467]. The successive approximations of scientific theories to these eternal, immutable regularities are responsible for global progress in science. For Hull, these regularities provide possibility conditions of nature as an object of nomological, scientific study, such that denying them makes both nature (as opposed to a chaotic universe) and science impossible.

Setting aside the controversial status of laws of nature in biology, Hull's claim about the goal-directedness of science and especially biology suffers from difficulties that arise from his view of science. Indeed, there are some internal tensions in Hull's understanding of scientific change – an understanding which seems to be supported by his history

of biology – and these tensions prohibit him from preserving the idea of global progress in science. In what follows, I argue that there is no sign of goal-directedness in the historical trajectory of biology depicted by Hull. This argument undermines the evidential basis for his claim. Then, by referring to his externalist conception of science, I will challenge the legitimacy of his claim about the goal-directedness of science.

1. In his account of the controversy between idealist morphologists and evolutionary systematists in the nineteenth century, Hull demonstrated that the controversy, rather than being about specific theories, was about the nature of genuine science and the notion of explanation. On the one hand, pattern-cladists such as Owen aimed to discover a proper pattern for description and classification of creatures, and virtually dealt with prototypes of groups of creatures. On the other hand, phylogenetic-cladists and systematists, including Darwin, found discovering patterns unfruitful and believed that understanding the mystery of life depends on understanding the process causing such a pattern (Hull, 1988: 44,73). In their attempts at classification, morphologists followed an inductive method based on the morphology of creatures, while evolutionarists appealed to the evolutionarily common ancestors. We should not be surprised by these kinds of cases in the history of science (for another example, see Grantham (1994)), if we take L. Laudan's (1978) claim seriously, according to which an essential aspect of change in the history of science is related to conceptual problems, which are associated with questions of methodology and the metaphysics of science, rather than empirical problems. Accordingly, a history of scientific change is not only a history of changing theories but also a history of changes in methodology, goals of science, evaluative standards, and the nature of proper science. In fact, not all scientific controversies are about better explanans for permanent phenomena, but a considerable part of them is about the formulation of explanandum in a discipline (Griffiths, 2000). As a result, the view that science has a unique goal and direction is supported, in a concrete way, by neither the history of science in general nor Hull's historical account of biology in particular.

In opposition to Hull's conservative approach, known as 'lock and key' (*ibid.*), other evolutionary epistemologists suggested an alternative approach to the practice of science. In this approach, while interactors in science attempt to meet their ends by

improving their adaptation, they alter their targets simultaneously to make them more adapted. Returning to Hull's terms, unconstrained regularities that science tends to uncover, rather than being stable, change to fit our ideas of science and nature further. To put it in biological terms, as creatures alter the environment to form their niches, scientists alter the nature they encounter to fit their scientific image.

2. Hull (1988) defined fitness of conceptual replicators or ideas *prima facie* in terms of their scope in a scientific society, i.e., the idea's domain of applicability by other scientists, its impact on other works, and the proliferation of references to it. Since he viewed scientific theories as historical entities, parallel to biological lineages, he employed the notion of *conceptual inclusive fitness* to reveal the part that scientists' cooperative activities play in adjudicating the fitness of ideas. Consequently, the fitness of an idea is determined by the extent to which it influences other scientists or scientific works. The environment of the selection process here is likely the scientists' social world, as the corresponding adaptation is determined in relation to the social environment. As Hull's externalist approach requires, features influencing an idea's reception by scientists go beyond epistemological features and include professional interests, discipline relations, education, and their genealogical associations. In fact, groups of scientists form demes that seek to extend the impact domain for their ideas, while borrowing and assimilating from each other, and consequently enjoy more credit in their conceptual environment. This is Hull's externalist view of science, according to which working virtues in science are not restricted to epistemic virtues.

To retain his connection with reality, Hull, as a realist, also included another part in his array of fitness: evidence; to be distinguished from social constructivists. However, since evidence is only one among many factors capable of influence, it does not necessarily take precedence. Other factors, which might be non-epistemic, may be more powerful than the evidence supporting a theory. Be that as it may, there is no guarantee that the spread of an idea results simply from its epistemic virtues. As Ruse argued, 'at least in theory it is possible that a simply false idea (judged against the best evidence) gets accepted and a true idea rejected, simply because the false idea is promoted by the scientist with the superior political skills' (Ruse, 1995: 125). Hence, this mechanism cannot secure the required move toward the ultimate goal of science i.e., progress in deter-

mining the regularities of nature. Again, this problem is rooted in committing simultaneously to two incompatible positions, i.e., specific externalism on science and scientific realism. Indeed, due to these dual commitments, Hull's position seems to waver between two opposite evolutionary images of science; it is Popperian, considering its commitment to theoretical progress in science, and it is Kuhnian, and even beyond, in considering how non-epistemic elements contribute to scientific change. A question which now arises is that, for the sake of coherence, which commitment is more reasonable to abandon?

Abandoning externalism, in one sense, is equivalent to leaving the evolutionary model, since it is not clear what an internalist evolutionary model of science would be like. Considering the close relationship between the environment and the process, especially in an adaptationist approach, and the variety of factors affecting the practice of science, including the scientific community's professional and social interests, taking an internalist approach to scientific enterprise is highly disputable. Moreover, a history of biology, as it is presented in *science as a process*, which provided the evidential basis for Hull's philosophy of science, is an exemplary externalist history of science. Hence, the price for abandoning externalism is too high, not only for Hull but also for every evolutionary model of science. This philosophical preference is not, however, free of methodological commitments. When the part taken by non-epistemic virtues is taken into account, the influence of epistemic virtues on one's choice of theory and the direction of scientific change will diminish, while the significance of pragmatic virtues in scientific change will be underlined. This influences, in turn, how progress should be understood in an evolutionary model of science, such that theoretical progress, as related to realism, is deemed subordinate to pragmatic progress. In this respect, science progresses not through the discovery of natural laws that bring clarity and unity to natural phenomena, but through the postulation of solutions to pragmatic problems that achieve practical success. This pragmatic progress is presumably in harmony with various views on the practice of science, from saving the phenomena as offered by instrumentalists, to Kuhn's puzzle-solving or Laudan's problem-solving. Then the evolutionary approach to science is not positively restrictive; it doesn't imply any particular approach to the practice of science. Nonetheless, it is negatively restrictive, as it rejects realism and any

purely theoretical approach to the practice and progress of science.

However, paying attention to non-epistemic virtues that contribute to scientific change provides us with a clue to solve the earlier tension within the analogy between the globality and locality of science. Pragmatic virtues, opposed to epistemic virtues, are hardly global by their nature, as they are usually discipline-related, history-related, and society-related. Thus, theoretical changes which were directed by these virtues can scarcely be understood as globally progressive. In fact, concentrating on Hull's externalism in science, philosophically and historically, motivates us to modify his idea of scientific progress in two respects: altering its nature from theoretical to pragmatic and narrowing its domain from global to local.

5. Conclusion

Applying evolutionary models to science requires a special understanding of both science and evolution, metaphysically and methodologically, the disregarding of which brings about tensions in an evolutionary account of science. Scientific change, in such an account, cannot be understood as globally progressive, while evolution is simultaneously believed to advance locally. In fact, it is implausible to talk about the same mechanisms operating in evolution and science, while having different beliefs about how progress occurs in each case. Moreover, Hull's externalist understanding of science prohibits him from believing in its theoretical progress. There are two ways to resolve this latter tension: either to give up externalism about scientific change, or to abandon theoretical progress in science. As I argued, choosing the first way undermines the comprehensive character of his project, respecting his historical account of biology. Therefore, my suggestion is to dilute the idea of progress in such a way as to make it compatible with his externalism. This new idea would allow for some kind of pragmatic progress. Regarding the first dilemma, i.e., that of global and local progress, however, as far as some global, either theoretical or pragmatic, idea of progress is retained, the first tension arises again. However, if the context-dependent character of pragmatic virtues is taken into account, the problem will be automatically solved and scientific progress restricted to some local domain or some particular respect.

References

- Brandon R N (1998) 'The Levels of Selection: A Hierarchy of Interactors', in Hull, D. and Ruse, M., *The Philosophy of Biology*, Oxford.
- Cain J A, Darden L (1988) Hull and Selection. *Biology and Philosophy* 3:165–181.
- Cohen L J (1973) Is the Progress of Science Evolutionary? *Brit. J. Phil. Sci.* 24: 41–61.
- Dawkins R (1998) Universal Darwinism. In *The Philosophy of Biology*, eds D. Hull D. and M. Ruse. Oxford: Oxford University Press.
- Futuyma D (2005) *Evolution*, Sinauer Associates Inc.
- Godfrey-Smith P (1996) *Complexity and the Function of Mind in Nature*. Cambridge: Cambridge University Press.
- Grantham T (1994) Does Science have a Global Goal?: A Critique of Hull's View of Conceptual Progress. *Biology and Philosophy* 9: 85–97.
- Griffiths P (2000) David Hull's Natural Philosophy of Science. *Biology and Philosophy*, 15: 301–310.
- Hull D L (1980) Individuality and Selection, *Annual Review of Ecology and Systematics* 11, 311–332.
- Hull D L (1988) *Science as a Process*, Chicago: University of Chicago Press.
- Kuhn T (1970) *Structure of Scientific Revolutions*, Chicago: University of Chicago Press.
- Hull D L (2000) *The Roads since Structure*. Chicago: University of Chicago Press.
- Laudan, L. (1978) *Progress and its problems*. University of California Press.
- Odling-Smee F J, Laland K N, Feldman F W (1996) Niche Construction. *American Naturalist*, 147: 641–648.
- Oldroyd D (1990) David Hull's Evolutionary Model for the Progress and process of Science. *Biology and Philosophy* 5:473–487.
- Popper K (1996) The rationality of scientific revolutions. in *The myth of framework*. Routledge.
- Popper K (1972) *Objective Knowledge*. Oxford: Clarendon Press.
- Popper K (1981) The Rationality of Scientific Revolutions. In *Scientific Revolutions*, ed. Ian Hacking. Oxford: Oxford University Press.
- Ridley M. (2004), *Evolution*, Blackwell.
- Rosenberg A (1992), Selection and Science: Critical Notice of David Hull's Science as a Process. *Biology and Philosophy*, 7:217–228.

Rosenberg A and McShea D W (2008) *Philosophy of Biology*, Routledge.

Ruse M (1986) *Taking Darwin Seriously*. Oxford: Basil Blackwell.

Ruse M (1989) The View from Somewhere: A Critical Defense of Evolutionary Epistemology. In *Issues in Evolutionary Epistemology*, eds K. Hahweg and C. Hooker, 185–228. New York: State University of New York Press.

Ruse M (1995) *Evolutionary Naturalism*, London: Routledge.

Sterelny K (1994) Science and Selection. *Biology and Philosophy*, 9: 45–62.

Spencer H (1857) Progress: Its laws and Cause'. *Westminster Review* 67:244–267.

Sober E (2008) *Evidence and Evolution*. Cambridge: Cambridge University Press.

POLITICAL ONTOLOGY OF ALAIN BADIOU AND SYLVAIN LAZARUS: TURNING TO THE SUBJECTIVE PRINCIPLES OF POLITICS

Irina Zhurbina

Department of History, Theory and Practice of Social Communications,
Institute of Social Communication, Udmurt State University, Izhevsk, Russia
e-mail: soloveyIV1@mail.ru

Abstract:

The article reviews the concepts of the French anthropologist and political theorist Sylvain Lazarus and the philosopher Alain Badiou, who suggest a new perspective on the subjective foundations of politics as thought. The focus on the subjective foundations of politics can be explained by the initial ambiguity in the works of the French theorists, who interpret the activities of the intellectual activist in different ways. The paper shows that Sylvain Lazarus is more concerned with the intellectual activity of political activists, whom he categorizes as political activists and politicians by the degree of intellectual activity. It was concluded that, according to Lazarus, politicians occupy a priority position. They are presented as professional lone thinkers with revolutionary consciousness, which allows them to think politics from the perspective of a probable revolution. In this regard, the politics, according to Lazarus, is a politics of revolutionary action. It was found that in Alain Badiou's theory the semantic emphasis is on the participation of intellectuals in politics. Based on Plato's thought on the development of a philosopher, Badiou formulates the idea of an exemplary subject of politics. The exemplary subject of politics is a philosopher-mathematician who is good at mathematical logic.

Key words: Politics Interiority, Metapolitics, Political Militant, subject of politics, intellectual activist, political actor.

In modern social philosophy, researchers link the emergence of a new ontology of politics with the

concepts of a politics in interiority by Sylvain Lazarus and metapolitics by Alain Badiou (Bosteels, 2011; Srnicek, 2008; Toscano, 2006; Žižek, 2009). The need to build a new ontology of politics on absolutely different foundations is associated with the crisis of Marxism as the theory and practice of public administration on the one hand and the tendency for neoliberalism to depoliticize the life of society and replace politics with state management technologies on the other hand (Žižek, 2008, p. 40). The situation of the crisis proves that politics can no longer be correlated only with the state as an objective basis for politics. The exhaustion of the objective foundations of politics in the form of a state allows the French theorists to argue that politics must be viewed based in itself as a reality capable of thinking about itself. Thus, they are making a radical turn towards the subjective principle of politics, which opens with thinking. From the perspective of subjectivity, politics is interpreted as thought. Politics as thought is possible since people have both the ability and the internal need to think; "people think gives an original meaning to the statement: 'politics is of the order of thought'" (Lazarus, 2016, p. 109-110).

The new ontology of politics effects the transition from the Aristotelian concept of man as a political being endowed with speech (Aristotle, 1885, p. 4) to people as beings capable of thinking politically. The turn of the French theorists to the category of "people" can be explained by the fact that the modern concept of politics cannot be described based on a person understood by modern philosophy mainly as an individual living in an individualized society where the priority is given to the private sphere of life

(Bauman, 2001). Hannah Arendt was one of the first to write that it is necessary to talk about people in the plural in order to implement politics as thought (Arendt, 1982, p. 40). The French theorists refer to “people” as a category that has no other meanings except marking the plurality.

Based on the principle of plurality, the French theorists build their own theory of political thinking, which differs both from the philosophical interpretations of human thinking and people’s thinking which functions at the level of common sense. This is due to the fact that human thought is viewed in philosophy as a process of contemplative activity that has no ultimate goal other than itself (Aristotle, 1893), or a stream of intellectual intuition (Schelling, 1978), or a process of rational doubt (Descartes, 1998). As a result, thinking is limited exclusively to the circle of intellectuals – philosophers, scientists and writers – who, by virtue of their professional activities, have chosen *bios theōrētikos* as a way of life. On the other hand, the natural ability of people to think, usually called common sense, abolishes the principle of plurality. By virtue of its obviousness, common sense sets a single order of thought, within the framework of which any difference is perceived as a self-evident given that does not require any further reflection. Alain Badiou writes:

“With ‘common sense’ one provides a norm that in actual fact is transcendent, because it suggests, not only plurality, but a subjective unity of this plurality, at least in principle. This concession to the One undoes the radicality of the multiple, which had allegedly been guaranteed. It opens the way for a doctrine of consensus” (Badiou, 2005, p. 18).

According to the French theorists, thinking is understood extremely broadly as an intellectual ability that many people possess. This means that, by its nature, thinking is multiple, but at the same time the intellectual ability of each individual person remains unique. Antonio Calcagno writes:

“There is a multiplicity of intellectualities. This derives from the fact that there are people who think. *Les gens pensent*. There are many people who think, and intellectualities thought by many people can be inherently multiple. Each of these intellectualities, though, is singular” (Calcagno, 2007, p. 14).

The theory of subjectivity, built on the principle of plurality, allows one to overcome traditional metaphysical ideas about both the Cartesian subject (Ego),

which is the center of thought, and the absolute subject of Hegel’s philosophy.

Politics as a thought appears when people begin to ask political questions as such. The questions “What to do in this situation?” and/or “What can we do ourselves?” become extremely revolutionary for people. The questions turn people to themselves as political activists capable of political thinking. The degree of activity of political thought determines the degree of political involvement of people in various kinds of meetings and organizations. Organized activity of people performs the function of an immanent boundary not only separating politics from non-politics, but also allowing politics to be considered as an organized process initiated by the intellectual activity of people.

The subjective basis of politics as thoughts are activists who are acting participants in the political process: “the point from which a politics can be thought – which permits, even after the event, the seizure of its truth – is that of its actors, and not its spectators” (Badiou, 2005, p. 23). An activist is the subject of modern politics. What makes the existence of the subject of politics peculiar is the fact that the subject exists in two realities at a time: an intellectual reality determined by thinking and the reality of politics as a sphere of action. Members of political organizations are intellectual activists who operate in politics and think about politics. “The central subjective figure of politics is the political militant” (Badiou, 2005, p. 122).

The ambivalent position held by intellectual activists in politics opens up different ways of interpreting their activities. The identification of various interpretations of the activities of intellectual activists allows the use of a philosophical and hermeneutic research method. The philosophical-hermeneutic method, on the one hand, allows us to show that in Lazarus’s theory, the subject of politics is interpreted through the intellectual activities of political activists, while Alain Badiou makes the semantic emphasis on the involvement of intellectuals in politics. On the other hand, it is philosophical hermeneutics that opens up the opportunity to fully reveal the ontological aspect of politics as thought, since it is based on the initial identity of being, thinking and language (Heidegger, 1968).

1. Intellectual activity of political activists in Lazarus’s theory

In his studies, Sylvain Lazarus starts from the situation of the crisis of Marxism, the onset of which allows him to speak of modernity as a “post-class”

stage, at which “classes”, “masses” and “parties” turn out to be absolutely inactive and meaningless concepts that do not reflect the modern reality of politics. In a “post-” situation, politics should be thought of based on itself as politics in interiority existing independently of the state and not aiming at achieving state power. “The conception of politics which rejects the one given in terms of party and State (that is to say, giving itself the State for object) will be called *politics in interiority*” (Lazarus, 2016, p. 108).

The intellectual activity of participants in a political process or organization is carried out in the form of militancy practice, which makes it possible to think of politics based on itself without the help of other doctrines: philosophy, history, economics, sociology, etc. According to militancy practice, politics is thought by the same people who make it. Using the concept of “activism”, Lazarus presents his own interpretation of political practice as the activity of political actors directly involved in self-organizing policy projects.

Based on the statement “people think”, Lazarus places semantic emphasis on the intellectual activity of political activists. The activity of thinking of intellectual activists manifests itself in their ability to think an action. Lazarus implicitly divides the subjective principle of politics as thought according to the degree of activity of the thinking of the participants in the political process. We can say that Lazarus’s politics in interiority appears as a *prescriptive thought of political activists* on the one hand and as an *analytical thought of politicians* on the other hand.

The prescriptive thought of activists is actualized situationally, depending on the conditions of their social existence, which initiate the feeling of discontent. In modern social philosophy, social discontent is read through Descartes’ thesis “ego cogito, ergo sum” (Descartes, 1998). However, the semantic emphasis in this case is on existence that initiates thinking: “I express dissatisfaction, therefore I exist” / “We discuss our complaints, therefore we exist” (Walzer, 1988). According to Lazarus, the semantic emphasis is on thinking, which is initiated by a feeling of social discontent. The Cartesian thesis “ego cogito, ergo sum” is transformed into the statement “people express dissatisfaction, therefore they think.” However, the activity of political thought is determined by the conditions of social existence (e.g., legal, economic) of people, which form an awareness of the need for changes. In his reasoning, Sylvain Lazarus follows Karl Marx, who wrote, “It is not the consciousness of men that determines their being, but, on the contrary,

their social being that determines their consciousness” (Marx, 1951, p. 329).

The essence of prescriptive thought is that, starting from what is in reality, it speaks about what can be done. In the prescriptive thought of activists, the action is presented in the perspective of the possible. Each statement is a single thesis that expresses people’s private attitude to the opening opportunity. The boundaries of the possible are set by a multitude of statements, each of which sets its own perspective of action. According to Lazarus, possibility (what may be) is a category of subjectivity that expresses the thought of action (Lazarus, 2001, p. 395).

Sylvain Lazarus directly connects the possibility of the emergence of modern politics with the activity of political thought of immigrant workers who are dissatisfied with the conditions of their legal status. A lot of statements by worker-activists of foreign origin set the situation in a state of conflict: “at the factory they call me a worker, on the street they call me an immigrant because they forgot that I am a worker”; “I am an immigrant and a worker. Inside the factory, we are mistaken for immigrant workers, and outside the factory, they mistake us for immigrants. French people forget that we are here to work and that we work”; “I am a worker who immigrated to France, but as an immigrant I work like all workers” (Lazarus, 2001, p. 394). When used by activists, the word “immigrant” is replaced by the problematic word “worker”, which opens up the possibility for the subsequent organized action of workers of foreign origin.

The prescriptive statements of the activists make the situation conceivable, and therefore accessible for further interpretation. Interpreting the meaning of the prescriptive utterances of political activists is an analytical process. While the prescriptive thought of activists is local in nature, the analytical thought of politicians is universal. The analytical thought of politicians moves in the “opposite” direction – it proceeds from “what may be”, i.e. from the new opportunity, and turns to “what is there.” In the analytic thought of the politicians, the prescriptive thought of the activists is tested for intelligence, that is, the ability to be implemented through action. The analytical thought of politicians is the thought of what needs to be done here and now.

Sylvain Lazarus substantiates the doctrine of the mode existence of politics. Each mode – revolutionary, classist, Bolshevik or dialectical – points to politics-there, produced by the analytical thought of politicians. Lazarus writes, “Politics in interiority does not appear and does not exist except in the form of a mode” (Lazarus, 2016, p. 112). The analytical

thought that produces politics-there belongs to lone figures, such as Saint-Just, Karl Marx, Lenin and Mao Zedong, according to Lazarus. At the same time, in order to engage in politics, politicians do not have to live badly themselves, that is, their participation in politics is not determined by the social, legal or economic conditions of their existence. Karl Marx's idea of the relationship between social being and consciousness is applicable to politicians in an "inverted" form: "it is not the social being of people that determines their consciousness, but, on the contrary, their consciousness determines their social and political being."

The analytical thought inventing politics-there achieves its full actualization in the works of political figures. In this case, Lazarus rejects concepts of the doctrine that refer to the objectivity of the state. Each mode of politics corresponds to a work of a theoretician who describes the ideas of politics-there in the categories that are characteristic of the given mode of politics. It is possible to understand the ideas of politics-there only on the basis of categories that have a local character, since their appearance is initiated by a specific situation.

The revolutionary mode of politics (1792–1794) is defined by the text of the Jacobin Constitution (1793), in which "virtue" is a prescriptive category that allows one to think of the possibility of revolutionary politics in terms of "depravity", "government", "power", "terror", etc. The classist mode begins with Karl Marx's ideas expressed in the *Communist Manifesto* (1848). The prescriptive category of the classist mode of politics is "history", which allows one to think of the possibility of establishing communism in terms of class struggle and class antagonism. "For Marx, history is above all a category of politics, that is to say, what I call a prescriptive category (I remind you that the prescriptive is what opens a specific possible)" (Lazarus, 2016, p. 120). The Bolshevik mode of politics starts with Lenin's *What is to be Done?* (1902). In Lenin's work, the prescriptive category is the "party", which sets policy in the categories of "professional revolutionaries", "proletariat", "class consciousness" and so on. The dialectical mode of politics (1928–1950) starts with the programmatic work of Mao Zedong *Why is it that Red Political Power can Exist in China?* (Lazarus, 2016, p. 123). In this case, the prescriptive category is "law", which allows thinking about politics in a situation of "revolutionary war", "people's war", "army", etc.

Lazarus has long associated the modern mode of politics with the activities he created together with

Alain Badiou and Natacha Michel – *Organization politique* (1984). Therefore, the article by Sylvain Lazarus *"Can Politics be Thought in Interiority?"* can be considered as a draft of a new policy, representing his main idea as an activist and politician, declaring the basic principle of the organization: "politics are the people who make it." The prescriptive category is "organization without a party", which makes it possible to think of contemporary politics in the following categories: "person without documents", "worker without documents", "worker immigrant", "the excluded", etc. Currently, the concept of Sylvain Lazarus is actively used as a methodological paradigm that allows thinking of modern politics in terms of subjectivity (Moraes, 2019, 2020; Neocosmos, 2009, 2012).

The mode principle of the existence of politics proves the absence of a single concept of politics. Sylvain Lazarus believes that politics should be considered only in the plural (*une politique*), which do not form a conceptual unity, but exist as some individual and unique events. The singular nature of modern politics suggests that politics lacks a unified history. Every politics-there (*politique-là*) refers to itself as a thought produced by politicians in a certain place and at a certain time.

It is necessary to pay attention to the fact that the common meaning connecting the various modes of politics is the word "revolution", which, according to Žižek, performs the function of the dominant signified or "*point de capiton*" linking together all works (Žižek, 1989, p. 95). The common thing is that many politicians understand the meaning of the "policy of revolutionary action" as the only possible practice in different ways. The degree of actualization of the analytical thought of politicians is determined by its ability to be the thought of revolution as an ultimate political action that changes the existing social order – the Great French Revolution (1789–1794), the Revolution in France (1848–1851), the October Revolution in Russia (1917), People's Revolution in China (1948). In a revolution, the thought of politicians organizes the actions of a faceless crowd of political activists – revolutionaries (Saint-Just), communists (Marx), party members (Lenin), popular masses (Mao), émigré workers (Lazarus). The revolution itself becomes a way of materializing the thoughts of politicians. In this case it is not "people's thought organizes action" as Lazarus believes (Lazarus, 2016, p. 114), but the thought of actors organizes people's action. At the same time, it should be noted that the presence of a guiding thought deprives political activists of the need to show their own intellectual activity.

In this regard, the simple name of a politician, which should retain the status of an unnamed (Lazarus, 1996, p. 81), is ultimately referred to by the name of a politician – Saint-Just, Marx, Lenin, Mao – by Sylvain Lazarus (Lazarus, 2016, p. 112). The name of every politician is a symbolic expression of the policy of revolutionary action. The completion of a certain mode of politics, initiated by a specific thought of a politician, leads to the objectification of a proper name, which is replaced by a common noun. In political theory, a common name is used to denote certain revolutionary doctrines: Marxism, Leninism, Maoism, etc. The common name indicates the boundary, or limit, of the representation of politics-there caused by the thought of a certain politician.

Thus, Sylvain Lazarus limits politics as thought to its ability to be a thought of revolution. The subjective principle of politics as thought proclaimed by Lazarus actually turns into extreme objectivism, that is, a project of a policy of revolutionary action directed and guided by politicians who perform the function of a political subject. By themselves, politicians are professional thinkers with revolutionary consciousness, which allows them to think of politics in the perspective of a possible revolution, rather than professional revolutionaries.

2. Political activity of intellectuals in Alain Badiou's theory

In Alain Badiou's theory, the question "Is it possible to think of politics as thought?" is actually a philosophical question that allows one to form an idea of the metapolitical position of philosophy (Badiou, 2018). At the same time, Badiou's theory differs from Zaibert's theory of metapolitics, which focused on the question of "What is the state?" (Zaibert, 2004) or the theory of Viereck, who uses the concept of "metapolitics" to study the origins of the Nazi mind (Viereck, 2003). Using the prefix "meta-", Badiou shows the need to reconsider the traditional relationship between philosophy and politics. According to Badiou, philosophical thought appears after (meta-) certain political events which, without the "intervention" of philosophy, will exist in a state of "vague non-existence" unaffected by thinking. As part of the metapolitics theory, Badiou considers the political principles of philosophy, and politics is presented as a non-philosophical condition of philosophy itself.

"Accordingly, what is at stake here is what I name *metapolitics*, or what, in philosophy, carries a trace of a political condition which is neither an object nor what requires production in

thought, but only a contemporaneity that produces philosophical effects" (Badiou, 2005, p. 55).

The need to reconsider the relationship between philosophy and politics arises primarily due to the fact that philosophy gives priority to thinking, which is detached from the sphere of political practice. This is initially determined by the specifics of philosophy itself, which implies that to be is the same as to think. In philosophy, thinking reaches its highest point in contemplative activity (*βίος θεωρητικός*), which has no external goal, but exists "for its own sake." Even Aristotle wrote that "the exercise of the reason seems to be superior in seriousness (since it contemplates truth), and to aim at no end beside itself, and to have its proper pleasure" (Aristotle, Peters, 1893, p. 340). A philosopher separated from political life can take either the position of a transcendental observer analyzing the empiricism of politics, or a theorizing position focused on defining the principles of "good politics."

"The philosopher would then have the triple advantage of being, first, the analyst and thinker of this brutal and confused objectivity which constitutes the empirical character of real instances of politics; second, the one who determines the principles of the good politics, of politics conforming to ethical demands; and, third, in order to meet these demands, the one exempt from militant involvement in any genuine political process. Whence the philosopher could keep the Real at arm's length indefinitely in the manner most dear to him: that of judgement" (Badiou, 2005, p. 10).

A return to the subjective source arises in the situation of the limit expressed in various variations of the theme of the end of philosophy, the end of metaphysics, the death of the subject, etc. Return is identical to the act of reflection, contributing to the formation of the subject. The formation of the subject of politics in Badiou's theory occurs by analogy with Plato's concept of the formation of a philosopher described in *The State* (Plato, 1976, pp. 227-234). We can say that Badiou not only creates his own *Manifesto* (Badiou, 1992), following Karl Marx, but also creates an analogue of Plato's *The Allegory of the Cave*. In this case, Badiou uses Plato's metaphor of the "turn" which personifies the stages of the formation of the subject of politics. The first turn is associated with liberation from the fetters of public opinion, when an individual begins to "look around" and tries to estab-

lish one's own position in the social world, independent of the opinion of the majority. The second turn is turning to oneself with the fundamental question "What can I do?" The third turn is turning away from reality "as it really is" to the reality of a hypothesis and/or idea. The turns involve deindividualization of the individual, who turns into a subject who has courage and experience of resistance with thought. Political activity is viewed as the practice of thought, which is available only to the subjects of politics rather than to individuals. Only an intellectual who participates in politics as an activist can be a political subject. Badiou writes, "the point from which a politics can be thought – which permits, even after the event, the seizure of its truth – is that of its actors, and not its spectators" (Badiou, 2005, p. 23).

The direct involvement of the intellectual activist in the collective forms of the movement allows the activist to think of politics in the element of politics itself. The politics of activism unites workers and the academic intelligentsia in such a way that they form a kind of union of proletarians and philosophical and/or scientific thought. In the politics of activism, the main thesis is those who do not "do" politics are not in it (Badiou, 2018). Intellectual activity is viewed by Badiou as a universal ability shared by both the academic intelligentsia and the workers, whose participation in assemblies makes them "an intellectual who is organic to the new politics" (Badiou, 2008). The universal capacity for intellectual activity becomes a condition for the axiom of equality of minds, equalizing philosophers and proletarians: "anyone can be a philosopher" (Badiou, 2011).

Politics begins with an analytical question "How do we live depending on what the world is like?"; "In what worlds do we live?", which causes a turn from spontaneous and incoherent reality into the sphere of the reality of thinking. The thought of activist intellectuals build on "what exists" in the present and refers to what may happen in the future. Due to its uncertainty, the future contains limitless possibilities. The analytical question opens up a field of normative statements, each of which interprets the reality in the hypothesis mode: "we need twenty million", "we need compensation for social contributions", "we need a monthly salary for each year of service" and so on. The essence of a normative statement is that it is an expression that asserts the need to perform any action. In this regard, the statements of intellectuals must meet the main condition "Demand the impossible!", which means "Hold on to new opportunities!" (Badiou, 2013).

In the field of normative statements, which opens with an analytical question, the universal ability of people for intellectual activity is fully manifested. Badiou shows that intellectual activity is a collective process, therefore, everyone can be activists of thought. "But 'militant' is a category without borders, a subjective determination without identity, or without concept" (Badiou, 2005, p. 142). The normative statement realizes the principle of freedom, equalizing activists in their right to be intellectually active and participate in the discussion of the future. "To tell it like it is, and to draw the consequences of this 'telling' situation, is in the first place, as much for an Auvergne peasant as for a philosopher, an operation of thought" (Badiou, 2005, p. 7). Thus, the analytical question opens up a promising discussion of the future of politics, in which the principles of equality and freedom of expression work. Each participant in the discussion can rightfully be considered an activist of thought. In turn, the statement of each activist has the right to exist due to the fact that it is the result of intellectual activity.

The normative question "What kind of world do we want to live in?" sets the process of discussion of hypotheses, which are confirmed or refuted retrospectively through a sequential analysis of the statements. The field of analytical statements opens up with a normative question. Hypothesis analysis limits freedom of expression by requiring participants to adhere to strict rules of discussion. The art of reasoning requires that intellectuals have more than just knowledge of philosophical logic, i.e. forms, methods and laws of intellectual activity. An intellectual who thinks about a multitude of possibilities must be proficient in mathematical logic and know the set theory. It is mathematical logic that will allow the intellectual to correlate the many hypotheses expressed with a specific situation in such a way that a new opportunity appears.

In Badiou's new ontology of politics, the philosophical logos (λογος) is replaced by mathematical logic as the system of formal proofs and refutations underlying the invention of modern politics. The logical and mathematical side of politics limits the number of intellectuals who can take part in the discussion of possibilities. An activist must be more than just an intellectual, i.e. a representative of the academic intelligentsia. An activist must first of all be a philosopher-mathematician. Justifying the logical and mathematical essence, Badiou deduces a policy formula: $\sigma, \varepsilon, \pi(\varepsilon), \pi(\pi(\varepsilon)) \Rightarrow 1$, showing that four pluralities are involved in the numerical order of politics: an infinite set of situations (σ), the superpower of the state (ε),

gaps between events $\pi(\epsilon)$, and activist statements and practices $\pi(\pi(\epsilon))$ (Badiou, 2005, p. 150). In this case, Jason Barker believes that building an ontology of politics on set theory limits the ability to understand it (Badiou, 2005, p. ix).

A philosopher-mathematician who is good at the methods of mathematical logic becomes an exemplary political subject for Badiou. The best example of the political activity of intellectuals are philosopher-mathematicians and philosopher-logicians participating in the movement of the French Resistance: Georges Canguilhem, Jean Cavaillès and Albert Lautman (Badiou, 2005, p. 1-8). An exemplary political subject, according to Badiou, should be resistant by logic, and the best resistance figure is a mathematician. Mathematical logic returns thought to politics. The political activity of intellectuals is interpreted by analogy with the movement of the French Resistance, which is regarded as an exemplary political event. Badiou writes:

“This results from the fact that a Resistance figure ‘by logic’ obeys an axiom, or an injunction, which he formulates in his own name, and whose major consequences he lays out, without waiting to win over other people, in the objective group to which he belongs” (Badiou, 2005, p. 5).

If we follow the logic of the reasoning of Badiou himself, it turns out that the absence of philosophers, philosopher-logicians and philosopher-mathematicians in politics leads to the disappearance of thought from politics. In this case, Badiou’s thesis that those who do not “do” politics are not in it sees a logical continuation: those who do not do politics do not exist in philosophy either. While Arendt raised the question of the relationship between philosophy and politics (Arendt, 1990), Badiou raised the problem of the political engagement of philosophy and philosophers, without which politics loses thought and turns out to be a meaningless activity, and philosophical thought becomes inactive. An example of the inaction of philosophical thought for Badiou pertains to Jacques Rancière (Rancière, 1999), who is not an activist of political movements, which casts doubt on his philosophical thought about politics.

“Admittedly, no one is obliged, in order to do politics, to deploy an underlying ontology <...> But Rancière doesn’t do politics. If, on the other hand, one does philosophy, there is an obligation to make use of explicit ontological categories and to argue their cohesion. However,

er, all things considered, Rancière doesn’t do philosophy either” (Badiou, 2005, p. 116).

Badiou’s metapolitics politicizes philosophical thought. Since the involvement of philosopher-mathematicians, as well as philosopher-logicians and simply philosophers in politics is a single fact, politics is a rare event.

Badiou goes beyond the concept of classical revolutionary politics associated with proper names: Leninism, Stalinism, Trotskyism, Castrism, Maoism, etc. Modern politics can be named only by the name of a generic set, which has no name and does not lend itself to classification. For Badiou, the “excluded” becomes the plural name for the generic set of modern politics. In this case, the “excluded” refers not just to workers of foreign origin, but also to illegal workers who are at the center of modern politics.

Thus, the French theorists are reviving the tradition of the ontology of politics, which they build based on the principle of plurality, which changes the idea of the subject of politics. The subject of modern politics can only be an intellectual activist who simultaneously acts in politics and thinks politics. At the same time, the French theorists interpret the position of an intellectual activist in two ways: Lazarus is more concerned with the intellectual activity of political activists, whom he distinguishes by the degree of intellectual activity, their ability to think about action. The subjects of politics are divided into political activists and politicians who think politics from different perspectives. The prescriptive thought of political activists moves from the present (“what exists”) to the future (“what may be”). The prescriptive thought of political activists makes the situation open to subsequent thoughts. The analytical thought of politicians moves in the “opposite” direction – it proceeds from “what may be”, i.e. from the new opportunity, and turns to “what is there.” In terms of the degree of intellectual activity, the priority belongs to politicians who are able to analyze the current situation and think of the revolution as the only possible political action. Lazarus’ politics in interiority is actually a concept of the politics of revolutionary action. In turn, actualization of the policy of revolutionary action always requires the participation of a professional thinker and/or intellectual with a revolutionary consciousness. In Alain Badiou’s theory, the subject of politics is considered in the aspect of the active participation of intellectuals in politics. While for Plato an exemplary politician is the philosopher, for Badiou an exemplary political activist and subject of politics is the

philosopher-mathematician, who is not only good at mathematical logic, but also knows set theory.

References

- Arendt H (1990). Philosophy and Politics. *Social Research*, 57(1): 73–103.
- Arendt H, Beiner R (Eds) (1982) *Lectures on Kant's Political Philosophy*. Chicago: University of Chicago Press.
- Aristotle (1893) *The Nicomachean Ethics of Aristotle*. (F. H. Peters, Trans.). London: Kegan Paul, Trench, Trübner & Co.
- Aristotle (1885) *The Politics of Aristotle*. (B. Jowett, Trans.). Oxford: Clarendon Press.
- Badiou A (1992) *Manifesto for Philosophy*. (N. Madarasz, Trans.). Albany: State University of New York Press.
- Badiou A (2005) *Metapolitics*. (J. Barker, Trans.). New York: Verso.
- Badiou A (2008) *Meaning of Sarkozy*. (D. Fernbach, Trans.). New York: Verso.
- Badiou A (2011) *La Relation énigmatique entre politique et philosophie*. Paris: Germina.
- Badiou A, Tarby F (2013) *Philosophy and the Event*. (L. Burchill, Trans.). Cambridge: Polity Press.
- Badiou A (2018) *Can Politics Be Thought?* (B. Bosteels, Trans.). Durham and London: Duke University Press.
- Bauman Z (2001) *The Individualized Society*. Cambridge UK: Cambridge University Press.
- Bosteels B (2011). *Badiou and Politics*. Durham – London: Duke University Press.
- Calcagno A (2007) Abolishing Time and History: Lazarus and the Possibility of Thinking Political Events Outside Time. *Journal of French Philosophy* 17(2): 13–36.
- Descartes R (1998). *Discourse on Method*. (D. A. Gress, Trans.). Indianapolis- Cambridge: Hackett Publishing Company.
- Heidegger M (1968) *What is Called Thinking?* (J. Gray, Trans.). New York: Harper & Row.
- Lazarus S (2001) Anthropologie ouvrière et enquêtes d'usine: état des lieux et problématique. *Ethnologie française* 31(1/3): 389–400.
- Lazarus S (1996). *Anthropologie du nom*. Paris: Seuil.
- Lazarus S (2016) Can Politics be Thought in Interiority? (T. Harper, Trans.). *Cosmos and History: The Journal of Natural and Social Philosophy* 12(1): 107–130.
- Marx K (1951) *A Contribution to the Critique of Political Economy*. Moscow: Foreign Languages Publishing House.
- Moraes A (2019) Apresentação do Artigo de Sylvain Lazarus: Breves Anotações Sobre “Pensar no Registro do Possível”. *Plural. Antropologías desde América Latina y el Caribe* 3: 195–229.
- Moraes A (2020) Podemos pensar a política politicamente? As respostas de Bruno Latour e Sylvain Lazarus. *Tempo da Ciência* 27(53): 46–65.
- Neocosmos M (2009) Civil Society, Citizenship and the Politics of the (Im)possible: Rethinking Militancy in Africa Today. *Interface: a Journal for and About Social Movements* 1(2): 263–334.
- Neocosmos M (2012) Are Those-Who-Do-Not-Count Capable of Reason? Thinking Political Subjectivity in the (Neo-)Colonial World and the Limits of History. *Journal of Asian and African Studies* 47(5): 530–547.
- Plato (1976) *The Republic of Plato* (F. M. Cornford, Trans.). London-New York: Oxford University Press.
- Ranciere J (1999) *Disagreement and Philosophy Politics* (J. Rose, Trans.). Minneapolis-London: The University of Minnesota.
- Schelling F W (1978) *System of Transcendental Idealism* (P. Heath, Trans.). Charlottesville: University Press of Virginia.
- Smneczek N (2008). What is to be Done? Alain Badiou and the Pre-Evental. *Symposium: Canadian Journal of Continental Philosophy* 12(2): 110–126.
- Toscano A (2006) The Bourgeois and the Islamist, or, The Other Subjects of Politics. *Cosmos and History: The Journal of Natural and Social Philosophy* 2(1–2): 15–38.
- Viereck P (2003) Metapolitics Revisited. *Humanitas* 16(2): 48–75.
- Walzer M (1988) *The Company of Critics: Social Criticism and Political Commitment in the Twentieth Century*. New York: Basic Books.
- Zaibert L (2004) Toward Meta-politics. *The Quarterly Journal of Austrian Economics* 7(4): 113–128.
- Žižek S (2009) *The Ticklish Subject: the Absent Centre of Political Ontology*. London-New York: Verso.
- Žižek S (2008) *Violence*. New York: Picador.
- Žižek S (1989) *The Sublime Object of Ideology*. London-New York: Verso.

INTEGRATION OF LEARNING AND NEUROSCIENCE THEORIES WITH AI-BASED TECHNOLOGIES IN INTELLIGENT LEARNING SYSTEM IN ACCORDANCE WITH WHITEHEADIAN TRADITION AND CONTEMPORARY PROCESS THEORY

Rossitza Kaltenborn

Smart Infomanagement & Consulting, Frankfurt am Main, Germany
e-mail: Rossitza.kaltenborn@gmail.com

Abstract:

The purpose of the article is to present the possibility of integrating basic learning theories into the Extended Intelligent Learning System for data processing, optimization, adaptation and decision making in learning, which is based on the combination of teacher and intelligent tutor with artificial intelligence implemented, which supports the target formation, learning strategy, pedagogy and control. As a framework for the creation of the integrative model of theories, process philosophy is used, which enables a better understanding and explanation of the different paradigms and their functional combination. The article explores the strengths and weaknesses of selected theories and focuses primarily on constructivism, as numerous studies on learning theories have found that a constructive approach to learning is at the heart of many models in both traditional and digital learning in the Era of Big Data. The article explores certain influential learning theories, including the AI methods, their advantages, flaws and fields of intersection with neurosciences in terms of their application in intelligent training systems.

The goal of developing the integrative model is to realize the learner's potential in personalized knowledge formation in an intelligent learning environment and to enhance the efficiency of learning.

Key words: Artificial Intelligence (AI), constructivism, integrative model, Intelligent Learning System (ILS), learning theories, process philosophy.

1. Introduction

In the development of AI models for education and in the area of human-machine research in general, great attention is paid to the concept of learning (Kao and Venkatachalam, 2021, p. 890). This is because it is assumed that the two approaches can find points of contact and build on each other conceptually, improving the quality of both human and machine learning, which would promote the realization of learners' potential and benefit the whole of society. Reaching this goal is not easy, as there is no one definition of learning that is universally accepted by theorists, researchers and educators (Schunk, 2020; Thomas, 2021).¹ One of the reasons for this can be sought in the different focus of theories. Different theories concentrate on individual components in learning, such as the teacher-student relationship, and may ignore other aspects. Thus, they often lack the necessary toolkit to study education as a whole, which consists of multiple relationships. Methodological and explanatory deficits are repeatedly detected (Harasim, 2017). Therefore, most learning theories have epistemological limitations, as learning is a complex phenomenon that is domain- and historically-dependent and requires improved mod-

¹ Thomas (2021) analyses some definitions of 'learning' and highlights nine major learning theories, which are presented and compared with their pedagogic properties in a table. He investigates the possibility for assessment of how useful a learning theory might be to inform the teaching practice in analog and digital environments.

els of explanation and understanding as to how people learn.

2. Epistemological exploration and categorization of Learning theories

Many learning theories are assigned to one of the two groups based on epistemology: objective epistemology (behavioral, cognitive and connectivity) or constructivist epistemology (Constructivism and Online Collaborative Learning) (Thomas, 2021; Harasim, 2017).²

There is currently also not only one discipline of educational research that deals with the term “learning” and explains it. The field has long been interdisciplinary. It is stated that neither behaviorism or constructivism nor cognitivism or humanism (which remain the most influential theories) sufficiently explain the learning process methodologically and conceptually. A new direction of exploration is offered by the neurosciences, which are also reflected in the concept of learning from the point of view of consciousness and the brain. Other theories are the theories of motivation and self-regulated learning, Connectivism, Online Collaborative Learning (OCL), Multimedia learning theories, and Cognitive Load Theory (CLT) (Thomas, 2021).

Evaluating learning theories is not always connected with assessing whether they are right or wrong (Lefrançois, 2019), but with weighing how useful they are in the educational practice and for teachers’ understanding of how knowledge is created, of how students learn, and of which technologies are most suitable for a given learning approach (Bower, 2017). Bower notes that the ratings assigned to these approaches are “coarse generalizations, and there may be quite a variety of possibilities for each pedagogy, depending on how the teacher and students engage in the task” (Bower, 2017).³

In identifying which ideas of the basic classical theories of learning can be used in different digital learning models, it is necessary to explore and specify with the help of existing literature and the discoveries of neuroscience and cognitive psychology, what are the benefits and disadvantages of constructivism, behaviorism, cognitivism, connectivism. This is the starting point for determining the strategy for their combined application in the digital learning

process. It should also theoretically justify the additional possibilities that artificial intelligence and big data procedures provide – e.g., separation of functions between teacher and robot, the combination of teacher and robot, guidelines for the development of social robotics. As a result of such a study, it would be important to establish the optimal contextual correlation between theory, technology and learning purpose. The framework for this gives us Whitehead’s learning philosophy, which contains in its concept essential elements of constructivism, behaviorism, cognitivism, connectivism, and also shows their relationship of complementarity and contradiction, with its attendant benefits and disadvantages.

For a new direction of exploration, it is crucial that learning theories are studied in their mutual influence and adaptation, contradictions, development, advantages and disadvantages. The role and intervention of the teacher and the student, as well as the technology in learning (Teacher vs/with student vs/with technology) is also a major aspect of learning theories. Other research criteria are: Possible Learning Activities (LA) – Active L., Problem-based L., Project-based L., Collaborative L., Self L.; the perspective of knowledge – objectivist vs constructivist; the perspective of information – information acquisition vs active task-based learning; interpersonal interaction – individual vs social; product output – achieved aim vs no achieved aim (Thomas, 2021). Various researchers explore different instruments to describe and compare the properties of these theories. Often, they are presented in tables, which refers to traditional and new research perspectives. This article takes into account (in part) Thomas’ (2021) table, which was adapted from a table by Bower (2017), who based his on one devised by Conole et al. (2004).

The proposed Riffert table (2018) presents “Paradigms of Learning: proponents (of different learning theories) in their historical and actual philosophical foundations.”⁴ Riffert noticed that “many more learning theories could be added to these categories. Moderate Behaviorist approaches, as well as Moderate Constructivist approaches are subsumed under the category of ‘Cognitivism’” (Riffert, 2018, p.76).

Riffert points to the source of the problem of defining learning theory in the context of myriad learning theories, or “the wilding field of learning theories,” by looking for the reason for this in the

² Thomas (2021) offers an overview of learning theories and cites Harasim (2017), who also engages in meta-research in this area.

³ Thomas (2021) refers to Bower (2017).

⁴ Riffert, 2018, p.76

improper use of paradigms that compare individual theories without looking for common parameters of intersection. He states,

So, this way of ordering just reduces the number of theories by allocating them to some major paradigms while leaving the problem of any interrelations between them untouched.⁵

He chooses to take a different approach in order to try to make sure that the similarities and differences between the paradigms are more visible.⁶ Riffert is seeking a solution by creating a new categorization in "a framework that allows for at least some coarse categorization of the different approaches".

3. Categorizations for systematic study of theories

Riffert contributes to his research on the structured study of theories precisely in that he adheres to clearly made categorizations, which he summarizes and integrates in a new way. "Note that any integration of these elements will necessarily lead to the modification of these same elements" (Riffert, 2018, S. 83). He is deliberately looking for the basics of the new integrative model in the organismic theory of training, which is developed within the framework of Whitehead's procedural philosophy.⁷

3.1. Riffert's categorization of learning theories and its reference to process philosophy

Riffert refers to Bertalanffy's (1967) and Dubs' (1995) research on learning theories and notes their focus when examining three main theories (behaviorism, cognitivism, radical constructivism) (Riffert, 2018, S. 74). This categorization contributes to clarifying the different approaches in competing learning theories that are the cause of ambiguity in categorization. "This categorization is helpful to distinguish particular Behaviorism from the other two aforementioned paradigms in learning research: Cognitivism and Constructivism (Riffert, 2018, p. 74). Indeed,

Bertalanffy's crude dichotomous distinction between a mechanistic worldview (which is characterized as static, passive, and elementaristic) and an organismic worldview (which

shows the features of dynamics, holism, and self-activity of organisms) gives us at least three basic criteria for a first differentiation and categorization of psychological and educational learning theories (Riffert, 2018, p. 74).

Riffert also cites Dubs' (Dubs, 1995) view, which uses the categories of subjectivity and objectivity. "Rolf Dubs (1995) has proposed to categorize learning theories along the continuum between the two paradigmatic poles of objectivism and subjectivism" (Riffert, 2018, p.74). The category of objectivity corresponds to the mechanical point of view and is characteristic of behaviorism, and subjectivity focuses on the mental interpretation and construction of the world and is therefore related to the constructive and cognitive understanding of the world and through this quality corresponds to the organismical dynamic point of view.

Riffert deliberately sought the integrative framework for the realization of the relationship on various issues. This framework with new categories deals with "Whitehead's Organismic Philosophy as an Integrating Frame for Behaviorism, Cognitivism and (Radical) Constructivism. I have thus far claimed that Whitehead offers a frame for integrating behaviorism, cognitivism and (non-radical) constructivism" (Riffert, 2018, p. 83). In accordance with Riffert, Whitehead's educational theories influence not only learning and knowledge acquisition in a classroom and research environment. They underline our actions. This doctrine "enables us to distinguish between [1] pure instinctive action, [2] reflex action, and [3] symbolically conditioned action" (Whitehead 1929 / 1967, 78; Riffert, 2018, pp. 84-85). These three "actions" are related to three states of learning: "romance" (excitement, emotional arousal); "precision" (rules); "generalization" (idea, the new knowledge) (Riffert 2018, 83). They are characteristically learning actions.

Whitehead's theory of learning is influenced by the idea of a cycle of the "romance," "precision," "generalization" phases, being rhythmically interrelated and related to each other according to the model of "freedom" and "discipline" (Whitehead A., 1929/1967, p. 31; (Riffert, 2018, p. 82). According to Whitehead, Riffert highlights the relatedness of the romance and generalization phases with the state of freedom and the relatedness of precision with discipline.⁸ In this model, he discovers innovation,

⁵ Riffert, 2018, p.73

⁶ Ibid.

⁷ Riffert, 2018, p.83

⁸ Riffert, 2018, p. 83

which consists in complementing traditional theories of learning with the knowledge that learning is a discipline (precision) characterized by freedom and innovation. Referencing the two new phases that introduce freedom into learning is an original contribution that makes it possible for Whitehead to be treated as a foreman of constructivism. Whitehead's concept is important because it describes the dynamic trio of elements that are freely interconnected throughout life and surpasses previous theories that focus only on static teaching and learning of "inert knowledge,"⁹ which he criticizes. The study of learning theories in terms of freedom and discipline could be reduced to the issue of teacher-led and non-teacher-led learning and of self-organizing and control. These are some of the main criteria for distinguishing between behaviorism, constructivism and cognitivism. In Whitehead's next choice of philosophical and pedagogical concepts, this one, as well as other criteria discussed so far in the article, will be analyzed. Finding constructivistic elements requires a deeper study of his philosophical concept of learning from the perspective of the question, "Is Whitehead a constructivist, and if so, to what extent?"

3.2. Exploring constructivist and non-constructivist ideas in Whitehead's organismic philosophy

In Whitehead's process-philosophical approach to studying development and achieving the goal of learning, some basic concepts of constructivism and cognitivism can be found, which are later developed in Dewey (1916), Piaget (1972), Vygotsky (1978), and Bruner (1990) and become dominant paradigms in the theory of twentieth-century learning. The cognitive-psychological approach to learning related to emotions, motivation and engagement has an intersection with the process philosophy of organism¹⁰ and constructivism, and has been explored by Woodhouse, who compares Whitehead's views with Bruner's (Woodhouse, 2017). These points of view – cognitive-psychological and constructivist – stem from the ontology of Whitehead's organismic process philosophy. These concepts of learning in Whitehead's process philosophy contain constructivist elements. They can be selected and summarized as Concepts (C) in the following order:

C1. Education is a process of discovery;¹¹

"Education should consist in a continual repetition of such cycles. Each lesson in its minor way should form an eddy cycle issuing in its own subordinate process" (Whitehead 1929/1967, p.19).

"[...] It is a process of discovery, process of becoming used to curious thoughts, of shaping questions, of seeking for answers, of devising new experiences, of noticing what happens as the result of new ventures. This general process is both natural and of absorbing interest" (ibid., p.32).

C2. Life is the only object of learning;¹²

"There is only one subject-matter for education, and that is Life in all its manifestations" (ibid., p.7) [...] you cannot put life into any schedule of general education unless you succeed in exhibiting its relation to some essential characteristic of all intelligent or emotional perception" (ibid., p.8); [...] the education at the university preserves the connection between knowledge and life" (ibid., p. 93).

C3. Whitehead's notion of 'organism' with respect to education;

"It is, of course, a process completely of its own peculiar genus. Its nearest analogue is the assimilation of food by a living organism: and we all know how necessary to health is palatable food under suitable conditions" (ibid., Ch. III, The Rhythmic Claims of Freedom and Discipline, p.33).

"[...] all mental development is composed of such cycles, and of cycles of such cycles. Such a cycle is a unit cell, or brick, and the complete stage of growth is an organic structure of such cells." (ibid., p.31).

C4. Education must stimulate the development of personality;

"The two principles, freedom and discipline, are not antagonists, but should be so adjusted

⁹ Whitehead, 1929/1967, pp.1-5

¹⁰ Whitehead, 1929/1967

¹¹ This statement could also be found in Bruner's concept (Bruner, 1990).

¹² This statement can also be found in Dewey's concept and shows that they are distinguished from radical constructivists.

in the child's life that they correspond to a natural sway, to and from, of the developing personality" (ibid., p. 31)

For the purpose of a good education must be overcome; "the neglect, arisen from two disastrous antitheses, namely that between mind and body and between thought and action" (ibid., p. 58).

C5. Education is cyclic; "The cyclic processes" (ibid., p.19);

"I am pleading that we shall endeavour to weave in the learner's mind a harmony of patterns, by coordinating the various elements of instruction into subordinate cycles, each of intrinsic worth for the immediate apprehension of the pupil" (ibid., p.21). "The education should consist in a continual repetition of such cycles" (ibid., pp 17–19); "threefold cycle of freedom, discipline and freedom; but all mental development is composed of such cycles, and of cycles of such cycles" (ibid., p.31).

C6. Education must be adaptive;

"It is this adaptation of freedom and discipline to the natural sway of development that I have elsewhere called The Rhythm of Education" (ibid., p.31).

C7. Self-education and self-assessment are the central problems of education;

"Now the natural mode by which living organisms are excited towards suitable self-development is enjoyment" (ibid. p.57).

C8. Education should be creative;

"Creative enjoyment is the outcome of successful effort and requires help for its initiation. Such enjoyment is necessary for high-speed work and for original achievement" (ibid. p.57).

C9. Education must be satisfying and enjoyable;

"There can be no mental development without interest. Interest is the sine qua non for attention and apprehension. [...] without interest there will be no progress. Now the natural mode by which living organisms are excited

towards suitable self-development is enjoyment" (ibid., p.31).

C10. Both correct and incorrect thinking are useful

C11. Teaching must not always progress from the simple to the complicated;

"The responsibility of the teacher at this stage is immense. To speak the truth, except in the rare case of genius in the teacher, I do not think that it is possible to take a whole class very far along the road of precision without some dulling of interest. It is the unfortunate dilemma that initiative and training are both necessary, and that training is apt to kill initiative" (ibid., p.35).

3.3. Whitehead's Concepts for education which do not correspond to constructivism

The following concepts address learning in several aspects that do not overlap with constructivist thinking, because a leading element of them is the teacher's leadership role. In this role, he is not only an adviser. This should not be understood as contrary to the thesis of self-education. The main elements with which they are characteristic are rhythm, conduction, practical usefulness, action, effectiveness, non inert knowledge.

C12. Education consists of three components:

— **General vs. specialist education;** amateur vs expert (ibid., pp 11-13).

"The subjects pursued for the sake of the general education are special subjects specially studied; and on the other hand, one of the ways of encouraging general mental activity is to foster a special devotion." (ibid. p.11).

— **Scientific education;** (ibid., Ch. IV. Technical education and science and literature (ibid., pp 43-61).

"A scientific education is primarily a training in the art of observing natural phenomena, and in the knowledge and deduction of laws concerning the sequence of such phenomena" (ibid., p. 49).

— **Technical education;** (Chapter IV, p. 43-61)

"A technical education is in the main a training in the art of utilizing knowledge for the

manufacture of material products” (ibid., pp 49-50)

— Religious education; “Religious education is an education which inculcates duty and reverence (ibid., p.14).

In this concept, the areas of learning are precisely set, which is not typical of constructivism.

C13. Education includes three stages: (ibid., pp 17–19)

— Romance (emotional), Precision; Generalization;

“I call the first period of freedom the "stage of Romance, "the intermediate period of discipline I call the "stage of Precision", and the final period of freedom is the "stage of Generalization” (ibid., p.31).

The "stage of Precision" is related to teaching and guidance on the part of the teacher.

C14. Education denotes conducting and must be guided

“The case is far different for a living **organism** which grows by its own impulse towards self-development. This impulse can be stimulated and guided from outside the organism, and it can also be killed. But for all your stimulation and guidance the creative impulse towards growth comes from within, and is intensely characteristic of the individual” (ibid., p.50).

C15. Theoretical interest and practical usefulness coincide;

“This is the aspect of university training in which theoretical interest and practical usefulness coincide” (ibid., p.26).

Abstract (theoretical) knowledge is associated with practical actions and is related to problem solving.

C16. Education is a union of thought and action

“The peculiar merit of a scientific education should be, that it bases thought upon first-hand observation; and the corresponding merit of a technical education is, that it follows our deep natural instinct to translate thought into

manual skill, and manual activity into thought” (ibid. p.51).

C17. Education is rhythmic: “The Rhythm of education”

“I am convinced that much disappointing failure in the past has been due to neglect of attention to the importance of this rhythm” (ibid., p.30).

“[...] the development exhibits itself as rhythm involving an interweaving of cycle, the whole process being dominated by a greater cycle of the same general character as its minor eddies” (ibid. p.27).

The rhythm is related to following rules and maintaining this rhythm.

C18. The learning process must be effective;

“There is no getting away from the fact that things have been found out, and that to be effective in the modern world you must have a store of definite acquirement of the best practice” (ibid., p.34)

C19. Inert knowledge must be eliminated;

“Education with inert ideas is not only useless: it is, above all things harmful – Corruptio optimi, pessimal [...] Education in the past has been radically infected with inert ideas” (ibid., pp.1-5).

C20. Freedom and discipline are interdependent aspects of education;

“The antithesis in education between freedom and discipline are not so sharp as a logical analysis of the meanings of the terms might lead us to imagine” (ibid., p.30);

“The two principles, freedom and discipline, are not antagonists, but should be so adjusted in the child's life that they correspond to a natural sway, to and fro, of the developing personality” (ibid., p.31).

Discipline is about obeying rules, and being led by a teacher or an intelligent tutor.

The article looks at constructivism as a theory of learning, because it is one of the most common

theories used in learning and is a driver of digital transformation (Dennick, 2016).

3.4. The benefits of constructivism consist of:

1. The development of cognitive qualities, in problem solving, innovation, critical thinking.
2. Development of personal qualities in the context of collaboration, communication, ethics.
3. Realisation of active learning in students, which improves motivation. It can be beneficial to combine with personalized training and to observe the personal maximum permissible load.
4. Develops the knowledge of how to acquire "knowledge" for oneself and not wait for knowledge to be given smoothly and fully as prepared lectures. The student learns "how to learn," which gives him knowledge of lifelong learning.
5. Mastering skills for creating habits and procedures for building new thought constructions with a higher degree of abstraction.
6. Convinces the student that collective learning in a small group is advantageous.
7. Improves the memorization of facts, procedures, methods already known to the student, as they are reconsidered and invested with meaning in a new context.
8. Due to the active participation of the student in clarification, dialogue, discussion, analysis, reasoning, formulation of the problem, his knowledge of the links between existing and new mental structures is strengthened in solving various problems.
9. Students can gain momentum for more efficient absorption of new ideas at a specific time, rather than according to a preset curriculum and calendar.

3.5. The disadvantages of constructivism consist of:

1. Not good for novice students who need structured learning activities.
2. Does not take into account the difference in permissible cognitive load (processing efficiency and working memory resources) of each student. There is no evidence that with a constructive approach permissible cognitive ingrowth will increase.

3. It is not associated with personalized learning, which is considered almost unanimously one of the main goals of the planned digital transformation of modern higher education.
4. There is no performance score (time, effort). If ideas need to be rediscovered with the help of the teacher through dialogue, reflection, formulation, this leads to an end to learning time.
5. It is opposed to curricula with clearly defined structures and project objectives.
6. There is no individual bias hazard assessment (inductive bias).
7. It partially rejects or ignores previous experience of Higher Education standards around the world which establish what is important, basic, secondary.
8. Demands innovation at every level that a student is not always capable of. Constructivism adheres to the 100-year search for innovation, which is inefficient.
9. In constructivism, there is no idea of optimization of knowledge accumulation, which is essential for neural network-based systems (NN). Supervisor training with a teacher is much more effective than learning without a machine learning teacher (ML) when the goal is innovative knowledge.
10. The intellectual level of students does not allow them to reach the highest achievements of knowledge acquisition. They should serve as a basis and build new structures of knowledge on them.
11. Prior knowledge is taken for granted no matter how it is received, how deep it is, what its scope is, or what differences exist between the students in the group.
12. For the efficient functioning of the teacher's constructivism, significantly more current information is needed for each student and for the group as a whole. This requirement may conflict with the right to the protection of a student's privacy.
13. Being forced to be behaviorally active does not necessarily mean more efficient learning.
14. Constructivism is not universally suited to all areas of higher education. There are serious objections to its application in engineering, mathematics, physics, chemistry,

- though in medicine, it has an application (Dennick 2016).
15. For the acquisition of procedural skills, only an unmanageable (but only guiding) constructive approach is insufficient.
 16. There is a danger that students will acquire misconceptions and incomplete or disorganized knowledge.
 17. Students do not have the necessary experience and time to compare their personal mental structures with the real circumstances, which is among the basic requirements of the constructive approach. Constructivism is ineffective in accounting for personal opportunities and challenges in diverse groups, and thus a divide between leaders, middle and underperforming students emerges.

Based on the analysis of the advantages and disadvantages of constructivism, it can be concluded that constructivism is a theory of learning, not a pedagogical theory, and should not be used as the only doctrine in training. It is suitable for students to varying degrees:

- a. For the strongest learners it encourages innovation and problem solving
- b. For the average learners it encourages learning through reproduction and optimization of knowledge with resource methods
- c. For weaker learners it offers a simplified understanding of ideas and methods; they find it difficult to cope with the average requirements because they need a minimum base of knowledge to build on. Other skills are of secondary importance.

Through these discussions, the ideas shared by constructivism and process philosophy become apparent, which, indeed, is the appropriate framework for understanding and applying constructivist methods. (Similar links between the basic ideas of other learning theories and philosophy can also be sought.¹³)

This point is relevant in terms of how, with these qualities, the ideas of constructivism and other theories could be integrated into Big Data-era (BD) learning and Artificial Intelligence (AI) (Honebein, 1996; Hung and Nichani, 2001; Koohang, Riley,

Smith; Robinson, Molenda, & Rezabek, 2007; Schreurs, 2009).

In discussions regarding educational technologies (Edutech), there is no consensus about best practice or the best learning theory to inform technology-based education (Tobias and Duffy, 2009; Bower, 2017; Lefrançois, 2006/2019). There is not only one learning theory. Constructivism is one widely used practice. But is this the only right one? Could it be the basis for designing a model for digitalized and machine learning supported processes? An expanded multi-perspectival model (incorporating constructivism, behaviorism, cognitivism, connectivism) could help to bridge the explanation gap between subjectivism and objectivism, as well as brain and mind, which underpin these learning theories.

4. The influence of neuroscience on learning theories

New research focuses on brain functions and their influence on learning and provides their findings to an entire field – the neurosciences – in which new methods and tools for understanding processes in the brain figure in an array of experiments (Boon, 2013; Borck, 2006). It is important to note that as with other classical learning theories (as shown in Chapter 2), elements and areas of overlap with process ontology can be found. Research such as that of Georg Northoff, who studies understanding as a feature of process philosophy in terms of neuroscience can be cited in this area (Northoff, 2016,¹⁴2016¹⁵).

Georg Northoff (2016) is looking for a basis for his neurobiological approach to learning and brain research within the domain of the philosophy of organism. According to Whitehead's process-ontological framework, Georg Northoff develops a specific view on the brain – that of spontaneous activity of the “spatial-time structure of the brain” (Northoff, 2016). Northoff interprets this brain model as a neuro-ecological model and makes reference to Hume's and Kant's brain model. He makes evident that in the process ontological view of Whitehead, this brain is structured as an addition of moments in time – a process leading to complementarity or “satisfaction”. Satisfaction is a subjective pro-

¹³ In this article, the advantages and disadvantages of behaviorism, cognitivism and constructivism as they affect learning outcomes are not focused on specifically. This is the subject of another study.

¹⁴ Northoff, G. (2016). Neuroscience and Whitehead I: Neuro-ecological model of brain.

¹⁵ Northoff, G. (2016). Neuroscience and Whitehead II: Process-Based Ontology of Brain

cess and could be constructed as a component of “data addition.” The brain as neural activity corresponds to mental activity. Georg Northoff tries to find correspondence between the two levels of the mechanical and the phenomenal, the outer and the inner reality, the brain and the mind, the world and mind. His notion regarding this is captured by the idea of “the synchronization,” an art of correspondence in the dynamic space-time relation.

Neurosciences usually use other explanatory models and terms to describe the brain and its activity during learning, which incorporate hypotheses to be tested and verified. Among these, the theory of the learning window is widespread (Fuchs, 2021).

Studying the relationship between learning theories in terms of applying them to technology would be under-studied if their connections and applications were not addressed through the concepts of neuroscience. The questions that arise through such study focus on understanding the functions of the brain and their implementation in learning (Boon, 2013). The knowledge of the brain that neuroscience has achieved, both in terms of information processing and emotional processing and control, has significant implications for advancements in education in terms of motivation, doubling and success in learning (Kaltenborn, 2019).

Despite such advances, it should be noted that there is also strong criticism of the neurosciences. From the point of view of neuroscience, the self and brain have not been adequately parsed. The empirical measurements of brain functions that neuroscience offers still cannot explain the mind on the basis of purely physical properties. According to recent neuroscientific studies, learning is a correspondence between mental and neurological processes that take place in the brain (Fuchs, 2021). For example,

The exploration of the brain is not an easy project, for example by creating some experimental conditions in the laboratory and registering the associated brain states and deducing from this or that way consciousness comes about¹⁶ (Fuchs, 2021).

Yet, the derivation and transfer of the empirical to the mental is not yet scientifically, methodologically and theoretically explained (Borck, 2006).

5. Correspondence between neuroscience and learning theory in the organismic philosophy of Whitehead. The terms of “mind,” “emotion” and “learning”

Alternatives to understanding learning not only as instruction, mediation, teaching of rules, generalization of content, frontal teaching with permanent exams and assessments, can be found in the learning model of Whitehead and the context of his philosophy of organism. Some investigations of Whitehead’s learning concept shows that learning is an embodied process emerging as an outcome of emotion arousal (Woodhouse, 2017; Kaltenborn, 2019). Learning is explained in Whitehead’s view from an ontological nature-related perspective that extends the epistemological approach to the term (Petrov, 2019). In this context, the epistemological proximity to the term “mind” is reinforced in the development of the study.

In Whitehead’s view, the mind is “a process of mental activity” (Whitehead A. 1929/1967, p. 151):

The philosophy of organism abolishes the detached mind. Mental activity is one of the modes of *feeling* belonging to all actual entities *in some degree*, but only amounting to conscious intellectuality in some actual entities. *This higher grade of mental activity is the intellectual self-analysis* of the entity in an *earlier stage of incompleteness*, effected by *intellectual feelings produced in a later stage of concrescence*. (Whitehead A. 1929/1967, p.56)

In this claim, the correspondence and explanation for the relationship *between “mental activity” (mind) and learning activity can be found*. The explanation of a mental activity as an intellectual act of “self analysis” and learning activity is included in his ‘stages’ model. Each of the stages – romance, precision and generalization¹⁷ – is characterized as a *mode of mental activity* (feelings) that differ from one another in terms of time (early, later) and *degree of completion*. The highest stage is the most complete. The process of *movement from uncompleted to completed feelings is a learning activity and analytical relation of the self*, which characterizes the mind (Whitehead A. 1929/1967, p.56).

Both constructivism and behaviorism theoretically underpin some of the intelligent learning technologies (Harasim, 2017; Robinson et al., 2007).

¹⁶ Translation by Fuchs (2021).

¹⁷ These concepts are discussed in section 2 of the article.

The neuroscientific findings explain some of the paradigms of constructivism and behaviorism and are very important for understanding the process of knowledge creation, including processes that occur in AI-based technologies.

Behaviorism underlies the epistemological aspect of learning reinforcement, characterized by feedback within the learning process. Learning reinforcement, as one of the most important learning procedures in AI (and machine learning), correlates with the topic of the brain's reward system and strengthening certain connections in the brain's structure, which are the basis of memory and the speed and productivity of processing in the brain. Behaviorism focuses on the phenomenon that can be described from the outside. The approach conceives above all a change that turns into modifying existing behavior. The disadvantage here is that the dynamics of learning circulate between known/predefined states, without expansion in relation to a goal.

Behavior research as a practical assessment of the ideas of behaviorism deal with different techniques of observation, which are developed in the AI-based EDU-research and are connected with gathering from voice, face and brain data processing and images from EEG and fMR, and with their recognition. AI-based Emotion recognition is also connected with behavioral research. Emotion and satisfaction in learning are part of a widely studied field, which includes interdisciplinary and assessment approaches of constructivism, cognitivism, connectivism. Conclusions drawn from this research depend on the learning theory under examination and on the measuring instruments (or combination of instruments) used.

6. Integrating some of the most influential learning theories, such as behaviorism, constructivism, cognitive impairment in the era of artificial intelligence and big data.

Integrating learning theories into a technological context may help inform learning design and teaching practice in higher education on the basis of AI. To achieve a truly integrative approach, it is essential to study the relationship between different theories by taking into account traditional and new concepts of process philosophy (Anderson and Petrov, 2019). The graph shown in Fig. 1 presents how traditional theories can be usefully applied in AI-based learning technologies depending on their power of influence on knowledge over time.

Figure 2 shows that knowledge cumulation is highest when using a holistic approach – the combi-

nation of traditional teaching, Learning Theories (LT) and AI.

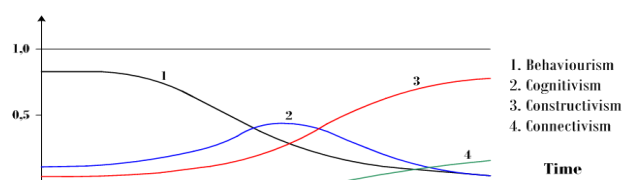


Figure 1. Development of knowledge over time according to the theoretical approach used in the learning process

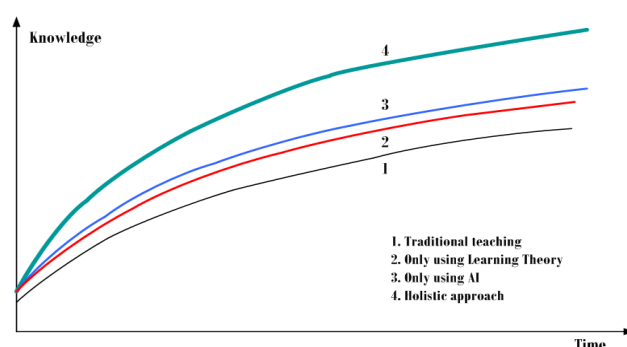


Figure 2. Knowledge generation considered under different learning approaches

In an effort to clarify deeper complex relationships in parallel ongoing multi-student learning processes, scientists are deploying artificial intelligence-based models to reinforce an individual student's understanding (Conole et al., 2004; Bower, 2017). These models can serve both to better support and understand human learning and to inform the ongoing training of intelligent tutors, leading to evidence-based processes of teaching, exercise, analysis, and evaluation that promote quality personalized learning, which aims to help realise the potential of the individual student.

The new concepts generated in this pursuit may correspond to philosophical concepts, such as Whiteheadian concepts based on his organismic learning theory. Appropriate explanation models can help to understand incentives for behavior and the achievement of learning goals, as well as answer the question of the reproducibility of brain processes during learning in mathematical algorithms, the difference between sustainable and unsustainable knowledge, and the value of gaining knowledge through supervised or unsupervised learning.

The combination between different learning theories and artificial intelligence is to be shown from the point of view of process philosophy. It enables the systematic point of view for different

aspects of assessment and offers a contextually approach method for investigating aspects of the learning process.

The integrative approach focuses on technological, dynamic elements and relationships that are not carried out substantially and statically, but as a transformation in the process of cognitive knowledge, since learning is based on the transition from one emotional state to another. The philosophy of organisms-based processes offers the framework for the convergence of these elements.

7. The implementation and application of the concepts of traditional theories for training in technology

The large number of surveys published in recent years shows the strong influence of constructivism as theory and practice (Dennick, 2016). There is a growing number of studies on learning theories which focus on the design of blended and online learning and on the learning outcomes that result from such approaches (Hung, 2001; Honebein, 1996; Murphy, 1997; Koochang, Riley, Smith, & Schreurs, 2009; Bates, 2014; Grabowski, 2009; Harasim, 2017; Anderson & Petrov, 2019). They present many of the positive qualities of constructivism, such as its beneficial influence on decision-making, self-learning, innovation, and scientific learning. Each examines the relationship between teacher and pupil differently, and a connecting element in all is the *principle of success and satisfaction in learning* (Whitehead, 1929/1978, p. 185; Grabowski, 2009; Dennick, 2016).

7.1. The development of learning theories in correspondence with the context of technological use

There are studies in which the development of learning theories is explored as a process that links the psychological states of optimism and disappointment to successful or unsuccessful applications of learning theory in communication technologies (Grabowski, 2009).

Thomas (2021) refers to Harasim (2017), who makes the point that each learning theory is associated with particular pedagogies and learning technologies. Bates (2014) also argues that “the choice of preference for one particular theoretical approach will have major implications for the way that technology is used to support teaching.” Schunk (2020) suggests that teachers need to ask the question,

‘How does learning occur?’ and that whatever answer they come up with will inform lesson design, teaching practices and student activities. Ultimately, teachers use learning theories (consciously or not) and their related educational approaches and technologies to design, develop and deliver effective learning experiences.

For this integrative model, it is important to find correct theoretical correlations for teacher-pupil relationships, so that the result – knowledge and learning achieved – is optimal for a particular learning behavior. In this model, it is important to manage the process dynamically during different phases of changing the behavior of the teacher and the student according to the changing ratio in the degree of autonomy of the student and the leading role of the teacher (Fig. 1). It is important in this model that the available (big) data from behavior, success and assessment of learning are interpreted as a result of the different activity of the student under the influence of the teacher in a given context and at a given time. Depending on this, the most appropriate procedure should be applied in accordance with a particular theory of learning.

Process philosophy enables the learning process to be considered as a *process with changes in emotional, cognitive states in the learner* (the student), and these conditions can be described by the different learning theories depending on the purpose the teacher pursues in the learning process. Within the framework of a process, these conditions can be adjusted in order to better apply them within a given learning context.

One of the reasons there is not a single very good model for digital learning is that there is no consensus regarding the use of a theory of learning in pedagogy that is embedded in a particular digital technology. This is partly due to the invisibility of the underlying theories embedded in the digital framework environment and the different contexts in which they are used.

8. The conformity between constructivist and AI point of view

One popular constructivist point of view on learning is: “The most important factor influencing learning is what the learner already knows” (Dennick, 2016). The learning focuses on the background knowledge of the students. *New knowledge* is therefore *interpreted* by existing knowledge *and then connected to existing knowledge*. The implications

of this conception are manifold and spread beyond conventional education. In addition, this phenomenon invites the question of how the process is conducted and manifested in the brains of individuals (Dennick, 2016, p. 201). Implications of this question can be sought both in the neurosciences and in the learning algorithms of artificial intelligence. Is there a functional proximity to a neural network?

An AI network (CNN) can also recognize patterns and assign them to known patterns by combining part and whole, so that in the end the desired shape appears as a whole.¹⁸ This comparison shows how the principle of constructivism is enshrined in the function of Convolutional Neural Networks and can be used for further research in relation to Goldstein's holistic approach to the mind in his "gestalt theory" (Goldstein, 1939).¹⁹

The investigating point of view of the synergy between learning theories and educational technologies can be detected in some research approaches (Grabowski, 2009; Koohang et. al., 2009; Dennick, 2016). Other studies have found with a similar approach the relationship between learning theories and learning analytics/big data (Greller & Al., 2012). This way of studying learning theories is expanded by neuroscience research that invokes the constructivist model (Dennick, 2016) and explores "learning as brain activity" (Boon, 2013), which becomes the basis for the emergence of educational neurosciences.

9. Proposal for Interaction Models between Learning Theories, Artificial Intelligence, Neurosciences, Cognitive Psychology

The philosophical approach to the integration of theories makes it possible to simplify the increased complexity of learning knowledge, the accumulation and measurement of which is partly unclear. The clarification process is key in terms of introducing theories in the development of new training technologies, such as AI-based reinforcement learning systems. For this we need to understand which is the optimal intersection of theories, learning analytics and neurosciences. (Fig. 3)

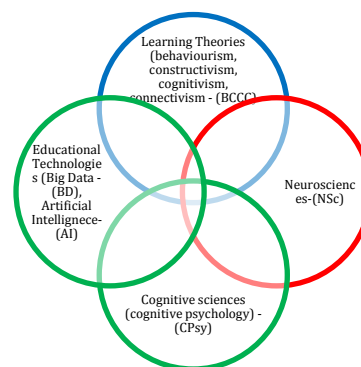


Figure 3. Spheres of interinfluence between different approaches to theories of learning, neuroscience and AI

Figure 4 presents the possibility of assessment of the efficacy of the integration of behavioural, constructivist, and cognitive components in one learning system with regard to their relatedness and correlations. In this extended learning system, AI control decision-making. Similar investigations in this area have been carried out by (Greller and Drachsler, 2012).

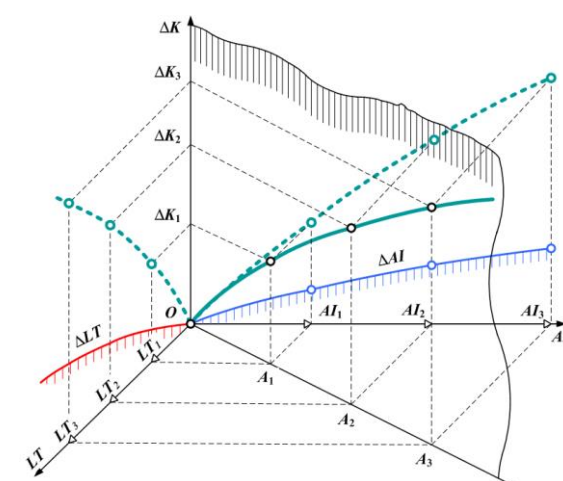


Figure 4. AI – Artificial intelligence, K- Knowledge, LT- Learning Theories integration model. Change of the accumulated knowledge (K) according to the degree of impact of different learning theories (LT) in combination with AI-based technology (AI). The goal of the system is the optimal realization of the potential of the learner.

10. Discussion

The article focuses on the proposal for combining elements of learning theories and AI in technological environments. Digital transformation in learning is impossible both without Artificial Intelligence and Big Data, and without the basic theories

¹⁸ This function may be related to the principles of "gestalt" theory, but this is the focus of another study.

¹⁹ Based on his holistic theory, Goldstein's point of view – as "Gestaltists" – can be explored in further research in accordance with the organismic philosophy of Whitehead and with constructivism.

of learning. This approach shows the advantages and disadvantages of influential paradigms (e.g. Constructivism) from the point of view of process philosophy. This result could be a starting point for studying other theories and their implementations in Edutech. The investigation carried out in this paper shows that the development of learning theories has also been influenced by the cognitive neurosciences and discoveries in the field of artificial intelligence, realized in educational technologies – e-learning, machine learning, reinforcement learning and learning analytics – and Big Data (Greller & Drachsler, 2012). The study demonstrates the benefit of forming new concepts and modeling their interactions, which expand the general and personal context of learning into different eras. Two models are shown that specify reasoning and provide a basis for creating an intelligent learning system with new methods of measuring learning outcomes. The aim of the presented model is to contribute to solving previously unresolved problems at the intersection of traditional and digitally personalised learning.

11. Conclusion

Learning theories are critical because they provide insight both into the ways knowledge is created and the ways in which individual students or groups learn. Using an integrative approach to applying learning theories serves as a challenge in improving individual learning methods. The conceptual development of new learning technologies based on artificial intelligence, supporting the application of the appropriate theory of learning according to the context of learning, is related to the aspiration, in the age of Big Data, of overcoming the ambiguity of decision-making involved in assessing student behavior and learning outcomes. The integrative development of both technology and theories would be supported with the help of a possible background. This article offers a (technological) way of evaluating, in which the discussion of whether technology and theories are right or wrong can be overcome by unification of the problematic and the potential into a single system. In order to achieve optimal realization of the student's potential in a particular learning situation, it would be very good to model a design based on combinations of cases that could be realized in a digital context. This can be achieved by developing a specific strategy to combine future research that will be strengthened or weakened according to the influence of artificial intelligence and big data technology.

AI without learning theory is not possible (without guiding), and learning theory without AI is very complicated, because learning is a complex process of achieving goals that comprise multiple components, relations and subrelations. The article demonstrates that it is possible to strike a balance between constructivist and behaviorist approaches to education that could prove an effective foundation for a successful AI education system. Some of the ideas underpinning this view have been borrowed from the computational sciences and their intersection with the neuroscientific understanding of mind and brain.

References

- Anderson K, Petrov V (Eds) (2019). *Traditional Learning Theories, Process Philosophy and AI*. Louvain-la-Neuve: Éditions Chromatika. <https://www.i6doc.com/en/book/?GCOI=28001100747010> abgerufen
- Bates T (2014) Learning theories and online learning. Retrieved from www.tonybates.ca/2014/07/29/learning-theories-and-online-learning/
- Bertalanffy L v (1928) *Kritische Theorie der Formbildung [Critical Theory of Form Building]*. Bortraeger.
- Bertalanffy L v (1967) *Robots, Men, and Minds Psychology in the Modern World*. New York: Braziller.
- Boon H J (2013) Neuroscience for old pedagogy. *Proceedings of Australian Association for Research in Education Conference*. Adelaide, SA, Australia: AARE 2013: Australian Association for Research in Education Conference, 1-5 December 2013.
- Borck C (2006) Lässt sich vom Gehirn das Lernen lernen? Wissenschaftshistorische Anmerkungen zur Anziehungskraft der modernen Hirnforschung. *Zeitschrift für Erziehungswissenschaft*, 5, S. 87–100.
- Bower M (2017) Design of technology-enhanced learning: Integrating research and practice. Emerald Publishing Limited. Von www.emeraldinsight.com/doi/book/10.1108/9781787141827 abgerufen
- Bruner J (1990) *Acts of meaning*. MA: Harvard University Press.
- Conole G, Dyke M, Oliver M, Seale J (2004) Mapping pedagogy and tools for effective learning design. *Computers & Education* 43(1): 17–33.

- Retrieved from
<https://doi.org/10.1016/j.compedu.2003.12.018>
- Dennick R (2016) Constructivism: reflections on twenty five years teaching the constructivist approach in medical education. *International journal of medical education* 7: 200–205. doi:10.5116/ijme.5763.de11. PMID: 27344115; PMCID: PMC4939219
- Dewey J (1916) *How we think*. MA: Houghton Mifflin Company.
- Dubs R (1995) *Lehrerverhalten. Ein Beitrag zur Interaktion zwischen Lehrenden und Lernenden im Unterricht [Teacher Behavior. A Contribution to the Interaction between Teachers and Learners in Classes]*. SKV/Steiner.
- Fuchs T (2021) *Das Gehirn – ein Beziehungsorgan. Eine phänomenologisch-ökologische Konzeption* (6 Ausg.). Kohlhammer.
- Goldstein K (1939). *The organism: A holistic approach to biology derived from pathological data in man*. American Book Publishing. doi:10.1037/10021-000
- Grabowski (2009) ICT as an enabler for effective learning design: Its evolving promise. *International Journal for Educational Media and Technology* 3(1): 12–23.
- Greller W, Drachsler H (2012) Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Educational Technology & Society* 15(3): 42–57.
- Harasim L (2017) *Learning Theory and Online Technologies* (2 Ausg.). Routledge.
- Honebein P (1996) Seven goals for the design of constructivist learning environments. *Educational Technology Publications* 17–24.
- Hung D (2001) Design principles for web-based learning; implications for Vygotskian thought. *Educational Technology* 41(3): 33–41.
- Hung D, Nichani M (2001) Constructivism and e-learning: balancing between the individual and social levels of cognition. *Educational Technology* 41(2): 40–44.
- Kaltenborn R (2019) Embedding the assessment of emotion in the learning process with AI-driven technologies. In *Traditional learning theories, process philosophy and AI*, eds V. Petrov, K. Anderson, 146–166. Chromatika.
- Kao Y, Venkatachalam R (2021) Human and Machine Learning. *Comput Econ* 57: 889–909. <https://doi.org/10.1007/s10614-018-9803-z> abgerufen
- Koohang A, Riley L, Smith T, chreurs J (2009) E-learning and constructivism: From theory to application. *Interdisciplinary Journal of E-Learning and Learning Objects* 5(1): 91–109.
- Lefrançois G R (2006/2019). *Theories of human learning: Mrs. Gribbin's cat* (Seventh Ausg.). Cambridge : Cambridge University Press.
- Murphy E (1997) Constructivism: From philosophy to practice. Retrieved June 2021 from <https://eric.ed.gov/>
- Northoff G (2016) Neuroscience and Whitehead I: Neuro-ecological model of brain. *Axiomathes: An International Journal in Ontology & Cognitive Systems* 26(3): 219–252.
- Northoff G (2016) Neuroscience and Whitehead II: Process-Based Ontology of Brain. *Axiomathes: An International Journal in Ontology & Cognitive Systems* 26(3): 253–277. doi:<https://doi.org/10.1007/s10516-016-9287-1>
- Piaget J (1972) *The psychology of the child*. Basic Books.
- Riffert F (2018) An Introduction to Whitehead's New View of Learning and Its Relation to Traditional Learning Theories. *Balkan Journal of Philosophy* 10(2): 73–88. <https://doi.org/10.5840/bjp20181029>
- Robinson R, Molenda M, Rezabek R (2007) Facilitating learning. In *Educational Technology*, eds M. Molenda, A. Januszewski, 15–49. Routledge.
- Schunk D H (2012/2020) *Learning Theories: An Educational Perspective*. Retrieved from www.pearson.com/us/higher-education/program/Schunk-Learning-Theories-An-Educational-Perspective-8th-Edition/PGM1996609.html
- Thomas H (June 2021) *What are learning theories and why are they important for learning design?* Retrieved from <https://www.mybrainisopen.net/learning-theories-and-learning-design>
- Tobias S, Duffy T (2009) Constructivist Instruction: Success or Failure?
- Vygotsky L (1978) *Mind in society*. MA: Harvard University Press.
- Whitehead A N (1929/1967) *The aims of education and other essays*. The Free Press.
- Whitehead A N (1929/1978) *Process and reality*. (Corrected Edition). The Free Press. Kindle Edition.
- Woodhouse H (2017) Contrasting Views of Emotion in Learning: Alfred North Whitehead and Jerome Bruner. *Interchange* 48:217–230. doi:10.1007/s10780-016-9299-1

LEARNING EMOTION CONCEPTS: FURTHER THOUGHTS ON EMOTIONAL GRANULARITY

Marina Bakalova

Institute for Philosophy and Sociology, Bulgarian Academy of Sciences,
13A Moskovska Str., Sofia 1000, Bulgaria
e-mail: marina.bakalova@gmail.com
<https://orcid.org/0000-0002-2847-1037>

Abstract:

This paper reveals the importance of learning emotion concepts due to the efficiency of emotional granularity during the categorization of emotions. There are two ways of learning emotion concepts that can contribute to emotional granularity. First, we can learn emotion words. Second, we can learn the implicit content of our emotion concepts, i.e. how emotions feel to us. In order to complete the second task, we need to acquire vivid awareness and vivid memory of the implicit content of our emotion concept. I claim that only after completing the second task can we learn emotion words in a way that is efficient for the categorization of emotions. The problem with that claim is that we do not know how to study the implicit content of our emotions, and how to obtain vivid awareness of it. In this article, I sketch a basic solution to this problem.

The article has three parts. In the first part, I outline Lisa Barrett's Conceptual Act View in order to reveal the functional role of emotion concepts in our brains. In the second part, I explain Anna Wierzbicka's classical attempt to define emotion concepts. In the third part, I suggest how it is possible to study the fine-grained details of our emotional experience in a scientific way.

The goal of developing the integrative model is to realize the learner's potential in personalized knowledge formation in an intelligent learning environment and to enhance the efficiency of learning.

Key words: emotion concepts, affective neuroscience, implicit content, emotional granularity, artistic expression.

We value emotional intelligence, since it matters in many aspects of our everyday lives. A prerequisite of being emotionally intelligent is understanding emotion concepts, not just basic ones, but also more complex ones that would enable a person to react properly in situations of greater complexity and to empathize with others. In recent decades, academic research on emotion concepts has been interwoven with research on emotions in general. Separately, there was a burst of interest in defining emotion concepts in the early 1990s among central European researchers such as Anna Wierzbicka (Wierzbicka 1992) and Zoltan Kövecses (Kövecses 1990) which set a standard in the linguistics of emotions. Today, it seems that the purely human-related interest in emotion concepts has been shifted and extended. Understanding them is needed for the sake of building intelligent machines. This would prove particularly useful in areas like social robotics, artificial empathy, etc. where scientists aim at building robot companions and empathizing robots.

Various features of emotion concepts make them particularly difficult to study. For instance, they are not susceptible to classical definitions (Wierzbicka 1992, 549), they are embodied and abstract at the same time (Winkielman P, Coulson S, Niedenthal P. 2018), it is hard to bridge their phenomenology and what we know about the brain, etc. In this article, I would like to address one particular feature of emotion concepts that, it seems to me, underlies many of these problems. Emotion concepts, especially complex emotion concepts, have *implicit* content basically related to how emotions

feel to us. By implicit content, I mean content that is not easy to verbalize because it is difficult to introspect and its production is automatic. It can be lexicalized or non-lexicalized, e.g. when a word for such content is missing. However, even when lexicalized, implicit content cannot be captured in detail by the meaning of a word. In some cases, words imply such content, but they are too vague to actually deliver it. The statement that we can conceptualize without words is largely recognized (see Piccinini, Gualtieri 2011, 186-7, Malt et al. 1999, 2003, Gennari et al. 2002; see also Malt and Sloman 2004). What has remained unrecognized, though, is that even when non-lexicalized, implicit content can be articulated and shared through various non-linguistic means. In the domain of art, there are refined tools for articulating such content, although it has been remarked upon only rarely that artistic tools developed over centuries can be studied as means of articulating the implicit content of our emotion concepts. Bakalova (2021), for instance, has attempted to explain how musical theory is useful along these lines. In this article, I will address and develop some of the ideas sketched there.

1. The role of emotion concepts in the human brain

In order to understand what emotion concepts are we need to take a look at their role in the human brain. Human emotional life is situated generally in two areas of the brain (Barrett, 2017a, 112-128). First, it is situated in the limbic system where the amygdala, hippocampus, etc. interact and prepare the body for action through feeling and affect. We feel an emotion, however, this occurs only after calculations are completed by the frontal cortex. Certain areas of the human frontal cortex calculate what would be the most adequate reaction to the current situation, sending information to the limbic system regarding the appropriate way to respond and act. This calculation depends on what emotion concepts we hold at that moment. This picture of brain activity characterizes emotion as dependent on cognition. It has been developed most prominently in Barrett's influential Conceptual Act View (Barrett 2017b) which we are going to outline here.

In her book *How Emotions are Made*, Lisa Barrett (2017a) explains that the main role of emotion concepts is to constantly regulate one's bodily economy. In short, emotion concepts prompt the most appropriate reaction to what is being perceived by predicting its cost. This means that the very emo-

tions we feel, how we are affected in a variety of situations, could be different for individuals undergoing the same experience, depending on how they perceive and categorize that experience.

Two things are salient in Barrett's analysis. First, emotion concepts help us to construct our emotions. And second, conceptualization of emotions is specific on different occasions, depending on the context. So, emotion concepts are dynamic and changeable entities. Although we have general ideas about certain emotions based on language conventions, stable valence, and arousal level, there are plenty of conceptual variations, degrees, and mixtures of emotional awareness that we construct upon each particular occasion of a relevant perceptual experience. If we have a word for a particular experience we can easily handle it. If not, it would be significantly more difficult, and we might fall into a prediction trap.

Let us take a look at the building blocks of emotional learning. If concepts regulate emotions, it is important to start building rudimentary emotion concepts from early infancy. Due to a lack of concepts, the infant brain struggles to make predictions and is awash in prediction error (Barrett 2017a, 113). In other words, infants do not have adequate, context-dependent emotions. First, babies must learn about the world from sensory input. In the process of that initial learning, their brains start to model the world. The first moves of conceptual learning are structural. The infant's brain must learn to ignore information, to select the relevant information, and to focus on it. By doing so infants obtain the most important tool of conceptual learning – a “spotlight” of attention, to use Barrett's metaphor (Barrett 2017a, 114), that illuminates some things while leaving other things in the dark.

Once this function has been enabled, the brain learns to make coordinated predictions that span the senses. As I have already mentioned, this happens in the cerebral cortex where concepts are represented as statistical similarities separated from sensory differences. The brain represents this separation of similarities as patterns of firing neurons. In the very beginning, all neurons of an infant's brain are linked with each other in a single big network. The structural changes are related to the formation of specialized neural networks, and the loss of a huge number of neural connections in the infant's brain. Conceptual efficiency entails minimal representation, i.e. representing a concept with as few connections as possible.

Perceptual experience, emotion concepts, and affects constantly interact with each other. According to Barrett's constructivist theory, each emotion concept is constructed for a particular situation by finding the most similar and efficient pattern applicable to the situation. The pattern "chosen" for the particular context is called the "winning instance." Winning instances are represented in different patterns of neural activation on each occasion.

To illustrate that, let us consider three instances of joy. Suppose that John entertains most prominently three particular concepts of joy:

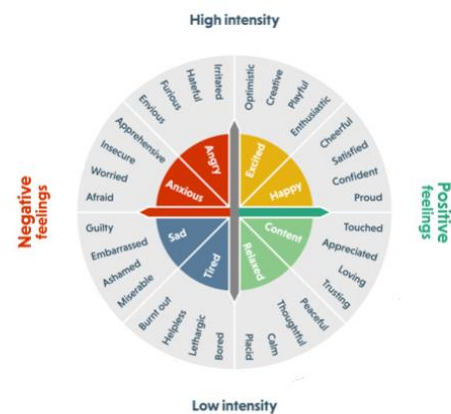
1. "Joy" related to winning a competition that opens the door for the realization of John's dreams for the future.
2. "Joy" of meeting an old buddy on the street.
3. "Joy" of seeing your parent improving after a serious incident.

These three instances of joy are represented as bits and pieces in John's brain. The next time John is in a situation that is likely to trigger joy, e.g. chatting with a girl he admires, his brain will use these previous instances as predictions, and together with predictions regarding other factors in his current situation, his brain will decide whether to categorize that situation as joyful. Suppose that he does so. Such a categorization is realized in virtue of mental similarities among previous instances of joy that his brain has learnt. Together with them, this newly constructed concept of joy will influence the next set of statistical probabilities governing the next prediction in that conceptual field, and so on.

The process of triggering and applying emotion concepts has two directions: bottom-up and top-down. Learning emotion concepts happens during the bottom-up circuitry, and categorizing happens during the top-down circuitry. In the bottom-up stage, sensory input is compressed into efficient, multisensory summaries to be used for categorization (Barrett 2017a, 312-13). These efficient summaries are, so to speak, our concepts. As learning proceeds from very specific to increasingly general concepts, the brain creates similarities that are progressively more efficient summaries of information, i.e. represented with as few neurons as possible. This process reduces redundancy and represents the information in a minimal, efficient form for future use.

Categorization is the next stage that runs top-down. It results in constructing an instance of a concept. Based on the multi-sensory summary from the

bottom-up stage, categorization is unpacked into detailed predictions. Barrett usefully calls this unpacking "concept cascade" (Barrett 2017a, Appendix D, 312-20). The goal of the cascade is to frame the appropriate affect, to categorize the emotion, and to prompt action. A related notion particularly important for the goal of this paper is the notion of "emotional granularity" which Barrett (2006) claims represents efficiency at the stage of categorization. Emotional granularity means constructing precise emotion concepts for one's experience at the end of a categorization chain. To illustrate what emotional granularity is, we can use the following circumplex arousal-valence chart, a version of the original chart offered by Russell and Barrett (1999):



On this chart, each emotion concept is represented by the location of a corresponding word along a structure with two axes representing valence (pleasure-displeasure) and arousal (activation-deactivation). The innermost circle, closest to the intersection point, represents the chart of basic emotions correlated with the two axes. At the outer circle we have more refined concepts that demonstrate emotional granularity, arising from the inner circle's basic concepts. Barrett's idea is that if we have words for very specific feelings, the brain will use less energy categorizing emotions. Having precise words for emotions has been reported to be beneficial for coping with emotional experiences because it allows us to label our emotions more accurately and to deal accordingly (Lindquist, Kristen A.; Barrett, Lisa Feldman 2008, 510).

In her article "The Epistemic Value of Music," Bakalova (2021) has defended a model of artistic expression (to be explained below) that bears similarities to this constructionist idea. In contrast with

Barrett's chart, it is supposed to capture and explain how we conceptualize our emotional experience based on purely implicit, artistically articulated, content. Bakalova claims there that by listening to music, we gain vivid awareness of how different emotions feel to us. This is to say that emotional granularity can work on an implicit level too. Of course the implicit content of our emotions goes hand in hand with language. Other provisions aside, vivid awareness of the details of our phenomenal experiences would be a natural trigger for introducing emotion words into our language. But words do not capture in sufficient detail that implicit content, which, I think, should be studied separately in its own right.

2. Semantics of emotion concepts

In the previous section we discussed the functional components of emotion concepts. From that functional point of view, an instance of a concept is an anticipatory guess about how one should act in the present moment and what her sensations mean. In this section, we will try to explain whether emotion concepts have universal semantic cores.

The main question is: If emotion concepts are so dynamic, how is it possible to maintain and use general concepts of emotions? As I mentioned earlier, emotion concepts are considered embodied and abstract at the same time. In philosophy and linguistics, the issue of the semantics of emotion concepts is usually treated under the assumption that they have abstract meaning, though not in the form of classical definitions. So, perhaps, we can frame the issue in the following way. The semantics vs. functionality of emotion concepts represents a tension between what is going on on the mental level vs. the embodied characteristics of these concepts. If emotion concepts have semantic cores, they are presumably encoded in the abstract part. Then, in what sense are emotion concepts abstract? Winkielman P, Coulson S, Niedenthal P. (2018, 1) claim that the abstract part of emotion concepts has multiple aspects. First, they refer to something created in the mind and second, they generalize to a variety of situations that differ in terms of their surface features. However, abstraction is also related to the embodied part. It captures common features that we encounter on the embodied and purely functional level.

In her pioneering article "Can Emotion Concepts be Defined?" Anna Wierzbicka (1992) provides a recipe for defining emotion concepts along

these lines. She claims that what we abstract on a mental level are complex prototypes. They are based on the similarities of situations provoking certain feelings. These similarities are grouped under so-called "prototypical mental scenarios." Wierzbicka shows in her study that emotion concepts from different languages can be very successfully modeled in terms of such scenarios. For the purpose of formulating her definitions she also assumes that emotion concepts can be defined based on universal semantic primitives such as 'good,' 'bad,' 'do,' 'happen,' 'know,' and 'want.' Just by using them, we can specify each prototypical mental scenario that stands for an emotion concept. She claims:

The small set of universal semantic primitives employed here allows us to show that even apparent synonyms such as sad and unhappy embody different and fully specifiable conceptual structures, and to reveal the remarkable precision with which boundaries between concepts are drawn: even between those concepts which at first sight might appear to be identical or only 'stylistically' different. (Anna Wierzbicka (1992), 578)

Wierzbicka's definitions are a bit long and unusual. But to put that aside, she really accomplishes what she promises. Being able to weave the delicate distinctions of our emotion concepts into her definitions, she accounts for the phenomenon of emotional granularity. I suppose that if we use these semantic primitives instead of the valence-arousal axes of the circumplex chart, we can illustrate a similar order of complexity on the level of mentalist semantic explanation. We will have semantic clusters within each field, and more fine-grained distinctions at the end. Tracking the idea backwards, Kövecses points out:

It was found that people judged fear and worry, joy and love, anger and envy, contempt and disgust, sadness and grief to be most similar to each other and that, consequently, these pairs form the bases of any larger clusters. Two superclusters emerged, one corresponding to what we usually think of as pleasant emotions (including love, joy, etc.), and one corresponding to what we usually think of as unpleasant emotions (including anger, fear, etc.) meaning. (Kövecses (1990), 10)

3. The Implicit Content of Emotion Concepts

What I want to do hereafter is to reimagine the circumplex chart, to see how it works for implicit

conceptualization. When we talk about biological efficiency, we usually have in mind examples in which our reactions need to be fast, such as when one sees a bear. But not all emotions are like that. We feel complex emotions like love, nostalgia, hope, melancholy, etc. The way they feel can be a question of fine-grained implicit conceptualization of events around us. I want to illuminate the fact that these fine-grained details can be scientifically studied. I will try to show briefly how we can model them and study them by using something similar to the circumplex chart.

As a preliminary step, I should clarify the relation between implicit content and implicit attitudes. Implicit attitudes are usually considered to be mental states or alternatively, they are taken to be traits (Machery 2016). Unlike propositional attitudes – beliefs, hopes, desires, etc. – they are simply likes and dislikes. Such likes and dislikes are automatic and hard to examine through introspection. Moreover, a person can maintain likes and dislikes that contradict their explicit claims. For example, one may state “I am a communist-leftist” while at the same time disliking poor people. Attitudes have three main features. First they have valence: They are either positive or negative. Second, they have strength, i.e. a like or dislike can be strong or weak. And finally, they have formal objects – i.e. a like or dislike is of or about something. Now, implicit content has wider reference than implicit attitudes. Implicit content is not determined by a hedonic tone alone. It spreads along multiple dimensions – for instance intensity-mildness, dynamicity and stativity, and so on. For example, the concept “lyrical,” arguably, involves implicit content, including one’s memory of how certain moods, poems, or music feel. These memories encompass more than just like or dislike. Implicit content is semantically richer than implicit attitudes.

Gualtiero Piccinini (2011), who distinguishes between implicit and explicit concepts, suggests that our implicit concepts operate on the basis of similarity relations between stimuli (including not only structural, but also causal and functional features). They encode statistical information about categories, and by doing that, implicit concepts explain typicality effects. Since emotion concepts have implicit (partially-encoding) typicality effects, it is not surprising that linguists tend to explain emotion concepts in terms of prototypes. For my purposes here, I do not need to sharply distinguish between implicit

and explicit concepts. My point is simply that emotion concepts have implicit content.¹

The implicit content of emotion concepts is a part of the conceptual content related to how emotions feel to us. Implicit content may be opaque and difficult to examine through introspection, but nevertheless, we have access to it by different means. For instance, artists can recreate their emotions, feelings, or moods in different media. Let me provide an example. Dan Hartman’s musical album *New Green Clear Blue* is supposed to be a journey within the subconscious. In Hartman’s own words:

The basis of this album was to use tones and shades in certain patterns so the door to a listener’s subconscious would be opened...There’s nothing placed in the music to tell you things. It’s merely my trying to tap the subconscious feelings of people who listen to it...I was unlocking my own subconscious. And the first two or three pieces I listened to I felt were planets away from fulfilling my concept. Little by little it began to flow...I sat at the keyboard and created some sounds that seemed interesting. When I listened back, I was amazed it came out completely as one piece...I played it to my neighbors. Some said it put them in a state where they felt very intimate with themselves. (Campbell 1989, 22)

In short, Hartman managed to trigger his listeners to explore their subconscious without a previously settled convention of how to do that. At first, the music he wrote did not correspond to what he had in mind – in his own words, “a concept”; but then, surprisingly, it came out exactly as he wanted it, and he presumably achieved his goal of communicating his message to the audience.

Along these lines, Green (2007, 192-207) claims that we can express a particular feeling of ours by producing a stimulus with similar characteristics to the feeling itself. Green develops a multi-dimensional model to illustrate this and to explain how such congruence is possible. He suggests that the phenomenology of our mental states may be described along a number of dimensions, including the following three: 1. intense/mild, 2. pleasant/unpleasant, and 3. dynamic/static (Green 2007,

¹ I even think that implicit concepts have underestimated characteristics that bring them closer to explicit concepts. For instance, it is arguable that implicit content can interact with syntax, in particular, musical syntax. See Bernstein & Leonard 1976, 53-119.

179). Now imagine a coordinate system with these three parameters. Consider emotions only, since the model is applicable to a larger class of states. The phenomenal character of each emotion would be represented as a region (or as multiple regions) in that coordinate system. Assuming that emotions have valence and arousal levels, each emotion will have to be represented along at least two dimensions: valence and intensity or valence and dynamicity. For instance, a joy which stands phenomenally in a particular intersection between the three parameters (for example: rather intense, pleasant, and dynamic) may be represented by a region on that coordinate system in an intersection between the three parameters. Alternatively, an emotion might only be two-dimensional – for instance, pleasant and mild but neither dynamic nor static. In this case, it will be mapped with two dots on two regions along the intense-mild dimension and the pleasant-unpleasant dimension.

This three-space model allows us to match an emotion with its sensitive parameters determined by the model, and thereby to make that emotion repeatable and recognizable. It also allows us to find a stimulus with similar characteristics, articulating the affective content of that emotion. Thus, we can create an extended memory of our emotional experience. Words are different – and probably more efficient – tools for determining our emotions at the cognitive level. But these extended memories can be used as referents of emotion words, either as illustrations of already existing emotion words or as referents of new words. I believe that this three-dimensional model would be useful for studying human emotional content in more detail and for programming emotional phenomenology easier within the sphere of artificial intelligence.

References

- Bakalova M (2021) The Epistemic Value of Music. *Organon F* 28(2): 303–325.
- Barrett L F (February 2006) Valence is a basic building block of emotional life, *Journal of Research in Personality* 40(1): 33–55.
- Barrett L F (2017a) *How Emotions are Made: The Secret Life of the Brain*. Houghton Mifflin Harcourt
- Barrett L F (2017b) The theory of constructed emotion: an active inference account of interoception and categorization, *Soc Cogn Affect Neurosci* 12(1): 1–23.
- Bernstein L (1976) *The Unanswered Question: Six Talks at Harvard*. Volume 33 of Charles Eliot Norton Lectures. Harvard University Press.
- Campbell M (1989) Dan Hartman Manages to turn a Career, *Mohave Daily Miner* 6, Mar. 7: 22.
- Gennari, S P, Sloman S A, Malt B C, Fitch W T (2002) Motion Events in Language and Cognition. *Cognition* 83: 49–79.
- Kovecses Z (1990) *Emotion concepts* Springer-Verlag.
- Lindquist K A, Barrett L F (2008) Emotional Complexity. In *Handbook of Emotions*, 3rd edn, eds Michael Lewis, Jeannette M. Haviland-Jones, Lisa Feldman Barrett. New York: Guilford Press.
- Machery E (2016) De-Freuding Implicit Attitudes. In *Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology*, eds Michael Broenstein and Jennifer Saul. Oxford Scholarship Online.
- Malt B C, Sloman S A, Gennari S, Shi M, Wang Y (1999) Knowing versus Naming: Similarity and the Linguistic Categorization of Artifacts, *Journal of Memory and Language* 40: 230–262.
- Malt B C, Sloman S A, Gennari S P (2003) Universality and Language Specificity in Object Naming, *Journal of Memory and Language* 49: 20–42.
- Malt B C, Sloman S A (2004) Conversation and Convention: Enduring Influences on Name Choice for Common Objects, *Memory and Cognition* 32(8): 1346–1354.
- Piccinini G (2011) Two Kinds of Concept: Implicit and Explicit. *Dialogue* 50(1): 179–193
- Russell J A, Barrett L F (1999) Core Affect, Prototypical Emotional Episodes, and Other Things Called *Emotion*: Dissecting the Elephant. *Journal of Personality and Social Psychology* 76(5): 805–819
- Wierzbicka A (1992) Defining Emotion Concepts. *Cognitive Science* 16: 539–581
- Winkielman P, Coulson S, Niedenthal P (2018) Dynamic grounding of emotion concepts. *Phil. Trans. R. Soc. B* 373:20170127

LEARNING WITH ANIMA

Rosen Lutskanov

Institute for Philosophy and Sociology, Bulgarian Academy of Sciences,
13A Moskovska Str., Sofia 1000, Bulgaria
e-mail: rosen.lutskanov@gmail.com

Abstract:

The paper develops a semi-formal model of learning which modifies the traditional paradigm of artificial neural networks, implementing deep learning by means of a key insight borrowed from the works of Marvin Minsky: the so-called Principle of Non-Compromise. The principle provides a learning mechanism which states that conflicts in the processing of data to be integrated are a mark of unreliability or irrelevance; hence, lower-level conflicts should lead to higher-level weight-adjustments. This internal mechanism augments the external mechanism of weight adjustment by backpropagation, which is typical for the standard models of machine learning. The text is structured as follows: (§1) opens the discussion by providing an informal overview of real-world decision-making and learning; (§2) sketches a typology of decision architectures: the individualistic approach of classical decision theory, the general aggregation mechanism of social choice theory, the local aggregation mechanism of agent-based modeling, and the intermediate hierarchical model of Marvin Minsky's "Society of Mind"; (§3) sketches the general outline of ANIMA – a new model of decision-making and learning that borrows insights from Minsky's informal exposition; (§4) is the bulk of the paper; it provides a discussion of a toy exemplification of ANIMA which lets us see the Principle of Non-Compromise at work; (§5) lists some possible scenarios for the evolution of a model of this kind; (§6) is the closing section; it discusses some important differences between the way ANIMA was construed here and the typical formal rendering of learning by means of artificial neural networks and deep learning.

Key words: formal theories of learning, multi-agent systems, principle of non-compromise, artificial neural networks.

§1. There is just one unavoidable question to ask in any real-world setting: "What is the right thing to do?" Thinking should purport to address this question, or at least to tackle some of its facets. It is evident that our ability to answer this question in a satisfactory way depends on the things that we know. Since the things that we know (or even can know) usually are severely limited (and so too our ways of processing what we know), we need to learn. Hence, learning is ubiquitous. Regrettably, learning is also hard. In many contexts we not only don't know what we know, but we don't even know what we *should* know in order to determine the right thing to do. Moreover, different approaches suggest themselves, different cues are salient, different data seem relevant, different concerns loom large. Thus each time we ask what the right thing to do is, we are prone to disagreeing with each other. This is usually seen as a major impediment to learning. If you and I disagree, then each one of us must figure out what this means for his epistemic position and even whether it means anything at all. This complicates immensely our initial situation: Instead of focusing on the available evidence, we must expand our view and ask ourselves what the relative reliability of our ways of processing information is. Thus disagreement can serve as a form of higher-order evidence that changes the standing of first-

order evidence.¹ Here I propose to view disagreement not as an impediment, but as a key ingredient of the mechanism of learning. It provides us with opportunities to assess, evaluate, and rearrange the division of cognitive labor and to weigh our evidence – not just inside our communities, but even inside our own heads.

Many illustrations for this broad thesis suggest themselves. I shall employ one that looms large in our time – the raging COVID-19 pandemic. Each government in the world was confronted by the question, “What is the right thing to do in order to overcome the pandemic?” It is obvious that the right thing to do was (and still is) to establish what the facts are, to determine the most important objectives in this setting, and then to design effective plans to be rigorously implemented. In addressing the pandemic, governments and individuals found the road ahead to be bumpy. Different actors tried their best to influence the answer to this question by participating actively in the deliberation process. The medical community, major economic organizations, and influential political actors wrestled with each other to get the upper hand. They had access to different information, they employed different conceptual machinery, and they were driven by different values; hence, it is unsurprising that their specific perspectives led to completely different suggestions. No common set of basic assumptions or aims was formulated during the course of this process. Moreover, the stance of each of these actors, engaged in a severe battle for the heart of the public, in general was not internally consistent. In the medical domain, epidemiologists, virologists, and immunologists had different takes on the subject, stemming from their divergent points of view. Finally, there were even virologists that disagreed with each other about apparently simple matters of fact: the basic reproductive number of the SARS-COV-2 virus; the virus’ stability under different conditions; the virus’ potential to evolve, etc. What seems evident in this context is that the actors at each level were not sure of the facts, of the right way to assign probabilities to different scenarios, of the scale of potential effects,

or even of the most important objective when devising a plan. Should we impose a strict lockdown in order to thwart the pandemic, or should we keep the economy going? Who should we vaccinate first: the most at-risk groups with the highest mortality rates, or the most active groups with the highest transmission rates? These are typical questions without commonly accepted answers, which continue to confound authorities.

§2. What is the right thing to do in this situation? By “right” we should obviously mean “rational,” since we have no direct access to the objectively right answers to the questions that we’ve discussed above. The classical stance on rationality renders it a form of maximization in a fixed individual setting. Each one of us should figure out the answers for ourselves and then try to help others understand our line of reasoning. When viewed as a trio (A, B, C) of acts A, beliefs B, and consequences C, decisions appear as functions that take a pair of act(s) and belief(s) (a, b) as arguments and return a specific consequence (c).² If we assume that we have a preference ordering on the set of consequences, then obviously the rational decision is the decision that provides the best chance of achieving the preferred outcome. When the preference ordering is specified by a utility function u so that x is preferred to y if and only if $u(x) > u(y)$, and beliefs are represented by probability distributions, the situation is utterly simple: The decision should pick the act that has the highest expected utility. Thus rationality becomes personified by the mystical TOTREP (trade-off talking rational economic person).³ The real question is how to take the rankings that represent the subjective tastes (or values) of each individual and aggregate them into a unified ranking which represents the right way to act for the society as a whole. Indeed, since no one can bring forth the desired consequences alone, we all need to work together in order to achieve the common good. If we turn back to our initial example, we can note that obviously we all need to wear masks, and we all need to vaccinate, if we want to achieve the desired outcome. Thus individual rationality should be conceived as a building block of collective rationality. When scientists started to study such aggregation mechanisms which

¹ This way of rendering the problem of disagreement is typical for a recently established domain in social epistemology called “epistemology of disagreement.” For disagreement as a form of “higher-order evidence,” cf. (Kelly, 2010). There are different points of contact between the approach to be presented here and this stripe of social epistemology. The analogy shall be pursued further in a follow-up paper.

² The notation employed here follows chapter 1 (“Revealed Preference”) in (Binmore, 2009).

³ This acronym was introduced by Mike Harrison and appeared in print in (Kreps, 1988).

reflect social preferences, based on individual rankings of alternatives, something utterly awkward happened. The story is well-known, so here we shall discuss only the essential facts.

Kenneth Arrow asked what the right way to aggregate individual preferences in a democratic society is. He formulated a list of natural requirements (“axioms” and “conditions”) which such an aggregation procedure has to satisfy. Then, he proved that there is no procedure that can achieve this when the situation is complicated enough (in fact, much less complicated than any real-world scenario that we can think of). His general impossibility theorem shows that each sensible aggregation procedure of this kind shall be dictatorial, which means that it is going to mirror the preferences of a particular member of the society.⁴ Since this individual is not guaranteed to be more intelligent or insightful than the others, the whole procedure reduces to a form of lottery, where you more or less randomly pick one individual and invest her with the authority to decide for everybody. (The “dictator” whose existence the theorem proves is not invested with such power on the basis of her intrinsic properties; she just happens to have the right sort of preferences). Thus, the effects of the “wisdom of the crowd” disappear completely, and the manifestations of this aggregation procedure (the voting and market mechanisms) have no added value.⁵ If each member of the society is rational (which is obviously desirable) and the society is fair (which means that it puts equal weight on the preferences of each of its members and does not exclude some preference rankings by convention), then the society collapses to just one of its members. Whether this means that “democracy” is impossible (as suggested by a popular (mis)interpretation of the theorem) is hard to say. What is obvious is that Arrow has exposed a severe limitation of simple aggregation mechanisms. This forces us to look elsewhere and to add some complexity to the picture.

Somewhat later, a different paradigm emerged. Its progenitor was Thomas Schelling, who imagined society as a group of interacting clusters of hetero-

geneous agents. The agents are heterogeneous since they process information in essentially different ways. They are clustered, since they interact locally – no grand aggregation mechanism is at play, but each group tries to manage the situation from its limited point of view.⁶ If successful, then obviously its impact on the society as a whole shall rise; if not, it shall become limited and can even vanish completely. This sounds closer to our running example with the COVID-19 pandemic. The different groups had to figure out for themselves what their priorities were before going public and declaring a stance. Thus, after some initial deliberation, the medical community as a whole started to advocate strict measures (despite some dissenting individuals that played an interesting role – a role that we shall have something to say about later), while different economic organizations and decision-makers actively opposed them. No general aggregation mechanism of the kind that Arrow has analyzed was implemented: We did not vote for a particular policy (as far as I know, no nation-wise referendum was implemented anywhere in the world).

On the other hand, the agent-based models (as they came to be called later) of Schelling and his followers don't provide a mechanism for social learning. Each agent learns from her own decisions by interacting with her own cluster of neighbors. New patterns of social behavior can emerge and stable policies can be enacted, but this stems from individual interactions alone. Hence, the system as a whole is always on the brink of disequilibrium – even small perturbations can push it in a completely new trajectory. In general we would like to think that society is not like this, that its history grants stability to the way it is functioning. Knowledge corresponds to a set of stable dispositions to act, hence learning should bring forth such dispositions. The responses to the COVID-19 pandemic were exactly like that. For example, take one of the most popular measures – the quarantine. It definitely was not a new invention, but a strategy that was employed on many different occasions and reflects a learned, top-down social behavior in the presence of infectious diseases. The name comes from “*quaranta giorni*” (forty days), and the practice was already established in Venetian policy around 1377 AD.

Here we have briefly reviewed Arrow's global aggregation model and Schelling's local aggregation model. Marvin Minsky's “society of mind” model

⁴ The classic reference is (Arrow, 1963). For an up-to-date discussion of this topic, cf. (Maskin and Sen, 2014).

⁵ Wisdom of the crowd is a well-known effect in psychology, which was (re-)invented by Galton. He noted that the average value of a set of independent individual assessments is usually closer to the real value. What is particularly interesting in the present setting is that the same effects can be established (diachronically) even with a single individual – hence we shall talk about “the wisdom of the inner crowd” (Herzog & Hertwig, 2014).

⁶ The *locus classicus* here is Schelling, 1978. A typical state-of-the-art publication is Epstein, 1996.

can be seen as the middle ground between these two extremes.⁷ Minsky envisaged a hierarchy of heterogeneous agents which have a clearly delimited view of the world. The fundamental principle determining the operation of this structure was the so-called “Principle of Non-Compromise,” which states that if the subordinates of a particular agent are in conflict, then the agent loses its standing and its “weight” is diminished.⁸ If I get different answers by employing two methods of the same kind, the information provided by both methods is degraded, since the disagreement is seen as a clear mark that these methods are not reliable. As far as our previous example is concerned, we can clearly see how this works: Since the medical community did not reach an early consensus about the proper way to handle the pandemic, because there were vociferous dissenters, the policy to be adopted was largely determined by other (largely economic and political) factors. The reason why Minsky’s model seems important here is that it lets us see disagreement not just as an impediment, but as one of the driving forces of social learning. The aggregation of different perspectives is stepwise and at each step disagreement (or conflict in general) has the same role to play: as a reliability cue, which shows what adjustments should be made so that this higher-level evidence is incorporated into the functioning of the system.

The problem is that the Principle of Non-Compromise is purely qualitative. It does not say how much the weight of the output of a set of agents (i.e. the weight of their superordinate) should go down if they are found to disagree, or even how to measure their conflict. Minsky’s idea was to provide a conceptual framework in which these questions could be posed; he did not strive to provide us with all of the necessary details. That is why we can perform a simple thought experiment and try to imagine how he would have tackled this problem from our present point of view. What would be the exemplary “scruffy” way of doing this, the way that is best suited to explicate the multi-agent way of thinking that Minsky has championed in all of his writings? Trying to answer these questions takes us away from

his original position.⁹ Minsky was not a big fan of artificial neural networks – in fact, some of his works were a big hurdle for their future development (in particular, his classical theorem that exposes some of the limitations of simple perceptrons¹⁰). What we are going to sketch below vaguely resembles an artificial neural network, although ideologically it is somewhat different. Our motivation is to try to imagine how a “society of mind” neural network would look. The most important differences between our sketch (the ANIMA model) and artificial neural networks (ANN) shall be outlined in the last section of this paper.

§3. ANIMA is an acronym that stands for Agent-Network Interaction Model of Adaptation. As the name aims to suggest, the basic building blocks of the model are simple entities called agents. Agents are things that act, or implement specific functions. They exhibit no intelligence, or purposeful behavior, although the system composed of these entities could be perfectly able to do that. Moreover, agents are related to each other in a complex hierarchical network with different layers that integrate and transmit information. Each agent competes with other agents at the same layer, integrates the agents below it and branches to agents above it. The crucial fact about agents is that they are heterogeneous – agents in one and the same layer with one and the same inputs can have completely different outputs. The system as a whole adapts to some features of the external environment, and can thus be said to learn about them. Treating learning as a form of adaptation suggests that the end result depends on the features to be learned: If the features are stable and the information about them is reliable, then the system is bound to reach an equilibrium state of maximal performance, which depends on its complexity. If they

⁷ The ground-braking approach of Minsky that diverged significantly from the orthodoxy was elaborated in (Minsky, 1986; a watered-down version was later presented in Minsky, 2006).

⁸ The original statement of the principle is: “The longer an internal conflict persists among the agent’s subordinates, the weaker becomes that agent’s status among its own competitors” (Minsky, 1986, p. 33).

⁹ Minsky believed that measuring things by numbers is the best possible way to forget about their meaning: “recall the popular children’s tale about three bears. The bears themselves are usually perceived as Two and One-Momma and Poppa Bear, plus Baby Bear. But their forbidden porridge bowls are seen as quite another sort of Three: too hot, too cold, and then just right; a compromise between extremes” (Minsky 1986, 192). Thus, numbers are nothing, but a fictional “currency of thought” that makes everything look the same. I would object that if numerical information is processed locally, then its meaning is also local, so the argument fails.

¹⁰ Cf. in particular Minsky and Papert, 1969, in particular ch. 13 (“Perceptrons and Pattern Recognition”)

are not, it can be perturbed and head in a completely different direction. This is not a defect but a feature, since learning by definition must be open-ended. In order to explain how ANIMA is supposed to work, let me start with the simplest possible case, and then successfully add complications that shall turn our initial sketch into a more realistic picture. Each level-1 agent has access to limited information about the environment and has to figure out its “preferences” for the future behavior of the system as a whole on the basis of this information. You can imagine that each agent at this layer assigns to each possible act a real number in the closed interval $[0, 1]$, which we can view as representing the utility of this act for the agent, or even better, as the probability that this act would be implemented if it were up to the agent to decide. Hence, the “utilities” assigned to different acts should obey the axioms of probability theory, in particular, the sum of the utilities assigned to different acts should add up to 1.

The utilities assigned to each act by different agents in general shall be different since the agents have access to different information, and even if they have access to one and the same information they can process it differently (as we claimed above, they are heterogeneous). Now, imagine that you have n level-1 agents which have formed their preferences concerning the feasible acts of the system. What should the system do? Of course, you can conceive this as a voting problem and render the aggregation of different preferences as a social choice rule, which picks a collection of individual preference orderings and assigns to them a social preference ordering, representing the preferences of the system as a whole. This is not a good way to proceed, for several different reasons. The most important reason is spelled out by Arrow's general impossibility theorem, as was briefly discussed above. Hence we are forced to add level-2 agents that integrate the information from level-1 agents. Each level-1 agent can be linked to several level-2 agents, hence branching occurs at this point. Level-2 agents should assign different weights to the agents below them, thus level-2 agents rank acts but they also rank level-1 agents. This can be spelled out in the idiom of decision theory by saying that level-2 agents have beliefs about level-1 agents and those beliefs determine the weights that they assign to their preferences. Now you have a multiplicity of level-1 agents that branch into a multiplicity of level-2 agents. Just as before, you can introduce a multiplicity of level-3 agents that integrate level-2

agents. Things this time are going to be even more complex, since level-3 agents can take into account not just the level-2 agents that they see directly, but also level-1 agents below them. The level-2 agents that share more level-1 agents below them shall be able to form “coalitions” (when integrated by a level-3 agent). Obviously, this is going to have an impact on the way level-3 agents assign weights to level-2 agents below them. Each coalition shall demand greater weight for its members and less weight for the members outside it. Thus, level-3 agents have much harder things to do: Along with acts they weigh not just single agents, but coalitions of agents.¹¹ In the same way, you can add more levels and get more complexity of interaction.

Things get interesting when all agents all the way up the ladder are initialized in this way and the system as a whole picks an act on the basis of the inputs received and the weights assigned by means of this process. The act has consequences that can be evaluated and employed in order to modify the system (this can be seen as an analogue of the back-propagation phase in artificial neural networks). Imagine that the act is picked by a single level-4 agent that oversees several level-3 agents. After this learning episode, the system compares the evaluation of the act with some predetermined aspiration level (hence its behavior can be seen as a form of satisficing). If the evaluation is above the aspiration level, then the level-4 agent awards (increases the weight of) the level-3 agents whose probability distributions on the set of acts were close to his own (obviously, since they were closer to the “right” answer) and punishes (decreases the weight of) all other agents. The higher the weight of an agent with the wrong probability distribution, the more severely it gets punished (since it contributed more for going in the wrong direction). Then each level-3 agent should distribute the effects to its subordinates: The awards and punishments should go all the way down to level-1 agents. If a single level-2 agent is linked to a pair of level-3 agents, then the effect on it should be determined by their weights: the greater the weight, the greater the impact. This is the basic structure of ANIMA. The model develops by adding

¹¹ Here I'm intentionally employing concepts from cooperative game theory, since it provides formal machinery by means of which you can calculate the weights of higher-level agents (e.g. the Shapley value). Because of the limited space available here, I'm not going to pursue this analogy further here. For more information on the theory of cooperative games, cf. Aumann & Shapley, 1974.

meta-agents that oversee hierarchies of the kind sketched above (synchronic pattern detectors), collections of hierarchies of this kind, undergoing sequences of learning episodes of a particular kind (diachronic pattern detectors), or even different kinds of learning, executed simultaneously by different systems (global pattern detectors). In this way, patterns that emerge in hierarchies with poor performance can be detected and modified and hierarchies with good performance can be selected for the task. Something like evolution by natural selection can be implemented by means of such meta-agents.¹²

§4. Before we continue, let's make things a little bit more precise (but not too much!). Take a set E , which represents all relevant data, all forms of evidence that may serve as a justification for the adoption of a particular policy, and a set P of all possible (practically implementable) policies. Of course, we can disagree what data should be treated as relevant, or what policies should be treated as practically implementable, but let us for the present moment put this difficulty aside. Now take a collection F of partitions of E which contains the trivial partition (with a single cell which coincides with E) and is totally ordered by the refinement relation (a partition p is said to refine a partition q if and only if each cell of p is contained in a cell of q). Thus, we shall have a finest partition at the bottom which corresponds to maximal fragmentation of data, and a succession of coarser partitions which ends at the trivial partition. Then take a collection of agents A and assign them to the cells of the partitions in F . The cell assigned to a particular agent corresponds to the information that it is able to respond to. It would be practical to assume that each cell of each partition in F is occupied by an agent, which means that each fragment of evidence is attended to. If we don't assume this, we shall need to make provisions that guarantee that if all cells subsumed by a cell of a coarser partition are occupied, then the subsuming cell is also occupied. This corresponds to the idea that all low-level information should be aggregated at a higher level. Obviously, if we assume that this is the case, and moreover assume that all of the cells of the finest partition are occupied, then the assumption that all cells of all partitions are occupied becomes super-

fluous. Finally, we shall assume that if there is a group of agents that occupy a particular cell (if the assignment is a surjective function), then there is a higher-level agent that aggregates their outputs only. This shall guarantee that information does not get mixed too early and some form of encapsulation is present. We can also postulate that the highest-level partition (the trivial one) is occupied by a single agent, which represents the aggregated stance of the society as a whole. The policies that it ranks highest on the basis of the input from its subordinates are those that are going to be implemented in practice. Thus, the society of mind becomes personified at the highest level.

Now assign a probability distribution to the set of policies P to each agent in A . This distribution reflects the take on the situation, which the agent bases on its limited view. The probabilities represent the preferences of the agent: If a policy a gets higher probability than b , then the agent believes that the probability that a is the best policy in this situation is higher than the probability that b is the best policy in the same situation. We shall assume that agents are heterogeneous, which means that even if they are assigned to one and the same cell of some partition, i.e. if they have one and the same evidence, they can have different preferences based on this evidence. Moreover, we shall associate with each partition in F a collection of weights for all of its (occupied) cells (for technical purposes it would be convenient to normalize weights so that the sum of the weights of all agents at some level is equal to 1; we shall see how this is supposed to happen). Those weights represent the power balance between the agents. The weights themselves can be based on the previous track record of agents (if agent x has ranked higher the policy a than the policy b and it turned out that carrying out a was good enough, or that it was better than carrying out b , then x receives higher weight than any agent y , if y ranked policies a and b in the opposite way). What we miss up to this point is an explicit relationship between the behavior of subordinate agents and the behavior of their superordinate agents. This is crucial for the approach sketched here, since it provides us with a way to implement the Principle of Non-Compromise in the present setting.

What the Principle of Non-Compromise says is that the divergence of the probability distributions assigned to the subordinates of a particular agent should affect the weight that this agent receives. Here we can discern between two different ques-

¹² Of course, this introduces features from the theory of evolutionary algorithms, cf. Holland, 1992. This is another topic that we are not going to pursue further, since we are focusing on elementary, entry-level systems.

tions: First, how the probability distribution of x should be formed on the basis of the probability distributions of its subordinates; second, how the weight of x should be formed on the basis of the weights of its subordinates and the divergence of their outputs. Before we see how this could be implemented, let us sum up what we have up to this point: (1) a set E of available pieces of evidence (data); (2) a set P of implementable policies; (3) a set A of cognitive agents that evaluate policies on the basis of evidence; (4) a collection F of partitions of E (fragmentations of data); (5) an assignment of agents to cells of partitions in F (i.e. to points of view which determine the data that each agent takes into account when calculating its evaluation of policies); (6) a collection of probability distributions on P , indexed by the lowest-level elements of A (evaluations of policies by agents, assigned to the finest partition, which are partially determined by their respective points of view, or by available information); (7) an assignment of weights to lowest-level elements of A (which respects the technical normalization requirement that was mentioned above); (8) an aggregation function which takes as arguments an agent x and the probability distributions of the subordinates of x and returns the probability distribution of x ; (9) an aggregation function which takes as arguments an agent x , the weights of the subordinates of x , and a measure of the divergence of the probability distributions assigned to the subordinates of x , and returns as value the aggregated weight of x . These are all basic building blocks of the model.

Let us consider a simple instantiation of this model. We shall assume for the sake of simplicity that the assignment is injective for all levels except the lowest one, i.e. that only low-level agents can have access to one and the same evidence and process the evidence differently. This assumption is nice, since it makes the arithmetic easier and puts an upper bound on the number of agents which depends on the number of cells in the lowest-level partition. We shall have just 2 pieces of evidence, e_1 and e_2 , and 2 implementable policies p_1 and p_2 ; 6 agents, a_1 , a_2 , a_3 , a_4 , a_5 , and a_6 , 3 partitions of $\{e_1, e_2\}$, viz. $\{f_1 = \{[e_1], [e_2]\}, f_2 = \{[e_1], [e_2]\}, f_3 = \{[e_1, e_2]\}\}$, an assignment of a_1 and a_2 to $[e_1]$ in f_1 , of a_3 to $[e_2]$ in f_1 , of a_4 to $[e_1]$ in f_2 , of a_5 to $[e_2]$ in f_2 and of a_6 to $[e_1, e_2]$ in f_3 . This means that we have a pair of agents seeing one and the same evidence (a_1 and a_2), an agent that aggregates their outputs (a_4), an agent that simply transmits the output of a_3 (a_5),

and an overarching agent (a_6), who has to make the final verdict. Now, assume that a_1 assigns 0,7 to p_1 (0,3 to p_2), a_2 assigns 0,6 to p_1 (0,4 to p_2), and a_3 assigns 0,5 to p_1 (0,5 to p_2). As you can see, all probabilities sum up to 1, which means that all agents treat the policies as both incompatible and exhaustive of all possibilities and obey the requirements of minimal rationality, imposed by the axioms of the probability calculus. Now we have to initialize the weights of the lowest-level agents. Since this is the first step in the learning process, the principle of indifference suggests to put equal weights to all cells in the finest partition so that their sum is 1 and then split the weight of each cell between the agents that occupy it. This means that we assign a weight of 0,25 to a_1 and a_2 and a weight of 0,5 to a_3 . We have thus passed through steps (1)-(7), and the last thing we need to do is to specify (8) and (9). For (8) the simplest possible procedure is to use the weighted averages of the probabilities assigned by the agents of the lowest level. Arithmetic mean seems a natural choice, although other variants are available (along with the other two Pythagorean means – the geometric and the harmonic – there are also lots of “generalized means” that can be applied¹³). Since the weights of a_1 and a_2 are the same, a_4 shall assign 0,65 to p_1 (the mean of 0,7 and 0,6) and 0,35 to p_2 (the mean of 0,3 and 0,4). Moreover, since the subordinate of a_5 is just a_3 , it simply assumes its distribution, so it still assigns 0,5 to both p_1 and p_2 . What about the weights of a_4 and a_5 ? The answer for a_5 is trivial: It has a single subordinate, so its weight is not going to be revised at this step. It remains 0,5. For a_4 the situation is somewhat more complicated. Different routes are possible, but again there is a natural way to handle this case (since there are just two subordinates; when there are more than two, different equally justifiable options are available). The weight of a_4 should obviously depend directly on the weights of its subordinates (more subordinates with greater weight means more weight for the superordinate), and inversely on the divergence of their evaluations (more difference means less weight). The number of the subordinates of a_4 is 2 (they are a_1 and a_2). Their weights are both 0,25, so their sum is 0,5. The mean difference between the values assigned by a_1 and a_2 to p_1 and p_2 is 0,1 (here we have just two pairs of values that sum up to 1, so the arithmetic is particularly simple). We shall

¹³ This is a broad and surprisingly interesting topic; for further information I suggest to the interested reader the comprehensive overview provided in Bullen, 2003.

employ 0,1 (or 10%) as a correction coefficient that is employed to modify the weights of conflicting subordinates. For a4, since 10% of 0,5 is 0,05, the reduced weight shall be 0,45. In the extremal case when a1 assigns probability 1 to p1 (0 to p2), while a2 assigns probability 0 to p1 (1 to p2), the reduction shall be 100%, so the weight of their superordinate becomes 0; this corresponds to our intuition that complete divergence of evaluations is a sign for complete lack of reliability. In this case the superordinate is just shut down and leaves the game, or it gets some preassigned threshold value. The second option seems better, since it lets us calculate weights without complications, in the way we have done so before. Moreover, in this way we can be sure that the whole structure is not going to crumble and no low-level agent gets complete control over it, since all its competitors are shut down.

So far, so good. Now we have reduced the weight of a4, which affects our arithmetic – the sum of the weights of a4 and a5 becomes less than 1. This suggests that the 10% percent of the weight of a4 that were reduced have to be redistributed to the other agents of the same level, so that the sum of their weights is still 1, as was the case at the lowest level. This means that we transfer these 10% to a5, which is the only other agent at this level, hence the weight of a5 becomes 0,55 ($0,5 + 0,05$). Now we come to the highest-level agent in our mini-hierarchy, i.e. a6. Its probability distribution again shall be determined by the weighted sum of the probabilities assigned by its subordinates: a6 assigns to p1 probability $0,65 \times 0,45 + 0,5 \times 0,55 = 0,2925 + 0,275 = 0,5675$ and to p2 probability $0,35 \times 0,45 + 0,5 \times 0,55 = 0,1575 + 0,275 = 0,4325$. The first step of the process is finished. Now let us see how close were the subordinates of a6 to these values, i.e. what was their impact on the end result.

Something interesting has happened. The distances between the distribution for a6 and those for a1, a2, a3, a4, and a5 are respectively 0,1325, 0,0325, 0,0675, 0,0825, and 0,0675 (the values for a3 and a5 are identical, as can be expected, since a5 is just a channel that transmits information from a3). Although the position of a4 was weakened because of the conflict of its subordinates, and a5 clearly dominates, a2 has come closest to the end result, i.e. its impact was largest, followed by a3 and a4. Agent a1 is farthest from the end result; it is an outlier whose impact was virtually neglected by the others, and this fact damages the position of a4. The conflict between a1 and a2, as well as the fact that their av-

erages are close to the assignments of a3, made a2 a winner.

Before we continue let us see what was going to happen if the values for a2 were shuffled, i.e. if it got a mirror evaluation of the policies p1 and p2. In this case the conflict between it and a1 is clearly larger. Just as before, a1 assigns 0,7 to p1 (0,3 to p2), and a3 assigns 0,5 to p1 (0,5 to p2), but this time a2 assigns 0,4 to p1 (0,6 to p2). The weights of a1 and a2 are still 0,25, and the weight of a3 is 0,5. Agent a4 shall assign probability 0,55 to p1 (0,45 to p2), but its weight is going to be reduced by 30%, i.e. now it is merely 0,35, while the weight of a5 goes up to 0,65 because of the redistribution of weights. The probabilities for a3 and a5 remain unaffected. Finally, we come to a6. The agent assigns to p1 probability 0,5175, hence 0,4825 to p2. Now the distances between the distribution for a6 and those for a1, a2, a3, a4, and a5 are respectively 0,1825, 0,1175, 0,0175, 0,0325, and 0,0175. Again there is something that deserves to be noted. The agent a3 dominates by almost completely obliterating the impact of a1 and a2. The reason for this is that their evaluations are grossly inconsistent. Moreover, exactly because their conflicting evaluations are close to the mean value, the standing of their superordinate is much better than theirs, and its distance from the end result is reduced more than two times. This shows that the Principle of Non-Compromise captures some of the truth (as suggested by the first case), but not the whole of the truth (as suggested by the second case). The conflict is not decisive; what is truly important is how each of the conflicting groups of agents on the same level stand with respect to one another.

Something like this can happen for various reasons, not just because of a major conflict between low-level subordinates. In order to see how this can happen, take the original case and exchange the evaluations of a1 and a3. Thus a1 assigns 0,5 to p1 (0,5 to p2), a2 assigns 0,6 to p1 (0,4 to p2), and a3 assigns 0,7 to p1 (0,3 to p2). Furthermore, a4 assigns 0,55 to p1 (0,45 to p2), and a5 assigns 0,7 to p1 (0,3 to p2). The weight of a4 is 0,45 (reduced by 10% because of the minor conflict between a1 and a2), and the weight of a5 is 0,55 (upgraded by redistribution). Finally, a6 assigns 0,6325 to p1 (0,3675 to p2) and the offset between this and a1, a2, a3, a4, and a5 is respectively 0,1325, 0,0325, 0,0675, 0,0825, and 0,0675. We get the same distribution of assets as in the first example, although the final values are different. What is decisive here is not the low-level

conflict between a1 and a2, but the higher-level conflict between a4 and a5. Finally, it is easy to see what is going to happen if in the last example we switch the values for a3. By going through the same arithmetic as before we get respectively 0,0875, 0,1875, 0,1125, 0,1375, 0,1125. Here the conflict between a4 and a5 is much larger, so the end result of the aggregation (the values for a6) does not come close to the evaluations of the lower-level agents (from a1 to a5). Agent a1 fares better than all the others (and a2 comes worse), but the fit does not come close to the values in the previous examples. This is a case where no low-level agent is an effective estimator of the end result. This can be taken to mean that the hierarchy as a whole has intrinsic value; moreover, no “dictators” are present.

Up to this point we have a completely static architecture, which is not able to learn, i.e. to adapt the weights of the low-level agents in such a way as to achieve higher ranking for those policies that are objectively better. On the other hand, we have a tool that can help with this task. In fact, a pair of tools. We have already seen that each agent has a different impact on the end result. This suggests that after we pick a policy from the highest-scoring group (ties are possible because of the way we have defined our framework) and execute it, we can evaluate the end result and see if it passes some previously defined benchmark. If the result is “good” (above the benchmark), then we need to award (upgrade the weight) of those agents that contributed most for achieving it and punish (downgrade the weight) of those agents that were opposed to it (ranked it lower than some other less productive policy). If the result is “bad” (below the benchmark), then we do exactly the opposite – we need to punish those agents that contributed most and award those agents that were opposed to it. Both the “distance” from the benchmark (how good or how bad was the result) and the “distance” from the end result (respectively, how big or small was the impact of the agent on the final evaluation) should matter.

Again there is a natural way to tackle this issue. Let us assume that a6 has systematically picked up a policy that ranks 20% above the benchmark (i.e. the performance of the system was good on average). The offset of a4 is bigger than the offset of a5, so a4 is going to be punished, while a5 is going to be awarded. Each of them passes this feedback to its subordinates. In particular a4 needs to distribute the punishment between a1 and a2, while a5 passes the award to a3. Since the result is 20% above the

benchmark, a3 has its weight upgraded by 20% which means that it becomes 0,6. Correspondingly, the accumulated weight of a1 and a2 should be 0,4. Since approximately 80% of the accumulated offset of a1 and a2 is associated with a1, it shall get 80% of the punishment, while a2 gets just 20%. Now the weights of a1, a2, and a3 are respectively 0,17, 0,23, and 0,6 and the iteration can proceed just as was sketched before. If things proceed in the same way, then a3 is going to gain dominance, while a1 shall gradually fade away. The same calculations can be performed for the other three cases that we have examined above.

Let us also note something interesting. In some cases the conflict of subordinate agents is counter-productive – when it downgrades the weight of the agent that integrates them and the agent is awarded, the award it gets is smaller. In other cases it is productive – when it downgrades the weight of the agent which is going to be punished, the punishment is smaller. In cases of counterproductive conflict the agents are motivated to minimize their conflict so that the premium they get from their superordinate is upgraded. If those agents have access to one and the same kind of information (i.e. are assigned to one and the same cell in the partition), they can be regarded as epistemic peers and according to a well-known doctrine (the so-called equal-weight view in the epistemology of disagreement) they should split the difference between them. This means that each one of them should move to some extent toward the average value. *Mutatis mutandis*, in cases of productive disagreement the agents are motivated to remain steadfast, i.e. to keep their initial positions.¹⁴ This seems pretty obvious: An organization which has less impact because of internal conflicts and remains relatively unaffected, even though it got things wrong, has less incentive to diminish its internal conflicts and try to upgrade its influence, in comparison to an organization which got things right but could not reap the benefits because of this fact. Being indecisive can be a virtue, especially when you are wrong. On the other hand, splitting the difference when the information you have is not identical in general is not a good idea: There is no intuitive way to justify it, especially since we have seen that conflicts, disagreements and divergent evaluations can be productive for the system as a whole. A system composed of different agents with different preferences and different ways to learn is at the same

¹⁴ For a standard exposition of the “equal weight” and the “steadfast” view, cf. ch. 3 & 4 in Matheson, 2015.

time robust and able to adapt, when things change. It is much better than a homogeneous system which has stable output but no internal resources to modify it when it is forced to adapt.¹⁵

§5. We have examined a collection of simple scenarios that show how a system of this kind can work and aggregate lower-level evaluations of policies by assigning weights to intermediate-level agents and modifying them internally on the basis of the magnitude of conflict between their subordinates. The performance of the system suggests possible scenarios for evolution of the system (since the weights of case 1 and case 3 are the same, we are going to explore them under one and the same heading). First, let us see what could happen if the system reliably picks policies that are good (i.e. passes the benchmark). In case 1 (and 3) the level of fitness (as measured by its offset from the final distribution, calculated by the overarching agent a6) is exceptionally poor for a1. Learning under such conditions can lead to significant reduction of its weight (which can boost the weight of a4, since a2 is going to have greater impact over it). In case 2 the level of fit is extremely good for a3 and extremely poor for a1 and a2. Since their evaluations balance each other, the standing of a4 is relatively good, hence no significant structural modification is needed. On the other hand, if a3 comes to dominate in the long run, the structure shall be simplified by weeding out a1, a2, and a4. In case 4 all agents perform poorly, although a2 is exceptionally bad. Since no one of them has a really significant impact on the right decision reached by the system as a whole, we should look for other pieces of information and exchange a2 with another agent which functions differently (we can imagine a pool of “spare” agents waiting for deployment). This shall transform the power balance between the different agents and can lead to significant changes.

Now, let us see what could happen if the system consistently fails to pass the benchmark. In case 1 (and 3) a1 is going to be rewarded most, while a2 is going to be punished, which can make a1 dominate and reduce a4 to the role of a channel that transmits information from a1 to a6. In case 2 a3 is going to be severely punished along with a4; if this condition

persists then a3, a4, and a5 can be eliminated since their weights are going to plummet and the higher-levels of the hierarchy are going to collapse. Thus a stable configuration can be reached where a1 and a2 are directly linked to a6. In case 4 agent a2 is going to gain dominance, together with a4, while the role of a1 is going to be reduced and it can be eventually eliminated. This is what we mean by “adaptation”: Not just the weights of the agents and the calculations that are based on them are going to change, but the whole structure mutates so that it matches more effectively the power balances inside the system. If we consider a collection of hierarchies, these structural changes, or “mutations,” can be rendered as a procedure that starts with one structure and substitutes it with another structure from the pool. Thus an artificial evolution by natural selection which is driven by goal-directed mutations can occur.

Of course, these scenarios don't exhaust the possibilities for the evolution of the system. Even at this low level of complexity it exhibits a behavior that can be described as deterministic chaos – small perturbations in the initial conditions can lead to completely divergent evolution and markedly dissimilar end states. What is important to keep in mind is that the described shifts of complexity are not irreversible: for example, if the system becomes simplified but its performance deteriorates, some previously discarded agents can be reused (even with their old default values) and their reattachment shall push the evolution of the system in a completely new direction. Thus in the long run the complexity of the system shall be balanced with the complexity of the task that it needs to solve: If the number of agents is below the minimal number of agents necessary for optimal performance, more agents shall be added; if it is above the minimal number, some agents are going to be discarded. Thus learning is not just weight adjustment of individual agents, it is first and foremost structural self-modification that does not reduce to but is driven by weight adjustment.

The same structural evolution is bound to occur in real-world decision-making. The governments around the world established task-force units to deal with the challenges of the COVID-19 pandemic. Those teams gradually evolved: More people got involved when the original teams ended up with mixed results and were judged to have performed inadequately. At the same time decision-makers that didn't have an impact on the choice of policies to be implemented gradually moved away. Finally, everybody watched what the others were doing and tried

¹⁵ I see this as a direct consequence of Minsky's dictum that “If you understand something in only one way, then you scarcely understand it at all – because when you get stuck you'll have nowhere to go” (Minsky, 2006, 6).

to copy the best practices and the best structural organization for this particular task. Of course, this is not something specific for this case: People have done this since the dawn of time. Both the social networks we are nested in and the neural circuits in our brain have changed over the course of time. Our artificial models cannot remain rigid if they are tasked with matching the performance of their natural counterparts.

§6. Up to this point we have sketched an outline of an ANIMA-based model of learning which differs from the now predominant paradigm of artificial neural networks (ANN) and deep machine learning. Here are the key points of divergence between these paradigms:¹⁶ (1) ANN is based on graph-theoretical concepts: Vertices are artificial neurons, edges are connections between them (both provided with weights). This affords great flexibility in the actual construction of the system, hence a plethora of designs, stemming from different engineering principles (rightfully described as a “zoo” by the practitioners).

The ANIMA framework is based on partition refinement, which effectively generates a tree-like structure (or a collection of trees with identical leaf-nodes if the assignment of agents to cells of the partition is surjective all the way up). Thus the complexity of the structure is bounded by the fragmentation of information at the lowest level of integration (and the maximum level of redundancy in the internal nodes). This is a marked contrast, since deep learning explicitly embraces the idea that the intermediate layers in the network can be much bigger (in fact, indefinitely larger) than the input layer; (2) ANN operates with a mechanism of weight adjustment which is completely external: When you initialize the input layer, you fix the values of the internal layers by forward propagation and then revise them after the result was computed by back propagation. This back and forth rhythm is supposed to lead in the end to a configuration that gets things right.

Compare this with ANIMA. Here we have both internal and external mechanisms: The internal mechanism is weight reduction in the presence of conflict (non-compromise), and the external mechanism operates in more or less the same way as in the

artificial neural network paradigm; (3) ANN does not supply a mechanism of structural evolution. The artificial neurons in the internal (“deep”) layers of the network can go on and off, but they are not supposed to come in and out of existence, which makes the structure rigid. ANIMA is provided with a mechanism of structural evolution and is supposed to undergo more radical transformations. Clearly, the most important defects of ANN are that (1) there is no justification for the complexity of its models (except their performance); (2) there is no explanation why these weights (of connections) are the “right” ones; (3) there is no modification of the system envisaged in advance. The ANIMA model strives to rectify them by (1) associating complexity with data fragmentation and integration; (2) basing weights on the relevance of the available data and the level of internal conflict; (3) envisaging a mechanism of performance-driven structural modification. Moreover, there are many different ways in which the framework can be augmented. In each step of the exposition of the ANIMA framework we have made more or less arbitrary decisions (which were justifiable in the context of the sample structures we were working with, but could turn out to be impractical in another setting). What is important are not the purely formal differences between ANN and ANIMA, but the ideological differences, which, although subtle, can effectively lead us to different ways of viewing the whole enterprise. The most straightforward generalization would be to implement more complex aggregation procedures that are better suited to the task than simple arithmetic means.

Many branching paths lay before us. Each of them shall lead us away from the present-day orthodoxy, which is focused on the end result – on getting a working model which we don’t understand (even if in principle we can’t understand why it is working). The vague similarities between the real neural networks in our brains and their artificial replicas are not enough to make this condition tolerable. We need a psychologically sound model that reflects some higher-level introspectively accessible features of our thinking. The framework that Minsky has sketched aimed to provide such a model. His Principle of Non-Compromise is thought to encapsulate such a feature which is pervasive and robust. What I’ve tried to do here is to put some flesh on the bones without deviating too much from the original insights that motivated the introduction of the model. The end result is a novel form of artificial neural

¹⁶ Here I’m stressing the differences, but there are (at least superficial similarities) between the two approaches that we compare: cf. the “committee machines” or “modular networks” in ch. 7 of Haykin, 1999.

network (or even better, a novel conceptualization of artificial neural networks), which (probably?) is able to perform better because it is able to learn more effectively. ANIMA (or something like it) is the ghost in the shell that makes things happen. The uniqueness of human learning stems from the structural features of the hierarchy (or, even better, the heterarchy) that implements it and the dynamics of its self-modification.

References

- Arrow K J (1963) *Social Choice and Individual Values*, second ed. New York: John Wiley and Sons.
- Aumann R J, Shapley L S (1974) *Values of Non-Atomic Games*. Princeton: Princeton University Press.
- Binmore K (2009) *Rational Decisions*. Princeton: Princeton University Press.
- Bullen P S (2003) *Handbook of Means and Their Inequalities*. Dordrecht: Kluwer.
- Epstein J M, Axtell R (1996) *Growing Artificial Societies. Social Science from the Bottom Up*. Washington, DC: The Brookings Institution.
- Haykin S (1999) *Neural Networks: A Comprehensive Foundation*, second ed. Upper Saddle River, NJ: Prentice Hall.
- Herzog St M, Hertwig R (2014) Harnessing the Wisdom of the Inner Crowd. *Trends in Cognitive Sciences* 18(10): 504–506.
- Holland J H (1992) *Adaptation in Natural and Artificial Systems*. Ann Arbor: The University of Michigan Press.
- Kelly T (2010) Peer Disagreement and Higher Order Evidence In *Disagreement*, eds R. Feldman, T. Warfield, 111–174. New York: Oxford University Press.
- Maskin E, Sen A (Eds) (2014) *The Arrow Impossibility Theorem*. New York: Columbia University Press.
- Matheson J (2015) *The Epistemic Significance of Disagreement*. Houndmills: Palgrave Macmillan.
- Minsky M (2006) *The Emotion Machine. Commonsense Thinking, Artificial Intelligence, and the Future of Human Mind*. New York: Simon & Schuster.
- Minsky M (1986) *The Society of Mind*. New York: Simon & Schuster.
- Minsky M, Papert S (1969) *Perceptrons. An Introduction to Computational Geometry*. Cambridge, MA: MIT Press.
- Kreps D (1988) *Notes on the Theory of Choice*. Boulder: Westview Press.
- Schelling T (1978) *Micromotives and Macrobehavior*. New York: W. W. Norton.

ON SOME ASPECTS OF THE EUROPEAN KNOWLEDGE ECONOMY

Petar Iliev

Institute for Philosophy and Sociology, Bulgarian Academy of Sciences,
13A Moskovska Str., Sofia 1000, Bulgaria
e-mail: petar.iliev@gmail.com

Abstract:

We survey some of the consequences of the managerialism of science and education as practiced in European countries and its main results: the Bologna declaration and the Bologna process. One notable feature of this process is the gradual introduction of currently fashionable management terms into school curricula. These linguistic changes are the direct result of the neoliberal philosophy behind the concept of knowledge economy, namely, that all sciences must justify their economic value; moreover, the introduction of such terms is via government “strategic documents” that are never the result of a democratic public debate. As a case in point, we examine two such strategic documents issued by the Bulgarian ministry of education.

Key words: Bologna declaration, education, educational competences, knowledge economy.

1. Introduction.

Here is a story of an 8th grade student I have met. The student had exceptional “raw brain power” manifesting itself in a very quick grasp of mathematical ideas among many other things, like being proficient in four languages and having, I would say, a somewhat unhealthy fascination with Hamlet. Moreover, the student had the looks and the class markers associated with old money families, like beautiful manners and extensive international experience acquired already by the age of 14. The student had studied in a famous and very expensive school and was transferring to another expensive school for family reasons. The problem was that the new

school was in a different country and one of its entrance requirements was passing a math exam. The parents had decided that it would be a good idea for their child to receive a lesson or two from someone familiar with the level of mathematics required by the new school and, through a mutual acquaintance, they contacted me. It became clear very quickly that, as far as problem-solving abilities required at present-day schools in the new country were concerned, the student was exceptionally gifted and was going to have no difficulty whatsoever passing the entrance exam (as, indeed, turned out later to be the case). The only problem was that the student had no real understanding of the meaning of the formula that we use to obtain the area of a triangle. The following questions were met by the student with something approaching fear and horror. Why do you measure the area of figures in square centimetres, square metres, etc., but you measure length in centimetres, metres, etc.? What is the meaning of the word “square” in “square centimetres”? What do “area” and “volume” mean? What does it mean to actually measure things like area and volume? It turned out that no one had told the student that to measure length involves taking some unit of length, like a centimetre, and then checking how many times we can fit this unit of length into the length of the object we are trying to measure; what is more, the fact that we can take a square of, say, side 1 centimetre because we find the square shape easy to work with and use the area enclosed by it to measure the area of a two dimensional object by, roughly speaking, counting the squares of this type that we can fit into the area of the object came as something

of a revelation. After that, I made the methodological suggestion: “Let’s develop some intuition about the area of two dimensional figures and how we can measure it using unit squares. What is the easiest possible figure whose area you can measure by directly counting the number of squares you can fit into it?” The answer – “a rectangle” – came almost immediately followed by the surprised look on the student’s face when the student suddenly realised why we multiply the lengths of the two sides of a rectangle to get its area. “So, it was simply to get the number of squares we use as a unit of measure for the area!?” To my question “Well, since you know how to compute the area of a rectangle, can you use this knowledge to obtain the area of a right-angled triangle?” The student made the astonishing discovery of using the diagonal of a rectangle, and thus splitting its area in half, to find the area of a right-angled triangle. Another discovery was made when the student finally understood “why on earth do we multiply the height by the base of a triangle and divide by two?” It is because we can use our knowledge of finding the area of right-angled triangles to solve the problem of finding the area of an arbitrary triangle. From then on, it was smooth sailing to the realisation that we use “cubic” in the units with which we measure volume because, intuitively, we are trying to see how many times we can fit the volume of a unit-of-measure-cube into the volume of the three dimensional shape whose volume we are trying to find.

To recapitulate, we have an extremely bright kid, who has passed all exams with the highest marks, who has travelled a lot, who is extremely (and very expensively) well-educated and “well-credentialed”, who has all the “competencies”, usually defined in managerial speak as

1. a set of individual performance behaviours which are observable, measurable and critical to successful individual and company performance,

2. individual characteristics of a person which result in an effective and superior performance in a job see, for example, (Management Study Guide),

demanding at school and who is doing simply what is expected, namely, performing at the highest level at exams as a preparation for performing “at the highest level in a corporate or business setting”, and who realised with the utmost astonishment that subjects taught at school can make sense and that they are not a painful list of formulae and pronouncements you should memorise and apply to mindless problems,

but rather something that can bring intellectual satisfaction and which revolves around understanding fascinating concepts that – oh, the horror – do not have immediate economic and business relevance and practical applications.

2. *Technical preliminaries.*

A caveat: I have no institutional training in pedagogy; I have no formal certificate saying that I have received the pedagogical training that is required to teach certain subjects, and I am somewhat doubtful that there is a well-defined, non-trivial (i.e., different from self-evident, common sense principles) and entirely separate-from-the-subject-we-are-trying-to-teach body of knowledge that we can call “pedagogy”, i.e., I am inclined to think that, for example, teaching students to play the piano is different from teaching mathematics, and that both disciplines require different teaching approaches and principles – but above all, they require mastery of the piano or mathematics, respectively. However, I suspect that my worst offense is that I believe that no amount of pedagogical courses, educational management training, and educational technologies can compensate for dull personality, and that students’ interest and success in a given discipline is too often the result of interacting with an interesting person teaching the subject combined with the not-always-pure-and-healthy desire to emulate and be like this mature person they find so interesting. In short, it seems to me that there is something very important and seemingly unquantifiable at present that goes on between a student and a teacher and that a student’s interest and success in a given subject is due, to an uncomfortably large degree, to the teacher’s personality traits that are completely unrelated to the subject they teach. Needless to say, I have not heard of (pedagogical or psychological) research on the subject of how to be an interesting and larger than life personality. Having said that, let us see what impression one is left with when one is trying to learn about contemporary European education and scientific institutions.

There seem to be several parties involved in the education process – with students, parents, and teachers being the passive party that seem to have no or negligibly little say in what happens to them; or, perhaps it is better to say that they have been conditioned into thinking that what is going on at the moment is the inevitable and universally correct response to the state of the modern world, which in

turn, is the only possible and sensible state of things. Then we have university professors of pedagogy, philosophy, sociology, psychology, neuroscience, and various experts that do research on aspects of learning and education, write papers and grant applications, and, occasionally, sit on various committees that are supposed to advise the government or different international bodies on (the national) schooling policy and strategy. Then we have the designated government or international organisations and agencies that actually produce and implement the (national) policies; above them all is the Bologna process that is the fountain of national policies that are actually not national, if by national we mean something that is the result of a transparent national debate that involves the active participation of the citizens of the country. It seems that one cannot arrive at an informed opinion about the state of education and research on education at a national level without consulting the huge pile of program declarations and communiqués that are the visible, often boring, and invariably maddening part of the Bologna process. Let us have a closer look at these elements.

3. The Bologna process.

Anyone interested in the history of the Bologna process can start by consulting Wikipedia and the sources cited therein, and the eye-opening article by Chris Lorenz (Lorenz, C. F. G. (2006)). Suffice it to say that

The Bologna Process is a series of ministerial meetings and agreements between European countries to ensure comparability in the standards and quality of higher-education qualifications. The process has created the European Higher Education Area under the Lisbon Recognition Convention. It is named after the University of Bologna, where the Bologna declaration was signed by education ministers from 29 European countries in 1999. (Wikipedia. Bologna Process)

The main guiding idea behind the Bologna process is the neoliberal dogma that the Market (whatever that means) is king. Everything should be efficient, mobile, innovative, disruptive, subject to structural reforms, and finally and most importantly a marketable commodity. Therefore, education, knowledge, and scientific research must be evaluated according to their immediate business and economic relevance, which entails the existence of knowledge economy and an education market where

1. scientists compete for grants by pitching the economic value of their research;
2. universities are trying to climb up international academic ranking tables (compiled by business entities with unclear objectives) and competing for attracting high paying international (mainly Asian) students;
3. academic hires and promotions are based on metrics like the h-index (Einstein's h-index would have been depressingly small; see for example (Gingras, Y., Khelifaoui. M., *Albert Einstein the mediocre*));
4. university students strive to build impressive CVs to become part of the human resources of big international corporations;
5. schoolchildren are subjected to ever increasing numbers of tests, exams, curricula changes; and finally, but most importantly,
6. there must be huge administrative bodies where real scientists are not allowed to enter and that are responsible for the external quality assurance process, i.e., imposing the next reorganisation of school education that is no doubt always motivated by new societal and business demands and everyday economic realities, and that always results in more testing (preferably in the form of multiple choice tests that can be automatically graded by the next wonderful and magical educational software so that the role of teachers is reduced in order to get rid of the pesky subjectivity factor when evaluating the educational attainments of students that are not heirs of rich families) and more orientation toward practical applications, punishing non-marketable research and science according to the prevailing administrative idea of marketable research at the moment.

Of course, the general public is unaware of the fact that the concept of marketable research entails the existence of hype cycles, among many other things, where, in order to get funding, scientists start to hype the importance and outcomes of their current research, which in turn leads to increased funding, media attention and inflated promises. After that usually comes the disappointment from the unfulfilled public promises followed by administrative cuts in research funding, which leads to researchers leaving the field. One of the well-known instances of such vicious cycles of boom and bust are the so-called AI winters whose low point was in the early

1990s when research in Artificial Intelligence was severely underfunded because of out-of-this-world promises made earlier which turned out not to be true. We are in the middle of another AI hype period at the moment when funding for AI is abundant and the area is fashionable again, but funding for research in cryptocurrencies has been quietly dwindling after the initial hype was followed by the realisation that maybe cryptocurrencies are not such a good idea after all. One can only wonder about the fate of, for example, all the new PhD students and postdocs hired to do research in the area.

The Bologna declaration was preceded by the Paris declaration (1998) and succeeded by a number of ministerial declarations and communiqués with the last one being the 2020 Rome Communiqué. They can be accessed via the web site of the European Higher Education Area (EHEA. *Ministerial Declarations and Communiqués*). The language of these declarations seems to evolve from blunt economic considerations in the beginning to nauseating corporatespeak in the more recent documents. While the earlier documents contain direct references to concerns about the competitiveness and the global attractiveness of European higher education, especially in comparison to USA, the 2020 Rome communiqué is a lesson in corporate verbiage containing pronouncements like

Higher education institutions will continue to diversify their learning offerings and innovate in contents and modes of delivery in order to respond to growing needs for innovative and critical thinking, emotional intelligence, leadership, teamwork and problem solving abilities, as well as enterprising attitudes.

Flexible and open learning paths, part of the original inspiration for the Bologna Process, are important aspects of student centred learning and are in increasing demand in our societies. In addition to full degree programmes, many higher education institutions offer or plan to offer smaller units of learning, which enable learners to develop or update their cultural, professional, and transversal skills and competences at various stages in their lives. We ask the BFUG to explore how and to what extent these smaller, flexible units, including those leading to micro-credentials, can be defined, developed, implemented and recognised by our institutions using EHEA tools.

We commit to the development of open science and education to facilitate the exchange of knowledge and openly licensed materials that can be easily shared among higher education stakeholders, who can adapt and repurpose them for their needs.

To make student-centred learning a reality, by: supporting higher education institutions in their efforts to start or maintain a structured dialogue on innovation and enhancement of learning and teaching, involving students, teachers and also relevant external stakeholders.

The issues addressed should include the development of curricula, learning outcomes, assessment and quality assurance, with due consideration for the skills needed to address current and future challenges of society; promoting, within disciplinary courses, the development of soft skills, and the skills necessary for the enlightened citizen of the 21st century, in order to understand the challenges of a complex world and respect others and the environment; exploring ways to stimulate the cross-border exchange of good practices for supporting the enhancement of quality learning and teaching, emphasizing active methods;...(Rome Ministerial Communiqué (2020) p. 6)

I have no idea what micro-credentials are and I have never heard of a high societal demand for flexible and open learning paths, part of the original inspiration for the Bologna Process. I did not know that the term *stakeholder* can be non-sarcastically used in the educational process and I am at a loss how promoting, within disciplinary courses, the development of soft skills is going to be implemented; however, I do think that promoting soft skills when teaching mathematics, or philosophy or physics or music, for example, sounds like a huge waste of time that should be spent on teaching science and art. I have no idea how to emphasize active methods nor what they are, and I openly admit that I had to look up the meaning of *transversal skills*. I was only mildly surprised to discover that there is no real agreement on the meaning of the term but that it is supposed to include *Critical Thinking, Innovative Thinking, Reflective Thinking, Reasoned Decision Making, Communication Skills, Collaboration, Self Motivation, Grit, Diversity, Intercultural Understanding, and Respect for the Environment* (Heron, C., *What are Transversal Competencies and why are*

they important). If this sounds like a list of corporate buzzwords, it is because it is a list of corporate buzzwords.

The quotes above are only marginally less offensive than the thousands of examples of corporate speak one hears and reads every day. I give a random example of such business language below in order for the reader to appreciate the lack of substantial difference between the way corporations and ministers of education speak. A friend of mine who works in a big international financial institution has informed me that she hears nothing but such pronouncements during business hours.

Only if you have the core competencies will you be able to auction the key deliverables for your relevant stakeholders going forward.

In short, it does seem that education and science are viewed mainly in business and corporate terms and that they should be subjected to strict corporate quality assurance; moreover, the original guiding idea behind the Bologna process is profit that should be obtained, among other things, by having the EU universities compete with the US universities for attracting part of the foreign students who contribute billions to the economy of the United States (see for example (IIE. *Economic impact of international students*)).

What I could not find while reading numerous documents stemming from the Bologna process are references to historical, sociological, linguistic, and cultural studies that answer the following questions: Why is American education considered so successful? Does its success have something to do with historical facts like, for example, the huge influx of European-educated scientists as a result of the two world wars? Are the persistent jokes in academic circles that American universities are institutions where Russian professors teach mathematics to Chinese students justified? Does the popularity of English have something to do with the success of American, British, Canadian, and Australian universities? Do European universities have to increase their tuition fees to match the huge tuition fees of the majority of elite US universities? What about the social costs of the unequal access to education that is the direct result of vastly increased tuition fees? Would professors be under pressure not to fail high-paying students (anecdotal evidence suggests that this might be the case but I am not aware of any serious systematic study of the issue)? Why are professors in European universities vastly underpaid on average in

comparison to their American colleagues? How do the ministers actually intend to motivate young and established scientists to not leave academia?

The current strategy seems to be based on “good practices” revolving around an increase in: workloads, competition for ever decreasing number of permanent academic positions, the number of years one has to spend as a postdoc if one is seriously trying to obtain a permanent academic post, the pressure to constantly apply for outside funding, and to work in areas that are considered fashionable at the moment, and to not waste time on problems that are too difficult to lead to a publication in a year. Of course, if salaries have to be raised (for political reasons like approaching elections, for example), then the increase should be as small as possible and dependant on “performance, accountability, quality assurance, and best practices” that are measured in terms of grant income, the number of publications, citations, h-indices, etc., which are nothing more than business-like criteria leading inevitably to wide-spread “best practices” designed to game the system of which the majority of scientists are not only painfully aware, but employ regularly simply to survive professionally. Would the ministers acknowledge that the increase of mental health problems in academia is a natural outcome of the quality assurance process (Shaw, C., Ward, L., Dark Thoughts: why mental illness is on the rise in academia)?

4. The role of the government in the national education policies.

Since one of the explicit aims of the Bologna declaration is

...the adoption of a system of easily readable, compatible and comparable degrees, also through the implementation of the Diploma Supplement, in order to promote European citizens' employability and the competitiveness of the European higher education system (EHEA. Bologna declaration (1999) p. 3),

one does begin to wonder whether European citizens, especially of the smaller and poorer European countries, have a substantial say in what is happening to their education systems. It is clear however, that the managerial language in which the various documents connected to the Bologna declaration are written is being adopted on the level of national governments, too. Needless to say, adopting this new language goes hand in hand with adopting the

political decisions and strategies that were expressed in this language in the Bologna declaration and the subsequent communiqués.

Since Bulgaria is a signatory to the Bologna declaration (despite the fact that the name of the minister Anna-Maria Totomanova who signed the declaration on behalf of Bulgaria is misspelled Anna Mmia Totomanova (EHEA. *Bologna declaration* (1999) p. 5), we are in many ways justified in assuming that the political currents shaping Bulgarian education are representative of the general developments in the whole of Europe. When one starts reading the Bulgarian “National strategy for scientific research development 2017–2030” (Ministry of Education and Science of Bulgaria. *National strategy for scientific research development (2017–2030)*), one is immediately convinced that the corporatespeak (with the underlying profit-oriented thinking) of the Bologna declaration has triumphed over standard Bulgarian. Said strategy makes for depressing reading mainly because of the enormous tension between the very competent and hard-hitting analysis of the current state of Bulgarian science and Bulgarian scientists on the one hand and the measures that are touted as solutions on the other. It is as if the strategy was written by two teams with completely different – some might say opposite – agendas: The first team gave ample evidence about the sad state of Bulgarian scientific institutions due entirely to the appalling politically motivated cuts to funding implemented by a number of successive governments, while the other copied all the buzzwords and corporate jargon of the Bologna declaration and disregarded all the findings of the first team. Of course, all the proposed solutions revolve around making an increase of salaries and funding (the amount of money and the timing are left conveniently unspecified, needless to say) depend on performance, competition, number of publications and citations while glancing over the glaring lack of political will for proper funding. We read the obligatory mantra that increasing funding for science is very good for business and will no doubt lead to an increase of foreign investment (Ministry of Education and Science of Bulgaria. *National strategy for scientific research development (2017–2030)*. p. 24). There is the ominous sounding idea that the funding for scientific research in underdeveloped subjects will be dependent on some unspecified strategic needs made explicit in ИЧИК (Ministry of Education and Science of Bulgaria. *National strategy for scientific research development (2017–2030)*. p. 27). I could not find

anywhere in the whole document what the meaning of the abbreviation ИЧИК is but I suspect that it may refer to Innovative strategy for intelligent specialisation 2014–2020 (Ministry of Economics of Bulgaria. *ИЧИК (2014)*) which sounds ludicrous in both standard Bulgarian and standard English; however, it is in complete agreement with the corporate jargon of the Bologna declaration. Nowhere in the whole document do we find references to peer-reviewed and published scientific studies that corroborate the idea that running scientific and educational institutions according to the neoliberal corporate philosophy is beneficial to science, scientists, and society as whole. In fact, we have peer-reviewed and published scientific articles that contain results which have been corroborated and duplicated by independent teams of researchers that imply quite the opposite. For example, (Pluchino, A., Biondo, A., Rapisarda, A.) is a study of the so-called 80:20 phenomenon, i.e., the observable fact that 80 percent of the wealth is owned by 20 percent of the people, which seems to apply to all societies. The meritocratic mantra, which is of utmost importance to the neoliberal thinking, explains this by positing that the rich 20 percent deserve all their riches because they are exceptionally hard-working, intelligent, and, undoubtedly, possess soft and transversal skills in extreme quantities. Of course, there is always some amount of luck involved but, surely, it does not explain why billionaires are so obscenely rich. There are numerous gaping holes in this argument that are never challenged in mass media publications and political discussions nowadays – the main one being that no one has observed a 20:80 distribution of abilities, like IQ for example, in human populations.

Pluchino, Biondo, and Rapisarda developed a computer model of wealth creation that accurately reproduces the 20:80 wealth distribution but it tells a completely different story. Namely, *The maximum success never coincides with the maximum talent, and vice-versa* (Pluchino, A., Biondo, A., Rapisarda, A.). Even more shockingly to human resources managers, the main explanation for having a lot of money is luck. This is of the utmost significance for science and education, too, as succinctly summarised in (MIT Technology Review (2018, March 1)).

Pluchino and co study this from the point of view of science research funding, an issue clearly close to their hearts. Funding agencies the world over are interested in maximizing their return on investment in the scientific world. Indeed, the European Research

Council recently invested \$1.7 million in a program to study serendipity—the role of luck in scientific discovery—and how it can be exploited to improve funding outcomes.

It turns out that Pluchino and co are well set to answer this question. They use their model to explore different kinds of funding models to see which produce the best returns when luck is taken into account.

The team studied three models, in which research funding is distributed equally to all scientists; distributed randomly to a subset of scientists; or given preferentially to those who have been most successful in the past. Which of these is the best strategy?

The strategy that delivers the best returns, it turns out, is to divide the funding equally among all researchers. And the second- and third-best strategies involve distributing it at random to 10 or 20 percent of scientists.

In these cases, the researchers are best able to take advantage of the serendipitous discoveries they make from time to time. In hindsight, it is obvious that the fact a scientist has made an important chance discovery in the past does not mean he or she is more likely to make one in the future.

A similar approach could also be applied to investment in other kinds of enterprises, such as small or large businesses, tech startups, education that increases talent, or even the creation of random lucky events (MIT Technology Review (2018, March 1)).

Would these findings result in a radical change of research funding and educational policies for example? I seriously doubt it. In Lorenz' memorable words

In summary, 'knowledge economy' nowadays does not mean the old idea that knowledge is systematically applied to economy or society, but the very new idea that all 'scientific' knowledge worthy of the name first has to prove its 'economic' value. (Lorenz, C. F. G. (2006)).

I could not find anything in the Bulgarian National strategy for scientific research development 2017-2030 that opposes this way of thinking. Quite on the contrary! This view of knowledge economy permeates the whole document. It also permeates the "Strategic framework for the development of educa-

tion, schooling, and study in the republic of Bulgaria 2021 – 2030" (Ministry of Education and Science of Bulgaria. *Strategic Framework for the Development of Education, Schooling, and Study in Bulgaria (2021-2030)*).

A clear executive summary of this strategic framework is provided in the quotes below.

More systematic effort is needed for implementing the transition from knowledge teaching to the acquiring of key competencies.

*The main goal of teaching is to make learning more attractive and oriented to practice and to build attitudes towards lifelong learning through the application of new methods and approaches. Optimization of standards, curricula and programs is carried out in order to connect more closely the knowledge and skills acquired in school and their practical applications (Ministry of Education and Science of Bulgaria. *Strategic Framework for the Development of Education, Schooling, and Study in Bulgaria (2021-2030)*p. 9).*

Unlike the authors of the above strategic framework, I subscribe to the old-fashioned view that the aim of education (which is usually very different from formal schooling) is a never-ending struggle to acquire discipline of intellect and mastery of knowledge in languages, fundamental sciences, and arts; what is more, discipline of intellect and mastery of knowledge should not be motivated by references to practice because they are worthy goals for their own sake. You cannot replace mastery of knowledge, appreciation of beauty and discipline of intellect with "competencies" for many reasons. Least of all, "competences" sound cheap. The kid about whom I wrote in the introductory section is a highly competent student who could competently solve geometric problems related to finding the area of a triangle without knowing the meaning of the formula for the area of a triangle. This kid was taught competencies at school and discouraged from searching for deeper understanding, but this very same kid was delighted to find that formulae are meaningful and not just tools with practical applications, chief among them being passing one's exam with the highest possible grade.

Many political leaders do, however, seem to think that competencies are much more desirable for the majority of people than mastery of knowledge and discipline of intellect. Woodrow Wilson was probably historically the most influential among

them because the educational reforms he initiated set the course of American education during the 20th century. In his address to the New York City High School Teacher Association on January 9, 1909, he said:

Let us go back and distinguish between the two things that we want to do; for we want to do two things in modern society. We want one class of persons to have a liberal education, and we want another class of persons, a very much larger class, of necessity, in every society, to forgo the privileges of a liberal education and fit themselves to perform specific difficult manual tasks. You cannot train them for both in the time that you have at your disposal. They must make a selection, and you must make a selection. I do not mean to say that in the manual training there must not be an element of liberal training; neither am I hostile to the idea that in the liberal education there should be an element of the manual training. But what I am intent upon is that we should not confuse ourselves with regard to what we are trying to make of the pupils under our instruction. We are either trying to make liberally-educated persons out of them, or we are trying to make skillful servants of society along mechanical lines, or else we do not know what we are trying to do. (High School Teachers Association of New York).

It seems to me that current educational strategies that are part of the Bologna process, especially those aimed at elementary and high schools are knowingly or unknowingly designed with the aim of applying Wilson's idea of making *skillful servants of society along mechanical lines* in practice. The only difference is that children are not trained to *perform specific difficult manual tasks* but to fit seamlessly into the knowledge economy. Note that there are so many implicit and explicit philosophical assumptions in the above quote for which no justification is offered. Why is it necessary for a large part of the society to forgo the privileges of a liberal education? The word *privileges* is very telling. However, do we really have anthropological, economical, historical, or sociological studies that corroborate the necessity for a society as whole to be split in two classes with the much larger class made to *forgo the privileges of a liberal education*? What is it that goes wrong if a society is not split like that? Why did Wilson make the brutal distinction between *liberally-educated*

persons and *skillful servants to society*? Does it have something to do with the fact that a liberal education has historically been the preserve of the ruling classes? Does it have something to do with the fact that by delegating the training of skillful servants to the school system, the big corporations are saving a lot of money that, otherwise, they would have to spend on professionally training their employees? Does it follow that a liberally educated person cannot be a skillful servant to society? Does it follow that a skillful servant to society stops being a skillful servant once he/she is allowed to have access to a liberal education? And finally, if liberally educated people are different from the useful slaves to society, what is their societal role? Obviously, they cannot be both useful and slaves, so one is left with the impression that they are either not useful or not slaves (or perhaps they are neither). What is the use of having a small class of liberally educated persons? I have no palatable answer to the last question but I do have the nagging suspicion that

1. We in the sentence

We are either trying to make liberally-educated persons out of them, or we are trying to make skillful servants of society along mechanical lines, or else we do not know what we are trying to do.

refers to rulers of society that come from the liberally educated class; and

2. skillful servants should not be taught knowledge but competencies (it is never mentioned though that the majority of the skillful servants are poor lest they make a connection between competencies and small salaries).

5. The role of university pedagogy, philosophy, sociology, psychology, and neuroscience

The role in public education played by these disciplines is extremely difficult to decipher. To be precise, it is unclear how much the prevailing political ideology shapes the research in these sciences and what the influence of these disciplines over the prevailing political ideology is. I am by no means a specialist in these extremely difficult matters. However, there does seem to be anecdotal evidence for two major currents although, no doubt, with some exceptions.

1. Writings in these disciplines that are in line with the prevailing political and social climate are subject to dissemination in academia and society at large and, thus, often

result in practical changes on the level of schools and universities (and, thus, society as a whole) whereas lines of thinking that go against the set of prevailing political ideas, and thus hurt scientists' career advancement, do not attract attention outside small circles of marginalised specialists.

2. It is very difficult to delineate convincingly the subject matter proper of the university major of pedagogy because a huge number (I would say the majority) of disciplines taught at pedagogy departments can be successfully labelled as courses in subareas of philosophy, sociology, psychology, economics or current government talking points.

One of the strongest pieces of evidence in support of the first current is the fact that we are still feeling the influence of the Prussian education system, which crystallised as a result of the Prussian Reform Movement (a political reaction to the defeat of the Prussians by Napoleon I at Jena-Auerstedt in 1806) of the nineteenth century and the huge role played by the philosophers Wilhelm von Humboldt and Johan Gottlieb Fichte in this reform and in the subsequent shaping of western education. There are many excellent scholarly studies of this phenomenon and the interested reader may want to consult the relevant pages in Wikipedia and the references cited therein (Wikipedia. *Prussian Reform Movement*).

As far as the second current is concerned, we read in "Strategic framework for the development of education, schooling, and study in the republic of Bulgaria 2021 – 2030",

The state requirements for acquiring the professional qualification "teacher" have been updated. The curricula of the higher schools are being updated in order to introduce the competence approach in the education of the students majoring in pedagogy. Courses such as "Competence approach and innovations in education", "Inclusive education", "Information and communication technologies in teaching and working in a digital environment" and others have been introduced (Ministry of Education and Science of Bulgaria. Strategic Framework for the Development of Education, Schooling, and Study in Bulgaria (2021-2030) page 6).

One does not have to be very paranoid to suspect that the courses mentioned in the above quote

will be used for propagandising the goals of the Bologna declaration. I would go even as far as to suspect that the lecture notes for these courses will be full of corporate jargon of the type we find in the Bologna declaration. I have not been able to find studies and experiments that justify the so-called competence approach. I am even at a loss as to the meaning of the notion of "competence approach." I have been unable to find any scholarly article providing a proper definition of the term competence as substantially different from knowledge along with some large-scale studies in school settings confirming the advantages of the former. I did read however, that, for example, a student must possess a mathematical competence and a social competence (whatever that means) but I still do not know what the difference between mathematical knowledge and mathematical competence is (although I do feel that the student I described in the introduction had a mathematical competence because the student's math marks indicated this, yet the student lacked mathematical knowledge and deep understanding). Every "definition" of the concept of competence I have been able to find stresses the fact that a competence is something that is measurable. This invites several questions. Do we want to go from *knowledge teaching to the acquiring of key competencies* because competencies can be measured whereas knowledge cannot be? Is this a tacit admission that school marks have been useless as a measure of knowledge all this time we used them but they are going to be very good indicators of competencies? Are there any large-scale scientific studies supporting this claim? Why not drop entirely the idea of using school marks as a measure of intellectual development? Is it because this strategy contradicts the puerile business notion that everything should be measured on a simple numerical scale so that we can have colourful corporate PowerPoint presentations in which every business success or failure is assigned a simple measure?

I openly admit that I feel the same way about many current pedagogical buzz words. Apart from the fact that suspiciously many of them seem to be part of the corporate jargon, they are never precisely defined but explained via dubious examples. A case in point is the notion of the *pedagogical model*. There is no universally accepted precise definition. If one has enough spare time, one can find on the Internet many such "definitions" that differ wildly in terms of language and aspirations. The examples

below are more than enough to illustrate the situation.

Pedagogical models are cognitive models or theoretical constructs derived from learning theory that enable the implementation of specific instructional and learning strategies. Examples of pedagogical models include anchored instruction, problem-based learning, cognitive apprenticeship, situated learning, and computer-supported intentional learning environments (What are Pedagogical models).

The pedagogical models are the different approaches to teaching that can be carried out by teachers in the classroom. Depending on the model they use, teachers will carry out a series of actions and focus on different parts of the learning process. Because there are different ways of learning, and because each student is unique, teachers must be armed with different pedagogical models to be able to adapt to different situations (Pedagogical models (traditional and current)).

What is the Pedagogical Model? The Pedagogical Model describes what effective teachers do in their classrooms to engage students in intellectually challenging work. It provides an overview of the learning cycle and breaks it down into five domains or phases of instruction: Engage, Explore, Explain, Elaborate and Evaluate (Department of Education and Training, Melbourne (2020)).

Do we have different pedagogical models for different students? The second “definition” seems to imply that for every *situation* there must be an appropriate pedagogical model. Do we have to assume that we have a very small number of different situations so that we have a small number of pedagogical models that a pedagogy student can memorise or is it possible for the situations to be too many for a human being to conceive of them all?

Not satisfied with these murky definitions, I tried to understand the meaning of some less general sounding concepts, namely, “deeper learning” and “deeper teaching” (see for example (Lampert, M. (2015))). I do not know if these are parts of a pedagogical model or a pedagogical process (the definitions of these terms are as varied as the ones of pedagogical models). Again, I could not find a single satisfactory definition but I did find examples of “deeper teaching” and “non-deeper teaching” in Lampert, M., 2015. Based on these examples, it was

concluded that deeper teaching teachers must fulfil the following requirements:

1. ... *be able to elicit students' current understandings and build from them toward the deeper learning competencies.*
2. ... *needs to know her subject well enough to decide to teach and be able to teach core content, critical thinking, problem solving, collaboration, communication, a disposition to learn, and academic mindset.*
3. ... *needs to choose and be able to use methods of instruction that link what the teacher is teaching with students' current understandings* (Lampert, M. (2015) p. 16).

When one reads the above requirements, one cannot help but wonder why we need a new term (deeper teaching) for something that most parents and students would call plain old proper teaching (by the way, teachers that engage in deeper teaching are called *ambitious teachers* at some point in the article). Surely, these are some – but definitely not all – of the requirements that proper teaching entails. In addition, would not the first requirement sound much better if it read... *be able to elicit students' current level of knowledge and build from it toward a deeper level of knowledge?* The second and the third requirement can be similarly reformulated and clarified (however, I do think that the second and the third requirements are immediate consequences of the first one and, therefore, are redundant). I could not find a single indication in the whole article that, maybe, the lack of *ambitious teachers* is due to the low salaries and poor working conditions and not to the lack of a proper terminology. What would happen if teachers' salaries were high enough to attract highly qualified scientists, who know the ins and outs of their respective areas, to teach at schools? My guess is that the majority of them would be very *ambitious teachers* without having to read articles on *deeper teaching*.

An academically paranoid person might suspect that the increase of new terms in pedagogy research articles is needed mainly to impress pedagogy students during lectures, funding committees when writing grant proposals, and politicians and business leaders in search of people who can dress their agenda in scientific jargon. A cynically minded person might suspect that pedagogy departments feel the need to emit such terms in order to claim their usefulness to the neoliberal economic thinking and thus to justify their existence.

6. Conclusion.

A very crude and undoubtedly incomplete picture of the current state of European education is that it is being refashioned along business and corporate lines. This is reflected most clearly in the language in which government decisions concerning education are written. One of the main outcomes of this process is that students and parents are gradually being convinced that the main purpose of education is the students' preparation for the job market (*skillful servants of society* according to Woodrow Wilson). Another more insidious outcome of current schooling policies is the inculcating of the belief in meritocracy. Small kids are taught at schools that their success in learning can be measured neatly along some grading system and that their marks depend solely on their effort. Their lineage, parental income, and parental social status are very much disregarded when evaluating their competencies (except when students with rich parents have their grades inflated of course). Later, their grades will play a decisive role in the university admission process and, hence, their professional and personal life. Thus, many people progress from the belief that school marks accurately measure academic achievement to the belief that one's worth as a person very much depends on one's financial and social status (which are measured neatly along another grading system: the amount of money one has). To sum it up in Alfred Whitehead's memorable words (whose *Aims of Education* (Whitehead A. (1985)) should be required reading for all who think that measuring competencies in terms of school marks is a good idea):

In particular, so long as we conceive intellectual education as merely consisting in the acquirement of mechanical mental aptitudes, and of formulated statements of useful truths, there can be no progress; though there will be much activity, amid aimless re-arrangement of syllabuses, in the fruitless endeavour to dodge the inevitable lack of time (Whitehead, A. (1985), p. 29).

References

- Department of Education and Training, Melbourne (2020) *The Pedagogical Model* Retrieved June 18, 2021, from <https://www.education.vic.gov.au/Documents/school/teachers/teachingresources/practice/pedagogicalmodel.pdf>
- EHEA (1999) *Bologna declaration* Retrieved June 18, 2021, from https://ehea.info/Upload/document/ministerial_declarations/1999_Bologna_Declaration_English_553028.pdf
- EHEA *Ministerial Declarations and Communiqués*. Retrieved June 18, 2021, from <http://ehea.info/page-ministerial-declarations-and-communiqués>
- EHEA (2020) *Rome Ministerial Communiqué*. Retrieved June 18, 2021, from http://ehea.info/Upload/Rome_Ministerial_Communique.pdf
- Gingras Y, Khelifaoui M (July 9, 2020) *Albert Einstein the mediocre: Why the h-index is a bogus measure of academic impact*. Retrieved June 18, 2021, from <https://phys.org/news/2020-07-albert-einstein-mediocre-h-index-bogus.html>
- Heron C (March 27, 2019) *What are Transversal Competencies and why are they important*. Retrieved June 18, 2021, from <https://medium.com/@chrisheron/what-are-transversal-competencies-and-why-are-they-important-72878778d80c>
- High School Teachers Association of New York* (1908–1909), Volume 3.
- IIE. *Economic impact of international students* (2019) Retrieved June 18, 2021, from <https://www.iie.org/en/Research-and-Insights/Open-Doors/Economic-Impact-of-International-Students>
- Lampert M (2015) *Deeper Teaching. Students at the Center*. Deeper Learning Research Series. Boston, MA: Jobs for the Future. Retrieved June 18, 2021, from <https://studentsatthecenterhub.org/wp-content/uploads/Deeper-Teaching-120315.pdf>
- Lorenz C F G (2006) Will the universities survive the European integration? Higher Education Policies in the EU and in the Netherlands before and after the Bologna Declaration. *Sociologia internationalis* 44(4): 123–151. Retrieved June 18, 2021, from <https://research.vu.nl/ws/portalfiles/portal/2172135/Sociologia+Internationalis.pdf>
- Management Study Guide. *Definition of Competency – What are competencies*. Retrieved June 18, 2021, from <https://www.managementstudyguide.com/what-are-competencies.htm>
- Ministry of Economics of Bulgaria. *ICHC* (2014). Retrieved June 18, 2021, from

https://www.mi.government.bg/files/useruploads/files/innovations/ris3_26_10_2015_bg.pdf

Ministry of Education and Science of Bulgaria. *National strategy for scientific research development (2017–2030)* Retrieved June 18, 2021, from https://www.mon.bg/upload/6527/SSstrategy_2030_BG.pdf

Ministry of Education and Science of Bulgaria. *Strategic Framework for the Development of Education, Schooling, and Study in Bulgaria (2021–2030)* Retrieved June 18, 2021, from https://www.mon.bg/upload/25571/Strategicheska-ramka_ObrObuUchene_110321.pdf

MIT Technology Review (2018, March 1). *If you're so smart, why aren't you rich? Turns out it's just chance.* Retrieved June 18, 2021, from <https://www.technologyreview.com/2018/03/01/144958/if-youre-so-smart-why-arent-you-rich-turns-out-its-just-chance/>

Pedagogical models (traditional and current) Retrieved June 18, 2021, from <https://www.lifepersona.com/pedagogical-models-traditional-and-current>

Pluchino A, Biondo A, Rapisarda A (2018) Talent Versus Luck: The Role of Randomness in Success and Failure. *Advances in Complex Systems* 21(3–4): 1850014

Shaw C, Ward L (6 March 2014). Dark thoughts: why mental illness is on the rise in academia. *The Guardian*. Retrieved June 18, 2021, from <https://www.theguardian.com/higher-education-network/2014/mar/06/mental-health-academics-growing-problem-pressure-university>

Wikipedia *Bologna Process*. Retrieved June 18, 2021, from https://en.wikipedia.org/wiki/Bologna_Process

Wikipedia *Prussian Reform Movement*. Retrieved June 18, 2021, from https://en.wikipedia.org/wiki/Prussian_Reform_Movement

What are Pedagogical models Retrieved June 18, 2021, from <https://www.igi-global.com/dictionary/student-perceptions-pedagogical-applications-learning/22156>

Whitehead A (1985) *Aims of Education*. Free Press

STRUGGLING TO ACHIEVE HIGH CREATIVE STANDARDS IN THE ROMANIAN ADVERTISING INDUSTRY

Mădălina Moraru

University of Bucharest, Romania
e-mail: madalina.moraru@fjsc.ro

Abstract:

The creative industry daily faces genuine challenges in its work when it comes to advertising and meeting clients' demands. Indeed, technology and the necessity of updating creativity resources urge on new approaches during campaigns, at least in the creative department. Extrinsic and intrinsic motivation stand for essential aspects in challenging new resources of creativity in a field where copywriters and art directors unfold incredibly sensitive messages based on strong and relevant insights.

The present paper aims to point out the difficulties and opportunities of creative work in any advertising agency, by exploring the purpose, the barriers and the prospects of this activity in the context of a complex relationship between client and agency, brand and consumers. One could perhaps say that copywriters are just gifted people able to simply follow instructions given by the planning department. Actually, they have their own psychological and social barriers, which represent real challenges. Therefore, we have investigated these issues by conducting semi-structured interviews based on their creative experiences in both cases, as juniors and seniors, respectively. The data collected via semi-structured interviews are investigated by using content analysis.

Key words: advertising, creativity, challenges, copywriting, success, brands.

Introduction

Creativity in advertising relies on a special dynamic of associations between data, information, cultural background, campaign purpose and, obvi-

ously, the personal touch of every person involved in the whole process. Designing the campaign from the initial strategic viewpoint to the final product can be seen as a creative approach, due to the fact that the connection between client demands, consumer needs, brand capital and agency resources should be very well thought out. One of the roles of creativity in advertising is achieving campaign effectiveness as scholars like Reid et al. (1998) have noticed. Over the years, scientists have identified two directions of defining creativity in advertising: divergence and meaningfulness. In the first case, the focus is on novelty and aesthetic (Till and Baack, 2005), whereas in the second it is on relevance and appropriateness, which have drawn Smith's and Yang's (2004) attention.

Certainly, divergence corresponds to a specific way of thinking that refers to the ability to absorb any source of inspiration arising from day-to-day life. With reference to this, the concept might be explained as follows: "Divergence is that component of creativity that differs from the expected; it provides the spark of the advertisement that separates it from ordinary communication."¹ Obviously, creative slogans, techniques or messages reveal authentically pleasant surprises both to customers and clients in order to sell, as people react more positively to that spark than to predictability. Novelty and conveying unexpected solutions to buyers' problems is the key to success in advertising.

¹ K. Lehnert, B. Till, and J.M. Ospina, Advertising Creativity: the Role of Divergence versus Meaningfulness: *Journal of Advertising*, 43 (3) (2014), 275.

Meaningfulness refers to the brand capital and its image because, despite the appeal of novelty and the surprising effect, every creative message must be related to the main goal of the campaign, no matter what. This goal is the campaign's engine and, therefore, creativity has to be strongly supervised by the planning department during every stage of campaign design. The term 'meaningfulness,' broadly associated with 'relevance' in the consumers' eye, can be unproblematically understood by reactivating the relationship between brand personality, brand trust and brand affect (as in Sung and Kim, 2010). According to them, the common consumer invests a brand with trust as long as this offers the proof of its reliability: "Indeed, the notion of <reliance> is crucial to the definition of brand trust, suggesting that there are two key components and characteristics essential to brand trust: trustworthiness and expertise."² Ignoring this dimension of brands while producing enjoyable advertisements generates neither awareness of every campaign goal, nor consumer loyalty and commitment.

Considering that professionals should pay attention to consumers' human nature and brand equity as well, creativity ought to establish its appropriate goal from case to case. Consumers trust practical and rational campaigns, but enjoy more emotional approaches. When analyzing the similarity between human personality and brand personality, Aaker³ identified five dimensions that should be discussed in terms of creative techniques welcomed by everyday consumers. These five dimensions are Sincerity, Excitement, Competence Sophistication and Ruggedness. Excitement and Sophistication stir deep emotions, giving copywriters freedom to evoke these feelings and kindle the special spark that encourages consumer loyalty. At this point, Ang and Lim (2006) note that hedonistic and symbolic products provide pleasure and excitement to consumers, rather than utility, as rational campaigns do. Under these circumstances, brand relevance is already well-known and recognizable, such that creative minds can be more divergent as they should be from time to time.

This dynamic of creativity in terms of divergence and meaningfulness stems from intrinsic and

extrinsic motivation on the one hand, but also from the availability of technological resources in an agency. The purpose of this paper is to examine barriers to creativity that exist in an advertising agency and illuminate the best opportunities to get over them. There are many studies focusing on the way creativity is perceived and developed, but in advertising these questions deserve additional attention given that both dimensions must be taken into consideration: the process and the product. Process refers to the entire design of thinking appropriately to the campaign type, while product refers to the successful advertisement.

1. Stages of creative thinking in an advertising agency

As Amabile asserted in 1996, the most important steps of the creative process involve the following aspects: (1) identification of the problem/opportunity, (2) gathering of information/resources, (3) generating ideas, and (4) evaluating, modifying, and communicating them. The problem or the opportunity is strongly connected with creative obstacles; collecting resources corresponds to the research phase producing ideas represents the brainstorming process; while assessing messages means analyzing their effectiveness in the campaign. This is finally assessed based on several criteria, such as efficiency, attractiveness and novelty. In 2010, Shimp created a model recognized as 'CAN', which is described according to these three issues: connectedness, appropriateness and novelty. The above features can be translated as empathy towards the target, then relevance to the brand, and last but not least, uniqueness.

Certainly, scholars have highlighted many other classifications of creative steps in advertising over the years. Revealing the creative opportunity means understanding the reasons for improving or updating an old-fashioned/inadequate idea that successfully achieves clients' goals and consumers' desires. Gathering information is frequently rooted in culture and society and is also related to the context of everyday life. Moreover, the extent to which emotional or rational involvement of professionals can lead to a specific allocation of resources is more than relevant. In terms of generating ideas, the process is, as expected, complex and can be associated with mental imagery according to Lubart (2003) and Amabile (1996). According to their views, creativity is the ability of people who generate new and useful ideas,

² Y. Sung and J. Kim, Effects on Brand Personality on Brand Trust and Brand Affect: *Psychology and Marketing*, 27 (7) (2010), 644.

³ J. Aaker, Dimensions of brand personality: *Journal of Marketing Research*, 34 (1997), 347–356.

while mental imagery refers to producing mental images or representations. At this point of our discussion regarding the process of generating new ideas, scientific literature (Torrance, 1976) has identified three categories of outputs: fluidity (the number of new ideas), flexibility (the variation or diversity of ideas) and originality (the unusual ideas). The last stage of the creative process highlights the effectiveness criteria of the whole campaign, which are usually not homogeneous, considering that clients' demands are quite subjective as compared to the agency ones. Each year, the Grand Prix awards amazing campaigns at the Cannes Lions International Festival of Creativity based on the following criteria: creative idea (25%), strategy (25%) and results and effectiveness (50%). These issues definitely underline that "Advertising effectiveness is now a significant issue within the advertising industry and creativity is a significant factor in how it is achieved"⁴

Considering that the design of the creative process demanded by an advertising task supplies a range of variety in terms of techniques and personal approaches, the synthesis of the above discussion praises the personality of every professional able to evaluate him/herself. Obviously, the problem of every creative brief is already established by strategic thinking, but the key message and the way a planning approach is interpreted by creative minds remains a mystery. When it comes to technology and artificial intelligence, for example, generating ideas builds up an algorithm based on data and all the information needed. Basically, the last two of Amabile's stages of the creative process typically emerge, regardless of the tools used either technologically or mechanically (artificial or human creativity). Nevertheless, the entire dynamic of both managing and recognizing campaign effectiveness in the advertising field brings to light a specialized profile of any professional employed in the creative department. Their main features are curiosity, innovative behavior, steady open-mindedness, the ability to pay attention to the 'big picture', and also detail-orientedness. Therefore, instead of gathering ideas, the process of 'digging' for fresh inspiration seems to offer a more appropriate perspective on the same stage of creativity.

2. A theoretical framework regarding the obstacles to creativity in advertising

What deserves further consideration in this study, due to the research goals, are the creativity barriers that certainly influence the success of every campaign. To better understand these hindrances to the creative process, we have decided to trace them through personal and subjective clues to external and objective factors, the ones being easily associated with the ways in which the creative context is organized. In terms of personal barriers to creativity, present research reveals significant data. Still, we should consider a subjective or qualitative approach to these kinds of barriers, which frequently arise from emotional attachment to ideas or issues related to motivation. As long as these two categories are not stimulated enough or strongly encouraged, creativity is jeopardized by routine or common, old-fashioned ideas.

Definitely, intrinsic motivation is related to increasing the quality of creative work in an advertising department, whereas external motivation depends on organizing a cohesive team and organization with high standards. Given the substantial research conducted on this topic, intrinsic motivation has been variously characterized by features such as openness to experience, self-efficacy, and perseverance (Prabhu, Sutton & Sauser, 2008), or strong cognitive skills. Somehow, cognition emerges from creativity and also the other way around as the respective skills generate more brainstorming and mental challenges. By contrast, when individuals passively accept routine or embrace ideas without analyzing or researching them, creative walls are built up and intrinsic motivation practically ceases to exist. As a result, professionals can neither stimulate their minds creatively nor accept compromises of their work, a rut which might lead to creative failure as a consequence of ignoring the strategic approach.

Another relevant and subjective matter that could represent a barrier to authentic creativity in an advertising department is the emotional attachment to one's work. This means either internalizing the advertisement as an artistic and aesthetic piece of work, or neglecting the relationship between brand, consumers' insight/needs and clients' demands. According to Thomson et. al., "Creative workers do frequently demonstrate affective attachments to their work, but this is different from industry contexts in which the requirement for 'affect' is initiated and

⁴ C. Fill, G. Hughes, and S. de Francesco, *Advertising, Strategy, Creativity and Media* (London: Pearson, 2013), 149.

imposed by management,”⁵ a fact that might, somehow, influence the strategic approach of their efficient message. Little is known about the methods of achieving success with advertising campaigns when people praise advertisements as creative products rather than examining the process of creation, which, in practice, would imply getting to the core of the campaign, i.e. the brand or consumers’ insight. Professionals understand their job as long as the driving force or the priority of their work is the brand and not their own point of view, because “a creative idea must be consistent with the communication strategy and brand position” (Percy and Rosenbaum-Elliott, 2016, p.324). When copywriters and art directors give birth to a creative product that bears their own distinctive human (psychological) signature too strongly, the creative process crosses the borders of self-representation with difficulty due to their emotional involvement. However, the line between professional creativity and personal creativity ought to be conscientiously drawn all the time. Moreover, one of the most important lessons taught by Fallon and Senn⁶ regarding the flexibility of the creative process in advertising is the ability to always start from the scratch if the strategic approach demands it. Otherwise, creative ideas are lost among stereotypes already used in other campaigns. A deep emotional involvement makes it difficult for copywriters to relate to their work.

Another hindrance calls upon the creativity team to ensure cohesion and top-down cooperation in order to reach a high level of performance. In case of conflicts and difficulties between a copywriter and his/her art director, creative standards are lower because professionals cannot develop or express their skills and abilities properly. Strong feelings of belonging and attachment to other members of the team help processes to run more smoothly, as Taggar (2002) states. The level of team engagement to achieve common purposes is characterized by interaction, dedication, shared experiences, and work-related motivational attitude (Costa et. al, 2014). In reaching this level of motivation and collective cohesion, a creative team ought to overcome personal pride and egotistical behavior. The core of creativity in a team consists in its steady interaction, mutual interests, mutual support and shared success. When

one of the tasks is separately or individually assumed and completed, team cohesion disappears and other invisible walls are built up within the entire creative department.

The above-mentioned barriers during the creative process can be seen as real challenges, but they are just a few of the creative obstacles that can arise, considering that technology, managerial context, health and psychological matters might also pose problems. The following research aims to unfold both problems and solutions to different kinds of creative hindrances in advertising.

3. Methodology

The qualitative approach enables us to assess different views expressed by professionals in the work environment, which gives them more credibility when discussing creativity issues. According to Davis, “qualitative research allows one to draw insights and explanations from respondents themselves, rather than having to predetermine areas of response or importance.”⁷ Therefore, the research based on direct interaction provides us with the necessary knowledge to reach relevant conclusions, even though they cannot be generalized.

We have carried out the current research in two stages: data collecting and data analysis. Firstly, data were collected by applying 14 semi-structured interviews, and secondly, information was assessed by using content analysis to generate outcomes. Payne (1986) considers the interview to be the most accurate method of exploring beliefs, personal preferences and idiosyncrasies in order to reveal nuances and differences meant to be qualitatively analyzed. In the present case, we interviewed 14 professionals working in creative departments as teams of copywriters and art directors, some of them being juniors, others seniors. Among them there were three creative directors we had the chance to talk to, whose experience was quite useful from the holistic and critical perspective. All of them work in the best full services agencies from Bucharest, Romania, such as Publicis, McCann Worldwide Group, Friends TBWA, Kubis Interactive and Next Advertising. The juniors and young seniors are between the ages of 21 and 25 and the senior copywriters and creative directors are between the ages of 28 and 39. For this

⁵ P. Thomson, R. Parker and S. Cox, *Interrogating Creative Theory and Creative Work: inside the Game Studio: Sociology*, 50 (2) (2016), 319.

⁶ P. Fallon, P. and F. Senn, *Creativitatea în publicitate* (București: All, 2008), 11.

⁷ J. J. Davis, *Advertising Research. Theory and Practice*, (Upper Saddle River, New Jersey: Prentice Hall, 1997), 196.

reason, we equally split the respondents into two categories, named accordingly: the group of young Millennials at the beginning of their career and the group of those already established in their career, sometimes still very young. In terms of the time span of interviewing the respondents, the data were collected between 1 of May and 10 of July 2018, in Bucharest, and were recorded either during face-to-face discussions or via Skype, and subsequently transcribed. All interviews lasted between 40 and 55 minutes and were conducted directly.

When it comes to the structure of the interview guide, we have focused on different questions, including the way creativity is understood in advertising; yet in this study, the main concerns are oriented towards daily challenges on the job, solutions to possible obstacles to creativity, and understanding creative successes. Each category was investigated by asking individual questions, and results were provided by the content analysis that emerged via the processes discussed above.

The main hypotheses of this research motivated the creativity analysis of the dynamics of the advertising field as follows:

H₁. Professionals deal with diversity in challenging creativity during their work all the time because they have to meet the consumers' and clients' needs constantly.

H₂. The solutions to obstacles discovered by professionals during their creative work do not follow a well-known recipe, but rather a personal approach (which could be considered creative as well).

H₃. Confirmation of successful creativity in advertising stems from both extrinsic and intrinsic motivation.

4. Results

In order to organize the results of the qualitative research, our intention is to analyze the interviews based on both age categories – junior and senior employees – in creative departments. This view provides the study with a comparative perspective on creativity challenges.

4.1. Creativity challenges in advertising agencies

Most of the juniors are firm in their conviction that external factors are the main barriers preventing them from expressing their creativity. They list various aspects such as short deadlines, bad organization of daily work, lack of time, clients' lack of market-

ing education, too many projects at the same time. One of them was extremely sincere and declared that the lack of experience could be another issue at a national level: 'you see something quite cool abroad and you want to do it as well, but you don't know exactly how to do it, what to start with or you can't find someone to do it for you.' Definitely, his opinion refers to innovative projects that require enormous budgets, additional resources, and, also, the courage to break the ice. Time management seems to be an essential reason for young copywriters' dissatisfaction, which is related to their need for better organization of their work. Only one of the respondents believed that personal problems affect the creative behaviour of the team – this being an answer focused on the collective approach of this complex ability.

By comparison, professionals having at least 7 years and a maximum of 18 years of experience in the field perceive internal factors as real challenges and barriers in their work. Most pay attention to the level of communication and empathy in the team, to intrinsic motivation, to lack of initiative, and last but not least, to the lack of reaction to negative feedback. Despite these perceptions, seniors do appreciate having ample time to work and consider short deadlines to be a disadvantage when working to achieve successful creative ads. They are also quite aware of the fact that personal weaknesses (exhaustion, fear of failure, inability to adapt to changes, conformism, lack of courage) may lead to a negative performance. The client's approach has also been taken into consideration as a potentially disruptive factor in the creative process.

Solutions to creativity obstacles

It was undoubtedly difficult to analyze homogeneously the way professionals discover solutions to their personal or collective obstacles in their creative work. Some of them answered these questions by invoking the need to update professional processes through various tools. Many of them expect to be helped by technology in the future – in terms of research and discovering new sources of knowledge. Among the answers provided, one emerges highlighting the importance of avoiding routine and deciding 'to approach less conventional ideas.' It is surprising that only one of the respondents relies on this concept of team creativity, given that the junior is deeply convinced that a better partner in the creative team will definitely improve the situation.

As for seniors, half of them strive to improve clients' knowledge and understanding in the relationship with the advertising agency and to ensure good time management to allow more freedom and creativity testing. The others strongly focus on their own progress in terms of creating ads that are effective both at the emotional and informative level. They consider new experiences as the best opportunity to fight against routine or other obstacles. Moreover, informal activities in their team or department may help them to face creative barriers. They are aware of the necessity of staying connected not only with their colleagues but also with the society the final consumer lives in. Perseverance becomes the solution to any kind of hindrances as one of them explicitly asserted: 'Finally, creativity is an extremely subjective process, and you need strength to keep doing your best even when not everybody applauds your <masterpiece>.'

4.2. Confirmation of creativity success

Getting over creativity barriers in their work allows professionals to achieve success and visibility, an important goal since they are aware of the necessity of being recognized in order to feel motivated in the future. Experience makes a difference between juniors' and seniors' perceptions regarding the success of creativity confirmation. Most young copywriters first aim to help their team gain validation, positive feedback and the enthusiasm of their colleagues. This is due to a desire to feel integrated, to be part of that team in order to be accepted. Just two of the individuals interviewed judge the success of their creativity predominantly by the visibility of their advertisements, clients' approval and creative awards. Their perspective on creativity confirmation is quite holistic. They speak more generally about their desired future, not specifically about the present stage of their career. What is relevant for all of them is the fact that they associate creativity success with the team, with the campaign's effectiveness and also with the personal joy of a job well done. When someone was asked about three aspects confirming his success as a copywriter, he named three significant moments, including aspects of his personal life, stating: 'I'm satisfied, when I receive positive feedback, when I sleep well at night.' This connection between daily work and a sense of inner satisfaction is a sign of creativity confirmation.

Regarding the seniors' category, the majority considers awards as the best way to recognize their creativity performance, because it means not only

achieving client's goals, but also being listed at the top of local Romanian (or even international) professionals. This promotes good working relationships, including those between mentor and mentee, and potential opportunities for promotion in the future. One of the interviewees pointed out that the confirmation of creativity success in advertising stems from consumers' assessment, because ads' humour, commercials' characters and innovative online ads are integrated into their life. This is, in their opinion, the highest level of personal achievement for a successful copywriter.

4.3. Discussion

The following table synthesizes the results of the content analysis of the respondents by using keywords for a brief comparison between generations.

Table 1
Creativity Insight in an Advertising Agency

Units of content analysis	Juniors	Seniors
Challenges of creative process	External pressure	Internal factors
	Time management and client's attitude	
Solutions to creativity challenges	Absorbing knowledge	New experiences
	Constantly updating	Emotional development
	Technology	Team cohesion
Confirmation of creativity success	Team integration	Being part of popular culture
	Positive assessment	Becoming a real mentor
	Being awarded	

Source: Author

The external factors are fully justified in being considered the main barriers for young copywriters who face integration into new teams, new management and high standards. They have to work on their own approach regarding the challenges of the creative process in an advertising agency. As for the seniors, they are more aware of the intrinsic motivation that cultivates the creative perspectives and emotional involvement of the whole team. Actually, this is the essential difference between artificial creativity produced by software and algorithms (artificial intelligence) and the human mind. In these situations, seniors cooperate with juniors and assess

their creative activities based both on empathy and objective landmarks. Issues related to time management and client needs depend on the relationship between client and agency and the agency's management style.

Analyzing the second element of the research, the above table highlights two main paths toward overcoming creativity obstacles: juniors' creativity usually relies on fresh knowledge and learning techniques from their mentors, whereas seniors look for new experiences and emotional development to continually rediscover themselves and to turn creativity into originality and uniqueness. Technology is an appropriate solution to save time and better manage routine tasks, and thus creative people keep up-to-date on the new tools available that could help them to implement their ideas.

Regarding the way the confirmation of creativity success is perceived by professionals, the analysis has identified heterogeneous outcomes in terms of respondents' experience. Juniors strive to be recognized by the creative team in their position in order to foster cohesive creativity in a department responsible for a successful campaign. Seniors, on the other hand, wish to become memorable with consumers due to their masterful execution of campaign messages, and their sense of personal fulfillment is less concerned with achieving marketing objectives and satisfying the client. They move forward and expect consumer loyalty and engagement. Prizes indicate international and national recognition and ensure their intrinsic and extrinsic motivation.

5. Conclusions

Coming back to the research questions which generated the study hypotheses, the first one is partially confirmed, because the interviewees have revealed various obstacles in their creative work, which are not always related to the consumer or client. Therefore, it is not fruitful to generalize that creativity in an agency may be hindered by consumers and clients, because this is an internal mechanism, not an external one.

The second hypothesis is entirely confirmed given that every interviewee, especially the juniors, tried to explain what the source of their creativity is and how it can be nourished or improved. The fact that they face internal or external pressure is also associated with the diversity of their work and with their level of experience. Seniors look for unconventional solutions that will be very helpful not only in

maintaining emotional intelligence at a very high level, but also in developing emotional creativity. This aspect represents a real advantage when understanding consumers' insights during a campaign and, even more so, when managing it creatively.

As for the third hypothesis of this research, it was not validated, because both juniors and seniors strive for team recognition and industry awards. The process of intrinsic motivation is supported by colleagues, and, frequently, creative teams made up of a copywriter and an art director, inspire each other a lot. Certainly, interviewees do not overlook personal requirements of motivating themselves intrinsically.

The main limitation of this research consists in interviewing a relatively small number of professionals, which still provided us with an unexpected amount of data. Even so, this aspect might confer some lack of credibility to the paper. From the qualitative viewpoint, information received is very insightful; therefore, we do not exclude the possibility of extending the study by investigating other professionals as well. A good idea would be to interview professionals working in agencies in-house or for the client who are nonetheless responsible for the creative part of the campaign. Nevertheless, a triangulation of the research method might be a smart decision, which would entail organizing a few focus groups on the same topic to analyze how members of the same creative departments interact.

To sum up, advertising creativity faces many challenges because technological progress is a steady pressure and originality now relies on innovation and innovative tools as much as possible. Generating new ideas that simultaneously appeal to clients, to consumers and to the creative director on an internal level involves taking into account many factors, such as various professional goals, agency creative performance, and also personal satisfaction. The difference between creativity used in other fields and in advertising derives from its purpose, as advertising relies on fostering a strategic approach for each campaign and team cohesion. It might be somewhat of a paradoxical fact that creative professionals want to be awarded and recognized for their work, which is an individual process; however, the success of the creative team depends on their contributions, and thus team and individual creativity and recognition can coexist. Therefore, one definition of being creative as a copywriter or as an art director in a team is a simple equation: Individual creative success enriches the cohesion of the creative team,

which leads to meaningfulness and efficiency of campaigns.

References

- Amabile T M (1996) *Creativity in Context*, Boulder, CO, US: Westview Press.
- Ang S W, Lim E A C (2016) The influence of metaphors and product type on brand personality perceptions and attitudes. *Journal of Advertising* 35: 39–53.
- Costa P L, Passos A M, Bakker A B (2014) Team work engagement: A model of emergence: *Journal of Organizational and Occupational Psychology* 87: 414–436.
- Fallon P, Senn F (2006) *Juicing the Orange: How to Turn Creativity into a Powerful Business Advantage*. Boston: Harvard Business School Press.
- Lubart T (2003) *Psychologie de la créativité* Paris: Armand Colin.
- Payne M S (1982) *Individual in-depth interviews can provide more details than groups. Marketing Today*. Chicago: Elrick and Lavidge.
- Percy L, Rosenbaum-Elliott R (2016) *Strategic Advertising Management*, 5th ed. Oxford: Oxford University Press.
- Prabhu V, Sutton C, Sauser W (2008), Creativity and certain personality traits: Understanding the mediating effect of intrinsic motivation: *Creativity Research Journal* 20: 53–66.
- Reid L N, King K W, Delorme D (1998) Top-level agency creative look at advertising creativity then and now. *Journal of Advertising* 27 (2): 1–16.
- Schimp T (2010) *Integrating Marketing Communication in Advertising and Promotion*, 8th ed. South Western: Cengage Learning.
- Smith R E, Yang S (2004) Toward a general theory of creativity in advertising: examining the role of divergence. *Marketing Theory* 4 (1/2): 31–58.
- Till B D, Baack D W (2005) Recall and persuasion: does creative advertising matter? *Journal of Advertising* 34(3): 47–57.
- Taggar S (2002) Individual creativity and group ability to utilize individual creative resources: A multilevel model. *Academy of Management Journal*, 45: 315-330.
- Torrance E P (1976) *Test de pensée créative*. Paris: Centre de Psychologie Appliquée.

Book reviews

Pepka Boyadjieva, Petya Ilieva Trichkova (2021). *Adult Education as an Empowerment. Re-imagining Lifelong Learning through Capability Approach, Recognition Theory and Common Goods Perspective*. Pallgrave Mcmillan, 354 pages¹

**[Reviewed by Nina Dimitrova, Institute of Philosophy and Sociology – BAS,
email: ninaivdimitrova@abv.bg]**

Others can teach us, but only everyone can educate themselves. It's not just a play on words. Educating yourself is really something completely different from being educated. However, when we are educated, we work to become something – we strive to exist in a certain way in the world.

Peter Bieri.

Wie wäre ist, gebildet zu sein? 2005

Pepka Boyadjieva and Petya Ilieva-Trichkova, scientists from the Institute of Philosophy and Sociology at the Bulgarian Academy of Sciences,² are an established creative team and their latest book builds on many of their individual or joint research in general on the topic of education and especially on lifelong learning and its most important form – adult education. This is a topic that has become even more relevant with the onset of the COVID-19 pandemic, which has revealed the need for more human soli-

arity, for a new balance between the individual and the collective. The authors share that the writing of the book itself also turned out to be a kind of learning and discovery process.

Empowerment – the concept signaled in the title – is thought of as the desired outcome of a learning process for adults. The authors cite Elaine Unterhalter's explanation that the neologism “empower” was first used in the mid-17th century, during the English Civil War (p. 109).

Thus, the joint work is dedicated to the many aspects, dimensions, applications, results, etc. of *lifelong learning*, a term that is preferred to *lifelong education*, because learning is a broader concept than education (it may refer to both knowledge and skills). The preliminary analysis (including that of etymology, their historical destiny and the current debates) of the concepts *education*, *lifelong education*, *lifelong learning*, *adult education*, *formal* and *non-formal education*, is exhaustive and precise.

In the pages of the book we can find out about the continual scientific debates on *lifelong learning* enshrined in European programs and policies. The authors cite a variety of opinions from leading researchers who emphasize its political commitment (as part of the neoliberal project); its ambiguity and uncertainty; its importance as a form of social control; as an instrument of globalization processes, etc.

In the book offered to the reader's attention, *lifelong learning* is considered as a concept of the societies of late modernity, or, according to the concept

¹ The book is part of the work on the project “Dynamics of the inequalities in higher and adult education: A comparative social justice perspective” (JustEdu), funded by the Bulgarian National Science Fund within National Science Program VIHREN, contract number KPI-06-IB-2/16. 12. 2019.

² During the last four years the scientists have been involved in the Horizon 2020 Project ENLIVEN (Encouraging Lifelong Learning for an Exclusive and Vibrant Europe).

preferred by the authors – the *second modernity*. The renaissance of the idea of lifelong learning can be understood only against the background of the changed sociality in the societies of the *second modernity*, for which the book offers an abundance of definitions (Anthony Giddens, Zygmunt Bauman, Ulrich Beck and many more. etc. are among the names whose ideas have been thoroughly analyzed). The choice of the authors falls on the synthetic definition of Ulrich Beck, who in turn used an article by the Russian artist Vasily Kandinsky, entitled simply “*And*”. This laconic “*And*” expresses the synthetic, holistic spirit of the age – the priority of conjunctive over disjunctive logic. The concept of “*And*” is becoming a major research attitude, and the definition of “*And*” a key feature of non-formal education. From this point of view, lifelong learning is seen as a sort of manifestation of the ties between the various kinds of education and the interpretation of the relationship between education and life. The advantages of lifelong learning are many and varied, and the authors have dedicated their efforts to conceptualizing them. Lifelong learning is thought of as a flexible activity that is adaptable to various individual and social circumstances, which the individual is free to practice at a time appropriate to him. The choice of educational institutions is also his, as well as the goals of his activity.

The proposed book is both theoretical – discussing in detail the approaches to lifelong learning (i.e. adult education) – and empirical, as it provides a wealth of data from various sociological studies on the subject. The rationale for writing the book is the need for (1) fresh theoretical ideas and insights for conceptualizing adult education and its role for both individuals and societies; and (2) new methodological instruments for studying the ways adult education has been implemented in different social contexts (p. 2).

The authors define the structure of the book, which may initially read as (overly) fragmented, as mosaic. I would suggest that it is “multifaceted” – each of the numerous “facets” illuminates the topic under consideration from a unique angle. This was the original intention of the authors, who invite the reader to choose freely which of the different parts of the book to begin with. I will not only take advantage of their invitation, but I will go further – in my review I will not offer a complete presentation of the book, but rather impressions of different discussions in it.

I am mainly interested in the philosophical basis of the proposed works, and the rich and valuable empirical part³ will be left to one side to be discussed by colleagues more competent than me in such matters.

The theoretical framework of the monographic study is a successful attempt to combine three approaches (as a philosophical basis) for the conceptualization of lifelong learning/adult education in order to make use of the heuristic potential of this union. One of these is the capability approach, used to rethink the essence and goals of lifelong learning, focusing on one of its types – non-formal education.

The authors explain in detail the essence of the approach proposed by the Nobel laureate in economics Amartya Sen, supplemented by the ideas of the American political philosopher Martha Nussbaum, as well as others. Preference is given, it seems to me, to the proposals of Amartya Sen, because his approach takes into account human diversity, due to the special emphasis placed on freedom and on quality of life. Attention is also paid to Martha Nussbaum’s proposal of ten central human capabilities and their political role. After carefully presenting and commenting on the criticism of the capability approach in its various versions, the authors of this book believe that the framework of the capability approach could be useful and may be applied in the field of education.

The reason for the generalized capability approach to look for opportunities to rethink lifelong learning is the fact that it takes into account the different learning abilities of people, the different context in which they implement this aspiration for education, and the fact that lifelong learning brings benefits to society itself, i.e. it has a positive social effect, not merely an economic one. The productivity of the capability approach in terms of lifelong learning is that it takes into account the activity of people who focus on continuing their education, as well as the social and institutional determinism of the participation of these people in non-formal education.

The capability approach is combined with the embeddedness approach, which assumes that all human actions are socially situated and that humans are not isolated atoms. Such a combination of these

³ Data from large-scale international surveys have been used: the Adult Education Survey, the Labor Force Survey, the European Social Survey, the European Values Study, the Programme for the International Assessment of Adult Competencies and the Continuing Vocational Education Survey.

two approaches can be the basis for building a comprehensive model of adult education.

A particularly important emphasis in the book is placed on the extremely close connection between adult education and recognition theory, leading the authors to seek an explanation of the role that recognition plays in unlocking human potential. The authors' thesis is that education is a separate area of recognition: "in modern knowledge-intensive societies, education should be regarded as a separated sphere of recognition (...), sociological analyses of adult education require taking into account both aspects of recognition – recognition as a matter of self-realization and identity formation and recognition as a matter of justice" (p. 146). The struggle for recognition (insofar as the latter is a *vital human need*, as one of the titles in the book states) and education are both considered lifelong learning processes.

Since the popularity of recognition theory has been significant in the last few decades, during which theories of mutual recognition have been intensively debated in social philosophy, the authors provide a brief historical overview of this theory, paying special attention to two main areas – the first is related to the views of Charles Taylor and Axel Honneth, and the second to that of Nancy Fraser.

The authors link Axel Honneth's theory, which has its applications in the fields of anthropology, ethics and social philosophy, to adult education and show how it can become a platform for the individual to acquire social visibility. Thus, education is seen as one of the ways to achieve an emancipatory effect – that of eliminating social invisibility, which, according to Honneth, can be understood as a kind of moral neglect. (The emblematic example discussed by Honneth is how the nobles undressed in front of their servants, because the latter were treated as if invisible or nonexistent: "Cultural history offers numerous examples of situations in which the dominant express their social superiority by not perceiving those they dominate. Most notorious, perhaps, is the fact that the nobility was permitted to undress in front of their servants because the latter were simply not there in a certain sense" [Honneth 2001: 112].) Citing Kant's idea from his *Groundwork of the Metaphysics of Morals*, Honneth describes the moral side of recognition in terms such as 'confirmation', 'affirmation' and 'according social validity' [Honneth 2001: 122]. Acquiring higher social status and prestige through adult education makes these adults socially *visible*. Adult education is seen as an inte-

gral part of *the struggle for recognition, or the struggle for social esteem*.

Nancy Fraser's alternative approach, which emphasizes recognition as a matter of justice, is also considered. Despite the debate between the two philosophers (*Redistribution or Recognition?* (2003)), in the book both versions of recognition theory are assessed as important for the sociological analysis of adult education. Recognition theory as a whole clearly shows, in particular, the *existential* meaning of education in general, that is, its *existential* significance for the individual and his self-esteem.

Among the main philosophical approaches to rethinking education is the view that adult education is a common good – the second part of the book outlines a theoretical framework for conceptualising and affirming lifelong learning/adult education as a common good. The "And" approach mentioned at the beginning of this review is also valid in the explanation of the role and the mission of adult education. The holistic principle inherent in the whole study explains the desire to go beyond the economic and instrumental perspectives on adult education.

The creative, innovative achievement of the authors is the development of a holistic model of adult education, consistent with both the level of influence – individual or societal – and the nature of this influence: intrinsic, instrumental and transforming/empowering (p. 190).

Numerous and diverse empirical data are also presented regarding the importance of adult education, which can be assessed by sociologists.

The book is published in *Palgrave Studies in Adult Education and Lifelong Learning*. The announcement for these series states the following: "Contributions to these series will contribute original knowledge and insights in adult education and lifelong learning, based on original empirical research and deep theoretical analysis, and stimulate debate on policy and practices." The book offered to the reader, dedicated to the conceptualization of adult education and its importance in various social domains, is exactly such a contribution – written competently and professionally, it offers deep theoretical insights and a rich empirical basis. The book contains many contemporary discussions on the issues raised.

With its honesty and realism, taking into account all the *pros* and *cons* of its own statements, with its scientific criticism and intelligence, it is undoubtedly an important step in the great debate on

education in its various forms. The book would be interesting and useful not only for professionals in the field of education, but also for all those for whom education is inseparable from life.

References:

Honneth, A. (2001). Recognition: Invisibility: On the epistemology of Recognition. *Aristotelian Society Supplementary Volume* 75(1):111–126.