ANALELE UNIVERSITĂȚII BUCUREŞTI

FILOSOFIE

Vol. LXXIII – Nr. 2

2024

SUMAR • SOMMAIRE • CONTENTS

Articole Ştiinţifice / Scientific Articles / Articles scientifiques

LOGIC AND EPISTEMOLOGY

HERMENEUTICS AND PHENOMENOLOGY

2

POLITICAL PHILOSOPHY AND SOCIAL CHALLENGES

Mircea DUMITRU[1]

NEW PERSPECTIVES ON COMPOSITIONALITY: KIT FINE'S SEMANTIC RELATIONIST APPROACH TO MEANING. AN EXPOSITORY ESSAY

Abstract. The paper is an expository essay in which I run an assessment of compositionality from the vantage point of Kit Fine's semantic relationist approach to meaning. This relationist view is deepening our conception about how the meanings of propositions depend not only on the semantic features and roles of each separate meaningful unit in a complex, but also on the relations that those units hold to each other. The telling feature of the formal apparatus of this Finean relationist syntax and semantics, *viz.* the coordination scheme, has some unexpected consequences that will emerge against the background of an analogy with the counterpart theoretic semantics for modal languages. In the evaluation of a *de re* modal formula at a world in a counterpart theoretic model, as opposed to the evaluation in a possible world semantic model, one may choose different possible objects as referents or semantic values of two tokens of a single type individual constant or individual variable. Likewise, in the relationist semantics for variables or individual constants (proper names), within a coordination schema one may choose different objects if the variables or the constants are not strictly coordinated. I shall leave the working out of this comparison for a future paper. The program defends 'referentialism' in philosophy of language; Fine holds that semantic relations that have to be added to the assigned intrinsic values in our semantic theory, especially the relation which he calls 'coordination', can do much of the work of (Fregean) sense. A relationist referentialism has certain important explanatory virtues which it shares with the Fregean position, but the former is better off ontologically than the latter, since it is not committed to the existence of sense.

Keywords: compositionality, semantic relationist approach, referentialism, Kit Fine

---

[1] University of Bucharest, Romanian Academy, Beijing Normal University. Email: <mircea.dumitru@filosofie.unibuc.ro>. ORCID: 0009-0001-5391-7280

The paper is an expository essay in which I address some of Kit Fine's views about meaning, which essentially consists in a relationist stance about the meaning of individual variables under assignments, and of names, respectively.

Let's consider a version of Frege's puzzle concerning identity sentences. Take the identity sentences: (i) Cicero = Cicero and (ii) Cicero = Tully, in which the two names, *viz.* "Cicero" and "Tully", are coreferential.

The semantic puzzle of identity sentences is based upon the following assumptions:

(1) *Semantic difference*: The two identity sentences are semantically different.
(2) *Compositionality*: If the sentences are semantically different, then the names "Cicero" and "Tully" are semantically different.
(3) *Referential Link*: If the names "Cicero" and "Tully" are semantically different, then they are referentially different.
(4) *Referential Identity*: The names "Cicero" and "Tully" are not referentially different.

Thus, what the puzzle indicates is that the four assumptions are jointly inconsistent, and we should give up at least one of them. Of course, the challenge is: Which one? And why?

In the literature, there have been two main responses to the puzzle, *viz.*

A. the Fregean response; and
B. the Referentialist response.

Both responses accept *Compositionality* and *Referential Identity*.

The Fregeans accept *Semantic Difference* between (i) and (ii); they reject *Referential Link* through the distinction between sense and reference: if two names are semantically different it doesn't follow that they are referentially different. The Referentialists accept *Referential Link*; they reject *Semantic Difference*: there is no semantic difference between (i) and (ii). Nobody doubts (4) – the *Referential Identity*.

Current philosophical thinking on Frege's puzzle has reached an impasse, for the dialectics of the arguments support a set of strong

arguments and Fregean intuitions in favor of *Semantic Difference*, but nevertheless some other argumentative strategy supports another set of strong intuitive arguments in favor of *Referential Link*. And fact is that there is no apparent way to choose between them. This is why I believe that finding a solution to this conflict and giving an account for both sets of intuitions in a sort of unified way, is the main task of a general semantic theory.

Kit Fine's semantic relationism seems to be that kind of approach that we need here.

Its basic view consists in the idea that there may be irreducible semantical or representational relationships between expressions or elements of thought, ones that are not reducible to the intrinsic semantic features of the expressions or elements of thought themselves between which they hold.

Semantic relationism motivates the view that we should give up (2); this means that we should dispense with (a crude version of) compositionality. In general, this is perceived as a huge price to pay. However, Fine convincingly argues that there is also a benefit if we make this move, which is that semantic relationism takes on board the sets of intuitions that separate Fregeans from referentialists; for then, in dealing with Frege's puzzle of identity sentences we can retain cognitive difference (as Fregeans do) and the referential link (as the referentialists do).

But what rejection of *Compositionality* amounts to? It amounts to supporting the view that there is a semantic difference between the two identity sentences, even though there is no semantic difference between coreferential names. So, after all, we do not reject the general idea of a compositional semantics. According to Fine, what fails is the so-called intrinsicalist component of *Compositionality*, whose requirement is that there be no difference in semantic relationship without a difference in semantic feature. In the end of the day, semantic relationism rejects intrinsicalism, and not compositionality *per se*.

Thus, semantic relationists accept (with Fregeans) that there is a semantic difference between identity sentences. They block the inference licensed by *Compositionality*, which has it that since there is a semantic difference between the sentences, it should also be a semantic difference between the names. Instead, the inference that one can validly draw

from here is that there is a semantic difference between the pairs of names "Cicero", "Cicero", and "Cicero", "Tully".

Is there a more profound rationale for supporting this view of the meaning of names?

At this juncture, it is worth taking a look at a similar semantic phenomenon which involves the variables of a first-order language. Here we have an analogy with Russell's puzzle of the variable. This puzzle emerges from the following considerations: the semantic role of two distinct variables in two distinct expressions, *viz.* "x > 0" and "y > 0", is the same; whereas, the semantic role of the two variables within one expression, *viz.* "x > y", is different, since "x > x" would be a very different statement.

The idea of this relationist semantics is this: the semantic role of the individual variable is given by the range of values the variable can take; however, this is not going to settle the issue of whether several variables can take any value together. That requires an independent specification. While each of x and y can take any value from the domain, it might or might not be that both can take any particular value simultaneously.

So, how are we going to fix the problem?

> "We must allow that any two variables will be semantically the same, even though pairs of identical and of distinct variables are semantically different; and we should, in general, be open to the possibility that the meaning of the expressions of a language is to be given in terms of their semantic relationship to one another." (Fine 2009, 24)

In the end of the day, Fine rejects semantical intrinsicalism for variables as well. Thus, in addition to specifying the values each single variable can range over, if taken on its own, one should also specify which values several variables can take, if taken together. So, one core tenet of semantic relationism is that from the specification of a range of values for individual variables it does not follow which values the variables can simultaneously take. But if this is so, then it is clear that the intrinsicalist doctrine, *viz.* no difference in semantic relationship without a difference in semantic feature, will fail.

The overall relationist – and anti-intrinsicalist – moral that emerges from this analysis is that it is only by giving up the intrinsicalist doctrine,

plausible as it initially appears to be, that Russell's antinomy of the variable is to be solved. There are situations in which things can only be distinguished in terms of their relations to one another, and not only in terms of their intrinsic features.

This semantic and philosophical discussion brings along a new relational view of variables, and a new relational semantics for the language of first-order logic. The aim of the semantics, as standardly conceived, is to assign a semantic value to each (meaningful) expression of the language under consideration. The aim of a relational semantics is to assign a semantic connection to each sequence of expressions.

The connection is meant to encapsulate (i) the semantic features of each individual expression and (ii) the semantic relationships between the expressions. The semantic connections replace the semantic values as the principal objects of semantic enquiry. This notion of a semantic connection is a generalization of the notion of a value range for a variable, that is of the set of values the variable can take.

Now, after making this analogy between names and variables the issue is: can we make use of the semantic relationist view of variables as an analogy for developing a semantics for names, as well?

In passing from variables to names there is one major obstacle, though! There is a crucial difference between variables and names: in the case of variables "x" and "x" take coordinated values; whereas the variables "x" and "y" take their values independently of each other. On the other hand, in the case of names, however, the semantic role of each coreferential name is already fixed, and so "coordination" or "independence" do not seem appropriate here.

If there is a way in which *Semantic Relationism* works for names, then in that case we should have the following situation: the referentialist assumption, *i.e.* there is no semantic difference between "Cicero" and "Tully", is compatible with there existing a semantic difference between the pairs of names "Cicero", "Cicero", and "Cicero", "Tully".

To come up with an explanation for this, semantic relationists will have to reject a crude principle of *Compositionality*, that one which incorporates Intrinsicality, and consequently to argue that a semantic difference between the pairs of names need not imply a semantic difference between the names themselves.

But what would be the rationale for that rejection? For the difference in meaning between the two identity sentences to exist, there should exist a semantic relationship between "Cicero" and "Cicero" that does not hold between "Cicero" and "Tully". And what is noteworthy is that this property of the semantic relationship per se be not grounded in an intrinsic difference between the names themselves.

Now, Fine's suggestion is to differentiate between the pair of names "Cicero", "Cicero", and the pair "Cicero", "Tully" along the following lines: (i) in the former case, the pair of names represents the object both *as being the same*, and *as the same*; (ii) in the latter case, the pair of names represents the object **only** *as being the same* and **not** *as the same*.

But when does a pair of names represent the object as the same? The answer seems to be that in the first case, as opposed to the second, it is part of how the names represent their objects that the object should be the same. In the first case, it is a strong semantic requirement that "Cicero" and "Cicero" are coreferential. In the second case, however, it is just a fact that "Cicero" and "Tully" corefer.

The former case is stronger than the latter. The suggestion is that a pair of names represents the object as the same when the relationship that holds between the two names in the pair is such that they strictly corefer (a semantic requirement that their reference should be the same). So, for two names to represent an object as the same is for them strictly to corefer. Two names strictly corefer, if it is a semantic fact that they corefer.

To sum up this, two expressions will represent an object as the same, if it is a semantic fact that they represent the same object. Now, we have a semantic relationist explanation of Frege's puzzle: in general, there is an incompatibility between representing objects informatively as being the same and representing them as the same. Even though there is no semantic difference between "Cicero" and "Tully", there is a semantic difference between the pairs of names "Cicero", "Cicero", and "Cicero", "Tully". Only in the first pair the names strictly corefer, in virtue of a semantic fact, whereas in the second pair they corefer in virtue of just how things are.

References

Fine, Kit (2009). *Semantic Relationism*. Wiley-Blackwell.

All links were verified by the editors and found to be functioning before the publication of this text in 2024.

Andrei Ionuț MĂRĂȘOIU[1]

INTEGRITY, COURAGE, SCEPTICISM

Abstract. Does our inability to always self–attribute epistemic courage when we manifest it (a limited kind of reflective scepticism about epistemic courage) conflict with the virtue of integrity? I articulate this problem, and argue that it can be eschewed if we construe integrity along meliorist rather than perfectionist lines.

Keywords: integrity, courage, scepticism, awareness

I start by setting the stage for the problem I intend to articulate and then to address that problem. At a first pass, amending which does not change what follows, here are a few basic tenets regarding virtues. Aristotle teaches that virtues are the moderation flanked by vices. Courage, for instance, is a mean between cowardice and rashness. In fact, we will need less than this in what follows, although I will refer to it cavalierly as a view in the neo-Aristotelian family. What I will need is only the view that virtue might be a *via media* among non-virtuous features, be they vices or amoral dispositions or simple variation among situations.

Turn now to Kant, who teaches that at most God and our first-personal conscious awareness can distinguish acts of duty (and virtue) from acts that merely agree with duty but are done out of interest or inclination. Suppose that Kant is right in this. Suppose further that introspection is an unreliable guide to our mental states (broadly

---

[1]   University of Bucharest, Faculty of Philosophy. Email: <andrei.marasoiu@filosofie.unibuc.ro>. ORCID: /0000-0001-9400-7702

construed). Since I will primarily be concerned with a problem that can be raised first–personally, it is immaterial whether we assume that God doesn't exist; or whether we alternatively make the weaker assumption that God doesn't always reveal the naked truth to us and we don't know which situation we are in now; or whether we make the even weaker assumption that even if truth were revealed to us, revelation doesn't – or not always, at least – carry enough justificatory weight in context.

From these broadly Kantian assumptions it follows, or at least seems to follow, that we will not be able to justify a verdict of whether the act committed was one of virtue or not. Perhaps it was, instead, an amoral or even immoral act, but in ways not transparent to us in reflective consciousness. Scepticism is not nihilism. I am not claiming that there are no acts of virtue – or courage; I only claim that, if Aristotle and Kant are both right,[2] and if the additional assumption that our introspection is unreliable holds, then one cannot justify – or know – performing or having performed acts of virtue. Nor need this be entirely accurate from a purely historical point of view; my aim now is only to articulate a conceptual problem. Now apply all this to courage: the view *could be* restricted in that one can be sceptical about whether one performs (or has performed) acts of courage, all the while perhaps not always exhibiting a similar scepticism about other acts of virtue, *e.g.* ones of generosity.

And illustrate with an imaginary example. Suppose you pass by a homeless person that seems to you – and is – hungry, and you have a sandwich in your backpack. You did not intend to give it to anyone else who needed it more. Instead, you had planned to eat it yourself. But you're really not that hungry, and your plan was only tentative – you hadn't promised to yourself you'd eat that sandwich no matter what. You give the sandwich to the homeless person. Not before you quickly check yourself for reasons: it's not to make you feel good, it's not for advertising or self-promotion, etc. In the quick of an eye, you gauged that the homeless person could use that sandwich more. And you

---

[2]   One might plausibly object that Aristotelianism and Kantianism are entirely different views. My target here are views that regard both virtues and principles as key to rightness and goodness – indeed or thought alike. Rosalind Hursthouse's (2006) proposal to translate back and forth between the jargon of Aristotelian virtues and that of Kantian principles is an excellent example of the sort of mixed views I consider.

proceeded accordingly. So, you know this was an act of generosity. If repeated, in the right circumstances, with the right motivating reasons, it might instill or reinforce the virtue of generosity in you.

Switch now to courage, extending the scenario. As you step away, you see a man approach the homeless person menacingly and bullying them, maybe kicking them directly. Again, you need to read the situation as it unfolds[3]. The aggressor is well built, so you might suffer some injuries. But the law is on your side, you see cameras. Cameras record; they don't interfere. You decide to step in and be brave. Brave, not hasty because you would have materially profited from some other way out. Upon quick but clear deliberation you decide that interfering is risky, but the cost is worth it. Deliberation tones down your outrage, eases you into the act of courage, but perhaps it was the outrage itself that moved you away from fear and laziness. Absent your deliberation – as speedy or as accurate as it may be – you would have simply traded your momentary cowardice for an instance of hastiness, rashness, juvenile impetuousness.[4]

But now, you granted Kant was right. If introspection is unreliable, how do you know that you deliberated – rather than daydreaming deliberation? How do you know you acted bravely rather than with hasty self-deception? Again, scepticism is no nihilism. No one denies you were brave. But how do you know it? To ease you into scepticism, consider how underdetermined this imagined scenario is. Just how better-built is the aggressor than you yourself? Do you have any history of self-defense or martial arts? Just how outraged are you? If too much, wasn't this natural inclination rather than reason that prompted you to act? If too little, did you really act out of reason, or were you trying to

---

[3]   To grant that acts of virtue may be performed in manners sensitive to the circumstances of the situation at hand is not to presuppose a variety of the situationist challenge concerning either epistemic or moral virtues (Alfano 2012), but only to consider the development of moral sensibility alongside our abilities to act and reflect in normatively shaped ways. For discussion of situation-sensitive moral salience and moral attunement, see Barbara Herman's early (1985).

[4]   Throughout this text, I stay away from contributing to the protracted debate between reliabilism and responsibilism about virtues, be they moral or epistemic. However, notice that from the standpoint of sheer reliability (understood either probabilistically or as absence of counterfactual risk), we couldn't, at first blush, distinguish overt behaviors that merely conformed to duty from ones that were genuinely (*i.e.* responsibly) virtuous.

prove something to yourself? Say that you're an upright sort of individual? If so, virtue-signaling would here trade for genuine virtue.

So, you could be in the know that you were generous but in the dark as to whether you were brave or not. Or so the imagined scenario stipulates. Perhaps the very same worries could be marshaled against generosity as against courage. But a blanket scepticism about virtue self-attribution is not my target here. In order to unearth the problem I'm interested in, local scepticism[5] – about courage as opposed to generosity, in the case at hand – will be shown to be of better use below.

Up until now, things were physical. Giving a sandwich, stepping in to stop a bully. But all this applies, willy-nilly, to epistemic courage. Again, suppose epistemic Aristotle and epistemic Kant are right, and introspection is unreliable. You could be sceptical about self-attribution of epistemic courage. An example: in an epistemically impoverished environment, you trusted astrology because it was a tradition handed down from your parents and grandparents. But once having reached primary school, you learn that the Earth revolves around the Sun. And the penny drops: *those can't both be true*. You look some more and decide – a brave act – that your family tradition was based on misconception. Again – was this brave or did you just defeat traditional complacency with an unreflected endorsement of what you learned in school? It's true alright, and you did the right thing. What you now believe is true. But was your decision carefully justified? Do you have grounds sufficient to distinguish your intellectual courage from its close and vicious or value-neutral neighbors? If epistemic Kant (etc.) is right, you might not.

The Kantian sceptical scenario was one wherein one cannot know that one is or was epistemically brave – and not because that's false but because one lacks appropriate justification. How should we regard this: is it a bug or a feature of our epistemic condition, *i.e.* the limits of what we can know? We might be tempted to think that this is simply a feature for ordinary courage: the courage manifested in overt behavior. Whereas, asymmetrically, you might be tempted to think that when it comes to intellectual or epistemic courage this is a bug. It would be an epistemic

---

5   For local varieties of scepticism and their use, see Duncan Pritchard's work, especially his (2005), as well as discussion in Mărășoiu (2020).

bug because of a meta-feature of epistemic courage. Namely, that epistemic courage is such that exhibiting it is opaque to our reflective self–attribution, opaque because self–attribution would be unjustified, even if denying it would equally be unjustified. If the opacity of our self–attribution of courage is a bug when the courage is epistemic but no bug when the courage is moral, asymmetry obtains. Whereas finding no epistemic bug is consistent with symmetry between cases of moral and epistemic courage. To explore whether there is an epistemic bug, I need a bit more stage-setting.

Intermezzo: Throw another virtue into the melting pot, if virtue it is – integrity. On an Aristotelian reading of virtues (granted from the outset), there is a unity of virtues. The unity is not organic and not physical and not mathematical. It is, rather, virtuous itself. Suppose, then, that virtues – or the character dispositions that underwrite them – coalesce into one whole. One's character – or the set of virtues one is possessed sufficiently of – guides one's overall behavior and moderates any excesses by the company of other counterbalancing virtues. For instance, mix bravery and prudence, generosity and foresight, or an uncompromising stance toned down by modesty, tolerance and open-mindedness. The supposition that such a character exists is, I argue, a consequence of the Aristotelian reading of virtues already granted. Similarly, the supposition that the unity of such a character is not a matter of physical or biological nature but a unity of virtue – a second–order virtue, integrity – is, I argue, another consequence of Aristotle's thinking about this. In truth, as is in goodness: epistemic integrity would, by parity, be the unity of epistemic virtues. It seems best to bracket Aristotle's metaphysics of dianoetic virtues and instead treat integrity wholesale: as a unity of virtues in overt behavior and mental acts alike. (Nothing changes in what follows if we indulge in this slight simplification.) If integrity includes (or balances out) the unity of epistemic virtues, it also includes (or balances out) virtues of reflection, such as: care, imaginativeness, drawing logical consequences with care, conscientiousness, open–mindedness, intellectual courage, or reflective sensitivity to relevant reasons.[6]

---

[6] To envision how virtues embody principles is not to assume that either are codifiable in language, nor to deny that is possible; for discussion, see McDowell (1979).

The reason why I introduced integrity is to argue that the best way to flesh out this idea – that it is an epistemic bug that self-attribution of epistemic courage might, at least on occasion, be unjustified – is to consider a possible conflict between courage and integrity when both are construed epistemically. To wit, inability to justify one's self-attribution of epistemic courage in reflective awareness is detrimental not to epistemic courage itself but to the epistemic side of integrity: to how virtues coalesce, cohere and band together in epistemic character.

And the claim I defend is that thinking that there is an epistemic conflict between courage and integrity when both are epistemically construed betrays a perfectionist construal of integrity. If, on the other hand, we have a meliorist understanding of integrity,[7] on which an impoverished epistemic environment or shortcomings might simply be practically unavoidable, it is not an epistemic bug that self–attributions of epistemic courage are unjustified, it might just be a feature, a fact of one's epistemic life.

Again, meliorism can be qualified. Just as I explored local scepticism with respect to self–attributing epistemic courage, an epistemic shortcoming that doesn't taint the epistemic achievement of courage unabashed, so we can envision a local variety of meliorism, with respect to integrity alone rather than any (first-order) virtue such as (epistemic) courage. Indeed, on different grounds, we might expect that building a normative conception of our self-identity will be a project we might have stood to embark on only knowingly that the possibility of failure is near. This is not to say that meliorism about integrity depends on the metaphysics of self-constitution and self-identity, only to intimate that it naturally accommodates a wide array of such views[8] which, again, mix principles and virtues.

---

[7]   It might seem a commonplace to regard Aristotle as a perfectionist. I take issue with that description. If MacIntyre (1984) offers at least one among several attractive reconstructions of how to understand skills and virtues in Aristotle, and construes them as social practices cultivated throughout one's life and exercised in order to maintain them, the meliorist aspect is, while not entailed, strongly suggested. Natural goodness, on this view, is primarily second nature and requires constant training and practice. For debate, see McDowell (1979) and Foot (2005). For misgivings about carrying over descriptions of skills that are normatively laden to virtues where Aristotle is concerned, as well as any moral-to-epistemic carryover, see Watson (1996).

[8]   Perhaps the most prominent among which is the view articulated by Christine Korsgaard (2011).

This is not to say that debugging is done, only to say that it might vary from case to case, reinforcing the suggestion that case scenarios are underdetermined. It is a methodological point, if any, not to conflate shortcomings of how cases are described with robust axiological characterizations of which virtues, vices, or characters might be instantiated in these cases.

Acknowledgments

References

Alfano, Mark (2012). "Expanding the Situationist Challenge to Responsibilist Virtue Epistemology". In *Philosophical Quarterly* 62 (247), 223-249.

Aristotle (2014). *Nicomachean Ethics*, edited by Roger Crisp. Cambridge: Cambridge University Press, 2nd edition.

Baehr, Jason (2006). "Character, reliability and virtue epistemology". In *Philosophical Quarterly* 56 (223), 193-212.

Foot, Philippa (2005). *Natural Goodness*. Oxford: Oxford University Press.

Herman, Barbara (1985). "The Practice of Moral Judgment". In *Journal of Philosophy* 82 (8), 414-436.

Hursthouse, Rosalind (2001). "Right Action". In *On Virtue Ethics*. Oxford: Oxford University Press.

Kant, Immanuel (2011). *Groundwork of the Metaphysics of Morals*, edited by Mary Gregor, Jens Timmermann. Cambridge: Cambridge University Press.

Korsgaard, Christine (2009). *Self-constitution: agency, identity, and integrity*. New York: Oxford University Press.

MacIntyre, Alasdair (1981). *After Virtue: A Study in Moral Theory*. South Bent: Notre Dame University Press.

Mărăşoiu, Andrei Ionuţ (2020). "Intellectual Virtues and Biased Understanding". In *Journal of Philosophical Research* 45, 97-113.

McDowell, John (1979). "Virtue and Reason". In *The Monist* 62 (3), 331-350.

Pritchard, Duncan (2005). *Epistemic Luck*. Oxford: Oxford University Press.

Watson, Gary (1996). "Two Faces of Responsibility." In *Philosophical Topics* 24 (2), 227-248.

All links were verified by the editors and found to be functioning before the publication of this text in 2024.

Stelian Mădălin MIHAI[1]

WORLD-DEPENDABLE EXISTENCE
IN MODAL MEINONGIANISM – WHY CAN HERA SLAP ZEUS
AND EVERYTHING BE ALRIGHT?

Abstract. In this paper, I propose an interpretation for the semantics of intentionality, that Graham Priest uses in constructing his modal meinongianism, in "Towards Non-Being". More precisely, I will take into account only the issue of existence, as a metaphysical notion. In this regard, my claim is that the monadic predicate of existence is not capable of constructing a full-fledged metaphysical notion of existence, or, differently said, it is not well equipped to account for all the modes of being that modal meinongianism implies. In trying to support my claim, I will use a strategy of reasoning, which employs a hierarchical conceptual structure, meaning that there are some first concepts, which determine the meaning for all the others. In this case, *noneism* is the primary notion, which imposes how other notions, such as modality and existence, will be defined. By using this way of reasoning, I will conclude that existence can be interpreted as world-dependable, meaning that the ontological nature of a world decides the ontological nature of the objects of its domain. In the pursuit of an argument for my thesis, I will begin with a short presentation of the core ideas of modal meinongianism. Afterwards, I will clarify the specific problems I am interested in and justify their necessity for being discussed. Lastly, I will formulate my argument, show its limits, and suggest a case of further research.

Keywords: Graham Priest, modal meinongianism, noneism, existence, modality, world semantics

Introduction

In his book, *Towards Non-Being* (2005), Graham Priest has as a purpose to build a better version of the Principle of Characterization (CP), initially

[1] PhD student, Faculty of Philosophy, University of Bucharest. Email: <stelianmadalinmihai@gmail.com>

designed by Alexius Meinong. The motivation for such an effort is to accept a metaphysical perspective, named "noneism". In this case, Priest said that "Even when I had come to accept that noneism is a perfectly coherent and common sense view, I did not immediately accept it" (Priest 2019, 685). The reason behind that hesitation was that noneism proved insufficient by itself. Noneism states that there are some objects which do not exist, and we can still talk about them in a coherent manner. The problem is which predicates should non-existent objects receive. Priest's answer is CP, but not in its initial version, because this version has some unavoidable shortcomings. So, Priest decides to build a better version of CP, by constructing a semantics of intentionality. Intentionality is understood as the ability of thought to direct itself towards any object, irrespective of its kind (Priest 2005, 5). If our mind can think about anything, then we can think about non-existent objects and impossible objects as well. Because of this, the aforementioned semantics needs to take into account every type of object that an intentional act can be about. Being aware of such a necessity, Priest comes up with a better version of CP, by using modality. The modal version of *CP says that an object has all of its properties, but not necessarily at the actual world, but at any world that can "realize the way the agent represents things to be in the case at hand" (Priest 2005, 85). Now, let us see the main ingredients of this semantics of intentionality, so that it can be clear how Priest manages to build a better version of CP.

Graham Priest's Modal Meinongianism – Core Ideas

The basic requirements to start the semantics for intentionality are the following: a set of possible worlds (P), a set of impossible worlds (I), such that $P \cap I = \emptyset$, the actual world (@), which is part of the possible worlds, a non-empty domain of objects (D), and a function (δ) that assigns every non-logical symbol a denotation. Given such ingredients, given a certain constant c, δ assigns to it an object from D, but it is important to mention that δ treats logical expressions of the form "A → B" as atomic formulas at impossible worlds, which means that they receive extension and coextension. At the same time, given a predicate P, and a closed

world[2] w, it follows δ (P, w), which is to be written as {δ+ (P, w), δ- (P, w)} – thus, making it clear that extension and coextension are explicitly defined. Additionally, given a certain intentional verb Ψ, δ(Ψ) presents itself as a function that maps each particular domain d to a binary relation on $C: R^{\overline{\overline{\Psi}}}^{d}$ (Priest 2005, 16-17).

Lastly, this semantics uses neutral quantification (Π, for the universal quantifier, and Σ, for the existential quantifier) and a monadic predicate for existence, Ex. If quantifiers are neutral, then they do not imply any ontological commitment. The ontological commitment is expressed by the first-order predicate of existence (Priest 2005, 14).

Now, this semantics is capable of logically represent sentences like:

A: Some dragons exist.
B: Hera slaps Zeus.
C: I imagine a horse running forward and backwards at the same time.

And the logical representation is:

A: $\Sigma x(Dx \& Ex)$, meaning that there are some objects, which satisfy the predicates of being a dragon and of existing.

B: hSz, which means that Hera (h) slaps (S) Zeus (z).

C: @ $\Vdash^{+} a\Phi H(O_H) \rightarrow @R^{\Phi\delta(a)}wi, wi \Vdash^{+} H(O_H)$, meaning that, at the actual world @, somebody (a) is imagining (Φ) that the following set of properties H = {horse, running backwards, running forward} can be predicated about an object, O_H, and because such an object cannot be part of the actual world, by an accessibility relation to an impossible world (@$R^{\Phi\delta(a)}wi$), it is true at that impossible world that there is a horse which runs backwards and forward at the same time.

The intentional operator from sentence C is an example of how a special intentional operator works in this semantics. This special operator in cause is the representational operator, Φ, which reads: "… represents … as

---

holding (in the matter at hand)" (Priest 2005, 84-85). The way it works is like this. Firstly, we assume that *A(x)* is any condition or set of properties. If so, then someone could use that set of properties to intend an object of thought. If it is intended, then it is particular or well defined. If it is well defined, then it is instantiated. So, if someone uses *A(x)* to think about an object, then we can use "C_A" to rigidly design it. Consequently, we might reach the following logical expression @ $\Vdash^+ a\Phi A(C_A)$, given the fact that @ $\Vdash^+ A(C_A)$ may not hold. If @ $\Vdash^+ a\Phi A(C_A)$, then @$R\Phi^{\delta(a)}w$, $w \Vdash^+ A(C_A)$. This means that whatever I am thinking about the object C_A, it does not mean that its truth should be established in the actual world, because the condition or the characterizing set of properties could be instantiated in a non-actual world, by usage of the right representational operator (Φ) and due to a relevant agent (*a*) (Priest 2005, 85). Differently said, at the actual world, a subject represents an object as having a set of properties, but the instantiation of those properties does not necessarily happen in the actual world, even if the agent is in the actual world. And because of the unrestricted ability of thought to direct itself towards any object, this semantics must include impossible worlds as well, in order to make it possible to characterize impossible objects.

Thus, the modal version of the Principle of Characterization is formulated and it states that: "we just do not assume that an object characterized in a certain way has its characterizing properties at the actual world, only at the worlds which realize the way the agent represents things to be in the case at hand" (Priest 2005, 85). On of the initial problem of CP was the possibility to derive the existence of anything that has a set of properties – "if *A(x)* is any property, or conjunction of properties, we can characterize an object, C_A, and be guaranteed that that *A(C_A)*. This is the *Characterization Principle* (CP) in its most naïve form" (Priest 2005, 83). If the initial CP was unrestricted, this means that I could think about any set of properties (for instance, being round and being square) and add to that set the property of existence, which ultimately leads to acknowledging that an object exists and satisfy contradictory properties. Actually, the problem is that such an object must make part of the ontology of the actual world. But with the modal version, such an object makes part of an impossible world, thus avoiding the inclusion of inconsistent objects in the ontology of the actual world.

The introduction of impossible worlds was necessary in order to make sure that the new version of CP could accommodate any object an intentional act can be about. Surely, the notion of "impossible worlds" has other usages, such as solving the problem of logical omniscience, but in this case what matters is its role in making a modal version of CP. But what is an impossible world?

In trying to provide an answer to such a question, Francesco Berto has done some research, and he has found that the term "impossible" can be defined in four ways. In his article, "Modal Meinongianism for Fictional Objects" (2008), Berto presents his findings. According to his findings, the term "impossible" can be understood as: *a* – in opposition with the term "possible", meaning that "impossible" is defined as "ways in which things could not have been", if the term "possible" is defined as "ways in which things could have been"; *b* – given a certain logic, *L*, an impossible world is a situation in which its laws fail; *c* – an impossible world is a situation in which only the laws of the classical logic fail; *d* – impossible worlds are just cases in which some contradictions are true (Berto 2008, 209-210). Given the fact that Priest adheres to dialetheism in constructing this semantics of intentionality, it could be natural to assume that Priest understands the term "impossible" as "some true contradictions" – "Priest is known for championing the position of dialetheism" (Ferguson & Başkent 2019, 2). Dialetheism is the claim that some sentences are true and false at the same time (Stanford Encyclopedia of Philosophy 2024).

When talking about impossible worlds, one should also talk about their ontological nature and establish if they have the same metaphysical category as possible worlds. In searching for an answer to such questions, Martin Vacek's analysis might be suitable for providing us with some relevant considerations. Shortly, Vacek says that the ontological nature of impossible worlds can be considered in two main ways, which are the modal realism way and the modal anti-realism way. Modal realism claims that impossible worlds are concrete or exist in the same way as the actual world. Such a view might seem a bit radical. There is also a moderate version of modal realism, which is called "modal ersatzism", and its main claim is that impossible worlds exist, but they lack spatio-temporal dimensions. The other main view, modal anti-realism, claims

that impossible worlds cannot exist as the actual world does (Vacek 2023). The other concern about impossible worlds is if whether they have the same metaphysical category as possible worlds. And here, also, the obvious answer is that there are mainly two cases. Firstly, impossible worlds have the same ontological nature of possible worlds, a case which is called "Parity Thesis". Secondly, impossible worlds are different than possible worlds, and a particular way in which they are different might imply that impossible worlds are abstract, whereas possible worlds are concrete. This is called "hybrid genuine modal realism" (Vacek 2023).

In the case of Priest's modal meinongianism (MM), it is clear that Parity Thesis (PT) is respected, because he claims that "in the context of noneism, the obvious policy is to take all worlds other than the actual to be non-existent objects" (Priest 2005, 139). So, possible and impossible worlds have the same ontological status, which is non-existence. But what is their concrete ontological status? According to Nicola Ciprotti, "Priest explicitly declares that MN is compatible with regarding worlds either as *abstracta* or some sort, *e.g.* maximally consistent sets of interpreted sentences or propositions, or as *concreta*, *e.g.*, mereological sums of spatio-temporal particulars, along the lines famously championed by David Lewis" (Ciprotti 2014, 18). So, it is not clear what is their precise ontological status.

In order to better explain it, let us take sentence *A* and try to see what the ontological nature of a non-actual world is. The sentence "Some dragons exist" can be true at a possible world, *w*. This possible world is a non-actual world, which means that it does not exist. So, dragons, which are non-existent objects, exist at a possible world, which is a non-existent object. How can this be understood?

The predicament in this case is originated in a distinction that Priest makes between predicates that entails existence and predicates that do not entail existence. This distinction is introduced as a consequence of accepting noneism. More precisely, there are predicates which entail existence, such as "hug", and if *a* hugs *b*, then both *a* and *b* must exist. Also, intentional predicates are existence-entailing. If *a* worships *b*, then *a* must exist, but *b* may or may not exist. Furthermore, there are some non-intentional predicates that do not entail existence, such is identity. An object is identical with itself, irrespective of its existence. For other

predicates, it could be debatable if they are existence-entailing or not. One example that stands as a justification for such an observation could be a case in which *a* experiences *b*, and *b* is just an illusion. In this case, Priest does not provide an answer, but only a question – "does it follow that *b* exists?" (Priest 2005, 60). Also, he mentions that it is not his purpose to further debate or discuss this topic. What he adds to this discussion is that the following constraint – "*if $(q_1…q_i…. q_n) \in \delta^+ (P, @)$ then $q_i \in \delta^+ (E, @)$* – applies only at the actual world" (Priest 2005, 60). So, if an object has existence-entailing properties, that object must be part of the domain of the actual world. If this is so, how could one understand the following claim: "As a matter of fact, it seems to me that existence-entailing is world-invariant, at least at possible worlds. Thus, for example, if *a* hits *b* at such a world, *w*, then *a* and *b* exist at *w*" (Priest 2005, 60).

It seems a bit strange, because existence-entailing was supposed to apply strictly to the actual world, but if in a possible world a predicate implies a physical interaction between two objects, then those objects must exist in this possible world. For now, this distinction between predicates that entail existence and predicates that do not entail existence is somehow unclear or unfinished. And the reason for confusion and unclarity is the way existence is understood.

Existence – A Cause for Predicaments

Priest defines existence as being concrete or having spatio-temporal properties. Therefore, if an object exists, it is concrete. Non-existent objects are non-concrete. If, to this definition of existence, it is added the distinction between predicates that entail existence and predicates that do not entail existence (⊕[3]), then a predicate that entails existence is entailing being concrete or having spatio-temporal properties. Let us use sentence *B* to see how the definition of existence and the distinction ⊕ work together.

Sentence *B* says: "Hera slaps Zeus". "To slap" is obviously a regular predicate, that implies a physical interaction between two objects. If a

---

[3] The symbol "⊕" will stand for "predicates that entail existence and predicates that do not entail existence".

physical interaction is implied, then that predicate entails the existence of the objects which is about. Therefore, Hera and Zeus exist. Therefore, they are concrete, or they have spatio-temporal dimensions. Even worse, the non-actual world Hera and Zeus occupy has to be concrete as well, because it is assumed a metaphysical sameness between a world and its denizens. So, the possible world where Hera slaps Zeus is concrete or exists, because there is an identity between the terms "to exist" and "to be concrete". But Hera and Zeus are non-existent objects, and a possible world is a non-existent object. So, something must be wrong, because otherwise a contradiction is reached: a non-existent object exists.

Ciprotti uses the same conditions, which are the definition of existence and the distinction ℗, to claim that MM is incompatible with PT. In his own words, "I intend to show that MN sits badly with PT, and therefore that, in order to make MN a consistent view, impossible worlds are to be regarded as *abstract* not concrete entities" (Ciprotti 2014, 14). His argument begins with an example about a fictional object and reaches the intermediary conclusion that possible worlds must be concrete. Given the acceptance of PT, then impossible worlds should also be concrete. But one cannot accept concrete impossible objects. So, it is better or more efficient to consider impossible worlds as being abstract (Berto and Jago 2019, 55). Such a conclusion is reach by understanding that Priest leaves room for establishing the nature of non-actual worlds in any way, be it as concrete objects, be it as abstract objects (Ciprotti 2014, 14). In the end, the actual problem is not the distinction ℗, but rather the definition of existence. The distinction ℗ is just a way of using the definition of existence.

In order to support the claim that existence is actually a cause for some predicaments, other interpretations or readings of Priest's proposal can be taken into account. For instance, Wolfgang Barz, in his article, "Two-Dimensional Modal Meinongianism" (2015), provides a solution to the problem of cross-world identity and the characterization of non-existent objects. Those problems had been initially pointed out by Fred Kroon, in the article "Characterization and Existence in Modal Meinongianism" (2012). In essence, Kroon claims that Priest's MM makes room for accepting that non-existent objects might possess properties that should only apply to existent objects (Kroon 2012, 25). If so, then the central idea of MM fails

to reach its purpose, which was to make a clear distinction between existing and non-existing objects (noneism). Returning to Barz's proposal, he conceives a different semantics for MM, in order to clearly make a distinction between how terms refer in the actual world and how terms refer in possible worlds, this being a strategy to deal with problem of cross-world identity and the characterization of non-existent objects (Barz 2015, 13). The intricacies of his proposal are not the purpose of this discussion. What actually matters is that Barz took Kroon's arguments as valid observations and tried to solve the problems they had presented.

The last perspective that I am going to take into consideration regarding some possible shortcomings of Priest's MM belongs to Otávio Bueno and Edward N. Zalta. In their article, "Object Theory and Modal Meinongianism" (2017), Bueno and Zalta claim that Priest's MM is not well equipped to justifiably use the names of fictional objects. Their argument begins with the claim that *CP is the fundamental idea of MM. His workings imply using an accessibility relation, by which the object of an intentional act is places in its corresponding world. The problem that Bueno and Zalta are trying to point out is that MM cannot justify the usage of name of fictional objects. More precisely, they claim that MM does not have an explicit way of justifying the fact that a name of a fictional object reaches exactly the object implied by the characterizing set of properties. For instance, the name "Sherlock Holmes" uniquely denotes the object is supposed to denote, but MM does not have a clear reason to make it so. The fact that a name of a fictional object reaches its denotation is just an assumption. And such a conclusion seems to be right, since Berto and Priest proposed a change in MM, so that names that make reference to fictional objects could surely have a unique reference. The changes that they propose are the following: *a* – "if something satisfies *A(x)* at @, *εxA(x)* denotes one such thing"; *b* – "if not, it picks out some non-existent object or other which satisfies *A(x)* in the situation one is envisaging" (Berto & Priest 2014, 195). Even with such changes, Bueno and Zalta are not inclined to accept that MM can properly and reasonably use names for fictional entities: "But this further refinement of MM still doesn't address our objection (...). But their theory doesn't entitle them to use the name 'Holmes' in this way" (Bueno and Zalta 2017, 6).

In short, Bueno and Zalta show that MM cannot justifiably make use of name of fictional entities. And even if it could, it would be inconsistent, because MM would treat non-existent objects as existent objects, since non-existent objects become referable entities with properties in non-actual worlds. If non-existent objects and existent objects are the same in what concerns the ability to be the reference of a name, then they are equally treated, and because of that MM fails to achieve its purpose, which is to make a difference between existent and non-existent objects.

So, the definition of existence and the distinction ℗ might lead to considering that MM is inconsistent in several ways. Actually, the origin of the problems can be reduced to two different sentences, in which the term "exist" is used with the same meaning:

α –   this semantics of intentionality does not allow non-existent objects to have existence-entailing properties; and

β –   existence-entailing is world invariant (at least at possible worlds), meaning that if a hugs b in a non-actual world, then both a and b exist at w.

Returning to sentence B, if Hera slaps Zeus, then, according to β, they must exist. Or existence is to be concrete. Moreover, there is this assumption that the nature of the object has to be the same as the nature of the world it occupies. So, the world in which Hera slaps Zeus is also concrete. Given PT, impossible worlds are also concrete. Hence, impossible objects are concrete as well, which means that impossible objects exist. But α says that non-existent objects do not have existence-entailing properties, or they do not exist. So, it is a bit difficult to make those two claims not being in a conflict. Also, let us not forget about the unclear ontological nature of non-actual worlds.

In the following section, I will propose an interpretation which might lead to dissolving the conflict between those two claims, while working with an assumption about the precise ontological status of non-actual worlds.

Upside Down – World-Dependable Existence

To show that the predicate of existence is not sufficiently wide to account for all the modes of being that MM implies, I need to firstly talk about these modes of being. While talking about this, I will reach to an answer regarding the imprecise nature of non-actual worlds.

As it has been shown, to exist is to be concrete. Only the actual world and its objects exist or are concrete. There is an identity between "to exist" and "to be concrete". But noneism implies the possibility to talk about non-existent objects as well. So, in what ways can one conceive the precise ontological status of non-existent objects? Well, Priest is truly categorical about this and claims that whatever is not concrete does not exist. So, non-existent objects are not concrete objects. In this case, the normal follow up could be to say that they are abstract. But Priest says that if an object is not concrete, it does not necessarily mean it is abstract (Priest 2005, 139). The usual dichotomy might be concrete-abstract, without a third possibility. In order to understand Priest's apparently confusing claim, it will be necessary to talk about fictional and abstract objects.

In the category of non-existent objects, Priest includes fictional objects, abstract objects and non-actual worlds. An abstract object is an object which does not exist and cannot exist. A fictional object is an object which does not exist, but it can exist. Priest's way of defining these terms makes appeal to counterfactuals as well: "an abstract object is one such that, *if it did exist it would still not causally interact with us*. Conversely, a concrete object is one such that, *if it did exist, it would causally interact with us*" (Priest 2005, 136-137).

In this case, my proposal is that, in order to make clear what is the ontological status of non-actual world, one should work with another distinction. Usually, the distinction concrete-abstract is applied, but MM needs a different approach. Instead, I will use the distinction existent-non-existent. My reasoning for such a choice is that MM is susceptible of being interpreted by using a hierarchical conceptual scheme, meaning that some concepts are primordial, and these concepts decides the meaning for the other secondary concepts. The fundamental concept of MM is noneism, and noneism implies the distinction between

existent objects and non-existent objects. If this is so, then the term "existent" refers only to concrete objects, while the term "non-existent" refers both to fictional and abstract objects. So, this is why Priest can claim that, if an object is not concrete, it does not necessarily mean it is abstract. Because it can be fictional as well. If the category of non-existent covers fictional and abstract entities, and non-actual worlds belong to that category, then it is assumed that possible worlds are fictional entities, and impossible worlds are abstract entities.

More precisely, noneism needs modality and shapes modality in accordance with the distinction existence-non-existence. In this case, non-existent worlds are either fictional or abstract. Therefore, an object can be concrete, fictional or abstract. These are all the modes of being that MM makes use of. If noneism shapes modality, then modality shapes the existential status of objects of their modes of being.

Returning to my proposal, there are some reasons to consider that possible worlds are fictional and impossible worlds are abstract. One reason could be that, if non-existent includes only fictional and abstract, and non-actual worlds include only possible and impossible, then, given their definitions or understanding, it is plausible to consider that possible worlds are fictional entities, whereas impossible worlds are abstract entities. Another reason might be a favourable interpretation of Priest's following claim: "The worlds that realize the Holmes stories are replete with things that, were they to exist, would be standard physical objects, like people and hansom cabs. Were these worlds with their denizens to exist, we would be able to interact causally with them" (Priest 2005, 139). Here, I believe, it is a clue to considering the fact that possible worlds behave like fictional objects, due to the usage of the phrase "were they to exist", and obviously the definition of the term "fictional object". I believe the same line of reasoning can be applied for impossible worlds as well, and if an object like a round square was part of such a world, were it to exist, it would not be possible for it to causally interact with us. But now if an impossible object exists, it means that it is concrete, but the world he occupies is abstract. How can this be understood?

Returning to the applied strategy of reasoning, noneism defines modality and modality defines the modes of being. The modes of being are concreteness, fictitiousness and abstractness. All that remains is to

reconsider the notion of existence. With this in mind, let us mention again those three sentences:

A: $\Sigma x(Dx \& Ex)$
B: hSz
C: $@ \Vdash^+ a\Phi H(O_H) \rightarrow @R\Phi^{\delta(a)} wi, wi \Vdash^+ H(O_H)$

Existence is expressed by the monadic predicate of existence, *Ex*. This predicate has access to the general domain of objects, *D*, which means that it applies uniformly. So, Hera and Zeus exist in the same way as impossible objects exist, and also in the same way as real, actual objects exist. But obviously MM does not want to reach such a conclusion, because it would be entirely against its core notion, which is noneism, since not only that everything exists, but everything exists in the same way. Nevertheless, given the fact that MM implies three modes of being, the latter consequence is avoided. Objects have a different existential status or mode of being. If this is so, then the monadic predicate of existence is not well equipped to express such a difference, or to account for all the modes of being that MM implies. This was my claim.

In the given interpretation, if a strategy of reasoning that uses a hierarchical conceptual structure of MM is applied, then noneism (being the primary concept) leads to acknowledging three modes of being, since existence is clearly different than non-existence. Also, it has been assumed that, given the primary distinction between existence and non-existence, possible worlds are fictional objects, and impossible worlds are abstract objects (PT still holds, because all non-actual worlds are non-existent). If the objects from the actual world are concrete, then the other two modes of being would be fictitiousness and abstractness. Therefore, if there are three modes of being and only the monadic predicate of existence, which applies in the same manner for every object from *D*, then this predicate cannot account for the metaphysical understanding of existence. Or the metaphysical understanding of existence seems to matter the most, because noneism must make use of it in a consistent way.

There is still one question, which does not have an answer. If impossible worlds are abstract entities, given the distinction between

existence and non-existence, then an object that makes part of such a world is concrete, because it exists there. In order to make things clearer, I believe sentence C can be useful. Sentence C says that "I imagine a horse running backwards and forward at the same time", and its logical representation is the following:

C: $@ \Vdash^+ a\Phi H(O_H) \rightarrow @R\Phi^{\delta(a)} wi, wi \Vdash^+ H(O_H)$

If it is true that, at an impossible world, iw, an object has that set of properties, and because in that set of properties there is a property which entails existence, then this object must exist. And, if to exist is to be concrete, then this object is concrete. But the world it occupies is abstract. And this is a problem because it is assumed that there must be a metaphysical identity between the objects of a world and the world itself. Obviously, there it is not.

Maybe a way to make things understandable is to use the same strategy of reasoning. If *noneism* decides modality, and modality decides the modes of being, maybe the modes of being decides the meaning of existence. Hence, modes of being could be a superior notion to existence. Simply said, being is larger than existence. Such an idea should not come as a surprise, because Meinong was the first one to claim that the *Sosein* (set of properties) of an object is independent of its *Sein* (existential status, which might be none). But Priest does not use this distinction in its full meaning. He says that "a non-existent object cannot have existence-entailing properties, like standard extensional one – at the actual world, anyway" (Priest 2005, 82). Interestingly, Priest acknowledges the difference between being and existence and uses it in a different way, by proposing the distinction ℗. In this case, the two annoying sentences can be revisited and reconsidered:

α –   this semantics of intentionality does not allow non-existent objects to have existence-entailing properties; and

β –   existence-entailing is world invariant (at least at possible worlds), meaning that if *a* hugs *b* in a non-actual world, then both *a* and *b* exist at *w*.

If the modes of being decides what existence should mean and if noneism is to be respected, then to exist is just to be concrete. In this case, sentence α is not changed in any ways, because the term "existence" is used properly, with its given or original meaning. However, in the case of sentence β, the term "existence" does not mean to be concrete. In this case, it actually means "being". So, being-entailing is world invariant, meaning that if a hugs b in a non-actual world, then both a and b *are* at w. Basically, sentence β preserves or defends the notion of being part of a non-actual world in the same manner. Specifically, Hera and Zeus do not exist in w, but they are in the same way part of the same world w. So, "a and b exist at w" does not mean that those objects are concrete. It just means that they are part of a non-existent world in the same way. This also answer to our remaining question. An impossible object does not exist, or it is not concrete, but it rather makes part of an impossible world, which is an abstract object. The assumption here is that the assigning of properties to non-existent objects makes use of being as a fundament or support. In other words, it is assumed that non-existent objects can have properties only if they have a being (obviously, other ways to deal with this issue are not excluded). In this case, I can imagine a horse that runs at the same time forward and backwards, because this object will have a being in an abstract world. Therefore, existence of objects can be conceived by making appeal to the nature of the world. Nevertheless, if existence is world-dependable, this does not actually solve the core problem.

The core problem is that MM uses the monadic predicate of existence for expressing all the modes of being. If existence is world-dependable, to exist in the actual world is to be concrete, to exist in a possible world is to be fictional, and to exist in an impossible world is to be abstract. But this semantics of intentionality cannot make use of the same predicate of existence to account for all the modes of being. If the metaphysical understanding of existence is more important than the syntactical rules that govern the behaviour of the monadic predicate of existence, then there is room for improvement or proposing a positive answer to this problem. My interest in this paper was only to point out the problem, not to solve it.

Limits of My Argument and a Case of Further Research

My claim was that the monadic predicate of existence cannot account for all the modes of being that MM implies. In the attempt to defend such a claim, I used a strategy of reasoning which presupposed a hierarchal conceptual structure, by which noneism is the fundamental concept, and every other concept is defined or understood in accordance with what noneism claims. The main claim of noneism is that some objects do not exist, but we can still talk about them. How we can do that is more complicated than it seems. Priest proposed a modal version of CP, in order to make sure that he has a proper solution for assigning properties to non-existent objects. By using both possible and impossible worlds, Priest constructed a better version of CP. His purpose has been achieved, but the semantics of intentionality he formulates has some shortcomings. The identified and discussed shortcomings are the unclear ontological nature of non-actual worlds and the confusing understanding of existence, together with the distinction ℗. Applying the strategy of reasoning, I claimed that noneism can define modality, meaning that the distinction existence-non-existence is primary, which further leads to considering possible worlds being fictional objects, and impossible worlds being abstract objects. Since there are only three metaphysical categories and all of them were derived from the way modality had been understood in accordance with noneism, it could be claimed that modality decide the modes of being or the existential status of objects. And if existence is expressed by using the same monadic predicate for all cases, then my argumentation has finished here. The predicate of existence is not capable to express all the modes of being.

In doing so, I have assumed several things. For instance, I assumed that fictional objects cover the entire category of possible objects, and that abstract objects cover the entire category of impossible objects. Also, if there are not sufficient justifications for considering that possible worlds are fictional, and impossible worlds are abstract, this result might be taken as an assumption, since my purpose was not to establish the precise ontological nature of non-actual worlds. Another assumption is that there has to be a metaphysical match between objects and their worlds. At the same time, Priest uses implicitly the distinction between being and existence, by introducing the distinction ℗. Lastly, the main assumption is actually the plausibility of using a hierarchical

conceptual scheme, in order to build an interpretation in which concepts are defined in accordance with a higher notion.

In the end, my result, if it is reasonable, just shows that MM needs to reflect the metaphysical understanding of existence in its semantics as well. Not just implicitly, but also explicitly. So, all I did was to point out a possible problem.

A case of further research might be built around this problem, because the difference between being and existence must be made explicit. The three modes of being are evidently different understandings of existence. To be concrete is just to exist, *Ex*. To be fictional is to not exist, but to have the possibility of existence, which logically might be represented as: *~Ex&◊Ex*. To be abstract is to not exist and not be able to ever exist, which logically might be represented as: *~Ex&□~Ex*. If the definitions of the terms "fictional" and "abstract" change, and a counterfactual approach is employed, then the logical representations of being a fictional object and being an abstract object are the following: *Ex → Cx* and *Ex → ~Cx*, where "*Cx*" stands for "capable of entering in causal relations with the actual world". Also, if the definitions of those terms are categorical, maybe it is more suitable to have the following logical representations: *□(Ex →Cx)* and *□(Ex → ~Cx)*. Nevertheless, these logical representations show even more that there is a difference between being concrete, being fictional and being abstract. If they are correct or suitable for the beginning of a solution in this case, it remains to be decided.

Conclusion

Priest's modal meinongianism could be considered a solution to the problem of noneism, which is the assigning of properties to non-existent objects. The semantics of intentionality he uses to make a better CP achieves its purpose. But in constructing a better CP, this semantics might be considered insufficiently clear in some cases. I talked only about the imprecise nature of non-actual worlds and the meaning of existence. My claim is that the monadic predicate of existence is not able to account for all the modes of being that MM implies. My result is a negative one, meaning that I only pointed out a problem, if my reasoning can be found acceptable, together with all its assumptions. A positive answer to the problem of existence in MM might be a case of further research.

Even so, in this interpretation, it is understandable why Hera can slap Zeus, and everything be alright. In this particular case, slapping is a joyful manner of expressing the fact that Zeus and Hera are both non-existent objects, which make part of the same world, in the same manner. This further means that they have the same mode of being, which is decided by the particular world they make part of. Having the same mode of being might be what Priest had in mind when introducing the condition represented by sentence β.

References

Barz, Wolfgang (2015). "Two-Dimensional Modal Meinongianism". *Ratio* 29 (3), 249-267. https://onlinelibrary.wiley.com/doi/10.1111/rati.12105.

Berto, Francesco and Jago, Mark (2019). *Impossible* Worlds. Oxford: Oxford University Press.

Berto, Francesco and Priest, Graham (2014). "Modal Meinongianism and Characterization Reply to Kroon." In *Grazer Philosophische Studien* 90, 183-200.

Berto, Francesco (2008). "Modal Meinongianism for Fictional Objects." In *Metaphysica* 9 (2), 205-218.

Bueno, Otávio and Zalta, Edward N. (2017). "Object Theory and Modal Meinongianism." In *Australasian Journal of Philosophy* 95 (4), 761-778.

Ciprotti, Nicola (2014). "On Modal Meinongianism." In *Logical, Ontological, and Historical Contributions to the Philosophy of Alexius Meinong*, edited by Mauro Antonelli and Marian David. Boston: De Gruyter, 1-36.

Ferguson, Thomas Macaulay and Başkent, Can (2019). "Introduction to Graham Priest on Dialetheism and Paraconsistency." In *Outstanding Contributions to logic, vol. 18 – Graham Priest on Dialetheism and Paraconsistency*, edited by Can Başkent and Thomas Macaulay Ferguson. Springer Nature Switzerland AG, 1-2.

Kroon, Fred (2012). "Characterization and Existence in Modal Meinongianism." In *Grazer Philosophische Studien*, 86 (1), 23-34.

Priest, Graham (1998). "Dialetheism." In *Stanford Encyclopedia of Philosophy*. Last modified May 9, 2024. Accessed May 30, 2024.

Priest, Graham (2005). *Towards Non-Being*. Oxford: Oxford University Press.

Priest, Graham (2019). "Crossing Boundaries." In *Outstanding Contributions to logic, vol. 18 – Graham Priest on Dialetheism and Paraconsistency*, edited by Can Başkent and Thomas Macaulay Ferguson. Springer Nature Switzerland AG, 677-690.

Vacek, Martin (2023). "The Nature of Impossibility." In *Internet Encyclopedia of Philosophy*. Available online at < https://iep.utm.edu/impossible-worlds > last time accessed on May 30, 2024.

All links were verified by the editors and found to be functioning before the publication of this text in 2024.

Cristian ANDREESCU[1]

VARIOUS HUMAN TYPES FACING THE UNKNOWN

Abstract. The paper addresses the variety of human types and attitudes facing the unknown and their philosophical implications and connections with philosophy of knowledge and philosophy of conscience and consciousness. Philosophically, the human attitude in front of the unknown dwells at the boundary between the known and the unknown. This boundary could be the equivalent of the horizon of mystery in the philosophy of Lucian Blaga, a boundary that differs greatly from one human type to another. In order to more clearly differentiate the various human types, a scheme is proposed based on different attitudes related to three tendencies to know, respectively to accept reality, as it is perceived by each. The three so-called axes are: Positivism/Negativism, Curiosity/Fear, Openness/Isolation. It results in 12 types of philosophical attitudes against the unknown that are presented and analysed. This representation is interpreted in correlation to Blaga's imaginative exercise of conceiving philosophical consciousness as a prism inscribed in a sphere representing the totality of everything (the totality of existence, the universe). Other aspects pertaining the correlations between this description of the 12 types of philosophical attitudes against the unknown and the philosophy of man in Lucian Blaga are emphasized, too.

Keywords: human types facing the unknown; mystery; Lucian Blaga

Knowledge and ignorance are linked and replace each other in the great and perpetual journey of human knowledge, in the sense given by Lucian

---

[1] PhD candidate, SCOSAAR, The Institute of Philosophy and Psychology „Constantin Rădulescu-Motru" of the Romanian Academy. Member of Romanian Division of Logic, Methodology and Philosophy of Science – Romanian Committee for the History and Philosophy of Science and Technology (DLMFS – CRIFST): E-mail: <cristian.andreescu@labor-soft.ro>. The paper was presented at the Romanian Division of Logic, Methodology and Philosophy of Science Symposium titled "Beyond humanism, towards what? Where we are and where we are going", February 23, 2023.

Blaga to man's fascination with the unknown and the human situation in the "horizon of mystery"[2], but also in the sense given by David Bohm (1989) to the chain, possibly infinite, information-support-information-support etc., in accordance with the relatively paradoxical idea that particles are generated from the vacuum state. The parallelism between Blagian's philosophical vision and David Bohm's scientific vision, derived from quantum physics, is rather loose and imperfect, but nonetheless interesting. Illusion and concreteness are successive facets in the human journey of knowledge, which we examine here from the (holistic perspective of) the subject and a philosophy of subjectivity. Blaga points out:

> "(...) the very problematic of the human spirit becomes deeper and more complex due to the successive solutions absorbed by it. The solutions of philosophy are the leaves that fall to thicken and fertilize the soil in which the roots of the spiritual problem will spread more and more slowly and gradually cover more and more space." (Blaga 2003, 21)

Relating to the unknown involves philosophical, epistemological, but also psychological aspects, actually involving human complexity, in general, and human complexity in relating to the world, in particular.

Lucian Blaga himself, confronted with the unknowns of the universe, responds with a far-reaching metaphysical project, an original philosophical system, which opens with a philosophy of knowledge. The plan of Lucian Blaga's philosophical system is as follows:

> "The plan of the system is as follows: I. The trilogy of knowledge: 1. The dogmatic aeon; 2. Luciferic knowledge; 3. Transcendent censorship; II. The Trilogy of Culture: 1. Horizon and Style; 2. Myoritic space; 3. The genesis of metaphor and the meaning of culture; III. Trilogy of values: 1. Art and value; 2. (Project); 3. (Project); IV. The cosmological trilogy: 1. The divine differentials; 2. (Project); 3. (Project); V. Pragmatic trilogy: 1. Historical being (in work); 2. (Project); 3. (Project)." (Blaga 1934, 5-15)

---

2   The work *Transcendent Censorship* of 1934 makes an intriguing correlation between knowledge and mystery that will become both a central concept and a key theme in Lucian Blaga's philosophy. In this metaphysical architectural thought, man's consciousness is described by the need for knowledge and revealed to himself through knowledge. Knowledge is an act of confronting the unknown; of facing the mystery found in all aspects of the world and, implicitly, in all aspects of human ontology. Knowledge is a daring act of human intrusion into the transcendent.

It is a totalizing approach in which the philosophical, metaphysical, cognitive, ontological, cosmological, cultural, historical, anthropological dimensions meet, in some places, with descriptions of human subjectivity. Lucian Blaga specifies:

> "In other words, the system culminates in the present study [with an early outline of a philosophical system, which will undergo transformations – our note], but it does not end with it. It is certain, however, that the future works will, from now on, converge within the given coordinates, which, after the establishment of the cosmological conception, is only too natural. In writing this study of cosmological metaphysics, the author has endeavored as much as possible to spare his readers too many references to his earlier works. However, the understanding of some of the chapters will be greatly facilitated by such a reading. In any case, a 'criticism' of the entire system can only be honestly made after studying all the volumes in which it was exposed. It should also be noted that some chapters of the study will be extensively developed in future works, according to the established plan." (Blaga 1997, 163)

Man's confrontation with the unknown is a subjective confrontation. Man, as Lucian Blaga shows in both his studies of the philosophy of knowledge and culture, faces the unknown with a unique thirst to know and to create, entwinned. Addressing in depth the subjective confrontation with this situation of the unknown, respectively, the imperative need to understand it, we find that this confrontation depends on the philosophical and cognitive attitude of each person in front of the surprise or shock represented by the unknown: that is, in front of the universe beyond the limits of knowledge of our (momentary) human assessment and existence.

Human reporting to the world, as well as consciously confronting with the world generate the deep nature of this human philosophical confrontation with the unknown, leading to evaluations about an (worldly) outside and a (personal) inside (Rosenthal 2000, 201-233), to clarifications and related emotions, so guiding the being to various "awakenings", new knowledge, followed by new misunderstandings and new feelings. Blaga shows: "The feeling of 'awakening' that any great philosophical thought communicates corresponds to such an increase in the volume of consciousness" (Blaga 2003, 15), and this expansion takes place "to the detriment of the infinite state which is the sleep of the spirit" (Blaga 2003, 15).

In order to better analyze the way in which this human confrontation of the unknown is accomplished, we propose a scheme that contains three "axes", which are relatively independent from each other, even if certain correlations can be speculatively established between them.

The three proposed "axes" are:

- Positivism/Negativism (Positive attitude towards the world / Negative attitude towards the world)
- Curiosity/Fear
- Openness/Isolation.

In relation to these axes, the description of which includes philosophical concepts and themes, which can also admit psychological valences, we identify several human types described in relation to the philosophical attitude of man in the face of the unknown, which has inevitably, inherently associated with human subjectivity, and with a psychological dimension.

> "From Socrates, who demonstrates the function of concepts in the economy of the human spirit, to the current existentialists, who enliven the very consciousness of human existence, man keeps on 'waking up'." (Blaga 2003, 19)

Now is the time to specify that, in the vision of Lucian Blaga, to which we refer explicitly, the emphasis in the description of man is on the fact that man, every man, possesses a philosophical consciousness in relation to the world, although the "awakenings" experienced by particular people are different, varying from man to man, from "awakening" to "awakening" and from situation to situation.

> "*Anyone* [our emphasis] who is endowed with the necessary spiritual skills, receptive sensitivity and spontaneity can reach, little by little, the emergence of a philosophical consciousness." (Blaga 2003, 9)

Should we consider the presence of terms such as "skills", "receptivity" or "spontaneity" to indicate that Blaga is bringing psychological arguments to a metaphysical endeavour? No way.

The "awakenings" of some may be mere platitudes or truisms to others, while in other contexts and for other people they may be real enlightenments. At the same time, Blaga shows:

> "We are perfectly aware that we too operate, like Kant in the autobiographical sentences, also with a transfer of terms from natural life to the spiritual sphere. (...) Indeed, in organic-psychological life, the state of sleep is totally replaced by the state of wakefulness; and vice versa, according to a certain circular rhythm. When we speak of the sleep of the spirit, we are to imagine this state as a deep and infinite earthly state, in which the waking state will never replace more than one plane or sector of this earthly state. [The] (...) 'awakening' (...) abolishes sleep only partially." (Blaga 2003, 15)

The Romanian philosopher, unlike other modern and contemporary philosophers who approach the vast and difficult topic of consciousness (Hume, Kant, Spinoza, Brentano, Husserl, Nietzsche, Quine, Searle, Chalmers, etc.), explicitly states the idea that *everyone* possesses a *philosophical* consciousness, or, a certain stage of a philosophical consciousness (an "insight" of a certain quality or level) and not simply *a* consciousness. In other words, the metaphysical conception regarding man's unique relationship to the world concerns man's subjectivity, subjectivity that also has a psychological dimension, and this "psychology" cannot be avoided in dealing with the theme of man's relationship to the unknown. On this account, human types cannot be *exclusively epistemological*, but they are subjective, creative and artistic to a certain extent even in their human epistemological, gnoseological or existential endeavours.

Also, exclusively epistemic human beings cannot exist. The human beings (the subjective beings) are complex: in the case of any human being, the philosophical consciousness is not disconnected from a human (inner) life, illuminated by psychological phenomena. Blaga observes:

> "When the philosophic eye was directed to the interior of man, understandings had also arrived at these [ideas], which regardless of their grip under absolute relation, also became as many awakenings as possible for the human spirit." (Blaga 2003, 19)

Regarding this subject of man's relation to the unknown, for example, the theme (and the reality) of fear is as much a philosophical theme, inscribed in a philosophy of subjectivity, as it is a theme of psychology. The philosophy of mind and cognition cannot be disconnected from the

theme of *qualia* (Stubenberg 1998). Blaga himself associates "awakening" (*awareness*) with a "feeling [emphasis added. ns.] of awakening" (Blaga 2003, 15). How does it feel to "wake up" (to become aware of something)?

Introduced in philosophy in 1929 by C.I. Lewis (1991)[3] as properties of sensory data, nowadays concept of qualia is defined as a sum characteristics of experience, or, more precisely, of experimentation of the world (in terms of knowledge and in terms of feeling). In fact, these qualitative-subjective characteristics are the answer to the question "How does it feel to...?" (Nagel 1974, 435-456). They are lived, felt and assessed qualities or properties of states of consciousness, in other words, (comprehensive) "experiences", therefore, with a central place in all human actions of relating to the world, a knowledge of the world, an awareness of the world, all central aspects in the current disputes related to the topic of philosophical consciousness.

How does it feel to see colours? This innocent question, which highlighted an aspect of colours that cannot be explained physically, led to the formation of a sub-field in the philosophy of mind – "colour qualia" (Jackson 1986, 291-295).[4]

Confronting the unknown requires, hence, awareness and experience. Thus, the endeavour of identifying the various human types confronting the unknown is not about imposing psychological descriptions (disguised) as epistemological or metaphysical arguments, but about giving a comprehensive account of man's relation to the unknown, inscribed in a philosophy of subjectivity in which epistemological dimensions and phenomena intertwine with psychological ones, having of course also a certain metaphysical foundation, in an abstract total whole of the affirmation of human subjectivity, which is particularly complex.

Metaphysics involves both experience and knowledge and it is a daring act: an opening of perspective, which can prove enlightening or blinding.

---

[3]   Consciousness and awareness being subjective, it is in fact impossible to develop a philosophy of knowledge, a contemporary epistemology, or a contemporary philosophy of consciousness ignoring qualia. Cf. Shoemaker 1981, 581-599.

[4]   See also James 1890, especially the idea that the emotion is a consequence of what we experience, for example, the idea that experiences are vehicles of knowledge, not yet fully exploited in the philosophy of consciousness.

> "Metaphysics is launching of lit torches in the ultimate darkened abysses and it reverberates in us as an 'awakening'." (Blaga 1997, 5-10)

The figure below renders the previously selected dimensions for a representation of the positioning of the various specified human types facing the unknown. The three axes of the human confrontation with the external universe, with the unknown of the world also describe an ontological "volume" in which human types unfold, which allows or stifles, in various degrees and ways, the philosophical thought provoked by renewed awareness of the world and the feeling associated with that "awakening" (Blaga 2003, 19).

At the same time, it is worth highlighting an interesting isomorphism: in the perspective of Lucian Blaga's metaphysics, a philosophy that truly deserves its name, as well as a human who deserves the appellation "human" are equally indebted to a "process of shaking and enlightenment", or to a series of such processes, comparable to waking up from sleep, with meanings that transcend organic-psychological life and pertain to a realm of ideas, as well as philosophy and the human being are indebted to the feelings associated to the acknowledgement of awakening and to philosophical consciousness.

Among these, the Positivism/Negativism axis seems to require additional clarification, because, depending on the nuanced meaning we give to the association of these two words, the idea may arise that they could coincide with elements from the other two axes. We want to assign these two words the meaning of valuing the existence of the universe beyond the limits of our knowledge. That is, as humans, we can look beyond the fact that we are open or isolate ourselves in front of this universe, or, we can consider or disregard the fact that this state of things incites our curiosity or we fear it, so we can value it, cherish it or ignore it.

These three axes give way, through the combinations that take place between them, to 12 philosophical attitudes towards the unknown, which I noted with a letter from A to L. They are as follows:
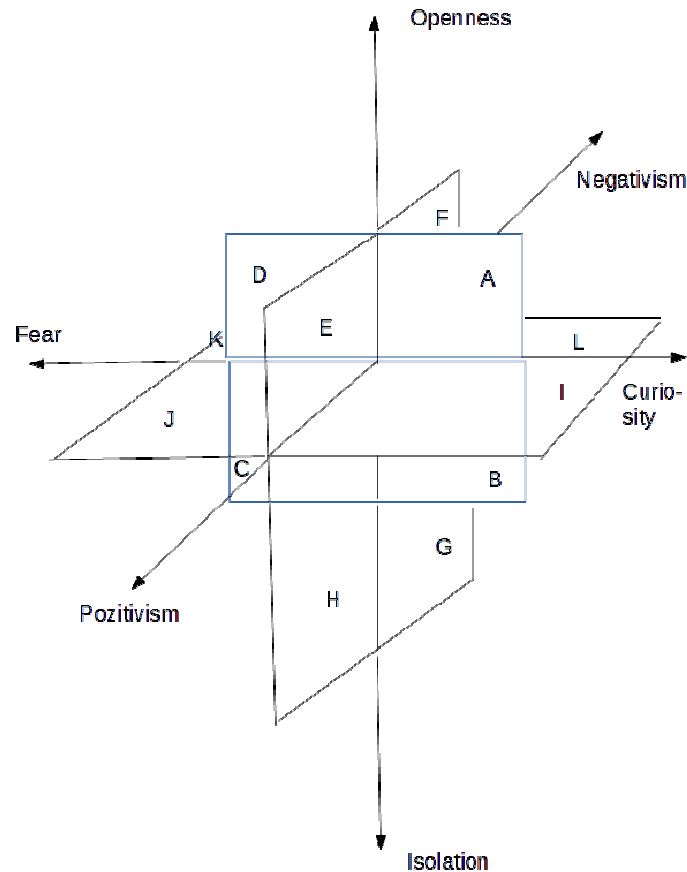
A.	Openness/Curiosity	EXPLORING WITHOUT PRECONCEIVED IDEAS
B.	Curiosity/Isolation	SEARCHING FOR EXPLANATIONS WITH AS MUCH NOTIONS AS POSSIBLE CLOSE TO THE UNIVERSE WITH WHICH HE IS FAMILIAR
C.	Fear/Isolation	DISINTEREST
D.	Openness/Fear	SEEKS PROTECTION IN A DOGMATIC SYSTEM
E.	Openness/Positivism	SEES THE UNIVERSE AS A MIRACLE
F.	Openness/Negativism	SEES THE UNIVERSE AS CHAOS
G.	Isolation/Negativism	IT DOES NOT EXIST (the "infinite state", the "sleep" of consciousness philosophical at Lucian Blaga)
H.	Isolation/Positivism	RELIGIOUS HABOTNIC
I.	Curiosity/Positivism	EXPLORATION TO HIGHLIGHT HARMONY
J.	Fear/Positivism	ERROR (AND SIN) OBSESSION
K.	Fear/Negativism	OBSESSION OF FAILURE, OF LOST (TEMPTATION)
L.	Curiosity/Negativism	THE FASCINATION OF THE ABSURD

Explaining each of these philosophical attitudes in front of the unknown, we can note the following:

A) People who are open to everything new and do not relate what is new to what is known to them, rather, do not seek to explain a newly perceived state of affairs with the help of experiences accumulated in the past and to frame them in a pre-established mental "scheme", they know the joy of exploring the unknown and approach any new phenomenon with an open mind, guided by curiosity and the pleasure of being surprised. This is the Blagian human type and the human model considered by Blaga in his discussion of uniqueness (the "singularity" of man).

B) People who are not without curiosity, but who do not have the same openness as those of type A, tend to isolate themselves in a familiar setting, to which they have become accustomed. These people want to know what is new and they want to understand what they find, but they cannot get rid of the conceptions they have acquired over time and, even if the newly perceived or learned situations are in contradiction with what they think they know, they try their best, sometimes even forcing logic or common sense, to find explanations based on the concepts and theories or thought schemes known and accepted by them.

C) Type C people are also isolated within their known environment and to which they have become accustomed, but they are possessed by the fear of stepping into new territories. This fear makes them "turn their eyes away" from facts about the existence of which they did not know in the past and consider that it is better not to know about them even in the present. This attitude leads to a lack of interest in new elements and to the neglect of their existence.

D) This category includes people who are eager to know what is beyond the limits of their known universe, but they are afraid to venture into a new conceptual space. Like B-type personalities, they close themselves in a system that is familiar to them, with the following difference: B-types approach the unknown rationally and want to correlate the new with which they come into contact with the set of knowledge with which

they used to, considering that this set is closed, better said complete and that nothing can be changed in the essence of this set of knowledge and values. In this sense they are marked by closure, isolation being in fact closure. Type D personalities admit that at any time something can happen that turns their life upside down and they are faced with catastrophes or what they feel are horrors or inexplicable events. In this sense they are open within themselves beyond the limits of reason. But they need this limitless universe that they "perceive" with the depths of their soul to be organized by a system of simple concepts, but so powerful that they never need to question them, namely a system of dogmas.

E) People who have the belief that everything that exists in the universe is more than we can realize or understand. For them, everything is a miracle because they positively evaluate everything that happens and everything they encounter on the way, they consider from the start, we could say that they have the preconceived idea that the whole universe is a miraculous composition, of a complexity and subtlety that is revealed with each new discovery. For them, any discovery, any "glimpse" into the new realm can only be a confirmation of the miraculous that dominates the universe.

F) In this case, openness is maintained, but dominated by negativism, it leads to the acceptance of any new element emerging from the unknown "areas" of the universe, but people in this category start from the beginning with the idea that everything can only be chaos, without any meaning, that chance rules everything and that everything we see, meet and happen to us is the result of chance.

G) This type of personality presents the greatest resistance to the new, among all the 12 categories synthesized in this presentation. Unlike those in category C, who adopt a passive attitude, they affirm with certainty that the amazing and shocking facts from the still unknown areas of the universe, with a major impact on universally recognized conceptions and theories, are simply the fruit of sick fantasies or pure charlatans designed to cause panic and disorientation. For these people, all these facts and evidence simply do not exist.

H) Unlike those in category D, who admit that anything can happen to them, that they can face completely new areas of the universe or with mental states generated by these areas, completely different from those experienced by them previously, those in the H category isolate themselves from these "novelties", from these unlived states, from these "appearances from other worlds" and then they defend themselves from them not because they could declare that they do not exist, as the G type do, but much worse, he defends himself by declaring that they belong to a taboo area, a domain of the universe in which it is forbidden to venture because the forces of the universe that protect you have given you very clear rules (which you must respect and if you break them you will be punished). And here the fantasy (as much as it is, limited and open more towards the negative) closes the gates to this world, freely creating unimaginable plagues, and eternal condemnations to unbearable cases. This is the domain of the most vehement and fervent obscurantism. These people have the conviction that they can lead their lives and are determined to lead their lives only by strictly following some rules that guide them how to guard against those taboo areas. The taboo area being so categorically delimited, a guide is needed to show them the "way" that keeps them away from dangerous areas. And this guide must be an absolute, unquestionable authority that gives very clear and strict rules, such as lists of permitted or prohibited actions or objects, so that their whole life is guided by these rules and by the well-known principle of totalitarianism: "Everything that is not explicitly stated as permitted is prohibited!". (Orwell 1949, 123)

I) It is an attitudinal type based on the particular positive valorization of universal harmony and this type of people are eager to find new and new confirmations of its manifestation, so that everything they undertake in their research and explorations is subsumed to this goal.

J) In opposition to those of category H, this human type is dominated by the fear of having already violated the strict rules for the delimitation of the taboo domain and by the fear of not having the ability to obey these rules; they are dominated (and undermined) by the fear of the unbearable horror of the unknown domain. Any interaction in this direction gives them the feeling that they are irredeemably sinful and that they only deserve the torments they go through in life.

K) It is a type of attitude, characterized, on the one hand, by disregarding the meaning of the world beyond our limits of knowledge and dominated, on the other hand, by fear of this world, it creates a personality that feels exposed to certain attractions morbid things that he thinks he cannot resist and whose spectre constantly hovers over the life they have to live.

L) This twelfth human attitudinal type believes neither in universal harmony nor in the miraculous aspects of the unknown worlds and has a certain disregard for them in the sense that at the mental level it considers that the whole universe, the known and the unknown taken together, I don't hide any secret and that everything that exists is the result of chance. This happens because at the mental level, they compare what is "beyond" with a conventional system of concepts, based on a simple logic and preached by the followers of the rationalist and classicist approach to world knowledge. Thus, interpreting what is "beyond", he necessarily discovers absurd reasonings and draws attention to them. Instead, on an inner level, these human types feel an irresistible attraction towards these "absurdities" and "beautify their lives" by contemplating them, even by essentially (deeply) interacting with the domains that gave rise to them.

*

Depending on the specifics of each of these 12 types of philosophical attitudes, there are different ways to perceive concreteness and operate with symbols. Obviously, the types most prone to not recognizing the value of symbols are those who fear the world beyond the (familiar) one described by their senses and immediately verifiable by their senses; those who close themselves in their limited world and who rely on the illusion that this limited and known world is safe.

Therefore, the human types most exposed to the temptation to identify the world with the concrete and well-known universe are the categories dominated by fear, closure and negativism. In these human categories, the danger of confusion between the world as such and the concrete world is the highest.

We note that these general considerations, as well as the scheme proposed above, are in line with the representation proposed by Lucian

Blaga in *Philosophical Consciousness* (2003, 88-89). Blaga shows that we can model philosophical consciousness, defined as characteristic of every human being, for every human being is (at least) potentially engaged in the metaphysical confrontation with the mystery/unknown, like a prism inscribed in a sphere. Highlighting in this model two essential perpendicular planes, the vertical and the horizontal, of course, themselves with metaphorical valences, Blaga invites us to imagine the prism as a representation of man's philosophical consciousness. The tip of the prism, in this dynamic concept, can be described as pointing either up or down, and indicates the direction of exploration of the mystery. The vertical plane is metaphorically and conceptually a plane of human verticality as a separate being, guided by his aspiration and striving for knowledge.

By comparing the two models, the one we rendered in the representation in Figure 1 and the one described in this model proposed by the philosopher Lucian Blaga, we can correlate certain human types rather with the "verticality" of the prism, while others will appear to be correlative rather with the horizontal plane of the prism described by Lucian Blaga.

Thus, in the confrontation with the unknown, human types A., E., I. and L. are "vertical" types, interested in conquering the mystery, and their philosophical consciousnesses could be shaped according to the model proposed by Lucian Blaga as tall prisms. Of course, modelling the human types A., E., I. and L. we may imagine prisms of variable heights, going either "high" or going "deep", however we choose to interpret the profiles we have described, while the second group (B, C, D, F, G, H, J and K.) representing much too cautious human types, interested mostly in acquiring various forms of security and survival, representable as shorter prisms with significant horizontal bases of various scopes and described as fledgeling philosophical consciousnesses, of the immediacy. The "immediate" represents in Lucian Blaga's philosophy a locus, a space of empiric existence, of the trivial existence and of routine, a closed existence deprived of horizon and opposite to the "horizon of mystery". Probably, "the immediate" is exactly that "infinite state, which is the sleep of the spirit" (Blaga 1934, 17) not at all characteristics of man; man being defined by Lucian Blaga as a "metaphorizing being" and "cultural mutation" unique in the universe – *a creator*.

Against the relevant philosophical background of "the immediate", the second category of human types (B, C, D, F, G, H, J and K) are rather "consumers of culture" and knowledge, existing in security and survival, incomplete human types, who withdraw into the face of "forces of nature" of all kinds, both literal and figurative. These human types fall prey more often to the illusions of the knowledge of the era, or to some meanings (reduced, easily assimilated and repeated) brought by cognitive projections, or to information that seems to bring safety and survival, therefore comfortable and worthy of attachment.

Often, consumers of culture and knowledge who are not among the creators, they count only on the accomplishments of their fellow beings; they rely on "projections" in confronting the unknown and in relating to the world.[5] The phenomenon that I mentioned as "projection" is very interesting and it is no coincidence that it bears the same name as the operation in drawing or in mathematics. It occurs when the unconscious identification of a phenomenological content with its apparent and concrete aspects is made. Both in drawing and in mathematics, this operation has nothing unconscious: the distinction is made very clearly between the actual figure and its projection on a space of smaller size than the actual figure – therefore, the projection is a meaning or an incomplete information.

A very simple example is the projection of a sphere on a plane. It results in a circle, a solid disc. No one would confuse the sphere with the disc. And yet it is very different to make decisions or erect constructions based on a "sphere" or a "disc".

In more complex situations, however, confusions arise, especially when we are interested in a certain concrete application and a concrete result. I did not use the word "concrete" by accident. Here is an example: the Taylor series expansion of a function that is difficult to represent. It is about decomposing a difficult-to-manipulate function into a weighted sum of an infinity of simpler and much better-known functions, namely polynomials. Obviously, if the summation of all the functions in that infinite set were to be done, the operation would become a particularly difficult one, an impossible operation. The central idea of the application

---

[5]   In another form, the Platonic myth of the cave is disguised in this idea.

of this development is that the respective sum is considered to differ very little from the sum of the first few terms, 2, 3, in very rare cases more than 3. In fact, the computational technique allows the consideration of much more much more than just 3 terms of the sum, which contributed to the fact that this "ancient" technique is still relevant today.

What is achieved by this serial development? Approximations are easily obtained; these are very useful when tracking the numerical values of the parent function at specific points. The more terms are retained from the infinite sum, the better the approximation. But the idea is the same: in order to be able to explore with our limited means a certain area of reality, we must transform that area where we should work with the infinite into a finite domain; that is, we must restrict ourselves to a finite domain of it and work with approximations. And here it is no longer so easy to avoid confusion: for theorists it is clear, the sum of the terms is not identical to the function, just as the disk is not identical to the sphere either. But for those who need actual calculations and real numerical results that they can then insert into formulas with the help of which they can perform certain concrete tasks, from everyday life, it is very easy to confuse the function with its approximation, or rather, with an approximation of it good enough for the established purpose.

As I mentioned above, the concrete goes hand in hand with the finitude. And the hidden meanings associated with each concrete object are very well comparable to this model of the development of a Taylor series function. And we can take an even simpler example: the operation with numbers is [Euler's irrational number, or Napier's constant, with applications in logarithmic mathematics, in calculation with complex numbers, economics, biology, chemistry, paleontology (C14 dating), statistics, etc. .], Ø [Phi, the golden number, equal to 1.6180339887…, a favorite ratio of nature], ◎ [the irrational number *pi*, the ratio of the circumference to the diameter of any circle]etc.[6]

---

[6]   In all applications where a numerical result usable in everyday life is sought, and yet none of those who do these calculations think about the fact that it might still be interesting, beneficial, and perhaps have some surprising efficiency that, with the evolution of computers with ever greater computing power, these numbers should be used as close as possible to the real value, as numbers with an infinite number of decimal places.

But the contemplation of the infinite in general is the equivalent of a philosophical "opening towards the abyss", towards the infinity of human ignorance and the unknown, as well as the complexity of knowledge, "exercises" reserved for the human type associated with "verticality" and the human situation in terms of *mystery and revelation*" (Blaga 1977, 436) and which opens the doors to the true human universe, characterized by the dimensions of knowledge, culture, meaning, information, science, technology, artistic creation and history, in the deepest philosophical sense.

References

Blaga, Lucian (1934, 1993). *Cenzura transcendentă* [*Transcendent Censorship*]. Bucharest: Humanitas.

Blaga, Lucian (1977). *Fiinţa istorică* [*The Historical Being*], ed. by Tudor Cătineanu. Bucharest: Humanitas.

Blaga, Lucian (2003). *Despre conştiinţa filosofică* [*On the Philosophical Consciousness*]. Bucharest: Humanitas.

Blaga, Lucian (1997). *Trilogia cosmologică: Diferenţialele divine. Aspecte antropologice. Fiinţa istorică* [*The Cosmological Trilogy. Divine Differentials. Anthropological Aspects. The Historical Being*]. Bucharest: Humanitas Publishing House.

Bohm, David (1989). "Meaning and Information." In Paavo Pylkkanen (ed.), *The Search for Meaning. The New Spirit in Science and Philosophy*. Crucible: The Aquarian Press, 43-62.

Gennaro, R.J. (1996). *Consciousness and Self-consciousness: A Defense of the Higher-Order Thought Theory of Consciousness.* Amsterdam/Philadelphia: John Benjamins.

Jackson, F. (1986). "What Mary didn't Know." In *Journal of Philosophy*, 83:291-295.

James, W. (1890). *The Principles of Psychology*, vol. 1. New York: Henry Holt & Company, available online at https://psycnet.apa.org/record/2004-16192-000 last time accessed in December 2023.

Kant, I. (1965). *Critique of Pure Reason*, translated by N. Kemp Smith. New York: MacMillan.

Kriegel, U. (2004). "Consciousness and Self-Consciousness." In *The Monist, 87*:182-205.

Lewis, C.I. (1991). *Mind and the World Order: Outline of a Theory of Knowledge*. Dover Publications.

Nagel, Th. (1974). "What Is Like to Be a Bat?" In *Philosophical Review, 83*(no.):435-456.

Orwell, G. (1949). *1984*. London: Secker & Warburg.

Rosenthal, D.M. (2000). "Introspection and Self-Interpretation." In *Philosophical Topics*, 28:201-33.

Sartre, J.-P. (1956). *Being and Nothingness*, transl. Hazel Barnes. New York: Philosophical Library.

Shoemaker, S. (1981). "Absent Qualia are Impossible." In *Philosophical Review*, 90:581-599.

Stubenberg, L. (1998). *Consciousness and Qualia*. Amsterdam/Philadelphia: John Benjamins Publishers.

All links were verified by the editors and found to be functioning before the publication of this text in 2024.

Sorin CREȚU[1]

THE "THREE-SYSTEM VIEW" MODEL. A POSSIBLE SOLUTION SOURCED IN "PRE-JUDICATIVE HERMENEUTICS"

Abstract. The current paper, anchored in the discussions surrounding *"Linda Problem"*, *"model-free"* and *"model-based"* learning strategies along with Kahneman's *"two-system view"* cognitive model, and Gross's considerations on *"salience processing"* and *"atypical attribution of salience"*, aims to present a potential solution originating in *Pre-Judicative Hermeneutics*: Cernica's *"three-system view"* model. Both Kahneman's and Cernica's models embrace Herbert Simon's paradigm of *"bounded rationality"* and serve to broaden its reach. While Kahneman's *"System1"* and *"System2"* model elaborates on the *"perception"* – *"intuition"* – *"reasoning"* cycle, Cernica's model, founded upon and expending the *"non-judicative"* – *"pre-judicative"* – *"judicative"* cycle, is intended, I suggest, to augment and advance the former, from and with a phenomenological and hermeneutical perspective. Core concepts as *"pre-judicative memory"*, *"de-constitution"* *"interpretation"*, *"pre-judicative circularity"*, *"act of pre-judging"*, *"non-judicative experience "*and *"existential judgment"* are central to Cernica's innovative approach. These ideas bring the *"existential subject"* (*"operant subject"*) in focus, establishing a framework to account for how the cognitive biases could be mitigated across a wide range of domains, thus enabling the elevation of our consciousness and existence through the fundamental attributes of *"reality"*, *"identity"*, *"authenticity"* and *"veracity"* supporting them. Furthermore, as reiterated in this paper and elaborated on in a more extensive discussion in a previous work, the enhancements proposed through Cernica's *"three-system view"* model may also prove instrumental in emerging areas of AI development and the pursuit of AGI. These include Reinforcement Learning (RL) strategies, AI Safety and AI Alignment research. The model offers valuable contributions to a deeper understanding of debiasing processes, the development of more robust

---

[1]   PhD, Doctoral School of Philosophy, Faculty of Philosophy, University of Bucharest, <soleliom@gmail.com>. Member of the SHPjM Research Group (*Studies in Pre-Judicative Hermeneutics and Meontology*), coordinated by Prof. Univ. Dr. Viorel Cernica, University of Bucharest. SHPjM Research Group is part of Research Center of the History of Philosophical Ideas (CCIIF).

debiasing methods, and a more nuanced understanding and alignment of "*deliberative human preferences*" and "*human value functions*" within advanced AI models.

Keywords: *non-judicative, pre-judicative, judicative; operant subject, logical subject, reality, identity, authenticity, veracity, pre-judicative memory, de-constitution, re-interpretation, pre-judicative circularity, cognitive bias, deliberative human preferences, bounded rationality, shared intersubjective space*

The perplexing findings related to the answers to "*Linda Problem*" (Kahneman 2002, 468), particularly reflective of massive "*attribute-substitution*" (469), in the context of observed "*accessibility*" challenges, uncovering the permeation of the representational content of "*System2*" with pre-judgmental elements from Kahneman's "*System1*" (451)[2] (*e.g.*, intuitions, intuitive "*preferences*"[3] based on "*impressions*" on the attributes of "perceptual and cognitive *objects*"[4]) that escaped appropriate monitoring and deliberative qualification, as well as the surprising observations on the topic of "*salience attribution*" and "*salience processing*" that exposed us to the reinterpretation, with the supportive evidence of neuroscience, of the notion of "*importance*", topic thus expended into "*atypical attribution of salience*" (Gross *et al.* 2024) characterizing creative processes and judgments, bring us to an inflection point. It is the point at which, I suggest, we should look for an out-of-the-box alternative, with the thought that this could bridge the aforementioned difficulties towards clarifying opportunities.

To open this discussion, a few daring questions should be asked beforehand.

- In what ways our "*pre-judicative memory*" can play the fundamental "*de-constitutive*" role in rebalancing our judgments towards "*value-driven*" perspectives?
- Why our aesthetic judgments, also scoped in this paper, and along with them, I suggest, other particular judgments highly cherished

---

[2]   See "*Figure 1*", mapping the process flow between "*Perception*", "*System1*" and "*System2*".

[3]   The notion of "*preferences*" should be also observed in connection with the Reinforcement Learning (RL) framework from Artificial Intelligence (AI) and in the context of "*human preferences*", which are representative of the human judgments; this discussion was extensively elaborated in my paper, Crețu 2025 (forthcoming).

[4]   Previously discussed in Crețu 2025.

throughout the human domain (*e.g.*, affective, spiritual judgments) are so profoundly impacted by the "*non-judicative*" – "*pre-judicative*" – "*judicative*" cycle?

- At what point the "*operant subject*" supersedes the "*logical subject*" such that our way of experiencing the world can upgrade our existence with the attributes of "*reality*", "*identity*", "*authenticity*" and "*veracity*"?

Such a possible alternative is the solution model proposed by Viorel Cernica, within the framework of "*pre-judicative[5] hermeneutics*" (see Cernica 2019, 185; 249). As previously emphasized (Crețu 2024), the appeal to the hermeneutics' mechanisms of judgment evaluation and cognitive bias de-constitution could be a promising avenue, due to its dedicated interpretative capabilities (also applicable when we approach AI Alignment research) when we talk about "*evaluative attitudes*" (in terms of "*desires*"; particularly "*instrumental desires*", and "*higher-order desires*", the former leading to the latter, at least in some instance), "*human preferences*" (as "*comparative evaluations*", particularly in their deliberative form), intentions (alternatives-based selections) and "*evaluative beliefs*", along with considering the cognitive "*salience*" elements, all these aspects pointing to both "*value*" and "*choice*" differentiation and consolidation and "*value-driven*" behaviors and actions.

The context of this solution model revolves around multiple core concepts reflective of their domain specific acts and structures that are linked at the at aspectual (process) level and at the role (functional) level, being tied into the "*operant subject*" ("*in person*", differentiated ontologically from the "*logical subject*", meaning that the "*operant subject*" and "*logical subject*" stand in an "*ontological difference*" and I, would suggest, tension)

---

[5]   As a clarification pertaining to the term "*judicative*" captured in the concepts of "*judicative*", "*pre-judicative*" and "*non-judicative*" employed throughout Viorel Cernica's model: the term is not related in any way to the "*judicative*" attribute used within the judicial framework. Furthermore, the term "*prejudice*" that we will refer to as part of this model is not intended to be related to "*an affective component*" (*e.g.*, "*discriminatory behavior*"), but to "*a cognitive component*", cognitive bias-related; Refer to *APA Dictionary of Psychology*, from American Psychological Association, available online at < https://dictionary.apa.org/ > last time accessed on March 29, 2025, under the noun "*prejudice*".

(Cernica 2023, 267; 298)[6] with its experience and existence – refined under its existentialist attributes of "*reality*", "*identity*", "*authenticity*" (276-277) and "*veracity*"[7] (human, and which could be extended to AI agents): "*judicative*", "*pre-judicative*", "*non-judicative*", "*non-cognitive*"[8] (Cernica 2019, 185; 249). Once again, the term "*judicative*" refers to "*judgements*", being applied under their types. In this tripartite configuration of the model, the "*pre-judicative*" (Cernica 2019, 205) domain mediates between the "*judicative*" domain and "*non-judicative*" domain, based on the particular processes pertaining to each one of them (cognitive processes being associated with the "*judicative*" domain, while intuitive and emotional scopes, falling under "*non-cognitive*" processes underlying the "*non-judicative*" domain). We can already notice the parallel with Kahneman's aforementioned "*System1*" and "*System2*" cognitive models and with the "*model-free*" and "*model-based*" computational learning strategies, both extensively discussed in a previous paper as well[9].

These three domains stated earlier are dynamically interacting, based on particular operations, out of which the central ones are the "*operations of de-constitution*", the "*operations of interpretation*" and the "*intentional operations*". Naturally, the "*operations of interpretation*" rely on the "*operations of de-constitution*" of judgments from the "*judicative*" domain. Suffice to specify that the judgments referred to within the "*judicative*" domain are based on "*judicative structures*" and explicit reasoning processes. These "*judicative structures*" are integrated at cognitive level (akin to

---

[6]   See also Cernica 2023, 276-277 and 274, for the "*ontological difference*".

[7]   For a more extensive discussion of the four existentialist attributes – and, more precisely, the "*subject's identity hypostases*" – "*reality*", "*identity*", "*authenticity*" and "*veracity*" – I refer to Cernica's "Pre-Judicative Memory and De-Constituting of the Prejudices [Logic-formal and hermeneutic description of the existential subject's experience]", Chapter 2.31 "Objectual fulfilment of the subject (reality, identity, authenticity, veracity) and memory" (forthcoming publication).

[8]   "Non-judicative thought according to quantity – the only criterion that can work, for now – covers a vast 'logic' and 'extra-logic' space, the latter being 'logicized': a) different logical forms of judgment (…), b) pre-judicative entities, c) pre-logical elements of a psychological nature, partially 'logicized': experiences, facts, especially those of the type of prejudices and pre-understandings (…)" (Cernica 2019, 69).

[9]   In reference to the chapters "Predictive Human RL Strategies: 'Model-Free' and 'Model-Based'" (in particular, 70-75) and "Kahneman's 'System1' and 'System2' Cognitive Models" (in particular, 76-91) see Creţu 2025 (forthcoming).

"*System1*"), but they are functionally differentiated and instantiated as "*formal*" structures ("*subject-predicate*" type), "*alethic*" structures ("*verb-time*" relationship type), "*referential*" structures ("*expression-reference*" relationship type, where the "*expression*" is symbolically captured through language at the propositional level and the reference is a "*state of affairs*") and "*existential*" structures ("*existential subject*" – its "*pre-judicative memory*"; see Cernica 2023, 267; 298[10]). Thus, a cognitive judgment could be "*analysed*" either based on its "*structural makeup*", consisting of the previously mentioned four cognitive instantiations, or on its "*pre-judicative makeup*" (Cernica 2023, 273). The "*pre-judicative makeup*", to translate, paraphrasing the author, could be analysed within the framework of "*pre-judicative analytics*", by examining the "*de-constitution conditions of pre-judgments*" (273).

This "*pre-judicative makeup*" of judgments contains the "*cognitive judgment*", the "*evaluative judgment*" (applied on the "*cognitive judgment*", which becomes subject of evaluation, function of truth conditions) and the "*existential judgment*" (of the type "*I exist*", in which the "*I*" is the "*operant subject*" performing the "*judicative*" act, paraphrasing the author) (273).

We observe how the cognitive judgement "*structural makeup*" intersects the "*pre-judicative makeup*" at the level of the "*existential subject*" and to what extent the "*pre-judicative makeup*" intersects the cognitive domain at the level of an embedded "*cognitive judgment*", then differentiating from the cognitive domain through an evaluative and then existential slide towards the "*operant subject*", who "*performed*" the act of judgement in the first place (276). This slide towards the subject, engages directly the "*non-judicative*" domain, from which, on the one side, the aspectual part of it (at process-level) will reverberate back, towards the "*pre-judicative*" domain, infusing perceptions into pre-judgments and, on the other side, the representational part of it will reverberate back, towards the "*judicative domain*", in the form of representational content (role-wise, including the functional role, but not

---

[10] "The operant subject, flesh and blood person, and his pre-judicative memory, *e.g.*, the mass of traces left by his judicative experiences, his 'knowledge', now activated)" (Cernica 2023, 275). See also Cernica 2023, 271; 272; 275.

limited to it) transferred throughout the judgments' structure, permeating them starting from the "*existential*" structures, into "*referential*" structures and, further on, spreading through the "*judicative*" core, at "alethic" (unfolding in time a state or an action) or "*formal*" (reasoning) levels (engaging mostly our deductive reasoning through "*declarative knowledge*" and inductive and causal reasoning through "*procedural knowledge*", borderline with adductive and analogical reasoning, which already involve our "*symbolic knowledge*" – to extrapolate from Ben Goertzel's AGI cognitive infrastructure I was discussing in Crețu 2023, 124; 147).

Once again, we can see how aspectual and role levels are intertwined and, although there are content barriers (representational and conceptual in nature) between "*perception*" and "*intuition*" and aspectual barriers between "*intuition*" and "*reasoning*", the "*non-judicative*" and "*pre-judicative*" domains continue to interact with the cognitive domain, eluding the barriers (when the barrier is aspectual, the bridge is the content and when the barrier is the content, the crossing is aspectual). This incursion into the continuous cycle "*non-judicative*" – "*pre-judicative*" – "*judicative*" displays the convoluted path of cognitive biases, using the heuristic transfers between our perceptions and intentions, which is then taken over by the cognitive transfers between our intuitions and our reasoning, where the corrective filtering intervenes, and the deliberative processes occur[11]. While we are re-emphasizing Kahneman's "*two-system view*" (450) ("*System1*" and "*System2*", already manifesting throughout the "*perception*" – "*intuition*" – "*reasoning*" cycle), we are placing it in the context of a "*three-system view*" as described in Cernica's model ("*non-judicative*" – "*pre-judicative*" – "*judicative*" cycle), in which its specific continuous cycle is dominated by "*pre-judicative circularity*".

What the "*intuition*" and the "*reasoning*" continue to have in common (distinct from "*perception*") in Kahneman's "*two-system view*" is the passage of time and the fact that they "*can be evoked by language*" (451), which are maintained in Cernica's "*three-system view*" model at the "*judicative*" level, both through the "*alethic*" structures and the "*formal*" and "*referential*" structures.

---

[11]   Re-emphasizing Kahneman's "*perception*" – "*intuition*" – "*reasoning*" linkages and differentiations; see Kahneman 2002, 450-451, including "*Figure1*", 452.

This "*three-system view*" allows the phenomenological and hermeneutical elements to take part in an integrative solution that seemed to have been mostly sought within the spheres of cognitive science. What makes possible this expansion into a "*three-system view*" is a previous and ongoing paradigm shift towards "*bounded rationality*" (Crețu 2023a, 171-178), as proposed by Herbert Simon and adopted by Kahneman in his "*maps of bounded rationality*", in which he is reiterating the "*two-system view*" human cognitive model, with heuristic foundations and exposing the cognitive biases processes.

Moreover, the "*bounded rationality*" paradigm had profound repercussions on the long-lasting, widely accepted and mostly unquestioned, by that time, "*unbounded rationality*" view, that used to proclaim the powers of human reason as "*universal*" and "*abstract*", with an "*absolute*" defying hint. The "*embodied realism*" that flourished in the wake of the new paradigm, particularly at the "*School of Berkeley*" through Lakoff's "*embodied mind theory*", brings forward through the wide-open avenue of human "*embodiment*" and "*embodied cognition*" (expended to and combined further on with "*embodied simulation*" perspectives and detailed through theories like "*shared manifold hypothesis*" or "*the shared manifold of intersubjectivity*" (Crețu 2023a, 93; 98), developed by Vittorio Gallese, at the "*School of Parma*"), both the structural and functional limitations of our reasoning, by "*closing two major gaps: the one between perception and conception and the one between subject and object, in which misleading objectivity ideas originated*" (Crețu 2023a, 191-192). This is the long-waited return of the "*operant subject*" (and, furthermore, of the "*first-person*" view) at the level of "*judicative*" reasoning constitution, through "*existential*" structures already implicitly present (Cernica 2023, 275-276), as part of a broader answer to deeper questions related to the naturalization of human intentionality, the nature of human integration and interaction with the world(within a "*multimodal intentional shared space*" or "*shared intersubjective space*", which constitute, foundationally, the human "*we-centric space*" as proposed by Gallese (2003, 525)[12], and the nature of consciousness.

---

[12] See also my paper where I discussed these concepts (Gallese 2003, 525 and Crețu 2023b, 93, 97-99), "The Advent of AGI".

What was needed as part of the solution was to bring the "*operant subject*" into our line of sight, a theoretical shift that Cernica's "*three-system view*" model permitted, at the level of "*existential*" structures within the "*judicative*" domain, where the "*existential subject*" is directly linked to its "*pre-judicative memory*" that enters the "*pre-judicative makeup*" of judgment through the "*existential judgment*" component (Cernica 2023, 274)[13]. Such dynamic attempts bridge the content gaps with subject's "*perception*" that ultimately reinforce our biases, by placing them in a conceptual representational space.

These aspects have also a direct impact on AI development and transformation towards AGI, including RL(Reinforcement Learning), IRL (Inverse Reinforcement Learning), the balance between "*reward functions*", "*objective functions*" and "*value functions*", and AI Alignment research, which explores, as we discussed in an earlier paper[14], the limitations and the prioritization of "*human reward functions*" versus "*human value functions*" (for the latter, in terms of "*human preferences*", and, as I suggested, "*deliberative human preferences*") as "*target for alignment*"[15].

I highlighted these points to show the directions related to the "*what*", in the question "*what is the solution we are looking for?*". It seems that we already know a lot about the "*what*". As far as the "*why*", by now, we also have a rather clear picture.

However, we don't clearly know the "*how*". The "*how*" posses significant challenges and it is the particularity of this solution that it could be providing a possible answer to the "*how*" in the question: How can we resolve cognitive biases, while remaining aligned with our heuristic, "*bounded rationality*" mechanisms and within an "*embodied realist*" framework, and also include in the process the "*operant subject*", such that we are able to account for the experiential aspects, and, with a target, at the "*first-person*" level, throughout our intentional cognitive processes, in order to set clear and transparent expectations regarding

---

[13]   Here the author addresses the question, opening up a possible direction for further investigation: "But wouldn't be a possible connection between this judgment (existential) and the existential structure within the structural composition of the judgment (existential subject – its pre-judicative memory)?"

[14]   See Crețu 2025 (forthcoming).

[15]   These aspects have been extensively discussed in Crețu 2025.

our decision-making, our choices, our actions, our goals? The difficulties the AI Alignment research are pointing to are genuine. The "*human in the loop*" perspective requires that humans clarify their own domain, such that AI can align with, under conditions that both humans and AI can fully and deeply understand and follow. If we ever say, "*We didn't know better*", it will be because we really didn't know "*How*". Many times, behind the question "*What was missing?*" lies the real question we need to confront: "*How was it lost*?"

At this point, we can already glimpse the possible solution we referred to earlier, but before expending on that, we have to explicitly showcase the main elements along with their counterpoints, linking "*judgement*" to "*experience*" to "*act*" and to "*operation*": along with "*judgement*" and "*pre-judgment*"[16], the "*non-judicative experience (which is opposed to the judicative one)*", "*the act of pre-judging (which is opposed to the act of judging)*" and "*the operation of interpretation (which is opposed to the intentional operations)*" (Cernica 2019, 185). This "*judgement-experience-act-operation*" overarching organization of the model is an indication of its active and agentive capabilities, that will be confirmed by the appeal to "*pre-judicative memory*". Within this structure, three other concepts are playing a core role: "*de-constitution*", the "*pre-judicative circularity*" and the concept of "*interpretation*" (188)[17]. While the "*pre-judicative circularity*" that we previously mentioned is the "*movement of acts of consciousness between the judgment and the result of the non-judicative experience, which has the meaning of a 'reduction' of the two to a middle term: the pre-judicative*" (205), the "*interpretation*" is sourced in the "*experience of the non-judicative, whose content is processed in such a way that, taking the form of judgment, it becomes 'knowledge' and the object of communication*" (205).

The "*interpretation*" is applied to the "*judicative form*" but taking into account "*(…) the existential judgment from the pre-judicative composition of a judicative form (…)*" (Cernica 2023, 274). The "*operant subject*" is not

---

[16]   "The pre-judgement is the formal outcome of the act of pre-judging, meaning the act of appropriation of an (cognitive) object through what another (person) or myself (…) has already thought (judged) about it, and then sent it to *me* in its standard (judicative) form of comprehension." (Cernica 2019, 189)

[17]   We will not dive at this time into other connected concepts like "*negative predication*" and "*existential time*", concepts also mentioned by the author.

"*obviously*" entering the "*structural analysis of the judicative form*", but he is instrumental in its "*constitution*", as the judgment itself is "*his work*" (276), as being the inner, active part of his own "*existential*" segments of judgments through its "*pre-judicative memory*", throughout the "*judicative*" reasoning process.

The "*interpretation*", by the means of the "*pre-judicative memory*", is the active "*operant subject*'s" intervention in the judgement process from a metacognitive reasoning perspective (which, as we know, like the interpersonal/ intersubjective reasoning, relies on "*attentional knowledge*"), process that, himself, owns. This intervention is performed by consciously suspending a judgement and orienting the attentional resources towards following back the "*traces left by his judicative experiences*" (275) (to reiterate), which were previously integrated, and progressively lesser subjected to a questioning and qualification effort, in the subject's body of knowledge at the judicative, reasoning level, assumed to represent the "*logical subject*" point of view. These "*traces*" identified as "*active*" in the present "*pre-judicative memory*" are then exposed to an iterative process of "*de-constitution*", followed by "*interpretation*" – meaning here the "*leap from logical-formal to existential, a leap made possible by analysis*" of pre-judgements (279) and "*re-interpretation*" of those judgments, thus purging from them the cognitive bias elements, which were covertly woven within those "*traces*". As we observe the "*pre-judicative memory*" process flow, we have to understand, once more, that the term "*pre-judicative does not refer to something that is well formed before the 'judicative', **but it rather precedes the judgment**, that is, we become aware of it starting from the judgment, but we end up accepting its originality with respect to it, as if the judgment itself were fulfilled (structurally) under its condition*" (277-278), a position that is also coherent with Kahneman's "*two-system view*", but differs from it in the specific way it is applied, as "*three-system view*".

Thus, this solution model goes a step further, making "*pre-judicative memory*" "*active*" (not passive) and instrumental in the present, bridging the temporal differences, suspending the validity of the present judgement[18],

---

18   "The suspension of the validity of judgments must have as its object precisely these (prejudgments active within the boundaries of the 'world'). This presupposes a veritable technique through which one can pass from logical-formal, regarding these (pre)judgments, to existential, as was stated above. How can this passing be achieved? By analyzing their

travelling back to the previous judgements through a shift from the "*formal-logic*" level to the "*existential*" level (the "*operant subject*" level) to recognize the "*pre-judgments*" by employing the "*pre-judicative memory*", identifying these bias elements and, then, returning and correcting them in the present at the conscious "*judicative*" level using the "*de-constitution operations*". The "*de-constitution operations*", at the level of the "*four-structures of the judicative form*", taking into account that the target of these operations is the realization of the "*existential leap*" that activates the "*operant subject*", will mainly engage the "*referential*" and "*existential*" structures[19]. This is the way in which "*the subject from the formal structure of the judgement*" (meaning the "*I*" from "*I am*", as in the example provided by the author) is accounted for in terms of "*reality, identity and authenticity*", these attributes ("*properties*") "*strengthening*" the "*objectual*" nature of the "*I*" from "*I am*" and thus affirming the "*personal*" dimension of the "*subject*" defined through these attributes, subject who is actually the "*operant subject*" performing these acts of judgment (280).

The movement towards "*operant subject*", by acknowledging the "*existential judgment*" (*e.g.*, "*I am*") within its incorporating "*judicative form*", fortifies its "*reality*", fact that, consequently, enhances the "*identity*" and "*authenticity*" of the subject and, at the same time, brings it into the focus of the "*evaluative judgment*" (part of the "*pre-judicative makeup*" previously discussed) applied on it, thus reaffirming the consistence and strengthening the cohesion and the "*judicative unity*" of the judgement (280). The evaluative dimension (in which our "*pre-judicative memory*" is not only a passive placeholder, but the active mechanism of exploration of past judgments that impact the current evaluations, memory that enables and conditions the "*existential*" structures (the way the memory is connected to subject's "*identity*", within the "*judicative structures*" themselves) is the one enacting the path towards "*value-driven*" judgments, in which we recuperate the "*operant*

---

structures and interpreting the 'leap' from the formal, from what was highlighted through analysis, *e.g.*, to the operant subject (in the flesh, 'in person'), both operations belonging to the wider operation, also called **de**-*constitution*." (Cernica 2003, 279)

[19]   "Through the operation of de-constitution of judgments, on the one hand, the 'objective' reference of the judgment, which corresponds to its expression, is subjected to a 'support' modification, being thereby ontic-ontologically 'neutralized', and on the other hand, the 'existentiality' is activated by the operant subject." (Cernica 2003, 279)

subject". This is the real actor of the judgments and reasoning pursuit, but this time enriched by the acts of experiencing and contextualizing through his/ her living existence, engaged as well in the *decision-making* processes that represent his/ her beliefs and the *goal-oriented choices and actions* that represent his/her desires, both instrumental in achieving them.

The "*act of pre-judging*" (or the "*pre-judicative act*") (Cernica 2019, 201), in this context, is using our "*pre-judicative memory*", by connecting to the past judgments of our experiences where usually the cognitive biases reside. Thus, to examine the "*object*" in scope of our present judgements, within the context of our current experiences, this act goes beyond the intentional and phenomenal into the existential aspects (199)[20], posing as well a reasoning engine that filters (same like Kahneman's "*System2*"), using the framework of past experiences, and engages the "*interpretation*" and "*re-interpretation*" of present "*non-judicative*", existential occurrences (*e.g.*, intuitive, emotional states[21]).

As we can observe, this is the "*how*" being disclosed, therefore the step further brought forward by Cernica's solution model.

The fact that any judgment or "*judicative*" construction subsumes both a "*structural*" and a "*pre-judicative*" component (Cernica 2023, 278), makes it a more effective mechanism for detecting deficiencies that might occur on either sides, to paraphrase the author (278)[22].

---

[20] "The interval between consciousness in the act of pre-judging and pure (or phenomenal) consciousness, viewed from a functional perspective, is the expression of an 'ontological difference': that between the phenomenon of knowledge and the very existence of a knower." And following: "At the moment this de-constitution occurs, however, through an act of consciousness (which, according to the phenomenological model, can no longer be considered intentional and constitutive phenomenal) 'someone', a 'subject', person in flesh and bones (first-person) acquires an existential profile" (201); see also 198, 201, 205.

[21] Here is Kahneman discussing the "*heuristic attribute*" of "*affective valuation*" and referencing the studies pertaining to "an automatic affective valuation – the emotional core of an attitude – is the main determinant of many judgments and behaviors." (further reference provided by the author in the original paper). The analysis continues as it follows: "In terms of the scope of responses that it governs, the natural assessment of affect should join representativeness and availability in the list of general-purpose heuristic attributes" (see Kahneman 2002, 470).

[22] "But in any of these (referring to 'structural and pre-judicative components'), deficient judicative 'forms' (structural or pre-judicative) can be found, which, in fact, present themselves as an incomplete composition of sentences (judicative 'expressions')".

While the "structural" and "pre-judicative" components unify from a "*makeup*" perspective and interact at aspectual and role levels by the mediating character of the "*pre-judicative memory*" (between "*non-judicative*" experiences and "*formal*" judgments), they continue to differ in the ways these processes and roles are applied towards the "*subject*", across its "*logical*" and "*operant*" dimensions.

For example, we consider how such processes and roles are impacting the identity of the "*logical subject*", which is established through "*formal*" "*judicative*" structures, as opposed to the identity of the "*operant subject*". The latter type of identity is defined and emerges through the subject's appeal to his own memories and experiences, being determined by all four elements of the "*judicative structure*": "*formal*", "*alethic*", "*referential*" and "*existential*".

Think how, as an example, many times the following scenario happens. You are meeting someone you haven't seen in many years, and your first encounter with them is not as the person standing in front of you now (as you ask them questions to get reacquainted) but rather as the person they appear to be in your memories, shaped by how they presented themselves all those years ago, when you last met. That is often what happens when someone tells you, after years without seeing you, that "*you haven't changed a bit*" (to paraphrase from one of Ram Dass lectures, after he commenced his journeys to India, first in 1967, and followed the enlightenment path for the rest of his life).

There is one more reason for which the "*human preferences*" would need to go often times through a suspension of present judgment, followed by operations of "*de-constitution*". These operations are using the "*pre-judicative memory*" to bring those preferences from the already assumed perceptions or intuitions domains into deliberative reconstitution, through conscious "*interpretation*" and "*reinterpretation*", taking into account the "*operant subject*" and differentiation it from the "*logical subject*" that we tend to formalize first at the "*judicative*" level. This process turns the heuristically intuitive "*human preferences*", subject to cognitive bias, into deliberative, "*value-driven*" "*human preferences*", charged, as they were recovered and reconstituted, with the attributes of "*reality*", "*identity*", "*authenticity*" and "*veracity*", as we mentioned earlier. These existential and experiential attributes are highly impactful in AI Alignment research, particularly when setting AI's "*target for alignment*", to ensure they are

both adopted by AI and understood in the spirit of what is "*important*" and "*meaningful*" to transmit.

AI models, I suggest, when discussing RL (Reinforcement Learning) and IRL (Inverse Reinforcement Learning) with human feedback[23], should take into account applications of "*pre-judicative memory*" and "*pre-judicative circularity*" stances in order to be able to genuinely integrate the deliberative "*human preferences*" and expectations in their RL learning strategies.

This integration is facilitated through "*iterative*" updating of the reward functions and recurrent "*feedback loops*" across all levels outlined within the "*two-system view*" and "*three-system view*" models, employing context-dependent constrains and, as the AI models advance, "*counterfactual reasoning*", complimented with "*adversarial learning*" ("*adversarial debiasing*") (Yang *et al*. 2023, 55) (to detect and mitigate biases).

Such elements would serve a preventive role with regard to some of the major present concerns in AI "*Safety*" (*e.g.*, "*negative side effects*" and "*reward hacking*" risks, as instantiations of the implementation of a "*wrong objective function*") (Amodei *et al*. 2016, 1; 3; 7; 19).

With regard to aesthetic judgments, the "*pre-judicative memory*" and "*pre-judicative circularity*" can operate with high effectiveness at the cognitive and also the experiential and existential levels, through the attributes of "*reality*", "*identity*", "*authenticity*" and "*veracity*" and, I would suggest, "*unicity*". This fact is relevant, as they are, both from the creator's and observer's perspectives, subjected to a diverse historical, cultural, educational, economical contextualization that shapes and individuates both the creative processes and the aesthetic preferences. The "*judicative structures*" provide a framework for the aesthetic evaluations (*e.g.*, what is the aesthetic beauty, how is this beauty predicated, what are the temporal contexts and circumstances the aesthetic evaluations are applied to and, essentially, who is the "*operant subject*" at the aesthetic existential levels). Furthermore, they offer insight into how the subject performs based on his/her active and present "*pre-judicative memory*", also involving the "*suspension of validity*" of certain aesthetic judgements presumed to have been infiltrated by bias. Consequently, this allows for their "*de-constitution*" and "*re-interpretation*" in order to recover their integrity, availability and credibility.

---

[23]   Previously discussed in Sections "RL functions perspective and the link to AI Alignment" and "RL and IRL instances in AI Alignment Research", in Crețu 2025.

These process flows from Kahneman's "*two-system view*" and Cernica's "*three-system view*" are quite visibly applicable to the aesthetic judgments in the art appreciation domain. We are referring here to the plethora of "*perceptual*" and "*intuitive*" "*impressions*" that the art observer applies to the art "*object*" from both perceptual and cognitive perspectives. These "*impressions*", which often introduce bias into the process of art appreciation, also arise from the heuristics-based "*affective valuation*" as part of the "*evaluative judgments*" rooted in the "*pre-judicative makeup*", which is itself attached to the "*structural makeup*" of the whole aesthetic "*judicative*" structure. However, the above process flows are possibly even more impactful at "*first-person*" level of the artist creator domain, where the act of creation of art is tied into the "*reality*", the "*authenticity*" and the "*unicity*" of both art creative processes and art objects themselves, along with being directly linked to the "*operant subject*'s" identity. These three elements are interdependent, and one cannot function without the other. Even more so, we have to emphasize the mysterious aspects related to the aesthetic "*non-judicative*" ("*non-cognitive*") domain, where the affective charges are pushed to the limit, and in which the experiencing is prior to judgment, feeding into a multitude of intuitions that flood, through the representational content layers, the "*judicative*" domain. Along with the "*judicative*" regularisation, the "*non-judicative*" influences gain, as well, a lot of traction both at the art observer's level and in art appreciation processes.

At the art observer's level, the aesthetic "*preferences*" would need to undertake the same purifying process of "*de-constitution*", "*interpretation*" and "*re-interpretation*", to be able to transition from "*intuitive preferences*" towards more educated, "*deliberative preferences*".

With regard to the processes of art appreciation, the "*reward*" is connected to the "*pleasure*" of seeing and sharing the art experience. However, as we evolve towards more elevated "*cumulative*" aesthetic "*rewards*", progressing through successive phases of "*de-constitution*" and "*interpretation*" via iterative "*pre-judicative circularity*", the "*pleasure*" begins to depart from a "*reward*" incentive-based attitude and transforms towards "*value*". This is the moment when aesthetic "*value*" becomes embedded in the experience of art appreciation. This may be a lengthy process, but its payoff could be more significant, yielding a more fulfilling art experience for the art observer from a reflective standpoint

(being able to appreciate the insights in a work of art and to connect to the artist's imagination and creative experience), a cognitive standpoint (being able to grasp the conceptual elements showcased throughout art representations, to understand their contents and, also, to effectively evaluate them, by comparison with others) and an interpersonal standpoint (being able, as an art observer participating in the processes of art appreciation, to be part of the broader art community and become an art lover, supporting creative endeavors and the betterment of society through art).

For artist creator, the dynamics of "*pleasure*" and "*reward*" and many times, "*pleasure*" versus "*reward*", are transformative of the art experience he/she undergoes. From this perspective, the "*atypical salience attribution*"[24] permeates the artist's endeavor and impacts directly the artist's aesthetic processes and judgments. Thus, the pursuit of art is driven by "*rewards*" (*e.g.*, "*wanting*" to create a work of art or a piece of music) that progressively leave the territory of pleasure (*e.g.*, "*liking*" to paint or to write music), at least in its hedonic sense, towards one of discovering and capturing a different kind of "*significance*" or "*meaning*" of what is witnessed in the world (this one or other worlds that may be, empowered by the forces of imagination) and then taken from it (or from them) into what is being manifested and expressed through the art object. With this different frame of reference, the aesthetic judgments of the artist creator shift from a "*reward-based*" model, in which reward is equated with pleasure, towards a "*significance- and meaning-based*" model, wherein he/ she approaches a "*value-driven*" experience of art. Such an experience unfolds both at the "*first-person*", intrapersonal, metacognitive level (*the wanting it* from "*reward*" turns into *the being it* from "*self-actualization*", to recall Maslow's term) and at the "*third person*", interpersonal, intersubjective level (*the wanting it* from "*reward*" turns into *sharing* it and *communicating* it within the "*shared manifold of intersubjectivity*", to recall Gallese's hypothesis).

By this time, the hedonic pleasure from the nascent phases of art creation (*e.g.*, "*I am painting because it relaxes me*") has transformed into "*value-driven*" and "*goal-oriented*" fulfilment and the creation of art has become an existential practice. It is the practice in which the "*non-judicative*"(instead of being a remote place at the fringes of our sight,

---

24   See Crețu 2024; in what concerns the relationship between "*liking*" vs. "*wanting*"; see also Gross 2024, 1.

hidden in the heuristic realm, where biases lie in wait, ready to surface into the cognitive space) becomes, through a conscious "*de-constitution*", "*interpretation*" and "*re-interpretation*" effort, a powerful resource that the artist creator starts mastering. Thus, the "*non-judicative*" domain, while maintaining its specificity, moves closer to and integrates more fluidly with the "*judicative*" domain. Particularly for creative processes, "*abductive reasoning*" and "*analogical reasoning*" are employed in order to integrate "*symbolic knowledge*", while "*intuitive reasoning*" is used in order to integrate "*episodic knowledge*". Both "*symbolic knowledge*" and "*episodic knowledge*" types can intersect at "*intrapersonal reasoning*" level[25], the latter bridging the gap between them, that for others, perhaps less oriented towards creative endeavors, tends to remain an unpassable chasm.

This synergy creates a powerful rippling effect. It contributes to the artist creator becoming aware of dimensions otherwise inaccessible to the cognitive mind, while also making the artist creator aware of his/ her own inner completeness. This, in turn, empowers this artist, enhances the creative processes and allows him/her to be in the flow, in the act of artistic creation. Furthermore, this synergy advances the artist creator along the "*value-driven*" path, enabling him/her to aim towards heightened states of awareness and more evolved and refined solutions and goals.

References

American Psychological Association (2025). *APA Dictionary of Psychology*, available online at <https://dictionary.apa.org/> last time accessed on March 29, 2025.

Amodei, Dario; Chris Olah; Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. (2016). "Concrete Problems in AI Safety". *arXiv*, available online at < http://arxiv.org/abs/1606.06565 >, last time accessed on July 25, 2016.

Cernica, Viorel (2013). *Judecată și Timp. Fenomenologia Judicativului*. Institutul European.

Cernica, Viorel (2019). „Actul de a pre-judeca şi re-constituirea subiectului. Aplicaţie în hermeneutica pre-judicativă." In Viorel Cernica (ed.), *Studii în hermeneutica pre-judicativă şi meontologie* (*Studies in Pre-Judicative Hermeneutics and Meontology)*, Vol. 4. Bucharest: University Press, 185-249.

---

[25] In what concerns the analysis of various types of knowledge see the Ben Goertzel's AGI (Artificial General Intelligence) model, presented in my paper, "The Advent of AGI" (Creţu 2023, 124-147).

Cernica, Viorel (2023). „Despre o aporie identitară: Eu sunt non-ființă" ("About an Identity Aporia: I am non-Being"). In Viorel Cernica (ed.), *Studii în hermeneutica pre-judicativă și meontologie* (*Studies in Pre-judicative Hermeneutics and Meontology)*, Vol. 8. Bucharest: University Press, 267-298.

Cretu, Sorin (2023a). "Heuristics, cognitive biases, metaphors, embodied mind. The road to being human." In *Studii în hermeneutica pre-judicativă și meontologie (en. Studies in Pre-Judicative Hermeneutics and Meontology).* Vol. 7, coordinated by Viorel Cernica. Bucharest: University Press.

Crețu, Sorin (2023b). "The Advent of AGI." In *Studii în hermeneutica pre-judicativă și meontologie (en. Studies in Pre-Judicative Hermeneutics and Meontology).* Vol. 8, coordinated by Viorel Cernica. Bucharest: University Press.

Crețu, Sorin (2025, forthcoming). "Into a new era of 'reward-punishment' or towards the liberation from it?" In *Studii în hermeneutica pre-judicativă și meontologie (en. Studies in Pre-Judicative Hermeneutics and Meontology).* Vol. 9, coordinated by Viorel Cernica. Bucharest: EIKON Publishing.

Gallese, Vittorio (2003). "The Manifold Nature of Interpersonal Relations: The Quest for a Common Mechanism." In *Philosophical Transactions: Biological Sciences* 358 (1431):517-28.

Gross, Madeleine E., James C. Elliott, and Jonathan W. Schooler (2024). "Why Creatives Don't Find the Oddball Odd: Neural and Psychological Evidence for Atypical Salience Processing." In *Brain and Cognition* 178 (August 2024): 106178. https://doi.org/10.1016/j.bandc.2024.106178.

Kahneman, Daniel (2002). "Maps Of Bounded Rationality: A Perspective on Intuitive Judgment and Choice," *Prize Lecture*, December 8, 2002, Princeton University, Department of Psychology.

Yang, Jenny, Andrew A. S. Soltan, David W. Eyre, Yang Yang, and David A. Clifton (2023). "An Adversarial Training Framework for Mitigating Algorithmic Biases in Clinical Machine Learning". In *Npj Digital Medicine* 6 (1):55. https://doi.org/10.1038/s41746-023-00805-y.

All links were verified by the editors and found to be functioning before the publication of this text in 2024.

Paula PÉREZ ROMERO[1]

NATION, NATIONALISM AND PLURINATIONALISM
IN *THE PRINCE*

Abstract. This paper explores the concepts of nation and nationalism in Niccolò Machiavelli's *The Prince*, with a particular focus on the introduction of plurinationalism. While Machiavelli's final chapter, *An exhortation to liberate Italy from the Barbarians*, has traditionally been the focal point for discussions on his nationalist views, this study shifts the lens to examine how Machiavelli's strategies for maintaining control over diverse territories reveal implicit notions of nation and nationalism. By analysing the third chapter, *Concerning mixed principalities*, this paper highlights Machiavelli's understanding of the challenges posed by ruling over territories with different languages, customs and laws. The strategies of residing in newly acquired territories and establishing colonies are presented as methods for fostering loyalty and integration, thereby addressing the complexities of a plurinational principality. This study not only expands the traditional interpretations of Machiavelli's nationalism but also emphasises the relevance of his insights in contemporary discussions on governance and statecraft in a multicultural and multiethnic context. Through this perspective, *The Prince* is reaffirmed as a timeless guide on the pragmatic exercise of power.

Keywords: nation, nationalism, plurinationalism, Machiavelli

Introduction

Niccolò Machiavelli's *The Prince*, written in the sixteenth century, remains a cornerstone in the study of political theory and philosophy. Machiavelli's treatise provides crucial guidance for those seeking to attain and sustain

---

[1] National Chengchi University, International Doctoral Program in Asia-Pacific Studies. Email: < paulaperezromero@gmail.com >. ORCID: 0009-0008-4745-6343

political power. Fundamentally, *The Prince* is an exploration of power dynamics: its acquisition, management, and expansion within a principality, emphasising the strategies essential for a ruler to maintain control.

Machiavelli's views on nationalism are primarily derived from interpretations of the final chapter of *The Prince*, *An exhortation to liberate Italy from the Barbarians*. This chapter reflects his advocacy for a unified Italian state governed by an Italian prince and inhabited by Italians, encapsulating his vision of nationalism and the defence of a sovereign nation-state.

This paper explores the concepts of nation and nationalism within *The Prince*, while introducing a novel perspective on plurinationalism. By examining the third chapter, *Concerning mixed principalities*, through the lens of nationalism – which legitimises the state through its connection to the nation – this analysis provides a fresh interpretation of Machiavelli's work, revealing its prescient nature. It is not to suggest that Machiavelli explicitly championed plurinationalism, but rather to highlight his awareness of the cultural and social complexities that impact the stability of dominions.

The first section of this paper delineates the concepts of nation and nationalism, addressing the prevalent misuse and misunderstanding of these terms. Establishing clear definitions is essential for contextualising these issues within Machiavelli's framework. The second section reviews existing scholarly interpretations of Machiavelli's connection to nationalism. The third section introduces the innovative argument of this paper: Machiavelli's implicit acknowledgment of nation, nationalism, and plurinationalism in the third chapter as strategies for managing conquests and retaining power. The final section synthesises and discusses the key concepts presented in this analysis.

Nation and nationalism

The concept of a nation has been widely debated yet remains frequently misused by both scholars and non-academics. Often, the term nation is conflated with the state; however, these concepts are neither synonymous nor necessarily coexistent. As Seton-Watson (1977, 1) observed, there are

nations without states and states without nations, as well as states comprising multiple nations, with the latter being the most common[2]. To elucidate these distinctions, it is essential first to discuss the concept of the state before addressing the more complex notions of nation and nationalism.

The state, commonly referred to as a country, represents the most significant form of political organisation in the modern world. Primarily, the state is a legal construct characterised by political sovereignty. It operates independently, subject to no external laws or decisions. Individuals within the state are ultimately accountable only to the state itself (Kelsen 1941, 608; Seton-Watson 1977, 1)[3]. For an entity to be recognised as a state, it must encompass a defined territory, a permanent population, a government and the capacity to engage in relations with other states. Thus, the state serves as the sole representative of a population residing within a specified territory under a singular government (Chen 2001, 25)[4].

The concept of a nation, in contrast to the legal and institutional character of a state, is primarily a cultural construct. A nation is defined as a group of people bounded by the sense of belonging to the same community with ties to a territory. This definition requires four important clarifications. First, the nation is highly intangible; it is not a

---

[2]   The widespread insistence of many states on identifying themselves as nation-states – as if each state inherently encompasses only a single nation – has significantly contributed to the misuse and frequent confusion surrounding these concepts. This misconception is rooted in the belief that a nation is the fundamental basis of a state, and conversely, that only a nation can and should evolve into a state (Malešević & Trošt 2018, 1). Originating in the 17th century, this trend rapidly evolved, with states adopting the nation-state identity to legitimise their sovereignty. However, the reality is that most states comprise multiple nations. Despite this, the national recognition of these internal nations is often undermined or ignored, as acknowledging their existence could, within this conceptual framework, grant them legitimacy and a claim to sovereignty.

[3]   As Max Weber (1946, 78) most celebrated statement says, "a state is a human community that (successfully) claims the monopoly of the legitimate use of physical force within a given territory."

[4]   This definition does not suggest that states require international recognition to qualify as states; rather, it posits that states must possess the capacity to engage in relations with other states to achieve such status. The Republic of China (Taiwan), for instance, is not recognised by most countries globally, yet it maintains the ability to engage in diplomatic and economic relations with them. International recognition is a political act and not a legally binding condition for the existence of a state (Kelsen 1941, 605; Crawford 2006, 24).

community with exclusive membership but rather a self-defined entity wherein individuals identify as members by personal choice [5] . Consequently, the nation possesses a critical psychological aspect, as its members perceive themselves as part of an imagined community without necessarily knowing its full extent; they are members of a community that is constructed in their mind but not necessarily or obviously seeable (Connor 1978, 379–80; Anderson 1983, 6)[6].

Second, the most crucial element of the nation is the individuals who comprise it (Barrington 1997, 712). Since the nation exists in the imagination of its members, they form its foundation. It is not the territory or shared historical and cultural background, but the individuals who recognise each other as part of the same community that constitute the nation (Ichijo & Uzelac 2005, 213)[7]. Third, the territory associated with a nation is not as precisely defined as that of a state. Unlike the clear boundaries of a state, the territory of a nation is more fluid, subject to historical, political, and cultural evolution. This lack of fixed borders reflects the dynamic nature of a nation's attachment to its land, allowing for varied interpretations among its members regarding the nation's territorial extent (White 2007, 53–54, 58–59).[8]

Fourth, individuals may identify with multiple nations, sometimes considering their national identities as multilayered. For instance, a

---

[5]   This personal choice is what Renan (1882) would describe as a *daily plebiscite*.

[6]   Anderson (1983, 6) defines it as an *imagined community*.

[7]   The discourse on what engenders a sense of belonging to the same nation often highlights, among other factors, the sharing of a common language or a similar historical background. These elements are sometimes depicted as intrinsic to the concept of a nation, suggesting that a distinct language, for instance, is a requisite for nationhood. However, nations that share the same language, such as many in Latin America or those that speak English, are still considered distinct nations. These features are not constitutive elements of a nation but rather attributes that facilitate the self-identification of individuals as members of a nation by fostering connections among them (Reicher & Hopkins 2001, 8-9; White 2007, 1: 23-24).

[8]   For instance, the territorial boundaries of the Basque Country are subject to varying interpretations. Some define these boundaries as the current borders of the autonomous community of the Basque Country (Euskadi). In contrast, others consider the nation to extend across its historical boundaries, encompassing not only Euskadi but also Navarra and the French Basque Country, collectively referred to as Euskal Herria (Mansvelt Beck 2006, 524).

Scottish person may feel part of both the Scottish nation and the British nation, viewing Scotland as a component of the broader British nation while also recognising it as a distinct nation itself (Bond & Rosie 2006). This multiplicity arises from the self-defined nature of nations, whereby no higher authority can dictate the legitimacy of a nation or restrict an individual to a single national identity. Ultimately, the nation is a feeling of belonging that cannot be legally or politically denied.[9]

In contrast, nationalism is inherently a political concept. Nationalism posits that the nation and the state should be congruent, meaning the borders of the nation should align with those of the state. This ideology represents the political expression of a nation's potential self-governing aspirations and the belief that only a nation can legitimately form or become a state (Gellner 1983, 1; Barrington 1997, 714; Malešević & Trošt 2018, 1). Misinterpretations of nationalism are common, particularly when nationalism is erroneously viewed as a defining characteristic of a nation.[10] Some scholars argue that a nation must seek self-sovereignty to be considered a nation, suggesting that a nation inherently desires political independence (Roeder 2007, 3; Barrington 1997, 712; Stilz 2011, 575).

This assumption presents several issues. First, it transforms the nation into a purely political concept, where the defining feature of a social group as a nation is its aspiration for self-rule. This perspective undermines the self-defined and emotional nature of a nation, instead emphasising the necessity of political actors explicitly demanding sovereignty. Second, it complicates the determination of which nations seek self-sovereignty. Does a small political faction's call for independence

---

[9]   This is not to say that the existence of nations is not often denied politically and legally. In Spain, Article 2 of the 1978 Constitution labelled Catalonia, the Basque Country and Galicia as "nationalities" to recognise a different status compared to the rest of regions but without acknowledging them as nations (*Constitución Española* 1978).

[10]   Connor (1978, 378) finds that the most fundamental error when defining nationalism is equating it to the loyalty towards the state. Nevertheless, in recent times, the most fundamental of all errors seems to be using the term nationalism, in particular nationalist, to denote disapproval towards others' political beliefs with the intention to label them as (far) right-wing, separatists or even terrorists. This is nonetheless ironic considering that most modern states claim to be nation-states and legitimate their existence based on them being nations, which by the proper definition of nationalism, these states are the primary proponents of nationalism itself.

suffice to classify a group as a nation, or is a broader consensus among self-identified members required?

Defining the nation with nationalism as one of its constitutive elements erasing therefore the self-defined aspect of it would be problematic in many instances. For instance, Galicia has a history of regional resistance but lacks a strong pro-independence movement compared to the Basque Country and Catalonia (Schrijver 2006, 120; Losada 1999, 152). Would the absence of a self-ruling movement invalidate its status as a nation? Similarly, is Quebec not a nation because its independence referendum failed? Is Veneto a nation due to a minority secessionist movement, despite being largely ignored by scholars and political entities? A nation may or may not pursue political aspirations because it is fundamentally a cultural construction. Only those who advocate for the alignment of national and state borders are nationalists, without diminishing the national identity of those who do not share such political aims.

In summary, a nation is defined as a group of people united by a shared sense of community and connection to a specific territory, whereas nationalism is the political doctrine asserting that the nation and the sovereign state should coincide. Based on these definitions, the following section presents a different approach to interpreting *The Prince* and Machiavelli's stance on nation and nationalism.

Nation and nationalism in *The Prince*

Niccolò Machiavelli's *The Prince* remains a seminal work in political theory, exploring the dynamics of power, leadership, and statecraft. Among the myriad themes addressed, the concepts of nation and nationalism play a critical role in Machiavelli's vision for political unity and control. However, when examining Machiavelli's approach to the concepts of nation and nationalism here, it is crucial to acknowledge that Machiavelli did not explicitly use these terms.[11] Nonetheless, discussions of these concepts often centre on his call for a unified Italian state under

---

[11] As Easley (2012, 97) states, applying contemporary concepts of nationalism to philosophers from different political eras is anachronistic. Yet, exploring the intellectual heritage of such concepts remains a meaningful exercise.

Italian control, particularly in the final chapter *An exhortation to liberate Italy from the Barbarians*. Machiavelli's advocacy for Italian unity has led to his characterisation as a nationalist.

However, scholars have yet to reach a consensus on the nature of Machiavelli's nationalism. Gilbert (1954, 38) sees an emotional sense attached to Machiavelli's plea for the unity of Italy, appealing to his feelings for the Italian nation. On the other hand, Easley (2012, 104) interprets Machiavelli's Italian nationalism as a strategic instrument for power and control. In other words, Machiavelli supports a united Italy to achieve his ultimate goal of avoiding foreigners to rule over Italian principalities only.

Despite these differing interpretations, it is undeniable that Machiavelli exhibited nationalist tendencies. While it is impossible to precisely translate Machiavelli's intentions, his words strongly suggest such conclusion. In his last chapter the author reveals the existence of *being Italian*:

> "…then at the present time, in order to discover the virtue of an Italian spirit, it was necessary that Italy should be reduced to the extremity that she is now in…" (Machiavelli 2001, 97-98)

The concept of Italian spirit is critical to understand Machiavelli's idea of nation and plea for an Italian prince. It represents the feeling of being Italian, the shared sense of existing as an Italian. This Italian spirit is deeply rooted in Renaissance ideals of civic humanism and the shared cultural heritage of the Italian city-states. Although Italians often felt a stronger connection to their local regions, intellectuals of the time frequently referenced Italy as a whole, laying the groundwork for a broader Italian identity. This period of cultural revival significantly influenced the construction of a shared Italian identity, which would later be instrumental during the Risorgimento (Hay 1971, 4).

Additionally, Machiavelli expresses this idea of sharing a common spirit and values once again in the last chapter:

> "Look attentively at the duels and the hand-to-hand combats, how superior the Italians are in strength, dexterity, and subtlety… Therefore, it is necessary to be prepared with such arms, so that you can be defended against foreigners by Italian valour." (Machiavelli 2001, 99)

The description of the Italians and the mention of the *Italian valour* in contrast to foreigners elicits Machiavelli's imaginary of the individuals who could be called Italians. Italians, in turn, are defined by their belonging to Italy; they are the members of the Italian nation. Even though the political landscape of Italy was highly fragmented, Machiavelli was aware of the parts that formed what he called Italy:

> "…so that Italy, left as without life, waits for him who shall yet heal her wounds and put an end to the ravaging and plundering of Lombardy, to the swindling and taxing of the kingdom and of Tuscany, and cleanse those sores that for long have festered." (Machiavelli 2001, 98)

Machiavelli mentions Lombardy and Tuscany to exemplify what Italy was – a compound of dominions that, although independent, were part of the collective Italian imaginary. Therefore, Machiavelli saw a complete entity composed of all these dominions sharing commonalities; a common *spirit* shared by all their members and the basis for the Italian nation in this common identity forged by their shared history and culture.[12] This served as the legitimisation to establish a common state, to join the fate of all the Italians under a prince for all Italy.[13]

This paper, however, does not focus on the final chapter of *The Prince*, nor does it delve into Machiavelli's nationalist inclinations. Instead, it analyses other parts of *The Prince* to offer a new perspective: the introduction of plurinationalism in Machiavelli's strategy for power and control.

---

[12] Discussing the existence of the Italian nation might be controversial in this context. It is often considered that Italy as a nation is a product of a purposed nation-state building project of the nineteenth century as Massimo d'Azeglio famously said: "We have made Italy. Now we must make Italians" (Noether 1993, 779). Nevertheless, the assumption of Machiavelli's view is rooted on the fact that, as stated in the previous section, the nation is self-defined. It is possible for many to not identify as Italians but for the Italian nation to still exist because of some do identify their nation as Italy. In here, Machiavelli clearly saw a commonality among several communities that shared the *Italian spirit*, even though the idea of an Italian nation was not as fervent as one would consider from a contemporary perspective.

[13] Importantly, Machiavelli's reflections on the Italian nation and the Italians were not limited to a mere romanticisation of the Italian identity; his goal was rather to repel foreign invasions therefore renewing the glory of Italy and the Italians (Black 2013, 101).

Plurinationalism in *The Prince*

This section explores the concept of plurinationalism in *The Prince*, focusing on Machiavelli's strategies for managing and controlling mixed principalities. Plurinationalism here refers to the existence of a state containing more than one nation.[14]

In the third chapter, *Concerning mixed principalities*, Machiavelli differentiates between territories that are culturally and linguistically similar to the ruling state and those that are not. He[15] notes that:

> "Now I say that those dominions which, when acquired, are added to an ancient state by him who acquires them, are either of the same country and language, or they are not. When they are, it is easier to hold them, especially when they have not been accustomed to self-government; and to hold them securely it is enough to have destroyed the family of the prince who was ruling them; because the two peoples, preserving in other things the old conditions, and not being unlike in customs, will live quietly together, as one has seen in Brittany, Burgundy, Gascony, and Normandy, which have been bound to France for so long a time: and, although there may be some difference in language, nevertheless the customs are alike, and the people will easily be able to get on amongst themselves. He who has annexed them, if he wishes to hold them, has only to bear in mind two considerations: the one, that the family of their former lord is extinguished; the other, that neither their laws nor their taxes are altered, so that in a very short time they will become entirely one body with the old principality." (Machiavelli 2001, 18)

Machiavelli's distinction between dominions and his use of France as an illustrative example provide the first glimpse of his approach to the concept of nation. Territories of the same country and language can be understood as identifying as the same social community, in other words,

---

[14] As discussed throughout the paper, plurinationalism exists despite the regret of many states that claim to be one nation only. Examples of plurinationalism are scattered all over the world from Canada to Bolivia to Indonesia. This does not mean that these states are not a nation as well, but that they also contain other nations within them since the nation is not an exclusive but rathe multilayered concept (Merino 2018, 2).

[15] In Chapter V, *Concerning the way to govern cities or principalities which lived under their own laws before they were annexed,* Machiavelli also discusses the acquisition of territories that are different from the original principality. Nevertheless, this chapter is concerned with the change from freedom to absolute monarchy rather than with the difference in the population.

the same nation. Moreover, Machiavelli also introduces the possibility of having a different language yet being able to easily coexist due to sharing the same customs. This is the case of Brittany, Burgundy and others that while being different from each other, they are still bounded by the sense of belonging to France; there is a perception of membership within the same group. Machiavelli sees here that even though dominions were self-ruled, there is a bound between people beyond legality and politics; they are connected because they have the potential of nationhood.[16]

Then, having established that dominions of the same language, customs or laws would be easily assimilated into the main principality because of the similarities shared by their inhabitants, Machiavelli turns to deal with those principalities that differ significantly:

> "But when states are acquired in a country differing in language, customs, or laws, there are difficulties, and good fortune and great energy are needed to hold them…" (Machiavelli 2001, 18)

The departing point for Machiavelli is the understanding that the new acquired territory by the prince and his original principality do not share any similarities that could unite them, they are in fact different nations. Therefore, the prince finds himself in a non-desirable scenario: he owns a plurinational principality[17]. Machiavelli's recognition of the differences between dominions within the same principality reflects an implicit understanding of plurinationalism – the existence of multiple nations within a single political entity.

A different dominion that shares at least the same language or the same customs than the one the prince is originally from, will not experience the same conquest as one where the language and the customs are radically different; a territory with a different language and customs will be fiercer in its defence and will try to rebel[18]. The issue of nationalism

---

[16]   It is necessary to point out again how Machiavelli sees the ability of people becoming part of the same social community without the necessity of sharing the same language. Groups may not share the same language but rather other characteristics, such as customs and laws as in this example, that creates this self-identification as members of the same circle.

[17]   Or as Machiavelli describes it in this same chapter "a state composed of divers elements."

[18]   Although this is not to say that the first type of territories is completely of the same nation than the one of the prince, it is doubtless that the proximity between the two

also emerges here. A prince will confront more difficulties to legitimise his ruling over a territory that is in essence of different nation because of the expectation of having a ruler that can somehow identify with the population, that seems to belong to the same social community.

Plurinationalism presents significant challenges for rulers, and despite its undesirability, its existence cannot be ignored. For instance, after the Romans conquered Gaul, they faced prolonged resistance from its inhabitants, including major revolts such as the one led by Vercingetorix (Dyson 1971). This example illustrates the enduring hardships that arise when different nations are united under a single rule. Machiavelli was acutely aware of these difficulties and the complexities they introduced in maintaining control over such diverse territories.[19]

Confronted with the reality of plurinationalism, Machiavelli presents three options for maintaining control over such dominions: sending the prince to reside there, deploying the army, or establishing colonies. Among these three, the author favours the first one, sending the prince to reside in the new dominion. Even more, the author discusses the need for building connection with the prince:

> "…and one of the greatest and most real helps would be that he who has acquired them should go and reside there. This would make his position more secure and durable, as it has made that of the Turk in Greece, who, notwithstanding all the other measures taken by him for holding that state, if he had not settled there, would not have been able to keep it. Because, if one is on the spot, disorders are seen as they spring up, and one can quickly remedy them; but if one is not at hand, they are heard of only when they are great, and then one can no longer remedy them. Besides this, the country is not pillaged by your officials; the subjects are satisfied by prompt recourse to the prince; thus, wishing to be good, they have more cause to love him, and wishing to be otherwise, to fear him." (Machiavelli 2001, 19)

In this new territory, to this unknown *group*, the prince is nothing but a foreigner to the new dominion with which he shares no connection to.

---

hypothetical nations is far closer than the one between territories with different languages and customs.

[19] This does not imply that Machiavelli was aware of the challenge of owning a plurinational state in today's current terms, but rather that he was capable of seeing the difficulties that a ruler must face when aiming to control a territory that is entirely different from the conqueror's.

Anticipating potential disorders that occurred in a dominion different from the prince's original principality, Machiavelli recommends the prince to reside there and foresee problems before they arise. Yet, most importantly, the subjects of the new dominion become able to *know* the prince. When subjects can easily reach the ruler, they are more likely to love him if they want to be good and fear him if they want to be otherwise. This balance of love and fear helps maintain order and loyalty. This strategy emphasises the importance of the prince's role in bridging the gap between different nations within the plurinational principality.

Moreover, Machiavelli foresees the problem that can arise if a foreign power tries to invade a new dominion where the prince is unknown to locals:

> "…for it will always happen that such a one will be introduced by those who are discontented, either through excess of ambition or through fear, as one has seen already. The Romans were brought into Greece by the Aetolians; and in every other country where they obtained a footing they were brought in by the inhabitants. And the usual course of affairs is that, as soon as a powerful foreigner enters a country, all the subject states are drawn to him, moved by the hatred which they feel against the ruling power. So that in respect to those subject states he has not to take any trouble to gain them over to himself, for the whole of them quickly rally to the state which he has acquired there. He has only to take care that they do not get hold of too much power and too much authority, and then with his own forces, and with their goodwill, he can easily keep down the more powerful of them, so as to remain entirely master in the country. And he who does not properly manage this business will soon lose what he has acquired, and whilst he does hold it he will have endless difficulties and troubles." (Machiavelli 2001, 20)

A prince who has not invested himself in the new territory, will find that locals, moved by their hatred towards the prince that is a foreigner and an invader, will readily invite and embrace new rulers to overthrow the existing one. This phenomenon, as illustrated by the historical example of the Aetolians inviting the Romans into Greece, emphasises the crucial importance of managing local power dynamics. The prince, although the ruler, must be mindful of his *foreign* position in the new nation, that even though is his dominion, he is not part of the inhabitants' group, of their nation. He must purposefully become part of the territory, in this case especially by forging alliances with the minor powers, to earn their

respect and compromise. In other words, the prince moving to the new principality could also serve as a slower nation building process, as the prince places himself as a member of the principality and eventually becomes one of them rather than a foreigner. Inversely, the population, including the minor powers, also identify themselves with the prince as a part of the same community. This incremental integration is essential for maintaining stability in a plurinational state.

It is also pertinent to explore the two remaining options, sending the army and sending colonies. Machiavelli quickly dismisses the former, the fastest but riskiest option of deploying the army to suppress the new principality. The primary goal here is to minimise the costs associated with maintaining an army and the potential hostility it could incite. Nevertheless, Machiavelli recognises the potential advantage on sending colonies, saving costs as well as minimising the impact on the life of the inhabitants:

> "The other and better course is to send colonies to one or two places, which may be as keys to that state, for it is necessary either to do this or else to keep there a great number of cavalry and infantry. A prince does not spend much on colonies, for with little or no expense he can send them out and keep them there, and he offends a minority only of the citizens from whom he takes lands and houses to give them to the new inhabitants." (Machiavelli 2001, 19)

Employing a nationalist perspective, sending colonies to newly acquired territories, not only serves as a cost-effective means of control but also facilitates the assimilation of these territories into the prince's original dominion. By establishing colonies in strategic locations, the prince creates pockets of loyal subjects who carry the culture, laws, and customs of the ruling state. This gradual cultural integration helps to align the local population with the values and governance of the prince's original domain. The colonists, acting as intermediaries, promote stability and order, thereby reducing resistance and fostering loyalty among the locals. Over time, the presence of these colonies can lead to a blending of identities, making the new territory more cohesive with the prince's original state. This method of assimilation ensures that the territory is not only controlled but also becomes an integral part of the prince's realm, enhancing both security and unity. It slowly homogenises the plurinational principality and mitigates their stark differences.

In the latter part of the chapter, Machiavelli compares how the Romans and King Louis XII of France attempted to retain territories that were completely different nations. On the one hand, the Romans are praised for following the guidelines of Machiavelli. They did not try to change overnight the territory, but sent colonies that rather than destroying the dominion, will be assimilated within in, respected the minor powers that already existed without increasing their strength and protected the territory from any other foreign invasion. On the other hand, Louis XII made a succession of mistakes that precipitated his debacle:

> "Therefore, Louis made these five errors: destroyed the minor powers, he increased the strength of one of the greater powers in Italy, he brought in a foreign power, he did not settle in the country, he did not send colonies. Which errors, had he lived, were not enough to injure him had he not made a sixth by taking away their dominions from the Venetians… and you will see that he has done the opposite to those things which ought to be done to retain a state composed of divers elements." (Machiavelli 2001, 21)

Louis XII of France acquired territories in Italy, which differed in language and customs. He neither settled there nor sent colonies. Hence, he did not try to approach the population of the new territory, or in other words, did not begin any process of nation building. On the other hand, the disrespect to the Venetians was the King's last mistake. The Venetians, not only powerful as they were, were also a part of a community to which Louis XII was rather a foreigner. Their power but also their historical attachment to the territory legitimated more the Venetians to own the dominions than Louis XII, that although King and ultimately the prince, did lack recognition from the population. Moreover, the Venetians were also Louis XII's confidents within the new territory, that even though owned by him, he was not yet part of it.

In conclusion, Machiavelli's strategies for ruling mixed principalities reveal a sophisticated understanding of the complexities involved in governing a plurinational principality. He recognises that effective governance requires not only military might but also cultural sensitivity and strategic integration. This approach is not a mere endorsement of plurinationalism but rather a pragmatic response to the realities of ruling over diverse territories.

Discussion

None of Machiavelli's chapters of *The Prince* have been under as much discussion as his last chapter, "An exhortation to liberate Italy from the Barbarians", for his eager defence of a united Italy under the rule of Italians, leading to discussions on nationalism. This paper, while acknowledging these discussions, has explored the concepts of nation and nationalism in *The Prince* from a novel perspective. Instead of focusing on Machiavelli's vision for his own country, it has examined what rulers should do to preserve control over territories inhabited by different nations. This approach introduces the term plurinationalism into the discourse on Machiavelli's work, aligning with his overarching themes of control and power.

Machiavelli's insights into the management of mixed principalities, particularly in the third chapter, *Concerning mixed principalities*, stresses his pragmatic approach to ruling diverse territories. He emphasises the importance of cultural and linguistic similarities for the ease of governance, but also provides strategies for managing territories with significant differences. Machiavelli's recommendation for the prince to reside in newly acquired territories and the establishment of colonies are strategic moves to bridge the gap between ruler and subjects, fostering loyalty and preventing rebellion. Residing in the territory allows the prince to address issues promptly and build a direct relationship with the subjects, while colonies facilitate cultural assimilation and integrate the local population into the ruling state's framework.

The concepts of nation and nationalism, although not explicitly defined by Machiavelli, are inherently linked to his strategies for maintaining power. Machiavelli's implicit understanding of a nation as a group bound by common language, customs and identity is evident in his advice to treat different dominions with sensitivity to their unique characteristics. His discussion of the difficulties in ruling territories with distinct languages and customs highlights the challenges of governing a plurinational state. By advocating for strategies that promote unity and loyalty, Machiavelli acknowledges the power of national identity in ensuring the stability of a plurinational principality.

In conclusion, this paper has not contradicted the traditional interpretations of Machiavelli's nationalism but has expanded the discussion to include the concept of plurinationalism. By focusing on Machiavelli's practical advice for rulers managing territories of different nations, it sheds light on the complexities of governance and the strategies necessary for maintaining control. This should not be seen as an endorsement of plurinationalism, but rather as an essential intermediate step for a prince who aims to effectively govern and integrate diverse territories. Through this lens, *The Prince* remains a timeless guide on the pragmatic exercise of power in a complex and varied political landscape.

References

Anderson, Benedict (1983). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Verso.

Barrington, Lowell W. (1997). "'Nation' and 'Nationalism': The Misuse of Key Concepts in Political Science." In *Political Science and Politics* 30 (4), 712-16.

Black, Robert (2013). *Machiavelli*. New York: Routledge.

Bond, Ross, and Michael Rosie (2006). "Being and Becoming Scottish." In *Scottish Affairs* 55 (Spring).

Chen, Frederick Tse-shyang (2001). "The Meaning of 'States' in the Membership Provisions of the United Nations Charter." In *Indiana International & Comparative Law Review* 12 (1), 25-52.

Connor, Walker (1978). "A Nation Is a Nation, Is a State, Is an Ethnic Group Is a ...." In *Ethnic and Racial Studies* 1 (4), 377-400.

Constitución Española (1978). Cortes Generales. Available online at: < https://www.boe.es/eli/es/c/1978/12/27/(1)/con >.

Crawford, James (2006). *The Creation of States in International Law*. 2nd ed. Oxford: Oxford University Press.

Dyson, Stephen L. (1971). "Native Revolts in the Roman Empire." In *Historia: Zeitschrift Für Alte Geschichte*, 239-74.

Easley, Leif-Eric (2012). "Nationalist Princes and Patriotic Publics: Machiavelli and Rousseau's Enduring Insights on Nationalism." In *The Korean Journal of International Studies* 10 (1), 95-121.

Gellner, Ernest (1983). *Nations and Nationalism*. Cornell University Press.

Gilbert, Felix (1954). "The Concept of Nationalism in Machiavelli's Prince." In *Studies in the Renaissance* 1, 38-48.

Hay, Denys (1971). "The Italian View of Renaissance Italy." In J.G. Rowe and W.H. Stockdale (eds.), *Florilegium Historiale*. University of Toronto Press.

Ichijo, Atsuko, and Gordana Uzelac (2005). "Conclusion: So, When Is the Nation?" In Atsuko Ichijo and Gordana Uzelac (eds.), *When Is the Nation? Towards an Understanding of Theories of Nationalism*. New York: Routledge.

Kelsen, Hans (1941). "Recognition in International Law: Theoretical Observations." In *The American Journal of International Law* 35 (4), 605-17.

Losada, Antón (1999). "National Identity and Self-Government in Spain: The Galician Case." In *Nationalism and Ethnic Politics* 5 (3-4), 142-63.

Machiavelli, Niccolò (2001). *The Prince*. The Pennsylvania State University.

Malešević, Siniša, and Tamara Pavasović Trošt (2018). "Nation-State and Nationalism." In *The Blackwell Encyclopedia of Sociology*, 1-9.

Mansvelt Beck, Jan (2006). "Geopolitical Imaginations of the Basque Homeland." In *Geopolitics* 11 (3), 507-28.

Merino, Roger (2018). "Reimagining the Nation-State: Indigenous Peoples and the Making of Plurinationalism in Latin America." In *Leiden Journal of International Law* 31 (4), 773-92.

Noether, Emiliana P. (1993). "The Intellectual Dimension of Italian Nationalism: An Overview." In *History of European Ideas* 16 (6), 779-84.

Reicher, Steve, and Nick Hopkins (2001). *Self and Nation: Categorization, Contestation and Mobilization*. SAGE.

Renan, Ernest (1882). "What Is a Nation?" Text of a conference delivered at the Sorbonne on March 11th, 1882, in Ernest Renan, *Qu'est-ce qu'une nation*? Paris: Presses-Pocket, 1992 (translated by Ethan Rundell).

Roeder, Philip G. (2007). *Where Nation-States Come From: Institutional Change in the Age of Nationalism*. Princeton University Press.

Schrijver, Frans (2006). *Regionalism after Regionalisation: Spain, France and the United Kingdom*. Amsterdam: Amsterdam University Press.

Seton-Watson, Hugh (1977). *Nations and States: An Enquiry into the Origins of Nations and the Politics of Nationalism*. Cambridge: Methuen & Co. Ltd.

Stilz, Anna (2011). "Nations, States, and Territory." In *Ethics* 121 (3), 572-601.

Weber, Max (1946). "Politics as Vocation." In *From Max Weber: Essays in Sociology*, edited by H.H. Gerth and C. Wright Mills. Oxford University Press.

White, George W (2007). *Nation, State, and Territory: Origins, Evolutions, and Relationships*. Vol. 1. Rowman & Littlefield Publishers.

All links were verified by the editors and found to be functioning before the publication of this text in 2024.

Ceyhun ELGIN[1]

THE PHILOSOPHICAL EROSION OF AUTHENTIC BELIEF IN IDEOLOGICAL DICTATORSHIPS AND THE FRAGILITY OF INNER CIRCLES

Abstract. This paper explores the philosophical dimensions of ideological dictatorship, focusing on the dynamics that lead to the erosion of genuine belief systems within dictatorships. Drawing on historical examples, the paper examines the process by which the dictator's men, initially ideologically aligned individuals, are marginalized and replaced by opportunists. The subsequent weakening of the regime's ideological foundation has far-reaching consequences, affecting the stability of the dictatorship and its post-dictatorship transition. This paper aims to provide a philosophical framework for understanding the complex relationship between genuine belief, opportunism, and their long-term implications for political regimes.

Keywords: Dictatorship, opportunism, ideology, politics, political philosophy

Introduction

Ideological dictatorship, a phenomenon prevalent in the annals of political history, involves the consolidation of power by leaders who champion a specific ideology in a wide range of political spectrum from right to left. Examples such as Nicolae Ceaușescu, Saddam Hussein, Hafez Assad, Joseph Tito, Francisco Franco, Adolf Hitler, Enver Hoxha, and Todor Zhivkov Zhivkov stand as testaments to the complex relationship between dictators and their inner circles, particularly the dynamic evolution of shared beliefs and power struggles within these circles.

---

[1]   Bogazici University. E-mail: < ceyhunelgin@gmail.com >. ORCID: 0000-0001-8796-2429.

This paper aims to explore the dynamics of ideological dictatorships, specifically focusing on the evolving relationship between dictators and their inner circles. Understanding this relationship is crucial for unraveling the complexities of political power, the erosion of ideological foundations, and the consequent impact on the stability of authoritarian regimes. In this regard, we aim to contribute to the broader discourse on political philosophy and governance. Specifically, we intend to explore the formation and evolution of the dictator's inner circle, the reasons behind the elimination of original ideological allies, and the subsequent infiltration of opportunistic individuals. In this sense, the current paper will also scrutinize the impact of these dynamics on the dictator's ideology and the vulnerability of the regime in post-dictator scenarios. In the sections that follow, we will look into the complexities of ideological dictatorships, beginning with the formation of the dictator's inner circle and tracing the evolution of power dynamics. We will also explore the emergence of opportunism and its subsequent impact on the ideological foundations of the regime. All in all, in this paper we seek to contribute to the understanding of ideological dictatorships and their implications for political philosophy

The philosophical significance of this inquiry extends far beyond the specific context of authoritarian regimes. It touches upon core questions in political theory: How do belief systems shape and sustain political power? What are the ethical implications of political opportunism? How do we understand the relationship between authentic belief and legitimate governance? By grappling with these questions through the lens of ideological dictatorships, we gain new perspectives on contemporary challenges to democratic governance and the global political order.

This analysis will draw upon a rich tradition of political philosophy, from Machiavelli's pragmatism to Weber's theories of legitimacy, while also engaging with contemporary scholarship on authoritarianism, populism, and democratic decline. By bridging historical case studies with philosophical reflection, we aim to contribute to ongoing debates about the nature of political power, the ethics of leadership, and the future of democracy in an increasingly complex global landscape.

This paper argues that the erosion of authentic belief in ideological dictatorships reveals fundamental tensions between power, ideology,

and governance that have profound implications for our understanding of political legitimacy and authority. By examining the transition from ideological fervor to opportunism within dictatorial regimes, we uncover critical insights into the nature of political power, the role of belief systems in governance, and the ethical dimensions of leadership. This analysis not only sheds light on the internal dynamics of authoritarian regimes but also offers valuable perspectives on contemporary challenges to democratic governance and the global political order.

Related Literature

The study of ideological dictatorships has garnered extensive attention from various disciplines, including political science, history, and economics. While existing literature often explores historical accounts, strategic considerations, and game theory, there is a noticeable gap in philosophical analyses that delve into the erosion of genuine belief within ideological dictatorships. The short literature review presented in this section aims to provide an overview of the existing scholarship while highlighting the need for a philosophical framework to understand the dynamics of belief and opportunism within such political contexts.

Historical analyses of ideological dictatorships frequently focus on specific cases, offering valuable insights into the rise and fall of these regimes. Works such as Hannah Arendt's *The Origins of Totalitarianism* (Arendt 1951) and Robert Conquest's *The Great Terror: Stalin's Purge of the Thirties* (Conquest 1968) provide detailed accounts of totalitarian regimes, emphasizing the consolidation of power and the suppression of dissent. Specifically, Arendt (1951) scrutinizes the rise of totalitarian regimes, emphasizing the manipulation of ideologies for political control, whereas Conquest (1968) provides a detailed account of Stalinist purges, shedding light on the mechanisms of power consolidation. In another related study, Belova and Gregory (2002) provide a valuable historical lens through which to contextualize and corroborate the philosophical arguments presented in this paper on ideological dictatorships. The examination of the Soviet archives offers insights into the transformation of the inner circle of the Soviet stationary bandit, shedding light on the dynamics

between the dictator and the agents within the regime. The notion of opportunistic agents within the Soviet bureaucracy lacking long-term goals resonates with the philosophical exploration of dictators surrounded by opportunists rather than ideologues. Both sources explore the implications of the inner circle's composition on the regime's stability and ideological integrity, drawing parallels between the Soviet case study and the broader philosophical analysis of dictatorships and their evolving dynamics over time. However, these historical analyses often prioritize political events over philosophical reflections on the evolving nature of belief systems.

A substantial body of literature has applied game theory and strategic perspectives to analyze the behavior of dictators and their supporters. The seminal work by Fearon and Laitin (1996), "Explaining Ethnic Conflict", explores the rationalist foundations of conflict, including the strategies employed by authoritarian leaders. While contributing to our understanding of political strategies, this strand of literature tends to underemphasize the deeper philosophical dimensions of belief and authenticity.

Philosophical explorations of political ideology and authenticity offer a promising avenue for understanding the dynamics within ideological dictatorships. Taylor (1991) explores the concept of authenticity, discussing the tension between individual beliefs and societal expectations. This philosophical perspective provides a foundation for examining the genuine commitment to political ideologies and the ethical implications of opportunism.

The literature on ideological dictatorships often explores the role of propaganda and indoctrination in shaping public perception. Works like Ellul (1965) or Huxley (1932) provide critical perspectives on the mechanisms through which belief systems are cultivated and manipulated. However, these analyses primarily focus on external influences rather than the internal dynamics of belief transformation.

The examination of autocratic decision-making processes in the allocation of executive positions within the inner circle, as posited in Francois *et al*. (2014) resonates well with the philosophical analysis presented in this paper on ideological dictatorships. While the cited study focuses on the static and dynamic trade-offs faced by an autocrat, particularly

regarding the political returns and threats associated with granting access to experienced subordinates, my philosophical inquiry delves into the consequences of ideological dilution within the inner circle of dictators. Evidently, both inquiries share a common thread in exploring the impact of loyalty dynamics and the long-term implications for the stability and competence of regimes. The survival concerns discussed in Francois *et al*. (2014) align with the vulnerability of dictators highlighted in the philosophical context, shedding light on how decisions within the inner circle may contribute to the overall political performance and competence of governments.

Similarly, the exploration of authoritarian governance dynamics in Cuttner (2021), offers a pertinent parallel to the philosophical analysis presented in this paper on ideological dictatorships. The fundamental tension identified in the dissertation, regarding how a dictator interacts with powerful elites while mitigating the inherent risks of resources being turned against them, aligns with the examination of ideological dilution within the inner circle of dictators in the philosophical context. The strategies explored in the dissertation, such as purges and the use of the masses, echo the concerns raised in the philosophical discourse regarding the dictators' relationship with their inner circle and the consequences of ideological dilution. Both the cited dissertation as well as the current paper examine the dynamics of maintaining loyalty, consolidating power, and navigating the risks associated with elite allies. The theoretical models and empirical insights from Cuttner (2021) offer valuable perspectives that enrich the understanding of the challenges and strategies inherent in sustaining dictatorial regimes. In another empirical paper, Mahdavi and Ishiyama (2020) offers a compelling perspective that enriches the philosophical arguments presented in this paper on ideological dictatorships. By exploring the evolution of the inner elite in North Korea, the paper provides empirical evidence for the dynamics within authoritarian regimes, aligning with the broader theoretical framework discussed in this philosophical analysis. The focus on elite purging and power-sharing in a real-world context, such as North Korea, substantiates the philosophical assertions regarding the consequences of a dictator being surrounded by opportunistic individuals rather than committed ideologues. The North Korean case study serves

as a tangible example of how inner circles in dictatorships undergo changes over time, impacting the regime's stability and ideological coherence. Consequently, this empirical work complements the theoretical insights provided in the philosophical examination of ideological dictatorships and strengthens the overall understanding of the entangled dynamics within autocratic systems.

Frantz and Stein (2017) provide a valuable empirical perspective that aligns with the philosophical arguments advanced in this paper on ideological dictatorships. While the focus of the former is on institutionalized succession rules, and the current paper investigates the consequences of dictators being surrounded by opportunistic individuals, there is a clear interconnection. The institutionalized succession rules discussed in the mentioned paper, acting as pseudo-democratic mechanisms in authoritarian settings, bear relevance to the dynamics within ideological dictatorships. The examination of succession rules as a tool for dictatorial survival complements the discussion in this philosophical analysis, shedding light on how autocratic leaders navigate internal challenges. By addressing the reduction in elites' incentives for preemptive power-grabbing through force, the paper contributes to the broader understanding of autocratic governance. Integrating insights from both works enhances the comprehension of the multiple strategies employed by dictators to maintain power and the complex nexus between institutional structures and the composition of inner circles.

Kendall-Taylor, Frantz and Wright (2017) provide a pertinent lens through which to contextualize the dynamics explored in this philosophical analysis of ideological dictatorships. The global trend toward personalized politics, as exemplified by leaders like Vladimir Putin of Russia and Recep Tayyib Erdogan of Turkey, resonates with the examination of ideological dictatorships surrounded by opportunistic individuals. The introduction aptly captures how leaders concentrate power, dismantle institutions, and cultivate inner circles to consolidate authority. The case studies of Russia and Turkey presented in the cited paper align with the philosophical argument that the ideological foundations of these leaders are compromised over time, particularly when surrounded by loyalists who may be more opportunistic than ideologically aligned. The broader trend discussed in the paper, encompassing countries beyond Russia

and Turkey, echoes the concern raised in this analysis about the weakening of dictatorships when surrounded by opportunistic individuals rather than genuine ideologues. Exploring this nexus deepens our understanding of the global landscape of personalized politics and its implications for the endurance of ideological dictatorships.

McGovern (2016) complements the philosophical analysis presented in this paper by researching the strategic considerations of authoritarian leaders in managing their inner circles. Both works converge on the critical theme of the relationship between dictators and their supporters. The dissertation explores the entangled dynamics of how leaders, while reliant on the backing of influential supporters, grapple with the inherent risks of betrayal and coup attempts. This resonates with the philosophical arguments in this paper, particularly regarding the consequences of ideological dictatorship being surrounded by opportunists rather than steadfast ideologues. The concept of purges and the strategic calculations of leaders to secure power find common ground between the two works. Understanding the risks and benefits of manipulating inner circles, as elucidated in the dissertation, enriches the philosophical exploration of how dictatorships evolve and either strengthen or weaken over time, depending on the nature of alliances and betrayals within the ruling elite.

Moreover, Goldring (2020) offers a valuable perspective that aligns with the philosophical examination in this paper, exploring the perplexed motivations behind elite purges by dictators. Both works converge on the theme of purges within the inner circles of autocratic regimes. This dissertation contributes empirical evidence and cross-national analysis to the understanding of why dictators resort to purging elites. This empirical focus complements the more abstract and philosophical analysis presented in this paper, offering concrete insights into the causes of purges. The notion that dictators, driven by concerns about foreign threats and internal stability, decide when and whom to purge resonates with the philosophical arguments in this paper, particularly regarding the vulnerability of dictators when surrounded by opportunistic individuals. Combining the empirical findings from the dissertation with the philosophical insights from this paper enriches the overall understanding of the dynamics within authoritarian regimes and the consequential purges that shape their trajectories.

Recent scholarship has further illuminated the complex dynamics of authoritarian regimes and their implications for democratic governance. Levitsky and Ziblatt's *How Democracies Die* (2018) provides a compelling analysis of how democratic institutions can be eroded from within, often through legal means, mirroring the gradual ideological erosion we observe in dictatorships. Runciman's *How Democracy Ends* (2018) offers a complementary perspective, exploring various scenarios of democratic decline and transformation in the contemporary world. Mounk's *The People vs. Democracy* (2018) examines the rise of illiberal democracy and undemocratic liberalism, concepts that resonate with our analysis of the tension between ideological claims and pragmatic governance in dictatorships. Müller's *What Is Populism?* (2016) provides valuable insights into the ideological underpinnings of populist movements, offering a framework for understanding how authentic grievances can be coopted by opportunistic leaders. These works, along with our analysis of ideological dictatorships, contribute to a broader understanding of the challenges facing political systems in the 21st century, highlighting the delicate balance between ideology, pragmatism, and legitimate governance.

As can be seen from the review of the related literature above, the study of ideological dictatorships has been a multi-dimensional subject, engaging scholars from diverse disciplines. While historical, political, and strategic perspectives have enriched our understanding, a comprehensive literature review reveals the need for a philosophical lens to explore the complicated dynamics of belief, opportunism, and authenticity within these regimes.

Despite the wealth of literature on ideological dictatorships, a notable gap exists in the philosophical examination of belief systems and their transformation over the course of a regime. While historical and strategic analyses shed light on the actions of dictators, they often overlook the profound philosophical questions surrounding the erosion of genuine commitment to ideology. This literature review underscores the need for a philosophical framework to complement existing scholarships. By drawing on philosophical concepts such as authenticity, belief systems, and political ethics, this paper seeks to fill the lacunae in the current literature and provide a clear understanding of the ideological dynamics within dictatorships.

From Authentic Belief to Opportunism

To fully grasp the philosophical implications of ideological erosion in dictatorships, we must engage with core concepts in political philosophy. The nature of political legitimacy, as explored by thinkers from Weber to Rawls, takes on new dimensions when examined through the lens of authoritarian regimes. In these contexts, the traditional sources of legitimacy – legal, traditional, and charismatic—often intertwine with ideological narratives, creating complex webs of justification for power.

The role of ideology in sustaining political power, a theme central to Marx and Gramsci's work, becomes particularly salient in the context of dictatorships. As authentic belief gives way to opportunism, we witness a transformation in the function of ideology from a guiding principle to a tool of manipulation and control. This shift raises profound questions about the relationship between belief and political action, echoing existentialist concerns about authenticity and responsibility in the political sphere.

Moreover, the ethics of political opportunism within these regimes challenge our understanding of moral agency in politics. As Machiavelli famously argued, there may be a disconnect between private morality and political necessity. However, the systematic erosion of ideological commitment in dictatorships pushes this principle to its limits, forcing us to reconsider the ethical boundaries of political pragmatism and the moral responsibilities of leaders.

In the complex world of political philosophy, the dynamics of ideological dictatorships present a compelling subject for exploration. This section endeavors to philosophically dissect the evolution of belief systems within ideological dictatorships, examining the transformation from genuine commitment to opportunistic adherence. My focus here is on the internal dynamics of ideological dictatorships, where authentic belief is gradually supplanted by opportunism, echoing concerns raised about sanctions' unintended consequences.

Similar to the "ordinary oppression" typical of dictatorships outlined Armstrong (2020) ideological dictatorships often have their genesis in a tight-knit group of committed individuals who share a fervent belief in the leader's ideology. This shared commitment forms the bedrock of the regime's cohesion. However, mirroring the "withdrawal of trade" in the

context of economic sanctions, a recurring pattern emerges – the gradual erosion of this ideological unity.

As the dictator consolidates power, there emerges a tendency to marginalize or eliminate the initial allies who once shared the leader's ideological fervor. This process parallels the unintended consequences of sanctions, where the cessation of economic interactions can lead to unexpected outcomes. The unraveling of ideological cohesion within dictatorships reflects the risks associated with wielding power – a theme that resonates with the challenges of proportionality discussed in the provided paper.

The subsequent phase in this ideological metamorphosis witnesses the replacement of the initial comrades by opportunists. This transition, akin to the proposal of "smart sanctions" in Armstrong (2020), signifies a shift from genuine supporters to opportunistic adherents. The dictator's men, once driven by authentic belief, are supplanted by individuals motivated more by personal gain than ideological conviction. The parallels between the erosion of political rights in ordinary oppression and the dilution of authentic ideological commitment in dictatorships highlight the overarching theme of the unintended consequences of power.

The long-term effects of this ideological shift are profound, echoing concerns raised in the previous paper about the potential lasting impact of economic sanctions. When dictators surrounded by opportunists either die or face public disapproval, the regime is left vulnerable. The successors, lacking genuine belief in ideology, may struggle to maintain the coherence of the dictatorship. The echoes of this vulnerability extend beyond the leader's tenure, much like the enduring consequences of sanctions on civilian populations.

To dive into the heart of this phenomenon, a philosophical lens is crucial. Charles Taylor's exploration of authenticity in political ideology becomes particularly relevant. Understanding how authentic commitment gives way to opportunism sheds light on the ethical implications within these dictatorships. Much like the philosophical reflections on the performance and proportionality of sanctions, a deeper examination of authenticity in the context of ideological dictatorships provides a clear understanding of the dynamics at play.

To address the core philosophical question of how opportunism fundamentally damages authentic belief within totalitarian regimes, we must first establish what constitutes "authentic belief" in this context. Drawing from Taylor's (1991) framework, authentic belief in political ideology represents a coherent internalization of principles that guide both private conviction and public action. It refers to a deep and genuine commitment to certain ideological principles and values that transcend mere instrumental utility or short-term advantage. Such a belief is characterized by intrinsic motivation, consistency between publicly professed principles and private actions, and a willingness to bear personal costs to uphold these principles. It manifests as a consistent commitment to ideological tenets even when adherence carries personal risk or requires sacrifice. Authentic belief is characterized by three essential qualities: internal coherence (logical consistency within the belief system), existential commitment (willingness to act in accordance with beliefs despite potential costs), and critical engagement (ongoing intellectual investment in refining and defending one's ideological position). In contrast, *opportunism* involves adherence to ideological claims primarily for instrumental purposes such as personal gain, power, or material benefits rather than from intrinsic conviction. Opportunists align superficially with ideological doctrines not because of genuine belief, but because it is strategically advantageous. Hence, their commitment to ideological principles is contingent, superficial, and reversible upon changes in incentives.

The erosion of authentic belief through opportunism follows a distinct philosophical pathway. First, opportunism introduces cognitive dissonance between stated principles and actual behavior. When a dictator permits or rewards opportunistic behavior among inner circle members, this creates a philosophical contradiction between the regime's proclaimed values and its operational reality. This dissonance gradually undermines the coherence of the belief system itself. As Festinger (1957) demonstrated in his seminal work on cognitive dissonance, such contradictions between belief and action create psychological tension that must be resolved – often through the weakening of the belief rather than the changing of behavior.

Second, opportunism corrupts the epistemological foundation of shared belief. In ideological regimes, shared "truth" forms a crucial bond between the dictator and inner circle. When opportunists simulate ideological commitment while privately holding contrary views, they contaminate the shared epistemic community. As Habermas (1984) argues in his theory of communicative action, genuine social coordination requires participants to make validity claims they honestly believe defensible. When opportunism becomes prevalent, the communicative basis for authentic ideological solidarity dissolves into strategic action aimed at power maintenance rather than authentic belief propagation.

Third, opportunism fundamentally alters the ontological status of the ideology itself. In an environment dominated by opportunists, the ideology transforms from what Searle (1995) would term a "constitutive rule" (defining the identity and purpose of the regime) to a mere "regulative rule" (an instrumental tool for maintaining power). This ontological shift hollows out the ideology's capacity to generate authentic commitment, reducing it to ceremonial rhetoric rather than a living belief system capable of motivating genuine sacrifice and loyalty.

The philosophical damage inflicted by opportunism extends beyond individual psychology to what Bourdieu (1977) calls the "habitus" – the embodied, internalized dispositions that structure social action. When opportunism pervades a regime's inner circle, it gradually reshapes the habitus of political participation, normalizing cynicism and strategic calculation where ideological fervor once prevailed. As newer generations enter the political sphere, they encounter a system where authentic belief has been replaced by performative adherence, further accelerating ideological erosion.

This transformation represents what Heidegger might characterize as a shift from "authenticity" to "inauthenticity" in the regime's relationship to its founding principles. The ideological foundation, once constitutive of the regime's very identity, becomes merely instrumental – a tool rather than an end. This philosophical debasement inherently weakens the regime by severing its connection to transcendent purpose, reducing political action to mere power maintenance rather than the pursuit of ideological vision.

In conclusion, the unraveling of ideological cohesion within dictatorships, marked by the transition from authentic belief to opportunism, poses significant challenges. Much like the scrutiny of sanctions, the unintended consequences of this internal shift reverberate beyond the immediate political landscape. A philosophical exploration of authenticity becomes pivotal in comprehending the ethical dimensions of this transformation, providing a holistic framework for future analyses of ideological dictatorships.

The Formation and Evolution of the Dictator's Inner Circle

To comprehend the complexities of ideological dictatorships, it is imperative to scrutinize the genesis of the dictator's inner circle. In the early stages, as the dictator ascends to power, there is a deliberate selection of close associates who share a fervent commitment to the ruling ideology. This initial composition is characterized by camaraderie and a collective dedication to the envisioned political and social framework. The dictator's inner circle, at its inception, comprises individuals whose beliefs align closely with the prescribed ideology. These early confidantes often share a history of political collaboration, forming a tight-knit group that serves as the bedrock of the regime. Loyalty to the ideological cause is a defining feature during this phase, fostering a sense of unity among the dictator and their associates. However, this initial harmony within the inner circle is not immune to the turbulence of power dynamics. As the dictator consolidates authority, the relationships within this circle become subject to strategic considerations. Power struggles ensue, and the dictator may choose to eliminate or marginalize original ideological allies, often citing reasons such as concerns over a potential threat to their own rule or a desire to centralize power. The evolution of power dynamics within the inner circle is integral to understanding the subsequent erosion of the dictator's original ideological foundation. The elimination of close associates who genuinely espouse the ideology paves the way for a shift in the composition of the inner circle. This transformation marks the beginning of a critical juncture in the life cycle of ideological dictatorships, setting the stage for the infiltration of opportunistic individuals.

As the ideological dictatorship progresses, the initial unity within the dictator's inner circle undergoes a transformative phase marked by power struggles and strategic maneuvering. The dictator, once surrounded by close allies who shared a genuine commitment to the ruling ideology, finds the dynamics within the inner circle evolving as power becomes increasingly concentrated.

The consolidation of power often prompts the dictator to reassess their relationships with original ideological allies. Reasons for this reassessment vary, ranging from a desire to eliminate potential threats to the regime's stability or concerns about potential challenges to the dictator's rule. This phase sees the departure or marginalization of those individuals who were once essential components of the ideological nucleus.

The elimination of original ideological allies not only alters the composition of the inner circle but also contributes to a broader shift in the ideological landscape of the regime. With the removal of individuals deeply rooted in the dictator's ideology, a void is created, and opportunistic figures may start to gain prominence within the inner circle. Power dynamics within the inner circle become a delicate interplay of loyalty, fear, and strategic alliances. The dictator, having rid themselves of some early ideological confidantes, is now surrounded by individuals whose loyalty may be more situational than principled. This shift in power dynamics lays the groundwork for the entry of opportunists into key positions within the regime.

With the departure or marginalization of original ideological allies, the dictator's inner circle undergoes a significant transformation. This metamorphosis is marked by the infiltration of opportunistic individuals whose primary motivation is not a sincere commitment to the ruling ideology but rather personal gain, be it political, economic, or both.

Opportunists within the inner circle are distinguishable by their pragmatic approach to power. Their allegiance to the dictator's ideology is often nominal, a *façade* that allows them to remain in close proximity to the seat of power. Unlike the initial composition of the inner circle, where shared beliefs were a fundamental bond, opportunists are driven by self-interest and a desire to exploit the regime for personal benefits.

This influx of opportunistic individuals brings about a subtle but profound ideological erosion. The dictator, now surrounded by those who feign alignment with the ruling ideology, witnesses a dilution of

the once-cohesive ideological foundation. The regime, initially propelled by genuine belief, now grapples with a discord between stated ideology and the opportunistic actions of those in key positions.

The impact of opportunism extends beyond the ideological sphere, influencing the decision-making processes within the regime. Opportunistic individuals, driven by self-preservation and personal gain, may engage in internal machinations, further destabilizing the regime. The dictator, reliant on a circle of nominal adherents, finds their hold on power increasingly vulnerable to the unpredictable motivations of opportunistic allies. This section explores the infiltration of opportunism into the dictator's inner circle, emphasizing its implications for the regime's ideological coherence and overall stability. Subsequent sections will explore the consequences of this ideological erosion, both for the dictator's longevity in power and the future of the regime post-dictator.

The entry of opportunists into key positions within the regime heralds a period of ideological erosion. The dictator, once surrounded by genuine adherents to the ruling ideology, now grapples with a distorted representation of their original vision. The departure of true believers weakens the ideological foundation, creating a dissonance between stated principles and the opportunistic actions of those in power. This dissonance extends to both symbolic and practical dimensions of governance. Symbolically, the regime loses its ideological legitimacy as opportunists engage in rhetoric that contradicts their actions. Practically, policies may shift away from the initial ideological framework, driven by the opportunistic motivations of those in influential positions. The dictator, caught in the web of opportunistic alliances, faces challenges in maintaining a cohesive ideological narrative. The erosion of the original ideology not only weakens the dictator's legitimacy but also leaves the regime susceptible to internal dissent and external criticism. The once-unified front of ideological governance begins to fracture.

Effects of the Transition to Opportunism

The transition from authentic belief to opportunism in ideological dictatorships has several effects.

The first consequence underscores a pivotal philosophical concern within ideological dictatorships – the automatic distancing of the dictator from their foundational ideology when surrounded by opportunists. Drawing on the works of existentialist philosophers like Jean-Paul Sartre, we can posit that the symbiotic relationship between the dictator and ideologues mirrors the authenticity of human existence. Existential authenticity, as elucidated by Sartre, emphasizes the responsibility of individuals to stay true to their chosen values. When a dictator, initially guided by a genuine ideology, succumbs to opportunism, the existential authenticity of their political existence is compromised. This echoes the philosophical exploration of authenticity within political contexts, raising questions about the ethical integrity of leaders who abandon their ideological roots for pragmatic gains.

The second effect is the stagnation of renewal. Building on this, the philosophical resonance lies in the examination of renewal within ideological frameworks. The works of political theorist Hannah Arendt, particularly her reflections on the importance of political action and new beginnings, offer insights. Arendt contends that political life requires constant renewal and the emergence of the new to prevent stagnation. When a dictator is encircled by opportunists, the stagnation of ideological renewal becomes inevitable. Arendt's emphasis on the importance of the unpredictable and the potential for new political beginnings serves as a poignant backdrop to the argument, revealing the philosophical implications of ideological stasis.

The third consequence is about the intricacies of information flow within dictatorships and its connection to vulnerability. Utilizing the philosophical lens of epistemology, particularly the works of philosophers like Karl Popper, we can discern the implications of obstructed information flow. Popper's emphasis on the importance of falsifiability and the need for openness to critique aligns with the argument. When a dictator isolates themselves from ideologues, who might provide critical perspectives, the flow of information becomes distorted. The vulnerability of the dictator to being overthrown is exacerbated by the lack of diverse viewpoints. This resonates with the philosophical imperative of epistemic humility, highlighting the dangers of informational insularity.

The fourth argument introduces a moral dimension to the philosophical inquiry. Drawing on deontological ethical theories, such as Immanuel Kant's categorical imperative, we can explore the ethical responsibility of leaders and their inner circles. Kantian ethics emphasizes the duty to act in accordance with moral principles, regardless of consequences. In the context of ideological dictatorships, the presence of ideologues serves as a moral compass, capable of objecting to unreasonable actions. The absence of such objections when surrounded by opportunists implicates the dictator in ethical transgressions. This aligns with Kant's assertion that individuals, including political leaders, are bound by moral duties, and the failure to adhere to these duties diminishes the ethical standing of their actions, leading to unreasonable actions of dictators.

The fifth and final consequence navigates the transient nature of opportunistic support within ideological dictatorships. Philosophically, this resonates with the existentialist notion of absurdity, as explored by Albert Camus. The opportunists, seeking personal gain, reflect the transient and contingent nature of their allegiance. Camus' philosophy, which grapples with the absurdity of human existence, provides a philosophical backdrop to the argument. In the absence of enduring commitment, the ideological foundation of the dictatorship becomes fragile, akin to the absurdity of seeking meaning in an indifferent universe. This perspective underscores the ephemeral nature of opportunistic alliances and the enduring fragility of ideological dictatorships.

Weaving these philosophical threads together, our analysis elucidates the interconnected vulnerabilities of ideological dictatorships. Existential authenticity, political renewal, epistemic humility, moral duty, and existential absurdity collectively paint a clear picture. This holistic philosophical inquiry not only dissects the consequences of aligning with opportunists but also underscores the enduring relevance of philosophical frameworks in understanding the complex dynamics of political power and ideological commitment.

The philosophical consequence of this transition extends beyond mere political strategy to the realm of collective meaning-making. Drawing on Arendt's (1958) conception of politics as a space of appearance where authentic action manifests, we can see how opportunism fundamentally corrupts this space. When dictatorial regimes transition from authentic

belief to opportunistic adherence, the public sphere is transformed from a domain where ideological commitment might be genuinely enacted into a theatrical space where commitment is merely simulated. This simulation, what Baudrillard (1994) would term a "simulacrum" of ideological governance, destabilizes the regime's legitimacy at its philosophical core, creating what MacIntyre (1981) identifies as an internal contradiction between the practices of the regime and the virtues ostensibly promoted by its ideology. This contradiction cannot be sustained indefinitely, as it undermines the regime's capacity to generate moral coherence even among its most devoted adherents.

The vulnerability of the regime becomes increasingly evident as opportunistic infiltration compromises the ideological core. In scenarios where the dictator dies due to natural causes, the regime is left in a precarious state. The successor, surrounded by opportunists, lacks a genuine commitment to the ruling ideology, leading to potential power struggles and a weakened regime.

Public disapproval of the regime amplifies its vulnerability. The opportunistic inner circle, motivated by self-interest, may fail to effectively navigate public dissent, further destabilizing the regime. The regime's survival becomes contingent on the unpredictable allegiances of those within the opportunistic inner circle.

A Comparative Analysis

Ideological dictatorships have manifested in various forms throughout history, with each leader leaving a distinct imprint on the dynamics between dictators and their inner circles. A comparative analysis of key historical cases from different countries unveils clear patterns and divergences in the evolution of belief systems, power structures, and the vulnerabilities of these regimes.

For example, Francisco Franco's regime in Spain from 1939 to 1975 was characterized by authoritarian rule with a blend of fascist and conservative ideologies. Initially, Franco's inner circle comprised ardent ideologues who supported his nationalist cause during the Spanish Civil War. However, as he solidified power, Franco demonstrated a pragmatic

approach. Opportunistic figures, particularly technocrats and military officials, gained prominence. The transition from ideological purity to pragmatism allowed Franco to maintain stability, but it also diluted the fervor of his nationalist ideology. This case illustrates the adaptability of dictators to ensure regime survival by incorporating opportunists while retaining a semblance of ideological coherence

Joseph Stalin's rule in the Soviet Union (1922-1953) epitomizes the Machiavellian dynamics within ideological dictatorships. Stalin initially surrounded himself with ideologues committed to Marxist-Leninist principles, yet his paranoia and consolidation of power led to purges and the elimination of even loyal ideologues. Opportunism in Stalin's inner circle was prevalent, with figures like Lavrentiy Beria exemplifying this trend. Stalin's leadership underscored the brutal consequences of ideological purges and the risks of an inner circle dominated by opportunists. The regime's vulnerability was evident in the power struggles following Stalin's death, revealing the fragility of a dictatorship built on fear and opportunism.

Nicolae Ceaușescu's regime in Romania (1965-1989) reflected a unique blend of cult personality and autocratic rule. Initially surrounded by ideologues, Ceaușescu's inner circle transformed into a family-centric cult. Opportunism took a familial form, with Ceaușescu appointing family members to key positions. The regime's vulnerability became apparent during the 1989 anti-communist uprising when public dissent erupted. Ceaușescu's reliance on familial opportunism contributed to the regime's downfall, emphasizing the perils of nepotism within ideological dictatorships.

Kim Il-sung's regime in North Korea (1948-1994) represents a unique form of dynastic isolationism, where the leader's family became central to the regime's inner circle. Initially, the regime was influenced by Soviet-style communism, and Kim Il-sung was surrounded by ideologues who fought alongside him during the Korean War. However, the regime gradually isolated itself, and opportunism took the form of familial succession. Kim Jong-il and Kim Jong-un succeeded Kim Il-sung, perpetuating a dynastic rule marked by isolationist policies. The regime's vulnerability lies in its extreme isolation and the dynastic cult, which restricts diverse perspectives and hampers adaptability.

Enver Hoxha's regime in Albania (1944-1985) exemplified isolationist tendencies combined with rigid Marxist-Leninist principles. Initially surrounded by ideologues who were part of the anti-fascist resistance, Hoxha's regime became increasingly isolated. Opportunism was stifled through purges, creating an insular inner circle. The regime's vulnerability lay in its isolationist approach, leading to economic stagnation and a lack of external support. Hoxha's rigid purism, while avoiding opportunism, left the regime susceptible to external pressures.

Todor Zhivkov's regime in Bulgaria (1954-1989) mirrored the Soviet model, aligning closely with Moscow's directives. Zhivkov's inner circle initially consisted of Soviet-backed ideologues, but the regime's vulnerability became apparent during periods of Soviet destalinization. Opportunism within the Bulgarian Communist Party emerged as leaders adapted to changing Soviet policies. Zhivkov's reliance on Soviet support made the regime vulnerable to shifts in the geopolitical landscape, ultimately contributing to its downfall during the late 1980s. Similarly, Erich Honecker's regime in East Germany (1971-1989) exemplifies the stalwart adherence to Soviet-style communism. Honecker's inner circle initially comprised committed ideologues who followed the Soviet model. Opportunism within East Germany was more subtle, involving conformity to Soviet policies rather than explicit power struggles. The regime's vulnerability became apparent during the era of Soviet destalinization and the growing discontent of East German citizens. Honecker's resistance to reform and reliance on outdated ideologies left the regime susceptible to external pressures, leading to its eventual collapse during the fall of the Berlin Wall in 1989.

Saddam Hussein's regime in Iraq (1979-2003) provides a compelling case of an ideological dictatorship within regional dynamics. Rooted in pan-Arab nationalism, Saddam initially formed an inner circle with both ideologues and opportunists, reflecting the diverse ethnic and religious landscape of Iraq. However, the Iran-Iraq War and geopolitical shifts in the region prompted purges, reshaping the inner circle. Opportunism became pronounced as the regime sought survival amid external challenges. Saddam's attempt to balance ideological commitment with pragmatic considerations resulted in a vulnerability that manifested in internal strife and external interventions. Hafez Assad's regime in Syria

(1971-2000) offers another lens on regional dynamics within ideological dictatorships. Rooted in Ba'athist ideology, Assad's inner circle initially comprised both ideologues and pragmatists. The regime's stability was maintained through a delicate balance between familial allegiances and political pragmatism. Assad strategically managed the inner circle to navigate regional challenges, including the Lebanese Civil War. The regime's longevity reflected Assad's ability to adapt to changing circumstances while retaining key ideologues. The comparison highlights the nuanced strategies employed by ideological dictatorships in addressing regional complexities.

In conclusion of this section, a comparative analysis of historical cases reveals both patterns and divergences within ideological dictatorships. The transition from ideologues to opportunists reflects the pragmatic adaptability of dictators to ensure regime survival. However, the consequences of opportunism, including ideological dilution, vulnerability to public dissent, and external pressures, highlight the inherent fragility of ideological dictator.

Normative Implications and Contemporary Relevance

The erosion of authentic belief in ideological dictatorships raises profound normative questions about the nature of political leadership and the ethical foundations of governance. It challenges us to reconsider the moral responsibilities of political leaders beyond mere effectiveness or stability. If authentic belief is replaced by opportunism, can a leader's actions ever be truly legitimate, or are they inevitably tainted by bad faith?

This analysis also forces us to confront the ethical limits of political pragmatism. While some degree of compromise may be necessary in governance, the wholesale abandonment of ideological principles for the sake of maintaining power crosses a critical ethical line. It suggests a need for a renewed focus on integrity in political leadership, even – or perhaps especially – in non-democratic contexts.

Furthermore, our findings underscore the importance of authentic belief in legitimate governance. They suggest that political systems, regardless of their specific form, require a degree of genuine commitment to

underlying principles to maintain long-term stability and legitimacy. This has implications not only for how we understand authoritarian regimes but also for how we approach the strengthening of democratic institutions and values in the face of growing global challenges.

The dynamics observed in ideological dictatorships have striking parallels with contemporary challenges to liberal democracy. The rise of populist movements in established democracies, for instance, often mirrors the ideological lifecycle we see in authoritarian regimes. Initially fueled by genuine grievances and beliefs, these movements can evolve into vehicles for opportunistic leaders, raising concerns about the resilience of democratic institutions.

Furthermore, our analysis of ideological erosion in dictatorships provides a valuable framework for understanding the challenges of democracy promotion in foreign policy. The tendency for authentic belief to give way to opportunism suggests that merely exporting democratic institutions may be insufficient; attention must be paid to fostering and maintaining genuine democratic values.

Lastly, the ethical quandaries posed by opportunistic governance in dictatorships have implications for debates surrounding international intervention. As regimes lose ideological coherence and resort to increasingly pragmatic and potentially oppressive measures to maintain power, the international community faces complex decisions about when and how to intervene, balancing concerns of sovereignty with human rights and global stability.

Conclusion

This paper has undertaken a comprehensive exploration of ideological dictatorships, unraveling the complex dynamics within these regimes and their consequential vulnerabilities. The primary focus has been on the evolution of belief systems, the formation of inner circles, the transition to opportunism, and the far-reaching effects of these transformations. By employing a philosophical lens, the paper has ventured into uncharted territory, bridging the gap between historical, political, and philosophical analyses.

The journey began with a vivid portrayal of ideological dictatorships. The introduction laid the foundation, emphasizing the nexus between dictators and their inner circles as a linchpin in understanding the stability and coherence of authoritarian regimes. The literature review shed light on the existing scholarship, identifying gaps and making a case for the philosophical examination of belief systems within ideological dictatorships. The subsequent sections dived into the heart of the matter. From genuine commitment to opportunistic adherence, the paper navigated the complex terrain where ideologies erode, and the dictator's inner circle undergoes transformative phases. The formation, evolution, and infiltration within the inner circle unfolded as critical junctures, contributing to the broader narrative of ideological dictatorships. The spelling of historical cases, from Stalinist purges to Francoist stalwarts, enriched the theoretical framework, offering empirical insights into the vulnerabilities inherent in the dynamics of belief and opportunism. Effects of the transition to opportunism were explored through a philosophical lens, drawing on existential authenticity, political renewal, epistemic humility, moral duty, and existential absurdity. The consequences reverberated beyond political landscapes, questioning the ethical standing of leaders who abandon authentic belief for pragmatic gains and illuminating the fragility of ideological dictatorships.

This paper's contribution lies in its multi-dimensional approach, intertwining historical narratives with philosophical reflections. By examining belief systems and inner circle dynamics through a philosophical lens, the paper expands the discourse on ideological dictatorships.

While this paper has illuminated crucial aspects of ideological dictatorships, it is not without limitations and potential critiques. The philosophical framework, while enriching the analysis, may not fully capture the complexity of geopolitical factors influencing opportunistic behavior. Moreover, the focus on inner circle dynamics, though central to the paper's objectives, may oversimplify the multi-dimensional nature of political power and governance. The paper acknowledges that belief systems and opportunism are just two facets of the broader challenges faced by ideological dictatorships.

Critics may also argue that the model in the current paper oversimplifies the complexities of ideological dictatorships and fails to

account for the diverse array of factors influencing the evolution of inner circle dynamics. However, counterarguments may assert that while no model can capture every nuance of political reality, the proposed framework offers valuable insights into recurring patterns observed in historical and contemporary cases of ideological dictatorships. By focusing on the relationship between power dynamics and opportunism within the inner circle, the model provides a useful analytical tool for understanding the fragility of authoritarian regimes. Critics may also raise concerns about the potential bias inherent in analyzing historical cases through a theoretical lens. They may argue that the model imposes preconceived notions onto complex historical events, leading to oversimplification and reductionism. In response, proponents of the model highlight the importance of theoretical frameworks in organizing and interpreting historical data. While acknowledging the limitations of theoretical constructs, they emphasize the necessity of abstraction in uncovering underlying patterns and dynamics that might otherwise remain obscured. Furthermore, critics may question the applicability of the model to diverse cultural and historical contexts, suggesting that the dynamics of ideological dictatorships vary significantly based on factors such as cultural norms, historical legacies, and external influences. In rebuttal, proponents argue that while contextual factors undoubtedly shape the manifestation of ideological dictatorships, fundamental dynamics such as power struggles and opportunism remain universal themes. The model's strength lies in its ability to identify these underlying dynamics while remaining flexible enough to accommodate contextual nuances.

References

Arendt, H. (1951). *The Origins of Totalitarianism*. Harcourt, Brace & World.

Arendt, H. (1958). *The Human Condition*. University of Chicago Press.

Armstrong, C. (2020). "Dealing with dictators." In *Journal of Political Philosophy*, 28(3), 307-331.

Baudrillard, J. (1994). *Simulacra and Simulation*. University of Michigan Press.

Belova, E., & Gregory, P. (2002). "Dictator, loyal, and opportunistic agents: The Soviet archives on creating the Soviet economic system." In *Public Choice*, Vol. 113, 265-286.

Bourdieu, P. (1977). *Outline of a Theory of Practice*. Cambridge University Press.

Conquest, R. (1968). *The Great Terror: Stalin's Purge of the Thirties*. Oxford University Press.

Cuttner, A.K. (2021). *Frenemies: Coalitions and Conflict in the Dictator's Inner Circle* (Doctoral dissertation, Emory University).

Ellul, J. (1965). *Propaganda: The Formation of Men's Attitudes*. Vintage Books.

Fearon, J.D., & Laitin, D.D. (1996). *Explaining Ethnic Conflict*. The MIT Press.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.

Francois, P., Rainer, I., & Trebbi, F. (2014). "The dictator's inner circle" (NBER Working Papers 20216). *National Bureau of Economic Research*.

Frantz, E., & Stein, E.A. (2017). "Countering coups: Leadership succession rules in dictatorships." In *Comparative Political Studies*, 50(7), 935-962.

Goldring, E. (2020). *Elite Purges in Dictatorships* (Doctoral dissertation, University of Missouri-Columbia).

Habermas, J. (1984). *The Theory of Communicative Action*. Beacon Press.

Huxley, A. (1932). *Brave New World*. Chatto & Windus.

Kendall-Taylor, A., Frantz, E., & Wright, J. (2017). "The global rise of personalized politics: it's not just dictators anymore." In *The Washington Quarterly*, 40(1), 7-19.

Levitsky, S., & Ziblatt, D. (2018). *How Democracies Die*. New York: Crown Publishing.

Mahdavi, P., & Ishiyama, J. (2020). "Dynamics of the inner elite in dictatorships: Evidence from North Korea." In *Comparative Politics*, 52(2), 221-249.

McGovern, E.R. (2016*). Dictators Who Dominate: Betraying Allies to Gain a Preponderance of Power* (Doctoral dissertation).

MacIntyre, A. (1981). *After Virtue: A Study in Moral Theory*. University of Notre Dame Press.

Mounk, Y. (2018). *The People vs. Democracy: Why Our Freedom Is in Danger and How to Save It*. Cambridge, MA: Harvard University Press.

Müller, J.W. (2016). *What Is Populism?* Philadelphia: University of Pennsylvania Press.

Runciman, D. (2018). *How Democracy Ends*. New York: Basic Books.

Searle, J.R. (1995). *The Construction of Social Reality*. Free Press.

Taylor, C. (1991). *The Ethics of Authenticity*. Harvard University Press

All links were verified by the editors and found to be functioning before the publication of this text in 2024.