BALKAN JOURNAL OF PHILOSOPHY

Vol. 7, Issue 2, 2015

The issue is co-sponsored by the Bulgarian National Science Fund, competition "Bulgarian Scientific Periodicals—2014," project No. DNP 04/51 from 12.12.2014.

CONTENTS

Articles

BALKAN JOURNAL OF PHILOSOPHY. Vol. 7, Issue 2, 2015

Articles

LEARNING HOW TO SEE

Urban Kordeš
University of Ljubljana, Faculty of Education,
Ljubljana, Slovenia
e-mail: urban.kordes@pef.uni-lj.si

Abstract:
The article discusses some of the consequences of choosing the radical constructivist epistemological point of view (as proposed by Glasersfeld and von Foerster). What happens when the role of the observer (author, philosopher, or scientist) is not ignored? What happens when observation is taken as interaction instead of one-directional copying? By explicating his answers to fundamental (or—according to von Foerster—"unanswerable") questions, the author makes an attempt at outlining the changes that science of the mind should adopt if the constructivist concept of interactional nature of observation is accepted in its entirety.

Key words: observation, constructivism, non-trivial, participatory research.

Cybernetics, cognitive science, and constructivism

The beginnings of the joint research project that later became known as cognitive science can be traced to the cybernetics movement that began in the early 1940s with the so-called Macy conferences. Organized annually from 1941 to 1960, the Macy conferences hosted a collection of scientists from a wide spectrum of disciplines, ranging from anthropology and mechanical engineering to biology. Arguably the most important results of these conferences were descriptions of a number of abstract patterns of dynamics, found to be capable of describing phenomena across the disciplinary spectrum. Margaret Mead, one of the movement's founders, described the "cross-disciplinary" nature of their findings in the following way: "I specifically want to consider the significance of the set of cross-disciplinary ideas which we first called 'feed-back' and then called 'teleological mechanisms' and then called 'cybernetics'—a form of cross-disciplinary thought which made it possible for members of many disciplines to communicate with each other easily in a language which all could understand" (Mead 1968: 10).

The power of such cross-disciplinary ideas (or metaphors) was tremendous. The resulting scientific collaborations brought into being several new scientific fields and approaches, such as communication theory, system family therapy, (partially) artificial intelligence, and—probably most prominently—cognitive science. In the 1970s, after the cybernetics movement had slowly begun to fade away and give way to the less broad and more concrete cognitive science (and the other aforementioned research endeavours), the feeling prevailed that cybernetics had completed its mission.

Yet a relatively small, loose group of scientists disagreed. They noted that the ideas of cybernetics had perhaps been rushed to concrete disciplinary realization a little too hastily, without being thoroughly considered. These ideas—especially about the workings of the mind—could (and should) be taken one step further. What this loose group of researchers had in common was the realization that "a brain is required to write a theory of a brain. From this follows that a theory of the brain that has any aspirations for completeness has to account for the writing of this theory. And even more fascinating, the writer of this theory has to account for her or himself" (von Foerster 2003: 289). In other words, they realized that they themselves are the cognizing subjects, and therefore their insights about cognition are not only about the systems they were researching, but also—in equal measure—about themselves. The acknowledgment of being "included in a larger circularity" (von Foerster 1992: 10) which connects the research subject and the researcher has reformed cybernetics into the so-called *cybernetics of cybernetics* or *second-order cybernetics*. The notion prevailed that this kind of cybernetics is "for all practical purposes" actually an epistemology (von Foerster 1974: 3). Keeney labelled it a *natural epistemology* (Keeney 2002) to denote the fact that it deals with the same subject as philosophical phenomenology, but relies heavily on the findings of the empirical sciences.

For researchers such as Heinz von Foerster, Paul Watzlawick, Humberto Maturana, Francisco Varela, Ranulph Glanville, and Ernest von Glasersfeld, accepting their inclusion in a larger circularity meant accepting the active role of themselves and the constructive nature of the process of cognition. The term *constructivism* is used to denote an epistemological credo and various underlying perspectives that emerged from the empirical, modelling, and philosophical investigations of the second-order cyberneticians. Broadly speaking, in this context constructivism stands for "the idea that the mental world—or the experienced reality—is actively constructed or 'brought forward', and that the observer plays a major role in any theory" (Riegler 2012: 237).[1]

Explicating one's epistemological position

The aim of the present article is to discuss the act of observation from a constructivist viewpoint. I want to explore the constructivist turn: from the understanding of perception as copying the external world into mental states, to perception as an act of construction of the phenomenal world. Since I am interested in exploring the consequences for (cognitive) scientific research, I will try to broaden the scope of standard constructivist epistemological discussion by focusing on *observation*, i.e. the mental act (or series of mental acts) aimed at grasping and ultimately understanding a phenomenon of interest. I will try to explore what happens if we reconsider the basic intuition about the role of cognition as a representation of the external world, and instead see its primary role as the creation of a stable and continuous phenomenal world (as suggested by von Foerster 1984; and others). My aim is to show where accepting the constructivist epistemological position can lead (I am especially interested in consequences for disciplines studying the human mind), even if the conclusions can sometimes be difficult to accept or counterintuitive.

I believe that most of the claims I will present on the following pages are well-grounded in theoretical and empirical insights of embodied cognitive science. But, as a constructivist, I also see any (philosophical) theory as a personal construction of the author. A personal construction based on a personal history—regardless of how well defended—is my personal story. "Any philosophical theory is (or was) important to its maker. (If it is not so, if a theory is just an existentially non-liable wordplay—then it is not worth discussing at all.) To a constructivist, philosophy is something like a testimony (one might even say a confession)—a personal account of the search for truth, sense, or essence" (Kordeš 2005a: 301). I believe that—from the constructivist point of view—it is necessary to approach any discussion from a modest starting point such as this.

According to Rorty (1991), there are parts of the lived world for which strong enough consent is prevalent that we can take them as objective and can thus afford to be "epistemological." However, outside this area we must act "hermeneutically." We can afford to act epistemologically in areas where we have already agreed upon premises and a practice of discourse, while we must act hermeneutically when we cannot be sure that such an agreement exists. Of course, for such a hermeneutic discussion to occur, it is necessary to be sincere in admitting a lack of epistemological common ground. Rorty (1991) says that the difference between "hermeneutic" and "epistemological" discourse is that in "hermeneutic" discourse one does not ascribe the possi-

[1] It must be noted that the above-described meaning of the term *constructivism* is not the only one. Von Glasersfeld proposed the term *radical constructivism* to separate it from other theories bearing the name constructivism. For a more thorough and historical overview, see Riegler (2012).

bility of reaching agreement to the existence of a common underlying ground or matrix, but instead to the discussion itself—as long as the discussion continues, there is hope that the participants will somehow align their meanings. Scientists (try to) stick firmly to "epistemological" ground, but philosophers often wander out into the "hermeneutic" area.

An idea similar to Rorty's distinction was expressed by von Foerster in his so-called metaphysical postulate (von Foerster 1995: 6): "Only those questions that are in principle undecidable, we can decide." According to von Foerster, explicating one's own beliefs is similar to explicating mathematical axioms. We cannot build a system without first articulating the foundations. He claims that by choosing answers to so-called *unanswerable questions*, we expound our epistemological position and thereby allow for an "epistemological" discussion (as opposed to "hermeneutic," according to Rorty 1991).

Simply by agreeing with von Foerster's metaphysical postulate, I have already chosen the constructivist perspective. From this perspective, there is not much difference between examining the nature of observation (the main purpose of the present article) and explicating the author's epistemological position. In this manner, let me continue by stating my answers to some basic epistemological (and methodological) questions. The answers clearly position the author within the domain of constructivism. Most of them are quite widely accepted, and thus do not seem too problematic. Despite this, some simple conclusions that can be derived from them appear to be almost completely unacceptable for most of the (cognitive) scientific community. Pointing to these problematic corollaries and discussing the possibilities of participatory scientific research is the principal aim of this paper.

Non-trivial systems[2]

Let us look at the first two question/answer pairs as a sort of a warm-up. They might seem nonessential, and are probably not among those which von Foerster would call undecidable (meaning the answer does not define the epistemological stance). I am stating them mostly to introduce the concept of *non-trivial systems/phenomena*—the concept that will hopefully hermeneutically become clearer as the discussion evolves.

Can all systems (which can be encountered in research) be treated as repeatable?

My answer is, of course, *no*. I believe we can encounter many systems (or phenomena) which are not repeatable, and thus are also not predictable, not even in theory. By this answer, I am obviously refusing Laplace's mechanistic concept of the universe. In accordance with von Foerster, I will call all systems that are in principle non-repeatable and non-predictable *non-trivial systems*.

Next:

Is it (always) possible to ignore the role of observer?

My answer: not at all. The position of the *view from nowhere* is always merely an approximation. It is sometimes useful or even necessary, but it is an approximation nonetheless. Further elucidation along the lines of constructivist perspective: observation is (inter)action. I am a part of the universe. Whenever I act (or inter-act for that matter), I am changing myself and the universe.

To express a similar idea, Glanville (1982: 6) uses the black box metaphor: the black box is a model, invented (by an observer, of course) to explain "how some perceived 'signal', observed to change, changes. […] The relationship between the inputs and outputs can be tested by the observer, who can modify the input in the light of the output by feeding the output back to be the next input." By studying this model we can find out many interesting things about the process of observing and communication. Among other things, the model helps us to become aware that the process of observation is all about searching for stabilities in interaction between the observer and "the thing out there"—the black box.

[2] This is a further elaboration of the ideas presented at the European Meeting on Cybernetics and System Research held at the University of Vienna, Austria, 18th – 21st April 2006, in the presentation entitled "Acknowledging the non-trivial."

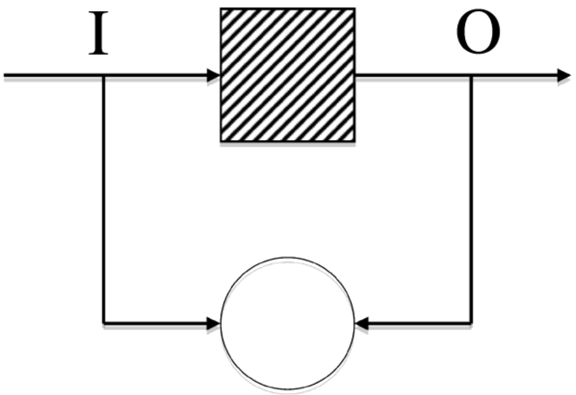


Figure 1: Input and output of black box are input for the observer (Glanville 1982)

During the process of observation (Fig. 1), the observer compares inputs (I) and outputs (O) of the black box (and tries to construct a functional description of its operations: O = f (I)). To put it more precisely: comparing what she calls inputs with what she calls outputs, the observer constructs a functional description.

In figure 1, the inputs and outputs of the black box are inputs for the observer. In order to test her functional description, the observer "controls" the input of the black box, and by that changes the direction of the arrow: the input of the black box is now the output of the observer (Fig. 2). We get a circle—circular mutual changing. The picture becomes symmetrical.

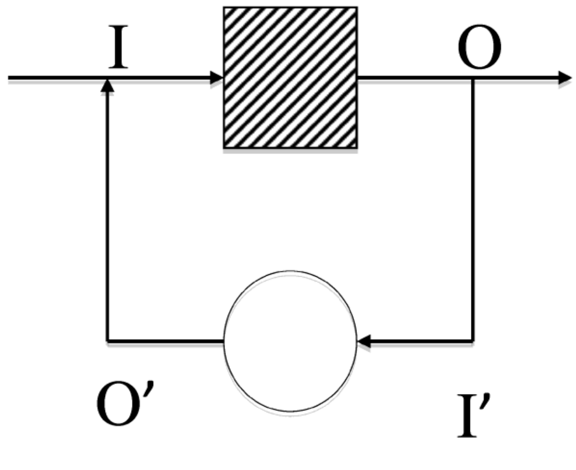


Figure 2: Observer and black box are in symmetrical positions (Glanville 1982)

When the observer constructs a working functional description, she claims that she "knows" the black box. "Knowing" the black box therefore does not mean that its inside was in any way uncovered, nor did the observer get assurance that the black box will behave the same way in the future. What we have is a stable interaction that exhibits a repeatable pattern. We have already named systems where such repeatability (and therefore predictability) is achieved: the so-called trivial systems. All stable entities that constitute our world belong to this category. Their repeatability makes them a ideal candidates for intersubjective (or better: inter-observer) validation. The complete methodological toolbox of (natural) sciences is based on the presumption that all its objects of study are of this nature. From this viewpoint, scientific methodology can be seen not so much as a method of study, but rather as a filter that selects what can or cannot be studied, one which only repeatable systems can pass through. But what about the systems where no stable interaction can be established? What if every change of (the observer's or the box's) output brings something new, unpredictable?

In cases where a stable interaction does occur, we can safely disregard the fact that all we can know are the properties of the interaction and pretend that we are dealing with a (real) thing "out there" that we are describing. In the terms of Rorty (1991), we are on the "epistemological" ground. This approximation can be, as I have mentioned above, quite useful. But when dealing with elusive, non-stable interactions, we have to admit that we are part of the observed system and that we are actually observing the process of observing. While observing we learn about the properties of interaction, not about the properties of (inter)actors. We enter into the realm of non-repeatable systems, the realm of Rorty's "hermeneutical" discourse. The process of observing is all about dialogue and co-creation or inter-change. The observer here is not a passive element, but an equal partner.

A similar understanding of observation (or, more broadly, perception) is described by Alva Noë (2004) as part of his enactive view of cognition. Regretfully, by not including himself in the equation, Noë follows the first-order cybernetics line of argumentation. Like most so-called paradigms of cognitive science, he fails to acknowledge that enactive rules of perception, as he describes them, are the same rules that govern the very investigation he is carrying. He describes a very convincing dynamic (dialogical) model of perception, while seemingly presupposing that his own is the (static) view from nowhere.

One could say that the dialogical view of observation goes one step further than Kant (and also one step beyond the standard enactive paradigm in cognitive science). It takes into account that observation—and therefore the phenomenal world—does depend on certain essential structural features of our cognition, but it also acknowledges its dependence on the individual history of interactions of a particular observer, philosopher, or scientist. Or, as Maturana and Varela (1980) put it, it depends on the history of structural couplings.

Now we can fine-tune the definition of non-trivial systems/phenomena. According to von Foerster (1993),[3] a trivial system is a system whose operations are not influenced by previous operations. It is analytically determinable, independent of previous operations, and thus predictable. For non-trivial systems, however, this is no longer true, as the problem of identification—i.e. deducing the structure of the system from its behaviour—becomes unsolvable. The history of interactions is a specific property of each system, so *de facto* all systems are non-trivial. Still, for practical purposes one can *treat* many systems as trivial (i.e. one can ignore the characteristics of interaction defined by personal history), but not all. With this in mind, non-trivial systems/phenomena can be defined as those in which the participatory and therefore interactive and constructive properties of the act of observation cannot be ignored.

Armed with this definition, I can proceed to the last of the basic question/answer pairs from my list:

Is intersubjective validation possible in the realm of the non-trivial?

It would seem that the answer must be *no*—if individual characteristics of the observer have to be taken into account, they can bring forth different patterns of couplings, and cannot necessarily be validated by a simple intersubjective comparison.

[3] In accordance with cybernetic tradition, von Foerster used the term (non)trivial "machines." For the purpose of this discussion I will replace "machine" with "system" or "phenomenon" (as explained in Kordeš 2005b).

When speaking of intersubjective validation, I do not mean to say that no agreement is possible. I just wish to emphasize that in cases where no agreement is found, such agreement (or rather lack of it) should not be the yardstick for validating the observation in question, i.e. we cannot rely on intersubjective validation as a yardstick for constructing a map of the world. But one might ask: if we lose intersubjective confirmation, what do we have left? Especially in the realm of scientific discourse, such a loss appears to be unacceptable. Science could probably tolerate many different radical paradigm shifts and even revolutions, but can it bear losing something that seems to be its very essence? Still, if we agree that there are realms of experience which the classical (trivial) way of study cannot grasp, it would seem reasonable not to just give up exploration of those areas.

It would seem clear that if reliance on intersubjective validation is out of the question, we cannot observe (describe, study, understand, etc.) non-trivial systems in the same way as trivial ones. The question remains: is there another way?

Dealing with non-trivial phenomena

Von Foerster (1972) writes: "When asked, all my friends consider themselves to be like non-trivial machines, and some of them think likewise of others. These friends and all the others who populate the world create the most fundamental epistemological problem, because the world, seen as a large non-trivial machine, is thus history dependent, analytically indeterminable, and unpredictable. How shall we go about it?" He saw three ways of how to deal with non-triviality: 1. ignore the problem, 2. trivialize the world, 3. develop an epistemology of non-triviality.

For the purposes of the present paper, we can skip the most popular one—willful ignorance. The second one is a little trickier. For instance, at the beginning of the twentieth century, physicists were faced with a surprising non-trivial system (the unignorable influence of the observer in quantum mechanics). They consciously decided for option no. 2—trivialization. They decided that it is beyond the scope of physics to deal with the observer's participation, and discovered that statistics is a great tool for escaping this annoyance. It was an epistemologically clear decision. It seems that the adherents of the Copenhagen interpretation knew what they were losing (understanding behaviour of individual particles as "solid" continuous realities, the opportunity to learn more about the role of the observer etc.), but the gain (being able to continue with the good old method) outweighed the loss. The case of quantum mechanics shows that trivialization is in many cases a good and sensible solution. Of course there are other cases where this method is far less useful. Almost the entire body of psychology and that of social sciences is trying to follow the spirit of the Copenhagen solution, but with infinitely less success.

We are equipped with very strong tools for trivialization, but in some cases the cost of such simplification is (too) high. What is lost by violent trivialization? Are there phenomena that cannot be squeezed into our sophisticated trivialization procedures at all?

In 1972 von Foerster published his so-called first theorem: "The more profound the problem that is ignored, the greater are the chances for fame and success" (von Foerster 1972: 1). However provocative this claim might appear, it is a fact that the incredible progress in (natural) science was reached exclusively by denouncing any questions concerning its epistemological foundations and by ignoring any mutual influence between the observer and the observed system. The step that ought to be taken in order to accept non-trivial phenomena into research is in direct contradiction with the fundamental requirements of established scientific methodology. How is it that we (those who share my answers to the above question/answer pairs) know this, and yet still try to do science the classical way? If we acknowledge the existence of von Foerster's "epistemological problem," we must either withdraw from science or change (broaden) its (methodological) foundations. It would seem that mere cosmetic changes will not do; a radical step out of what is now known as scientific methodology might be necessary.

Learning how to see

Non-trivial systems have a naughty habit of eluding stable interactions, and thus algorithmic description. On the other hand, stabilities of different kind may be found in such interactions—stabilities that cannot be described propositionally but can still be learned, like dance or conversation. Exploration of non-trivial systems therefore requires a shift from propositional to dispositional knowledge. It is of course not impossible to produce propositional descriptions of non-trivial phenomena, but they cannot be applied as universal mappings, valid (or even meaningful) for everyone. In a non-trivial world, studying the system means learning how to co-operate with it rather than learning *about* the system.

To learn how to co-operate with the system, we first have to acknowledge it. We have to acknowledge that our observation is (inter)action. While observing non-trivial systems we mutually change each other. The so called Heinz von Foerster's Aesthetical Imperative suggests that "If you desire to see, learn how to act." Dealing with non-trivial phenomena seems to require a skill of dialoguing and co-existing.

Employing phenomenological reduction

I believe that the modern history of attempting to learn how to see begins with Goethe's ideas about the science of observing (which were quickly forgotten and replaced by the rapidly developing objective Newtonian science). After that, there were other more or less successful attempts, among which the work carried out by the first-person investigation group gathered around Francisco Varela in 1996 (Varela and Shear 1999; Depraz et al 2003) certainly stands out. They speak of disciplined approaches to exploring the "act of becoming aware."

If *learning how to see* sounds too poetic, then the phrase *phenomenological reduction* might do more justice to the above-mentioned approaches. Referring to phenomenology, I do not necessarily mean the methods deriving from Husserlian phenomenology—I primarily mean a particular frame of mind where first-person experience is considered as irreducible and of primary importance (or even as the only venue). The praxis of phenomenological reduction itself is paradoxical. Its main component is *bracketing* the observers' natural attitudes. Still, it seems that the more one manages to adopt a position of wanting to see the phenomenon "as it is," the more space for change, construction, and transformation is opened**.** In other words: one has to put oneself into the position of wanting to see things "as they are," and at the same time know that it is precisely due to this act of observation that things are very likely to change. Seeing is constructing. Phenomenological reduction is an intrinsically dialogical process, no matter in which setting it is performed. It is about learning how to give space (by bracketing). By a disciplined praxis one can become skilled in putting his or her judgments aside, but the observer does not have to be a master of suspension and letting go—each process of observation is a learning session in which one learns about the system and about oneself through a dialogue with an equal partner. It is a continuous process of learning (how to act) as we go along.

All that has been said so far is empty epistemological debate until one starts practicing epistemological reduction. Since such praxis seems to elude algorithmic description, one has to be willing to learn the skill of observing. First and foremost, this means that one has to accept the fact that observing is a skill that can be practiced. One must be interested in finding possibilities for allowing for spaces where process inter-action can unfold in a reflective manner. Then and only then can phenomenological reduction become a venue which can be used for exploration of areas that are not accessible to standard scientific methods.

According to van Manen (1990), Heidegger once wrote something like: "Don't ask what you can do with phenomenology, see what phenomenology can do with you." What can the study of non-trivial systems do with us? (How) Can it change us? Can it help us acquire the skills to navigate within our own cognitive landscapes, as well as landscapes inhabited with other cognitive systems? From the philosophical point of view, what we are aiming at are guidelines for the development of an epistemology of non-triviality—an epistemology that would be in accordance with the constant fluidity resulting from the blurred lines between observation and construction and observer and observed, an epistemology that might serve as a foundation for the science of the non-trivial.

But could it be that our ambition to systematize and rationally frame the realm of the non-trivial is unrealizable? Could it be that the sensitivity for the non-trivial (and therefore for the unique, the unrepeatable, and the individual) cannot help us at all? Perhaps all we can expect to find is awe for the strange place we live in and the strange creatures that we are?

References

Depraz N, Varela FJ, Vermersch P (eds) (2003) *On Becoming Aware: A Pragmatics of Experiencing*. Philadelphia: John Benjamins Publishing Company.

Glaniville R (1982) Inside every white box there are two black boxes trying to get out. *Behavioural science* 12 (1): 1–11.

Keeney B P (2002) *Aesthetic of change*. NY: Guilford Press.

Kordeš U (2005a) A peaceful coexistence of epistemologies – philosophy from the constructivist's point of view. *Kybernetes* 34 (1/2): 295–305.

Kordeš U (2005b) Participatory position. *Interdisciplinary Description of Complex Systems* 3(2): 77–83

Maturana H R, Varela F J (1980) *Autopoiesis and cognition: The realization of the living*. Dordrecht: D. Reidel publishing company.

Mead M (1968) Cybernetics of Cybernetics. In: *Purposive Systems*, ed. Von Foerster H et al., 9–10. Proceedings of the First Annual Symposium of the American Society for Cybernetics. Spartan Books, New York and Washington,

Noë A (2004) Action in Perception. *A Bradford Book*. Cambridge: MIT Press.

Rorty R (1991) *Philosophical papers, vol. 2: Essays on Heidegger and others*. Cambridge: Cambridge University Press.

Rorty R (1992) *The linguistic turn*, 3 rd edn. Chicago: University of Chicago Press.

Van Manen M (1990) *Researching lived experience: Human science for an action sensitive pedagogy*. New York: SUNY Press.

Varela F, Shear J (eds) (1999) *View from within: First-person approaches to the study of consciousness*. UK: Imprint Academic.

Von Foerster H (1972) Responsibilities of competence. *Journal of cybernetics* 2(2): 1–6.

Von Foerster H (1974) *Cybernetics of Cybernetics*. Urbana Illinois: University of Illinois.

Von Foester H (1984) On constructing a reality. In: The invented reality: How do we know what we believe we know? ed. P. Watzlawick, 41–62. New

York: W.W. Norton.

Von Foerster H (1992) Ethics and second-order cybernetics. *Cybernetics and Human Knowing* 1(1): 9–19.

Von Foerster H (1995) Ethics and second order cybernetics. *Standford humanities review* 4(2): 308–327.

Von Foerster H (2003) *Understanding understanding: Essays on cybernetics and cognition*. New York: Springer-Verlag.

BALKAN JOURNAL OF PHILOSOPHY. Vol. 7, Issue 2, 2015

COHERENTIST JUSTIFICATION AND PERCEPTUAL BELIEFS

Anna Ivanova
Faculty of Philosophy, St. Cyril and St. Methodius University of Veliko Turnovo,
Veliko Turnovo, Bulgaria
e-mail: ivanovabg@gmail.com

Abstract:
A common objection to coherence theories of justification comes from belief revision processes: in a system of knowledge, perceptual beliefs seem to bear more importance than other members of the coherent set do. They are more stable in the face of confronting evidence, and may be preserved despite their degrading effect on the coherence properties of the system. This appears to be inconsistent with coherentism, according to which beliefs cannot possess independent credibility. In order to abide by the coherence theory, one must explain the stability of perceptual beliefs in belief revision in a manner that does not rely on foundationalist premises. A suggestion about the personal justification of perceptual beliefs in terms of coherence is presented in the paper to explain their stability in belief revision processes. The coherence of perceptual beliefs and a network account of knowledge are advocated in order to avoid weak foundationalism and to provide a new perspective to the normative problems of epistemic justification.

Key words: justification, coherence, perceptual beliefs, belief revision.

One of the few shared features of different coherentist theories is the commitment to the idea that all proper justification is inferential. The notion of coherence and mutual support in conceptions is reduced to some of the following inference structures—probabilistic, systematic, explanatory, or combinations of these. Beliefs in a system justified by coherence do not possess initial credibility or any degree of independent justification; the justification of a single belief is a function of its place within the system.

There are two consequences of this view regarding the status of individual beliefs in processes of belief revision. The first is that any change within the knowledge system affects the justification of all its members, and since some beliefs are fallible, from the structure of justification it follows that all beliefs are fallible. In contrast to foundationalism, coherentism rejects the idea that there are basic, irrefutable pieces of information in our theory about the world. Everything is a subject of revision; everything can be challenged in the face of new evidence.

The second consequence of adherence to inferential justification is that under the principle of mutual support alone, it is impossible to ascribe different justificational merit[1] to different members of a belief system. This is not evident without some inspection and has been often neglected in discussions. The reason is that the number of relations that sustain coherence may differ for different members in the knowledge system, thus suggesting a different justification status for particular beliefs. This is especially true if a complex notion of coherence, one that includes many kinds of relations, is employed. However, these variations are irrelevant in belief revision if coherence is the only principle that stands for justification. Different rearrangements in the belief system must be equally eligible if they result in equally coherent sets. To the end of truth, it is irrelevant whether the result of revision would be a smaller or a bigger system of beliefs, whether the new system is more similar to the one previously held or less. Without any further considerations outside justification (such as conservativeness or system expansion[2]), no rule of

[1] The expression "justificational merit" is preferred instead of "degree of justification" to include the comparative aspect: the degree of justification is measured with respect to maximal and minimal values, while the merit is established with respect to the relative importance of the belief in a system of many beliefs.
[2] These epistemic principles are sustained by pragmatic reasons of economy and effectiveness; they do not have a bearing upon epistemic coherentist justification.

preference in belief revision can be decreed upon any member or part of the subject's belief system.

A tendency that comes from belief revision processes challenges these results: in a system of knowledge, perceptual beliefs seem to bear more importance than other members of the coherent set do. They are more stable in the face of new input that requires rearrangements in the set to ensure conformity:

- They are less likely to be abandoned when they compete with non-perceptual beliefs.
- They may sometimes be accepted despite the alleged degrading effect on the coherence properties of the system.

The stability suggests that perceptual beliefs are better justified than their co-members in a knowledge system, which appears to be inconsistent with the idea of coherentist justification. In order to explain their stability in belief revision processes, the consistent coherentist has to show how the difference in the justificational merit is possible on the basis of coherence alone.

This paper is an attempt to explain the stability of perceptual beliefs in belief revision processes in consonance with a coherence theory of personal justification. To this end, existing conceptions about perceptual beliefs in contemporary coherentism are analyzed from the perspective of their relevance to actual belief revision. The conclusions drawn justify an atomistic approach toward the evaluative processes that determine the structure of justification and knowledge. The notion of coherence that underlies the proposed account is in agreement with Keith Lehrer's view of internal coherence and personal justification (Lehrer 1990, 112-131). It is in the tradition that sees coherence as explanatory inference. Keith Lehrer's idea that personal justification is performed with respect to the subject's "acceptance system" underlies the approach. However, the notion of an acceptance system is modified to reflect the advocated atomism in justification.

The status of perceptual beliefs in probabilistic accounts of coherence is not discussed. As Charles Cross argues, probabilistic measures may be fruitful in composite theories of justification[3] where coherence is only a part of the justification conditions (Cross 2006, 459). Probabilistic modeling has led to the impossibility results in Huemer (1997), Olsson (2000), and Bovens and Hartmann (2005), which show that probabilistic coherence on its own is not truth-conducive, even when supplemented by conditions like independency of belief acquisition or the possession of initial credibility for some beliefs. A more important reason for preferring the explanatory account is that it captures the coherence relation better than probability. The idea that the notion of coherence can be reduced to probabilistic relations has been successfully criticized in recent research by Mark Siebel (Siebel 2005).

The Role of Perceptual Beliefs in Coherentism

The lack of individual credibility for beliefs is both the best and the worst thing about coherence theories of justification. On the one hand, as in any fallibilist epistemology, coherentism secures openness of thinking and is in accordance with the finest critical traditions. On the other hand, while it seeks routes of certainty within the conceptual domain of the mind, it thus fails to show what the role of reality is in belief acquisition and knowledge. The isolation and the circularity objections against coherentist theories emerge directly from their assumption of a lack of individual credibility. When applied to the entire system of human knowledge, the isolation objection can be viewed much like a version of the "brains-in-a-vat" skeptical argument, designed not for foundational but for coherentist justification. Its counter-arguments are similar to the ones designed to overcome the "brains-in-a-vat" situation. However, when applied locally, the isolation objection may raise some serious concerns. Suppose someone gave you a story to read on a sheet of paper and asked you to guess whether it is true or not, whether it comes from a newspaper or from a fiction book. You could not figure out whether the story is true or false solely based on the coherence of its sentences. If a system is justified entirely through the coherence of its beliefs, then it becomes "trapped" in its immanent content with no reliable connection to reality.

It is this particular local application of the argument that is of interest to the problem of the stability of perceptual beliefs in belief revision. The coherentist conceptions address the isolation objection by suggesting that coherence proves to be selective of true beliefs. In order to evaluate the plausibility of a story through the lens of coherence, the subject has to test it within the context of *all* beliefs, including ones that are acquired through experience. To be justified, the story has to manifest

[3] Such as the one offered by Paul Thagard (Thagard 2000, 43)

a certain sort of coherentist relation to all elements of the subject's overall knowledge system, and the relation (as most coherentist theories suggest) has to go beyond mere logical consistency. Since nothing in the coherence theories of justification precludes the possibility of acquiring true beliefs, all coherentism needs to do is show why more coherence would lead to holding more and more true beliefs while leaving false beliefs behind. There has to be some sort of difference in the properties relative to coherence such that in time true members will remain in the process of belief acquisition and revision.[4] Different coherence theories of justification may be viewed as an attempt to provide an answer to this question. To answer how coherence properties are selective of true beliefs is to show how exactly coherence is truth-conducive.

One way to do this, without leaving the internalist framework, is to argue that perceptual beliefs—which are most often true—are selected and retained under the rule of coherence more often that the rest of knowledge. Coherentist accounts of perceptual knowledge developed by Bonjour (1975, 1985), Lehrer (1974, 1990), and Lycan (2012) aim at explaining the stability of perceptual beliefs with internalist arguments, and in this way provide an answer to the isolation objection.

Bonjour's Coherence Theory of Empirical Knowledge

The most detailed version of perceptual coherentism belongs to Laurence Bonjour, who sets out to refine and "work out in detail" Sellars' point of view (Bonjour, 1985, 117). After providing many reasons to consider perceptual beliefs to be inferential, and with their justification resting on other beliefs, Bonjour creates and develops his Coherence Theory of Empirical Knowledge (CTEK). He presents two conditions in which beliefs can be referred to as perceptual: in terms of their origin (the way in which they are produced or how they are caused), and in terms of their justification (Bonjour, 1975, 290). This distinction splits the problem into separate problems of origin on the one hand and justification on the other, allowing Bonjour to argue that beliefs—which are produced by sensations, although cognitively spontaneous regarding the way in which they are acquired—are not immediately justified, nor are they justified solely on the ground of their non-inferential status (Bonjour, 1985, 122). According to Bonjour, the justification of perceptual beliefs is an inference with the following structure:[5]

"(1) I have a cognitively spontaneous belief that P, which is of kind K.
(2) Conditions C obtain.
(3) Cognitively spontaneous beliefs of kind K in conditions C are very likely to be true. Therefore, my belief that P is very likely to be true.
Therefore, (probably) P." (Bonjour, 1985, 123)

The kind "K" indicates the type of perceptual belief—visual, auditory, etc. The conditions "C" are the normal conditions for accurate perception. The third premise is an empirical generalization based on previous experience. Bonjour argues that the three premises are themselves justified by reasons of coherence (Bonjour 1976, 1985). However, upon a more careful exploration of this suggestion, he comes across a difficulty that makes him revise his ideas (Bonjour 2010, 194). It resides in the differentiation of perceptual and non-perceptual beliefs within the system of knowledge. How do we know about a certain belief that it is of kind K? A rule of distinction is obviously necessary for the formulation of premise 1, and for the validity of premise 3. The criterion of distinction suggested by Bonjour is "origin" (process of acquisition, source, or cause). Claims about origin must be justified as well, but if coherence is the "bottom line" justification, then regarding a belief as perceptual is acknowledged from within the system and the input from reality is lost again. We may easily imagine a coherent system in which the categorizing rule is misapplied and beliefs from imagination and wishful thinking are classified as perceptual. When conformity to the rule is evaluated in terms of its coherence with the system, the result itself becomes arbitrary. Bonjour concludes:

"Thus, there is no way consistent with coherentism to distinguish genuine observational input from this counterfeit variety. And, in consequence, there is also no way on the basis of the only sort of 'observation' that is internally recognizable to answer the isolation and alternative

[4] This difference can be defined in an externalist manner: one may argue that the structure of the world is such that facts are best represented with coherent propositions, but since it is facts that we aim at elucidating through knowledge, an assumption like this would provide no justification of the principle of coherence itself.

[5] A modified version is presented about negative observational beliefs too. (Bonjour, 1985, 121)

coherent systems objections or to argue from coherence to likelihood of truth." (Bonjour 2010, 195)

His solution is to look for a foundationalist account—the justification for perceptual beliefs lies in the "conscious awareness" that accompanies perception as a belief-like state. This awareness about the content of beliefs is not distinct from them; it is imbedded in them in a "partly constitutive" manner. It is also neither purely non-cognitive, nor reflective (Bonjour 2010, 197).

Bonjour's concern with the correct method for recognizing perceptual beliefs does not take into account the coherentist connection between the premises themselves. His inferential schema of justification resolves the isolation objection after all. It is doubtful that premise 3 would continue to hold if the distinction rule about perceptual beliefs is substituted by a wrong but equally coherent rule. It is unlikely that random beliefs classified as perceptual would prove to be often true in the subject's former experience. Alternatively, if premise 3 in fact remains valid with a substitute rule, then the entire inferential justification scheme is successful again. As long as beliefs of a particular kind K, whether it be perceptual or not, prove to be often true, then the inference about the reliability of a particular belief is not violated. In contrast to what Bonjour believes, the alternative systems objection has no bearing upon such occasions if the truth-condition remains valid, a validation which must be done outside coherence to avoid circularity. The difficulty lays not so much in the classification of a belief as being of the kind K[6] as in the question of what it means for such a belief to be knowingly true (as opposed to merely plausible) under the justification principle of coherence. The rule that distinguishes truth is necessary for the third premise, and it has to be in turn justified if it is to be accepted in a non-arbitrary manner.

This difficulty is a consequence of the attempt to combine a coherence theory of justification with a correspondence theory of truth. In conceptions where "truth" is a more basic concept than "justification," a coherence theory that wishes to avoid circularity or infinite regress must always be reduced at some point to foundationalist premises to secure the epistemic value of beliefs. But this step is not necessary if a pragmatic view of truthfulness replaces correspondence, for one may quite successfully argue that we accept perceptual beliefs in general to be true or worthy of our trust precisely because they are far more justified than other beliefs, not the other way around.

Keith Lehrer's Conception of Personal Justification

The theory of justification provided by Keith Lehrer resolves the problem of perceptual beliefs by a novel understanding of coherence. Lehrer makes a shrewd distinction between belief and acceptance.[7] Beliefs may arise unreflectively, but their justification is a critical work of the mind, which compares and rules out alternatives and objections before acceptance. A belief *p* is justified if it coheres with the agent's evaluative system of accepted beliefs at time *t*, a formulation which Lehrer calls the acceptance system:

"The evaluation of all claims to truth, whether those of our senses, of reasoning, of memory, or of the testimony of others, must be based on our acceptance system, which contains our conception of the world and our access to it. There is no exit from the circle of what we accept." (Lehrer 1990, 113)

The aim of justification is to determine if it is more reasonable to accept a certain belief or not. Coherence is the positive principle of meeting objections and ruling out alternatives toward the acceptance of *p* through them being overcome or neutralized by the acceptance system. The beliefs that are more reasonable to accept than to give up are therefore justified (Lehrer 1990, 118).

The notion of coherence employed is relational, not holistic. Coherence is a relation between a single belief and a system, not a relation among beliefs, though the outcome of the justification process would suggest that all beliefs accepted by the subject would cohere among each other.[8] Bonjour's schema seems to be a special case of justification in Lehrer's theory, one that shows that the plausibility of the questioned belief can be inferred from beliefs in the subject's acceptance system. But there is no

[6] The classification may be viewed as application of a norm, as a decision rather than as just another belief in the system of knowledge. The rule stating that certain belief is perceptual may be constituted without regard of origin or cause but with respect to its propositional content.

[7] "I distinguish between belief, a first order attitude, and acceptance, a functional state based on metalevel evaluation." (Lehrer 1999, 243)

[8] A thesis that Lehrer himself abandoned in his later works (Lehrer 1999, 246), but without showing convincingly how inconsistencies could appear in the process of justification as described in his theory.

limitation to the number and the kind of procedures that can be involved to meet objections and produce alternative explanations in the evaluative process.

Does Lehrer's account permit a better justificatory status for perceptual beliefs? It does not. According to Lehrer, "Any claim that we see, remember, or introspect that something is the case immediately confronts the skeptical objection that the belief does not emanate from a trustworthy source but arises in some untrustworthy manner" (Lehrer 1990, 145). In his account, competitive statements like this are addressed by the acceptance system in the same way as all other statements: on the ground of what we already accept as known. It is then a contingent feature of our acceptance system that beliefs of this certain (perceptual) kind prove to be more trustworthy than others are. A conclusion like this leaves room for foundationalist explanations. Lehrer's "foundational coherentism" explains the trustworthiness of acceptance processes with external factors (Lehrer 1990, 151). The same approach could be applied to explain the stability of perceptual beliefs in belief revision processes.

Both presented accounts are a bit too conservative with respect to the evaluative process of personal justification. Because of the coherentist relation, justification procedures are symmetrical in nature—they could increase the credibility of a questioned belief, or decrease the one of another belief that has been previously accepted. This is what happens in belief revision. Bonjour's inferential schema does not reflect the possibility for revisions in the premises or explain how contradictory perceptual claims are considered if similar conditions C obtain. In Lehrer's account, the idea of acceptance also poses some restrictions to the development of the acceptance system. The model of relational coherence does not cover the possibility for changes in the acceptance system. Lehrer's personal justification should allow belief revision that goes beyond acceptance or rejection.

Another problem is that both accounts do not allow the subject to accept a belief that leads to incoherence in the system of knowledge. The principles of reasoning differ from the ones of logic in this respect. Gilbert Harman present examples (Harman, 1986, 15-16) of what we may call a case of "accepted inconsistency." It may be reasonable to accept inconsistent information at a given moment while expecting to resolve the conflict with further input in the future. These processes of doxastic acceptance do not contradict directly epistemic principles of justification.

William Lycan on the Status of Perceptual Beliefs

The account offered by William Lycan undertakes a route out of the circularity of coherentism by admitting a starting point of justification with perceptual beliefs and beliefs from memory. He agrees with the common coherentist claim that any proper justification (including the one that aims at accepting perceptual beliefs) is committed only by reference to the subject's entire system of knowledge, but he also admits the possibility of preliminary conditional acceptance for non-inferentially acquired beliefs. Lycan argues that perceptual and memory beliefs are accepted with minimal justification provided that, without regard to other evidence, they merely "seem to be true" (Lycan 2012, 9). Lycan calls this "The Principle of Credulity – Accept at the outset each those things that seem to be true." Perceptual beliefs are thus "very slightly justified" (Lycan 2012, 9); in order to be significantly justified they must be well-explained by other beliefs in a later stage of reasoning. To accept what seems to be true is not to refer to the faculties for belief acquisition (as in reliabilism) or to cognitive states (as in internalist foundationalism). Lycan holds that these first beliefs are not basic in the sense that they cannot by themselves justify other beliefs—they work only in context. Also, the initial justification is not epistemic justification. Epistemic justification for perceptual beliefs is sustained by four kinds of coherence relations: "consistency with other spontaneous beliefs; consistency and fit with the wider belief system; being explained; and holding metabeliefs about the source of the belief" (Lycan 2012, 9).

Unlike previous accounts, Lycan's conception is much better-suited to meet the isolation and the circularity objections. But does it explain the stability of perceptual beliefs? Changes in the system of knowledge are guided by the conservativeness rule which underlies the Principle of Credulity (Lycan 2012, 8). Conservativeness is responsible for abduction as the way to formulate and accept beliefs that cohere or fit best with what is previously known. Some might argue that this conception leads to scepticism, since it is simply a matter of luck that claims of perception "seem to be true" more often than claims of free imagination. And then it is a matter of luck that they become better-justified. Suppose a subject accepts under the Principle of Credulity beliefs from wishful thinking in a quantity that supersedes perceptual and memory beliefs and changes the overall coherence balance of the system. The four kinds of coherence relations would lead to quite different results in the justification of perceptual beliefs. A different way to put this issue is that it is at least in principle possible that a perceptual belief contradicts a large share of what is previously known, and still is held with reasonable confidence. Even if it conforms to the fourth, most important kind of coherence (holding metabeliefs about the source of the belief), it is not clear under what epistemic principle (if not foundationalist) it is reasonable to override a larger quantity of coherence relations elsewhere in the system. If justification is guided by conservativeness and not the other way around, then the stability of perceptual beliefs has nothing to do with their credibility.

Justification cannot be placed entirely on representation, as the isolation objection shows. The origin of a belief is not ultimately useful since false beliefs also have causes, and to distinguish proper causes from improper ones requires well-founded judgment. It seems that either coherentists should dispense with fallibility, or they should spiral endlessly into the circle of beliefs without being able to show how they were originally created.

The Network account of Knowledge and Justification

To explain the stability of perceptual beliefs in belief revision, one has to prove in consonance with the coherentist account of justification that they are fallible but still better justified. To do so it is necessary to resolve a difficulty: how does one establish degrees of justificational merit for different beliefs in a system justified by coherence?

In coherentist conceptions so far, it has been assumed that knowledge constitutes a single coherent system to the extent to which its beliefs are interrelated, and consequently all such beliefs take equal part in the evaluative processes of justification. The accounts analyzed above have succeeded in being sufficiently holistic to explain fallibility while avoiding skepticism, but have been insufficiently atomistic to secure at least partial independency in the justification status as a condition for varying justification merit. The independency condition is necessary to guide the transmission of justification in processes of belief revision. An inconsistent set of four beliefs that form two equally coherent but mutually exclusive pairs can be revised by accepting only one of them. Can coherentism alone provide a rule of preference in such occasions? In order for the transmission of warrant to even begin in some direction, individual beliefs must have initial difference in the degree of justification. This is possible if they are warranted by outside factors. The outside factors need not be obscure justifiers, different from beliefs. Justification may still rest on inference if the outside beliefs that provide higher justificational merit are members of an independent belief set.

The solution proposed here challenges the systematic view of knowledge in favor of a network account and atomism in the evaluative processes of belief revision. It is based on three modifications of the structure of coherentist justification that are designed to integrate the independency condition:

1. *Only relevant beliefs are employed in the justification process.* The acceptance system responsible for justification is formed with regard to a certain belief; it is a *relevant acceptance system*, not the whole body of knowledge.
2. *Perceptual beliefs form the most coherent subsets within the network of knowledge*. Perceptual beliefs are justified by their concurrent perceptual beliefs, all members of coherent subsets. Their higher justificational merit can be traced back to these prior justifiers.
3. *The subject's entire knowledge is a network of subsets that need not all be interrelated or even consistent.* Separate occasions of justified acceptance create subsets with different coherence properties. The employment of a single belief in different subsets secures network relations between the subsets in the body of knowledge.

Thesis (1) ties justification to a finite set of beliefs and explains away the infinite regress difficulties. The evaluation of a belief's justification is a process of resolving a problem—the problem of its credibility. When the problem-solving process is initiated, the relevant beliefs form a subset which is designed to address the question of coherence. The acceptance system against which coherence is tested is a set of *relevant* beliefs. Belief revision processes initiated by new input also take place not in the whole network of knowledge but in the relevant acceptance system. It would be impossible for the subject to use *all* accepted beliefs in evaluating new input not only because of memory constraints, but also because the system of knowledge is never a finite set of coherent propositions. Beliefs may be infinitely combined to produce new ones, one can have beliefs about beliefs, etc. In addition, the nature of experience is such that it constantly provides new information to the subject.

In addition to the difference in justificational merit, the idea that the subject does not employ all knowledge for justification resolves some accessibility problems for justified reflective beliefs. The justification of a reflective belief about a subset's coherence need not be within the subset itself, but in another subtset entirely, thus avoiding circularity at the level of personal justification.

The higher justificational merit for some beliefs is relative to a particular occasion of justification, not to the overall network of knowledge. A large amount of context is suppressed during the evaluation process. One does not need to recall beliefs that have been previously confirmed if they are not questioned by this new information. Therefore, it is possible that there are beliefs in the relevant acceptance system that are not justified by its members, but their justification lies outside the cognitive attention of the subject. The knowledge that they possess some degree of justification that may be traced to the occasion of its initial evaluation upon demand is sufficient for them to take part in the evaluation process. In this way different justificatory statuses are allowed for different beliefs in a the acceptance system. These justificatory statuses depend on their relation to the rest of the system (which is present but suppressed as context), rather than solely on the coherence relations among them. Consider the following example: If S doubts that it is going to rain, in the relevant acceptance system there are beliefs about the weather conditions, the forecast, general assumptions about the frequency of rain in a season, etc. They are supported by different contexts because each accepted belief was previously justified by a different relevant acceptance system. And if relevant acceptance systems tied to particular occasions of evaluation and belief revision differ with respect to their coherence properties, then beliefs will possess different justificatory merit: the ones that result from highly coherent sets will be better justified than the ones resulting from less coherent ones.

Thesis (2) explains the stability of perceptual beliefs in belief revision. The high justificational merit of perceptual beliefs can be explained by considering that their justification originates from highly coherent subsets. The relevant acceptance systems that justify perceptual beliefs contain other perceptual beliefs and are likely to be more coherent than the ones justifying theoretical claims. Perceptual beliefs are typically tested against other perceptual beliefs. What one sees in the end is confirmed by what one touches or smells. And it is because perceptual beliefs are members of highly coherent subsets in the network of knowledge that they are more stable in belief revision processes. Their justificatory merit need not even be considered during the evaluation process for this rule to apply. The subject considers them better justified because beliefs of a perceptual kind are typically corroborated by acceptance systems with a high level of consistency and coherence.

Thesis (3) explains some atomistic features of personal justification[9] without referring to foundational warrant. The attempt to resolve inconsistencies in the network of knowledge is more likely to change beliefs that are justified by less coherent acceptance systems. The rule to avoid inconsistencies explains the stability of perceptual beliefs in consonance with Quine's idea that changes in the total system of knowledge are guided by considerations of economy (Quine 1951, 40). In the case of an agent deciding to give up a perceptual belief about the concurrent weather conditions, she would have to give up a number of accompanying beliefs about her sensations and memories. The beliefs about the reliability of forecast reports are sustained by much less coherent subsets, ones that include more theoretical claims and isolated evidence. Gilbert Harman further explains that giving up what may be easily acquired again is avoided in belief revision.[10] If perceptual beliefs do occur spontaneously, then their acquisition will not be a one-time event if one persists in her distrust.

Why adopt such a complicated view if foundationalism provides a simpler explanation about the epistemic role of experience? One reason is that foundationalism would fail to explain the systematic features of our knowledge without a metaphysical commitment. The causal process that

[9] Like the possibility of being aware of and rationally holding inconsistent beliefs on the ground of their independent justification, while expecting to resolve inconsistencies later.

[10] Harman calls this rule of reasoning "Get back principle: One should not give up a belief one can easily (and rationally) get right back." (Harman 1986, 58)

produces a perceptual belief is independent of the rest of knowledge. Spontaneity and immediate justification would lead to a chaotic picture of perceptual beliefs. Justified acceptance must be responsible for the systematic connections.

The network account aims to show that inferential justification alone is sufficient to provide guidance in processes of belief revision, and that the stability of perceptual beliefs is a result of the structure of knowledge. The initial credibility of a single belief in a coherent set need not be derived from a "given element in empirical knowledge,"[11] but from coherentist properties outside that particular set in the network. Perceptual beliefs are highly credible because they are supported by the largest and most coherent subsets in the knowledge network. The only given element in experience is that it is most often coherent upon inferential justification without its beliefs being produced by inference.

A possible objection to the proposed account is that the high level of coherence among perceptual beliefs is insufficient to explain the high justificational merit. It is possible to establish a highly coherent subset of non-perceptual fictional beliefs, one that exceeds in coherence all others and thus becomes the most stable part in a knowledge system. Its stability would refute a number of perceptual beliefs on grounds of inconsistency. The isolation objection remains valid, but the question of whether or not we are systematically wrong in our beliefs or not is a skeptical objection that can address with equal stridency systems of beliefs as well as single beliefs with foundational justification. All it takes is the assumption that the processes of belief acquisition systematically provide us with false information. This objection does not remove the coherentist features of perceptual knowledge as a base for the systematization of beliefs.

References

BonJour L (1975) Coherence Theory of Empirical Knowledge. *Philosophical Studies* 1976-Nov 30(5): 381–312

BonJour L (1985) *The Structure of Empirical Knowledge.* Cambridge, MA: Harvard University Press.

BonJour L (2010) *Epistemology: Classic problems and contemporary responses*. Lanham: Rowman & Littlefield.

Bovens L, Hartmann S (2005) Why There Cannot Be a Single Probabilistic Measure of Coherence. *Erkenntnis* 63: 361–374

Harman G (1986) *Change in View. Principles of Reasoning.* Cambridge, MA: The MIT Press.

Huemer M (1997) Probability and Coherence Justification. *Southern Journal of Philosophy* 35: 463–472.

Lehrer K (1974) *Knowledge.* Oxford: Clarendon Press.

Lehrer K (1990) *Theory of Knowledge*. Boulder: Westview Press, Inc.

Lehrer K (1999) Justification, coherence and knowledge. *Erkenntnis* 50: 243–258

Lewis C I (1952) The Given Element in Empirical Knowledge. *The Philosophical Review A* 61 (2): 168–175

Lycan W G (2012) Explanationist Rebuttals (Coherentism Defended Again). *The Southern Journal of Philosophy* M 50(1): 5–20

Olsson E J (2005) *Against Coherence. Truth Probability and Justification*. Oxford: Clarendon Press.

Quine W V O (1951) Two Dogmas of Empiricism. *The Philosophical Review*. J 60 (1): 20–43

Siebel M (2005) Against Probabilistic Measure of Coherence. *Erkenntnis* 63:335–360

Thagard P (2000) *Coherence in Thought and Action*. Cambridge MA: MIT Press.

[11] See Lewis C 1952, 168–175

BALKAN JOURNAL OF PHILOSOPHY. Vol. 7, Issue 2, 2015

YET ANOTHER LOOK AT THE CONCEIVABILITY AND POSSIBILITY OF ZOMBIES

Maja Malec
University of Ljubljana,
Ljubljana, Slovenia
e-mail: malecmaja@yahoo.com

Abstract:

Since 1996, when David Chalmers introduced the zombie argument against physicalism in *The Conscious Mind*, numerous works of ever-increasing technical complexity and nuanced argumentation have been written on the conceivability and possibility of zombies. In this paper, I focus on the main points of the argument. First, I discuss the conceivability of zombies. I briefly outline three other thought-experiments in order to determine what is expected of a good thought-experiment and its workings. Next, I turn to Chalmers' defense of their conceivability, where the key consideration is to present conceivability as a credible a priori method that can entail metaphysical possibility. I conclude that Chalmers does not manage to create a credible link between conceivability and possibility, thus failing to show that zombies are not only conceptually possible, but also metaphysically possible. The most problematic idea is the identification of logically possible worlds with metaphysically possible worlds. Chalmers' main aim is to defend conceivability as an a priori method of acquiring modal knowledge, but by limiting it to the rational domain, the acquired knowledge is not knowledge of the objective reality, but of the content of our thoughts.

Key words: modal knowledge, conceivability, a priori, materialism, Chalmers.

The zombie argument was introduced by David Chalmers in his *The Conscious Mind* (1996) as his main argument against physicalism. If physicalism (or materialism)[1] is true, then all positive facts about the world are supervenient on physical facts, including facts about consciousness. However, if we can conceive a world, a physical duplicate of ours, in which there are no conscious experiences at all, and if we can, furthermore, show that such a world is not just a figment of our imagination, but a "real" possibility—that is, a "true" alternative to our world—then consciousness does not supervene on the physical, and physicalism is therefore false. Such a world would be a zombie world, inhabited by zombies who are our physical duplicates and function and behave exactly like us, but who are entirely devoid of conscious experience.

The main question here is whether zombies are conceivable as well as possible in such a way that their possibility refutes physicalism, a metaphysical thesis concerning the nature of our world. Chalmers answers affirmatively to both questions, and in his detailed and comprehensive argumentation focuses especially on the once-taken-for-granted step from conceivability to possibility which came under fire after Kripke's persuasive defense of a posteriori necessities (Kripke 1980). To disarm the danger that necessities which are knowable only through sense experience pose to conceivability as an a priori method of acquiring modal knowledge, Chalmers developed epistemic two-dimensional semantics, which he further improved over the years (Chalmers 2004; Chalmers 2010).

Numerous works have been written as a response to the zombie argument, works which are often very detailed and technical. Critics usually attack the thesis that conceivability entails possibility as Chalmers predicted, but there are some who question whether or not zombies are really conceivable. Chalmers has more clearly formulated the argument and provided answers to the main objections in *The Character of Consciousness* (2010: Ch. 6). The topic is obviously very popular, and it is literally impossible to read everything that has been written about it. It is also quite an expansive since it concerns the nature of the world, the nature of consciousness and its connection with the physical. The metaphysical aspect is closely

[1] Here I will follow Chalmers in understanding the terms "physicalism" and "materialism" as being interchange-able.

interwoven with the epistemological, conceptual, and methodological aspects (e.g., the nature of explanation). On top of this, the formulation of the argument within the two-dimensional semantic framework makes it very technical. In this paper I will try not to get entangled in technicalities and details—I will only present the main points of the argument. First, I discuss the conceivability of zombies. I consider Chalmers' thought-experiment in which he invites us to imagine our zombie twins and compare it with three other well-known thought-experiments. Next, I turn to Chalmers' defense of the conceivability of zombies, which relies on demonstrating that conceivability can be a respectable method of acquiring knowledge. The key issue here is that the method must be natural, but it must also be a guide to the right kind of possibility—metaphysical possibility. In the end, I assess how successful Chalmers is in this, and whether he manages to show that the conceivability of zombies entails their metaphysical possibility.

1. The Conceivability of Zombies

A zombie world is a physical duplicate of our world in which there are no conscious experiences. While physicalism claims that everything globally supervenes on the physical, i.e., that all the microphysical facts in the world (including fundamental microphysical laws) entail all other facts, it is reasonable to assume that consciousness is supposed to supervene locally. Thus, we can characterize the zombie world as a world which is inhabited by zombies, and in order to determine its conceivability we need to determine whether we can conceive a zombie: “someone or something physically identical to me (or to any other conscious being), but lacking conscious experiences altogether” (Chalmers 1996: 94).[2]

To make the notion more vivid and compelling, Chalmers proposes a thought experiment: whatever one is doing at the moment, they should try to imagine their zombie twin doing exactly the same thing but lacking conscious experience. For example, my twin zombie is at this point typing, listening to music, but unlike me is not feeling an aching sensation in her back. Everything will be the same, says Chalmers, imagining his zombie twin, only “[t]here will be no phenomenal feel. There is nothing it is like to be a zombie” (ibid: 95).

Let me try to imagine my zombie twin, then. By definition my zombie twin is my physical duplicate, embedded in an identical environment, functionally and psychologically identical to me. As it is a habit when conducting a philosophical thought experiment, I start with a pretty general idea—I imagine an environment identical to the one in which I am situated and put in my place a creature that is identical to me molecule for molecule, looks exactly like me and is doing exactly the same things as me, but lacks conscious experiences. I have no problem imagining such a situation, so I agree with Chalmers that the zombie hypothesis is prima facie coherent and imaginable (Chalmers 2010: 154), but does it persuade me that physicalism is false? No, it turns out that I am not quite sure how the fact that I can imagine a zombie relates to the falsity of physicalism. How does something I can imagine refer to what is the case in the world? Some further information is needed for the thought-experiment to work, but before I turn to Chalmers' exposition, here is a short reminder of how thought-experiments are being used in philosophy.

2. Comparison with Other Thought-Experiments

Testing theories against imaginary examples is an old and respected philosophical method of research. Philosophers develop them to support theories, but also to refute them. Probably the most famous thought-experiment in support of a theory is found in *Meditations* (1641/1996), in which Descartes, sitting by the fire, asks himself what he can be sure of, and by abstracting away things that do no pass the test, ends with the only thing he cannot doubt—that he is a thinking thing (*res cogitans*). Since this is distinct of his body (*res extensa*), he proceeds by developing an argument for substance dualism. In contemporary analytic philosophy, Gettier's imaginary counterexamples to the analysis of knowledge as justified true belief undoubtedly present the most successful refutation of a theory by use of thought-experiments (Gettier 1963). Also very influential was Kripke's use of modal arguments in his criticism concerning the descriptivist theories of proper names and natural kind terms (Kripke 1980).

Gettier and Kripke proceed in such a way that

[2] Chalmers adds that to disprove physicalism we actually only need conceivability of a world with one zombie amidst other normal conscious beings, or even a world with one invert, an individual who has a slightly different experience from the corresponding individual in our world (Chalmers 1996: 124 and 2010: 142). However, these details are irrelevant for the purposes of this paper.

they first present the theory they want to criticize, then they develop examples that the theory wrongly classifies. Gettier describes a situation in which a subject infers a proposition from another proposition. While the subject is justified in believing that the latter proposition is true, it is in fact false. However, the inferred proposition which the subject believes on this basis happens to be true, and the subject possesses justified true belief. According to the traditional analysis of knowledge, this means that the subject possesses knowledge as well, but Gettier's intuition, shared by most of philosophers today, is that it should not count as knowledge. From this it follows that justified true belief is not sufficient for—let alone equivalent to—knowledge. Similarly, Kripke, criticizing the descriptivist theory of proper names, invites us to imagine a situation in which the referent of a proper name does not satisfy the definite description by which he was picked out as a referent in our world. Would we in such a case think that he is the referent of the name, or would we pick out another individual who happens to satisfy the definite description in the counterfactual situation as a referent? Intuitively, argues Kripke, we would choose the first option, which shows that we understand proper names as rigid designators that pick the same objects in all situations. Since the descriptivist account suggests the second, counterintuitive option, it must be false. Admittedly, Kripke was not as successful as Gettier, but he persuaded quite a lot of philosophers.

What do these two examples have that the zombie example lacks? All are trying to show that some theory is wrong, but the theories in question are very different. Gettier is criticizing the analysis of knowledge that is expressed in one very short sentence, and it is thus very easy to comprehend. Kripke is assessing a semantic theory, namely a theory of naming, which he manages to expound on one page. It is true that some object that he misrepresents it, but this is a topic for some other time. What matters for this discussion is that the criticized theory is clearly stated, and we know precisely what the target is. In the case of zombies, criticism is directed towards physicalism, a metaphysical theory about the nature of our world. The main idea can be put very succinctly: all positive facts in our world supervene on physical facts. But when we start to fill in the details, matters quickly become complicated. What does the physical basis on which everything supervenes consist of, and how does it exactly relate to all the rest? What kind of relation is supervenience? Why must there be a conceptual link between the two classes? Maybe, because the theory is so complex, it is more difficult to decide whether or not the presented counterexample refutes it. But an even more important difference between the first two cases and the case against physicalism is their subject matter. Physicalism is a theory about a reality independent of mind and language, while in the case of knowledge and naming we are concerned with our dealings with this reality, so they are in part also dependent on us. As a result, we are more competent in judging whether the proposed thought-experiments reveal difficulties for the particular theory. What would I count as knowledge? How would I use a proper name? In the case of physicalism we are not so sure—is the zombie twin I am imagining really a physical duplicate of me? Of course I can imagine something along these lines, but is it just a figment of my imagination, or does the fact that I can imagine it really say something about the nature of our world? What is the nature of our world? Is everything in our world physical or is there also another ingredient—the phenomenal? Can the answer really come from the zombie argument?

Related to this is also the issue of how far-fetched the situations we are asked to imagine are. Chalmers admits that the idea of zombies is a strange one, and that it is unlikely that zombies are naturally possible, so the situation we are supposed to imagine is quite removed from our actual world (Chalmers 1996: 96). Accordingly, we are less sure about our imaginative abilities. Kripke's examples ask us to imagine counterfactual situations, some quite ordinary, like a situation in which Aristotle did not become a philosopher, or a situation in which Gödel did not prove the incompleteness of arithmetic. However, later in the book, when he discusses essentialism or develops the account of natural kind terms, counterfactual situations become less familiar. We are asked, for example, to judge whether Queen Elizabeth could have been born of different parents than the ones from whom she actually came, or whether water could have been composed of something other than H_2O. Judging these examples is more difficult because it is not enough to rely on our naming practices. We also need to consult our beliefs on the nature of things, what we consider as a legitimate change in a thing and what would bring about its destruction. Gettier's examples are on the opposite end of this scale of closeness to the actuality. We are not asked to imagine anything that could not have happened in our world, and the situations are very easy to imagine as a result.

The faculty of imagination is often accused of being too closely connected with our perceptual faculties, which are directed toward detecting happenings in the world, mostly limited to the level of middle-sized objects. In this way it is restricted in its abilities. Already Descartes has pointed out that it is unable to visualize extremely complex geometrical shapes (like chiliagons), while in assessing metaphysical issues it is completely useless. Descartes' solution is to distinguish between conceiving and imagining. Intellect, not the imagination, grasps the natures of things that according to him are the primary subject matter of metaphysics. These are what the thoughts of possibility and necessity are about, and something is judged to be possible if the mind clearly and distinctly conceives it.[3]

Thought-experiments, however, do not rely only on imagination. In constructing scenarios, words and concepts are used in addition to (sensory) images. In fact, a scenario often just means a verbal description of a situation. So, in conducting a thought-experiment we use imagination in a rationally constrained way. In fact, philosophers of a more rationalist bent treat imagination more like an occasionally helpful illustration, while the real work is done by conceiving a scenario and determining whether it is conceptually coherent. In this manner, Chalmers first invites us to imagine a zombie to "fix ideas," but then points out that it is crucial to determine whether the notion of a zombie is conceptually coherent or not (Chalmers 1996: 94–96).[4] So my ability to conjure up an image of my zombie twin serves as an illustration, while the majority of work is done through the analysis of concepts which are used in presenting the situation. It is conceivability which is the true guide to metaphysical possibility. So, the true question that I should be asking myself is not whether I can imagine my zombie twin, but rather whether I can conceive her. Is my zombie twin conceivable?

Chalmers suggests that, in order to determine this, I need to carefully describe the situation and then check that "there is no hidden contradiction lurking in the description." The procedure is no different than when we consider a mile-high unicycle. Undoubtedly, no such thing as a mile-high unicycle exists in a real world, but its description is coherent. Similarly, the analysis of a zombie scenario does not reveal any contradiction, thus showing that the described situation is coherent and the idea of a zombie conceivable (Chalmers 1996: 96).

I will admit that the zombie scenario seems conceivable, but it is not entirely clear to me how I reached this conclusion. The description is so short—a creature physically, functionally, and psychologically identical to me, embedded in an identical environment, minus conscious experiences—that there is not much chance of discovering any contradiction. Does this quick procedure suffice for me to confidently claim that my zombie twin is conceivable? And surely even if my zombie twin is indeed conceivable, it does not automatically mean that it is metaphysically possible? Does it?

3. The Characterization of Conceivability

Chalmers tackles this issue by further explaining the notion of conceivability. It is an epistemic notion, defined in epistemological and maybe also psychological terms. There are three dimensions of difference between the notions of conceivability: prima facie vs. ideal conceivability, negative vs. positive conceivability, and primary vs. secondary conceivability.[5] A sentence *S* is prima facie conceivable for a subject when *S* is conceivable for that subject on first appearances: after some consideration the subject finds that *S* passes the tests which are criterial for conceivability. This is the subject's initial reaction to the scenario, but this kind of conceivability is too weak, since it is tied to contingent cognitive limitations. A worry is that the person did not conceive the scenario in sufficient detail to discover a well-hidden inconsistency—as in the case when I worried that I came too quickly to the conclusion that my zombie twin is conceivable—and that maybe further reasoning would disprove this. What is needed here is an ideal conceivability for which one needs justification that cannot be rationally defeated, i.e.

[3] For more on the distinction between imagining and conceiving, as well as on Descartes' position, see "Introduction" in Gendler & Hawthorne (2002).

[4] In Chalmers 2010, he speaks of both conceivability and imaginability, but it is clear from the exposition that the outcome is achieved by rational reasoning.

[5] The three distinctions and consequent varieties of conceivability are stated in Chalmers (2010: 143–148), but for a detailed discussion, see Chalmers (2002).

ideal reasoning is required that cannot be defeated by any further reasoning.

However, the distinction between prima facie and ideal conceivability does not help me allay my doubts. My worry was that my reasoning was not good enough and that it could turn out that I did not conceive a real possibility after all. How can I know that my reasoning cannot be defeated by better reasoning? Have I done enough, or did I settle for prima facie conceivability of my zombie twin? According to Chalmers, although we are nonideal beings we still know many things that are ideally inconceivable (e.g., that 0 = 1), and that are ideally conceivable (e.g., that someone exists). We also know that certain things about the world are knowable a priori, but others cannot be known a priori even by an ideal reasoner. Similarly, in discussing consciousness we have good reason to think that there is no a priori entailment from physical to phenomenal truths, and that a zombie is conceivable even for an ideal reasoner (Chalmers 2010: 155).

Even so, one cannot simply infer from the fact that there are some or even many cases in which we do not have difficulties in judging that they are ideally conceivable or inconceivable that the case of a zombie is also one of them. It could be one of those cases in which it is difficult to determine that prima facie conceivability is not going to be defeated on better reasoning. Moreover, it is true that a distinction between the a priori and the a posteriori can be drawn, but can it really be claimed that we can know a priori that physicalism is false? Chalmers' example of an a priori knowable truth is that all philosophers are philosophers, and his example of an a posteriori knowable truth that there is the table in this room. To me, the issue of the truth or falsity of physicalism more closely resembles the issue of the presence or absence of the table in the room than the determination of the truth of the analytical sentence on the basis of its meaning.

Chalmers additionally distinguishes two ways in which we can determine whether something is conceivable, to which correspond two kinds of conceivability: negative and positive. For negative conceivability of *S*, it suffices if *S* is not ruled out, or if there is no (apparent) contradiction in *S*. For positive conceivability of *S* a subject must be able to coherently imagine a situation in which *S* is the case. Chalmers goes on to explain that “[a] situation is coherently imagined when it is possible to fill in arbitrary details in the imagined situation such that no contradiction reveals itself” (Chalmers 2002: 153). To positively conceive my zombie twin, then, I first need to be able to imagine a coherent situation in which everything is physically identical to a situation in which I am in at this moment, the only difference being that I have conscious experiences and my zombie twin does not. Second, I need to fill in as many arbitrary details as possible and try to see if the situation remains coherent. That is, I need to provide a more detailed description of my zombie twin's physical state to see if it is really identical with my physical state. This surely cannot just be assumed to be the case, since according to physicalism the description of my physical state should also describe my conscious experience. Accordingly, it is possible that my zombie twin must be missing some physical properties in order to be without consciousness, or maybe the whole idea is simply incoherent. Moreover, given that today science is not yet able to explain consciousness, maybe there will be further scientific discoveries and we will acquire new physical concepts, or the old ones will be improved and it will turn out that zombies are not conceivable.[6]

Chalmers responds to some objections to the conceivability of zombies in his *The Character of Consciousness* (2010: 154–157). It seems that in assessing the zombie hypothesis he does not rely so much on the comprehensive description of microphysical facts and on proving that there is no contradiction lurking around, but on his intuition that the physical and the phenomenal concepts are in principle so different that there is no way that the phenomenal concepts could apply to the same entities as the physical concepts.

Chalmers' rejection of physicalism and his zombie argument are informed by two ideas. One is the idea that modal knowledge is acquired through a priori means (modal rationalism), specifically by analyzing concepts (conceptual analysis). The second is that the metaphysical supervenience relation enables the reductive explanation which, to a great extent, proceeds by a priori means.

Now, if physicalism is understood as a metaphysical supervenience thesis, then to be true it must fulfill the following requirement: “for any logically possible world *W* that is physically indiscernible from our world, all the positive facts

[6] For example, Patricia Churchland claims that we are at this point unable to explain consciousness because neuroscience is still underdeveloped, and not because it is in principle unexplainable (Churchland 1997).

true of our world are true of *W*" (Chalmers 1996: 42). This also means that everything is reductively explainable in terms of microscopic physics, and that there is an entailment from the basic physical facts (fundamental entities and properties of completed physics, as well as the fundamental laws of physics) to all other positive facts, including phenomenal facts. This means that, in principle, if we possess phenomenal concepts, we should be able to read off phenomenal facts from the specification of microphysical facts (including the microphysical laws). In order to see how the application conditions of a phenomenal concept can be satisfied by fixing physical facts, we only need to be able to analyze what it takes for an entity to satisfy the intension of a phenomenal concept. But this we should be able to achieve, according to Chalmers, mostly a priori if we know the intensions of phenomenal as well as physical concepts, since intension is "a function specifying how the concept applies to different situations" (Chalmers 1996: 54). If we know both, we can determine their connection even without empirical information about our world. However, Chalmers claims that between the phenomenal and the physical there is no such conceptual link, and that there is an explanatory gap between the physical level and conscious experience; namely, physical properties are best explained as structural or dynamical properties, but phenomenal properties are not suited for such explanation, and therefore physics in principle cannot explain consciousness. Physicalism must be false.[7]

This conviction of Chalmers is also behind his denial that it may turn out after some new scientific discoveries that zombies are not conceivable after all. He points out that it is highly unlikely that the structural-dynamic character of physics and physical concepts would be changed by new scientific developments, so the phenomenal concepts should already today have a character that would link them appropriately with physical concepts, but this is not the case. The only vaguely possible candidate would be functional concepts. This brings me to my second objection that one cannot just assume that my zombie twin's physical state is identical to mine, but needs to be argued for. The view that explains consciousness in wholly functional terms is called analytic functionalism.[8] According to this view, to say that someone is conscious only means that they have a state that functions in an appropriate way in their cognitive system and in their behavior (Chalmers 2010: 155). This means that by definition, any functional duplicate of mine is also conscious. On this view zombies are inconceivable, and more generally, there is no epistemic gap between physical and phenomenal truths. It also means that explaining consciousness simply means explaining the various functions. But, as Chalmers argues, there is a strong intuition that there is something further to explain, an intuition which is shared by the majority, so functional analysts owe us very strong arguments to oppose it. They do not provide them, so their view should be rejected (ibid: 111–114).

Analytic functionalism is, in my opinion, suffering from the same disease as Chalmers' project: it relies too heavily on conceptual analysis. How do we get from knowledge of our concepts to knowledge of mind-independent reality, from conceptual to metaphysical? Our concepts are often inadequate to the true nature of things; just take the example of water. Moreover, we sometimes have concepts of things that do not exist after all, for example phlogiston. How can conceptual analysis show that physicalism, the substantive thesis about the nature of our world, is false? How can conceptual analysis on its own tell us what is metaphysically possible and what is metaphysically necessary? Surely this depends on the very nature of things in the world and not only on our conceptual scheme.[9] To what degree do our concepts mirror the objective reality? Moreover, there is a persistent feeling that our cognitive limitations crucially determine what we are going to count as possible. By distinguishing between prima facie and ideal conceivability, Chalmers tries to reduce the dependence of the method on the ability of a thinker, but the distinction seems to me to be too abstract to produce correct results. Concerning the issue of the application of concepts, Chalmers denies that they are provided by a definition (at least not in a majority of cases), but rather by intension, which gets determined from the detailed consideration of scenarios. We presuppose that a certain scenario obtains and then ask ourselves what we would say in that case. By considering a certain number of scenarios, we notice certain regularities and are thus able to determine the extensions of concepts. But what kind of scenarios—a hypothetical situation that

[7] For the exposition on why physics in principle cannot explain phenomenal consciousness, see especially Chalmers 1995.

[8] See, for example, Lewis (1966) and Armstrong (1968).

[9] For a similar position, see Lowe (2013).

might turn out to be true (a conceptual or epistemic possibility), or that might have been true (metaphysical possibility)?

4. The Link from Conceivability to Possibility

The most common objection in the literature is the objection to the conceivability-possibility entailment, and it draws especially from the existence of a posteriori necessities that were persuasively argued for by Kripke (1980). The identity statement “Water is H_2O” is the most often cited example of a posteriori necessity, namely metaphysically necessary truth that is knowable only a posteriori. A posteriori necessities present a threat for modal rationalism, for the a priori route to modal knowledge. Take the example of water. If we know the identity of water with H_2O only a posteriori, i.e., only after the scientists discovered its chemical structure, then first of all, we cannot come to know this necessity only by a priori means, and moreover, we can coherently conceive many situations concerning water which do not represent metaphysical possibilities. For example, we can conceive a situation in which water has some different chemical structure, say XYZ, which makes it conceivable that water is not H_2O, yet it is not metaphysically possible. From this it follows that conceivability is not a very good guide to possibility. So why would we think that from the conceivability of zombies follows their metaphysical possibility?

Chalmers tackles the problem of a posteriori necessities by distinguishing two kinds of conceivability—the primary one and the secondary one, which are linked to two different kinds of possibility. The distinction is based on the fact that within two-dimensional semantics there are two ways in which we can evaluate a sentence across possible worlds. In nontechnical terms, the distinction in the case of positive conceivability could be expressed as follows: one can imagine a situation as actual—as a way the world might actually be—and on this basis consider if it follows that *S* is the case. If it does, then *S* is primarily conceivable. The extensions of the involved concepts are fixed through their primary intensions. On the other hand, one can imagine a situation as counterfactual—as a way the world might have been. If the imagined situation is coherent and *S* is the case, then *S* is secondarily conceivable; concepts are associated by secondary intensions (Chalmers 2002: 156–158).

Primary conceivability is a reliable guide to primary possibility, and secondary conceivability is a reliable guide to secondary possibility. Primary possibility is a familiar conceptual possibility, but Chalmers also calls it epistemic possibility since we consider a particular situation solely on the basis of our a priori knowledge, suspending our empirical knowledge as it were. Accordingly, primary conceivability is always an a priori matter (what might conceivably be the case) and is the relevant type for modal rationalism. Secondary conceivability is linked to secondary possibility; Chalmers also calls it subjunctive possibility (what conceivably might have been the case). This is a familiar metaphysical possibility. When we evaluate a particular situation from the perspective of secondary possibility, we acknowledge that the character of the actual world is fixed, and ask: if the situation had obtained, would *S* have been the case? Since the application of words in situations considered counterfactually often depends on their reference in the actual world, secondary possibility is often a posteriori.

If we apply this distinction to the case of the sentence “Water is not H_2O,” we get the following result: a priori determined primary intension of a concept fixes reference in the actual world. The primary intension associated with *water* is something like “watery stuff” or “the clear, drinkable liquid in our environment.” The secondary intension is determined a posteriori and picks out reference in counterfactual worlds. The secondary intension associated with *water* is “ H_2O.” Chalmers does not say specifically, but H_2O seems to have the same primary and secondary intension: “ H_2O.” Now, if we evaluate the sentence according to the primary intensions involved, the primary intensions of *water* and H_2O differ, so it is primarily conceivable and primarily possible that water is not H_2O. However, when evaluated according to the secondary intensions, the sentence is not secondarily conceivable or possible since the secondary intensions of *water* and H_2O are the same.

For modal rationalism secondary possibility is clearly not acceptable, since in many cases the evaluation is possible only after a priori knowledge is supplemented with empirical information about the actual world. The central role is therefore given to primary conceivability, which is entirely determined a priori, and as a result constitutes a reliable guide to the primary possibility. The primary intensions of involved concepts are used in this way in the zombie argument also, and the demonstrated possibility of a zombie is therefore the primary possibility. But aren't secondary intensions required here, since they are the ones that ultimately lead to metaphysical modality? As we have seen in the case of water, if evaluated according to the primary intensions the result is not metaphysical, but rather an epistemic possibility.

Chalmers argues that although primary conceivability does not entail metaphysical possibility, there still is a link between the two, since when evaluating sentences—whether associated with primary or secondary intensions—we are evaluating them across the same space of possible worlds (Chalmers 2010: 146). As he says, the distinction between primary/epistemic and secondary/metaphysical possibility "is not a distinction at the level of *worlds*, but at most the distinction at the level of *statements*" (Chalmers 1996: 68). Thus, the world that is primarily conceivable and therefore epistemically possible (a way that a world might actually be), is also metaphysically possible (a way the actual world is not but could have been). For example, when we imagine a Twin Earth situation in which watery stuff in the oceans and lakes is XYZ, we have access to a metaphysically possible world. It is the same case when we conceive a zombie world using the primary intensions—it is a metaphysically possible world. Furthermore, "the primary intension determines a perfectly good property of objects in possible worlds." For example, there is nothing wrong with the property of being watery stuff, as well as with the property that is picked out by the primary intension of a phenomenal concept. Consequently, argues Chalmers, "[i]f we can show that there are possible worlds that are physically identical to ours but in which the property introduced by the primary intension is lacking," then physicalism is still proven to be false (Chalmers 1996: 132).

Chalmers explains his solution in more detail, especially in *The Character of Consciousness* (2010: "The Two-Dimensional Argument against Materialism"), but this basic outline should suffice. Despite careful exposition of the argument within the framework of two-dimensional semantics, my main question is still unanswered: how can conceiving of something be sufficient to establish the possibility of that something? As we have seen, Descartes believed that when conceiving, the intellect is revealing the essences of things which are external to our minds. If we can clearly and distinctly conceive something, that thing is metaphysically possible. While he too rashly judged the intellect to be infallible, I believe he was on the right track when he identified the essences of things as the source of metaphysical necessity and possibility.[10] Conceivability can be a guide to metaphysical possibility only if it manages to become something more than just knowledge of our concepts. Chalmers' suggestion that we determine the extension of concepts by considering their use in various hypothetical situations could be a way to latch on to something more objective, and in this way to shake off the curse of seemingly superficial and hopelessly mind-dependent concepts. However, his epistemic two-dimensional semantics—in which epistemic possibilities correspond to metaphysically possible worlds—is actually an attempt to subsume metaphysical modality under the conceptual.

The idea that epistemic and metaphysical modalities are being evaluated across the same space of logically possible worlds is counterintuitive.[11] These are two distinct kinds of modality that derive from different sources. Epistemic modality—which is usually referred to as conceptual, but is also logical modality in a broader sense—is concerned with what is possible or necessary given our concepts and the connections among them. Chalmers focuses more on the apriority of such truths, the fact that we can know them without consulting empirical information, and accordingly calls them epistemic. An epistemic possibility is, as Chalmers puts it, "a specific way the actual world might turn out to be, for all one can know a priori" (Chalmers 2010: 216). Metaphysical modality, on the other hand, is not modality concerning our concepts, but the mind-independent reality. It is about what could or could not have been the case, given how our world in fact is. Modality concerns the world itself, which we take in its totality. We do not limit ourselves to the content of our concepts and wonder what kind of worlds could go with them, if we varied empirical data. We start with all possible knowledge we have

[10] In contemporary metaphysics, the view that modality is grounded in essence and that metaphysical modality should be characterized in terms of essences or identity of things and not the other way around is persuasively defended by Kit Fine (1994). In modal epistemology, Hale (2013) and Lowe (2012) develop accounts of modal knowledge as being derived from essentialist knowledge.

[11] Vaidya similarly claims that the identification of metaphysical modality with logical modality leads to the counterintuitive view that the sources of modality are the same (Vaidya 2008: 198).

about our actual world and wonder in what ways it could have been different and in what ways it could not have been. If in the first case we ask ourselves, for example, what is our concept of water, then in the second case the question is what is the essence or nature of water, that which makes it the thing it is. Clearly, to conceive a world on the basis of our conceptual knowledge cannot mean that we determined one way our world could have been. Intuitively, we see a clear distinction between the two possibilities. Chalmers says that what connects both possibilities is the fact that they are both "ultimately grounded in what it is rationally coherent to suppose" (Chalmers 2010: 187). Sure, we certainly employ our reasoning when we are determining both kinds, but as I said, they are grounded in two distinct domains, and the metaphysical kind cannot be limited to the rational domain, as Chalmers suggests, since it concerns the nature of the independent reality.

Chalmers goes on to argue that all modal phenomena are best analyzed within one modal space, and that there is no reason to postulate two separate spaces of worlds. He considers also the possibility that one could try to introduce metaphysically possible worlds as a subset of logically possible worlds, and proposes a few counterexamples to the strong necessities which would go with such a construction of a modal space (Chalmers 2010: 166–184). There is another option he does not consider: a unified modal space of metaphysically possible worlds and conceptual modality as a restricted kind of metaphysical modality. Such a view is proposed by Kit Fine, who identifies the source of metaphysical necessity in the identity of all objects and sees conceptual necessity as a restricted kind of metaphysical necessity, since the source of its truth is the identity of concepts, a more restricted class of objects (Fine 1994: 9).

As we have seen, the thought-experiments against a certain theory are successful if they persuade us that the theory wrongly explains a certain example by appealing to our intuitions. The main problem with the zombie argument is that the idea of a zombie is counterintuitive. A zombie world is a complete physical duplicate of our world, but instead of conscious beings it is inhabited by zombies. A zombie is a complete physical duplicate of a conscious being, but lacks conscious experience; it nevertheless functions and behaves exactly like the latter. Now, if a zombie and a zombie world are indeed metaphysically possible, then this is a problem for physicalism as well as dualism. Physicalism is false and one needs to adopt some kind of dualism. However, while Chalmers may be proven right and a nonphysical explanation of consciousness is needed, there is a new awkward thing to be explained: why do we have consciousness at all? If a zombie functions and behaves exactly like us, and if it does not reveal any deficiency, then what is the point of having consciousness at all? It seems to be something we can easily do without, which goes against what we firmly believe.

References

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

Chalmers D J (1995) Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies* 2: 200–219.

Chalmers D J (1996) *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

Chalmers D J (2002) Does Conceivability Entail Possibility? *Gendler & Hawthorne* 2002: 145–200.

Chalmers D J (2004) Epistemic Two-Dimensional Semantics. *Philosophical Studies* 118: 153–226.

Chalmers D J (2010) *The Character of Consciousness*. New York, Oxford University Press.

Descartes R. (1641/1996) *Meditations on First Philosophy*. Edited by J. Cottingham. Cambridge: Cambridge University Press.

Churchland P S (1997) The Hornswoggle Problem. In: *Explaining Consciousness: The Hard Problem*, ed. J. Shear, 37–44, Cambridge, Mass: MIT Press.

Fine K (1994) Essence and Modality: The Second Philosophical Perspectives Lecture. *Philosophical Perspectives* 8: 1–16.

Gendler T S, Hawthorne J (eds.) (2002) *Conceivability and Possibility*. Oxford: Clarendon Press.

Gettier E. (1963) Is Justified True Belief Knowledge? *Analysis* 23: 121–123.

Hale B. (2013) *Necessary Beings: An Essay on Ontology, Modality, and the Relations Between Them*. New York: Oxford University Press.

Kripke S. (1980) *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.

Lewis D (1966) An Argument for the Identity Theory. *Journal of Philosophy* 63: 17–25.

Lowe E J (2012) What Is the Source of Our Knowledge of Modal Truths? *Mind* 12: 919–950.

Vaidya A (2008) Modal Rationalism and Modal Monism. *Erkenntnis* 68: 191–212.

BALKAN JOURNAL OF PHILOSOPHY. Vol. 7, Issue 2, 2015

WHAT IS LEARNED FROM CONDITIONALS?

Borut Trpin
University of Ljubljana, Faculty of Arts,
Ljubljana, Slovenia
e-mail: borut.trpin@gmail.com

Abstract:
Some of the information that we learn comes to us in a conditional form. This has proven to be a problem for philosophers, who try to explain how probabilistic beliefs change when one learns from conditional sentences. The problem is that a straight-forward solution is not possible: the partial belief in the antecedent and the partial belief in the consequent either increase, decrease, or remain the same. Two existing approaches to learning from indicative conditionals are considered: an explanatory one, and another that builds on relative information minimizing with regard to the causal structure. A novel method based on epistemic entrenchment is proposed to overcome the drawbacks of the competing approaches. The method solves all the standard examples and some other examples for which existing approaches have failed to provide adequate solutions.

Key words: indicative conditionals, formal epistemology, epistemic entrenchment, Bayesian networks, explanation.

1. Introduction

Some of the information that we learn comes to us in form of conditionals. For instance, we may learn that a picnic in the park will be cancelled if it rains. We learn, at the same time, that the picnic will also probably be cancelled if the wind will be very strong, or if it will rain heavily just before the picnic. Although we have no problems learning from such conditionals, it is very hard to describe how learning from conditionals proceeds in general.

In this paper, I propose an approach to the problem of learning from conditionals. My proposal may be summed up in three main ideas:

(1) When an agent learns something from a conditional, her posterior degree of belief in either the antecedent or the consequent remains the same as her prior degree of belief (*). Posterior degree of belief in the other does, however, increase or decrease. (*) is thus epistemically more entrenched.

(2) A conditional may convey implicit information which needs to be made explicit during the analysis.

(3) After taking (2) and supporting beliefs into consideration, a simple entrenchment test can determine:

a) whether the posterior degree of belief in the antecedent or the posterior degree of belief in the consequent remains unchanged, and
b) whether the posterior degree of belief in the other increases or decreases.

The most pleasing analysis of learning from conditionals would be to apply the standard Bayesian procedure for belief updating—conditionalization. Let me formally sketch the idea behind standard conditionalization (SC):

An agent starts with some degree of belief in A and a conditional degree of belief in A given B, represented by conditional probability $Pr(A|B)$. After learning B, i.e. $Pr'(B) = 1$, her posterior belief in A changes to her (prior) conditional probability $Pr(A \mid B)$:

$$Pr'(A) = Pr(A|B)^{1} \qquad \text{(SC)}$$

Jeffrey's conditionalization (JC; also known as probability kinematics) provides a generalization of (SC) for cases when new information is not (necessarily) learned with full certainty. Formally:

$$Pr'(A) = Pr(A|B)Pr'(B) + Pr(A|\neg B)\, Pr'(\neg B) \qquad \text{(JC)}$$

It is obvious from the equation that (SC) is just a special case of (JC) when $Pr'(B)=1$.[2]

To paraphrase David Lewis (1976), alas, the most pleasing solution cannot be right. I will discuss why (JC)—and hence (SC)—cannot provide the

[1] Prior degrees of belief are denoted by Pr(-), and posterior degrees of belief by Pr'(-).
[2] Suppose $Pr'(B) = 1$. Then $Pr'(\neg B) = 0$ and: $Pr'(A) = Pr(A|B)\, Pr'(B) + Pr(A|\neg B)\, Pr'(\neg B) = Pr(A|B)*1 + Pr(A|\neg B)*0 = Pr(A|\mathrm{B})$, which is exactly the (SC).

correct results (at least not generally) in the next sections. But let me first clarify the main concepts that will be used throughout the paper.

2. Theoretical background

2.1 Types of conditionals

When I refer to *conditionals*, I refer to standard non-nested indicative conditionals. The indicative conditional operator will be represented by an arrow (→). In the case of other conditionals, i.e. material, subjunctive, left- or right-nested conditionals, the distinction will be explicitly mentioned.

But what exactly are indicative conditionals and how do they differ from other types of conditionals?

Broadly speaking, an indicative conditional is a conditional that has both the antecedent (the if-clause) and the consequent (the main clause) in indicative mood. As noted, I only consider standard indicative conditionals for present purposes. The (indicative) mood is a grammatical feature**,** and as such also includes some non-standard types of conditionals. I will not consider the following non-standard types of indicative conditionals:

- speech act conditionals
- biscuit conditionals
- Dutchman conditionals
- even if conditionals
- independent conditionals
- counterfactual conditionals

Speech act conditionals are the cases in which the antecedent conditionally modifies not the contents of the consequent, but the speech act that the consequent carries out (Dancygier and Sweetser 2005). The typical examples are conditional tips, conditional threats, conditional questions, conditional promises, and conditional commands (Krzyżanowska 2015, p. 61). Let me list a few examples which will clear up why speech-act conditionals are excluded for present purposes.

Conditional threat: “If you don't shut your mouth, I will smack you.”

Conditional question: “Does the light turn on if I press this button?”

As the dependence between the antecedent and the consequent is non-inferential, there is nothing to be learned from them. When Tom says to Sarah: "If you don't shut your mouth, I will smack you," Sarah learns that Tom is angry, but she doesn't learn anything about her continuing with her talking or about her being smacked, at least not in the standard way. The mechanics of learning from speech-act conditionals thus remain an open issue.

Biscuit conditionals are another non-inferential type of indicative conditionals. They are called “biscuit conditionals” after a typical example from Bennett (2003): “There are biscuits on the table if you are hungry.” The biscuits are on the table regardless of the recipient's hunger. Biscuit conditionals are therefore non-standard, as they could easily be analysed in an equivalent non-conditional form, i.e. “Feel free to eat the biscuits,” and as such are not important for our analysis of learning from conditionals.

Dutchman conditionals (Hájek 2012, p. 146) are named after their typical example: “If you manage to do this, I am a Dutchman.” They are used to express a high degree of doubt in the antecedent and use an even more doubtful (or impossible) consequent as the means to convey the doubt in the antecedent. They are not conditional in their nature as they express non-conditional content, i.e. disbelief or doubt in the antecedent, and not the dependence of the consequent on the antecedent.

Even if conditionals are a special type of conditional which is easily recognized by the adverb “even.” They behave in very specific ways and will therefore be excluded.[3]

Independent conditionals (Bennett 2003, p. 149) are conditionals in which there is no real dependence between the antecedent and the consequent, e.g. “If I go outside, Africa is a continent.” Nothing may be learned from such examples, although strange conditionals in which we do not notice any connection between the clauses may suggest that there is some unexpected connection between them (Douven 2008, p. 25).

The last type of indicative conditional that will be excluded is the counterfactual conditional. Although counterfactuals are often thought to coincide with the subjunctive mood, it may be better to think of them as of the conditionals that have an antecedent that is assumed to be false. As (Krzyżanowska 2015, p. 10) pointed out, one could think Denmark is likely not ruled by a king and still assert: “If Denmark is ruled by a king, then it is a kingdom.” Learning from counterfactuals is a very interesting problem, but it is too complex for this paper.

---

[3] The adverb *even* is not always present in the even-if conditionals, which can lead to some puzzling consequences. For example, „(Even) If she dyed her hair, she didn't dye it blue.” Suppose the consequent is false. Modus tollens then does not hold:
She dyed her hair blue.
∴ She didn't dye her hair.

2.2. Semantics of conditionals

But what is it that makes indicative conditionals so special? One of the main problems is that it is very hard to pin down the exact semantics of indicative conditionals. But the problems do not end even when a system of semantics is clearly described, because all systems of semantics for conditionals lead to epistemological problems (Douven 2013, p.4).

There are three main candidates for a semantics of indicative conditionals: the material analysis (MA), the Stalnaker-style possible worlds analysis (PW), and the non-propositional (or, in the weaker form, the tripartite) analysis (NPA or TPA).

(MA) is the traditional approach to conditionals.[4] It is historically attributed to ancient Greek philosopher Philo of Megara (Sanford 1989, p. 14–25), who defined the semantics of (MA) in 4–3 BC.

The semantics of (MA) are very simple: a conditional ($p \supset q$) is false if and only if the antecedent (p) is true and the consequent (q) is false.

This view was prevalent throughout most of the history (in classical logic) and is still prevalent in mathematics and even in some traditions of psychology of reasoning (e.g. mental logic and mental models). While (MA) is very useful in deductive proofs in mathematics, it has some very unintuitive consequences if we try to asses everyday conditionals by it. The conditional is equivalent to the following disjunction:

$$p \supset q \leftrightarrow \neg p \vee q$$

This means that the conditional is true if its antecedent is false or the consequent is true, which leads to paradoxical examples, e.g.:

1) If Europe is not a continent, then Europe is a continent.
2) If Italy is a country, then 2+2=4.
3) If Berlusconi never existed, then Italy is not a country.

1) is true because both the antecedent is false and the consequent is true.
2) is true because both clauses are true, although there is no connection between them.
3) is true because the antecedent is false (Berlusconi exists), although it is, again, highly unassertible.

A Stalnaker-style possible world semantic of conditionals (PW) is a much better attempt to capture the epistemological relation between the antecedent and the consequent. The main idea of (PW) is that the conditional (regardless of being indicative or subjunctive) is true if and only if the consequent is true in the closest possible world to the one in which the antecedent is true *and* given that there is a possible world in which the antecedent is true. Although (PW) has some major advantages over (MA)—i.e. it escapes the typical paradoxes of (MA)—it has its problems.

For instance, Gibbard's river-boat paradox (Gibbard 1981) demonstrates that two conditionals with the same antecedent and opposite consequents ("If p, then q" and "If p, then ¬q") could both be true in the closest possible *p*-world. Further, the conditionals in which both clauses are true—despite there being no connection between them—are also considered true in (PW). For instance, "If 2+2=4, then Italy is a country" is true, because Italy is a country in the closest world where 2+2=4 is true (the actual world).

Let me now turn to the non-propositional analysis of conditionals (NPA). (NPA) is the most popular semantics amongst theorists of conditionals. According to this view a conditional is true if both the antecedent and the consequent are true, and false if the antecedent is true and the consequent false. In a case in which the antecedent is false, the truth of the conditional is undecided (Table 1).

	p	***q***	***p → q***
1	1	1	1
2	1	0	0
3	0	1	1 or 0 (undecided)
4	0	0	1 or 0 (undecided)

Table 1: Truth table for conditionals according to (NPA)

Although (NPA) captures many of our intuitions, the non-propositional nature of conditionals has a few undesired consequences. If conditionals do

[4] The material conditional operator will be represented by the horse-shoe (⊃).

not express propositions, how can we believe[5] them? I very strongly believe that salt will dilute *if* I mix it with water. Furthermore, if the semi-truth functionality (table 1) of conditionals is accepted, then conditionals could not occur in Boolean combinations, as their truth cannot always be determined (Douven 2013).

I embrace (NPA) in this paper regardless of its limitations. The proposed analysis of learning from conditionals does not require any Boolean combinations. Although I will inspect partial beliefs,[6] (NPA)'s semi-truth functionality still provides valuable guidelines. With some appropriate simplification, if an agents thinks a proposition is more likely to be true (false) than false (true), her degree of belief in this proposition is over (below) .5.

Supposing the newly learned conditional is not misleading, an agent should thus either have degrees of belief in both the antecedent and the consequent over .5 (probabilistically generalised line 1 of Table 1) or both under .5 (to prevent probabilistic cases of line 2).

3. Learning from conditionals

It may be clear from the previous sections that the problems with learning from conditionals are at last partially triggered by the very complex nature of conditionals. There are almost no common assumptions among philosophers researching conditionals. As Krzyżanowska (2015, p.2) noted, "it seems that any time a new account arises, a refutation arises sooner or later."

The situation is even worse when it comes to learning from conditionals. The problem is not just that all approaches have counterarguments, but also that learning from conditionals has not been given much attention. Only a few (somewhat general) approaches to (uncertain) learning from conditionals exist.[7]

3.1. Updating on standard material conditionals

Let us start with updating on standard material conditionals (UMC). One of its main strengths—the fact that it is very straightforward—is also its main limitation. After all, according to (UMC), after the agent learns a conditional her posterior belief in the antecedent should always *decrease* if she is uncertain about the antecedent and the consequent:

(UMC PAR)

Suppose $Pr(antecedent) > 0$ and $Pr(consequent) > 0$. According to (UMC) it then follows that $Pr'(antecedent| \ antecedent \supset consequent) \leq Pr(antecedent)$.
If additionally $Pr(antecedent) < 1$ and $Pr(consequent) < 1$, it follows that $Pr'(antecedent| \ antecedent \supset consequent) < Pr(antecedent)$.[8]

It is very hard to believe that my degree of belief in tomorrow's rain should decrease after learning: "If it rains tomorrow, the picnic in the park will be cancelled." (UMC) is thus not adequate to explain learning from conditionals.

3.2. Adams conditioning and explanatory analysis of learning from conditionals

Another approach is based on the so-called Adams conditioning, as proposed in Douven and Romeijn (2011). They developed their view in response to the so-called Judy Benjamin problem (originally in van Fraassen 1981). They started with a general rule for updating conditional degrees of belief—dubbed Adams' conditioning (AC)—and adapted it for conditionals. As (AC) is just a special case of Jeffrey's conditioning (JC), their approach is actually just a special case of (JC), with which we can analyse learning from conditionals of the form: "If A, then the odds for $B_1, \ldots, B_n$ are $c_1, \ldots, c_n$".

One of the biggest drawbacks of this account is that the posterior degree of belief in the antecedent always remains the same. While this seems fine for specific cases like the Judy Benjamin problem (which will be discussed later), it is very limited and gives the impression that it is tailored for specific examples.[9]

This is also the reason why Douven later (2012) provided a more general account of learning from conditionals based on the explanatory relation between the antecedent and the consequent (EXP). The main idea is quite simple. When an agent learns a

5 The standard definition of "belief" is an attitude towards a proposition.

6 It should be noted that the exact relation between truth and probability is not clear. Rational agents believe tautologies fully (with $Pr(\top) = 1$) and disbelieve contradictions ($Pr(\bot)=0$), but it is not clear how to assess partial beliefs.

7 More research has been done in the realm of AGM theory and computer science. These analyses are, however, only concerned with categorical beliefs (Douven 2012, p. 239).

8 For proof, see Douven and Romeijn (2011), p. 645.

9 Refer to Lukits (2015) for further refutation of their approach.

conditional $p \to q$, she sets her conditional probability close to 1, i.e. $Pr'(q|p) \approx 1$. If there is no change in the explanatory status of the antecedent, Adams conditioning is applied. If the explanatory status goes up, $Pr'(q) \gtrsim Pr'(p) > Pr(p)$, and if the explanatory status goes down, $Pr'(q) \gtrsim Pr'(p) < Pr(p)$.

One of the main drawbacks of (EXP) is that is not clear how to determine whether the explanatory status of the antecedent goes up or down (if at all). Research of explanatory status in Bayesian epistemology is, however, increasing, and it is reasonable to expect that (EXP) will have firmer ground once more research on explanatory status is done.

There are, however, some other crucial drawbacks of (EXP). There are some cases where (EXP) excludes results which are intuitively acceptable. The same problems occur for relative information minimizers (INFOMIN). I will next examine how (INFOMIN) approaches the problem of learning from conditionals.

3.3. Minimising Kullback-Leibler divergence and causal representation

When van Fraassen introduced the Judy Benjamin problem in 1981, he demonstrated that (INFOMIN) leads to intuitively strange results.[10] I will discuss the Judy Benjamin problem later, but first I will examine what (INFOMIN) actually is and how a combination of (INFOMIN) and Bayesian causal nets provides a very robust approach (INFOMIN+CAU) (Hartmann and Rad, unpublished). (INFOMIN) is based on Kullback-Leibler divergence $D_{KL}(P'||P)$, which measures how much information is lost when moving from a prior probability distribution to a posterior probability distribution. Updating then proceeds through finding the minimum of divergence.[11]

Although (INFOMIN) yields correct results in many cases, it leads to very strange, although not necessarily wrong (Lukits 2015) results in some cases—for instance, in the Judy Benjamin problem. Hartmann and Rad combined (INFOMIN) with Bayesian causal nets and developed a technically complex but very robust model of updating on conditionals (INFOMIN+CAU). For instance, they provide scenarios with additional disabling conditions and demonstrate how their model still gets the expected results.

Their approach, however, has at least two problems:

1) Their application of (INFOMIN) depends on properly represented Bayesian causal nets. They do not provide any rules about how to determine when a representation is proper. As I will demonstrate later, a slight (and intuitively correct) change of structural representation may result in wrong results.

2) Similarly to (EXP), (INFOMIN+CAU) runs into problems in cases where both a change in degree of belief in the antecedent and the consequent appar feasible.

4. The examples

Let me first list the standard examples[12] of learning from conditionals. This will show the strength of both (EXP) and (INFOMIN+CAU). I will then show how both accounts fail to provide (all) possible results if the examples are slightly modified.

(1) The Sundowners Example

Sarah and her sister Marian are planning to have sundowners at the Westcliff hotel tomorrow. Sarah thinks there is some chance it may rain tomorrow, but thinks they could still enjoy the view from inside. To make sure, Marian makes a telephone call to the staff at the hotel. They tell her that in the event of rain, a wedding party will occupy the inside area. So she tells her sister: "If it rains tomorrow, we can't have sundowners at the Westcliff hotel." Sarah now thinks there is no chance for sundowners at the hotel in rain. Her probability for tomorrow's rain, however, remains the same. Learning the conditional thus leaves the probability of the antecedent unchanged.

(2) The Judy Benjamin Problem

Judy Benjamin, a soldier in training, and her platoon are dropped in a swampy area which they have to patrol. The war games area is divided into the region of the Blue Army, to which Judy Benjamin and her fellow soldiers belong, and that of the Red Army. Each of these regions is further divided into Headquarters Company Area and Second Company Area. The patrol has a map, which none of

[10] Van Fraassen's conclusion is ambiguous. it is not clear whether he embraces or ridicules the unintuitive consequences of (INFOMIN).

[11] $D_{KL}(P'||P) := \sum_i \frac{P'(i)\ln(P'(i))}{P(i)}$. For further technical details of (INFOMIN), refer to Hartmann and Rad (unpublished) or van Fraassen (1981).

[12] Examples are adapted from: example 1 from Douven and Romeijn (2011), example 2 is from van Fraassen (1981), examples 3 from Douven and Dietz (2011), example 4 from Douven (2012) and example 5 from Douven and Romeijn (2011).

them understands, and they are soon hopelessly lost. Using their radio they are at one point able to contact their own headquarters. They are told by the duty officer “I don't know whether or not you have strayed into Red Army territory. But if you have, the probability is 3/4 that you are in their Headquarters Company Area.” At this point the radio gives out. Supposing Judy accepts this message, how should she adjust her degrees of belief?

It seems intuitively that her degree of belief that she is in the Red Army Territory (the antecedent) should not change.

(3) The Ski Vacation Example

Harry sees his friend Sue buying a skiing suit. He is surprised because he did not know she had plans to go skiing. He knows she had an important exam a few days ago and thinks it is unlikely she passed. Then he meets Tom, his best friend and also a friend of Sue, who is just on the way to Sue to ask whether she passed the exam. Tom tells Harry: "If Sue passed the exam, her father will take her on a ski vacation." Harry recalls that Sue was buying a skiing suit, and now comes to think that it is more likely that she passed the exam. Learning the conditional thus increased his probability of the antecedent (Sue passed the exam).

(4) The Driving Test Example

Betty knows that Kevin, the son of her neighbours, was to take a driving test yesterday. She has no idea of his driving skills; she thinks it's as likely as not that he passed the test. She notices his parents have just started spading their garden. Sue's mother, who is friends with Kevin's parents, later tells her: “If Kevin passes the driving test, his parents will throw him a garden party.” Sue recalls that spading has just begun and comes to think that it is very unlikely (although not fully excluded) that they will have a garden party. As a result, she thinks it is less likely that Kevin passed the test. Learning the conditional thus decreased her probability of the antecedent (Kevin passed the driving test).

(5) The Deadly Robber Example

A jeweller has been shot in his store and one of his golden watches was stolen. However, it is not clear at this point whether the two events are related. It is possible that somebody first shot the jeweller and another person took the opportunity to steal the golden watch. Kate knows her friend Henry has financial troubles and frequently walks past the store. She thinks it is likely that he is the thief, but she highly doubts he would shoot the jeweller. After the investigation of the scene, policemen conclude that a single person committed the crime. A detective comes to Kate and tells her: "If Henry stole the watch, then he also shot the owner." As a result, Kate now thinks it is very unlikely he stole the watch. Learning the conditional thus decreased her probability of the antecedent.

The examples thus show that learning a conditional may influence the antecedent in three different ways: one's degree of belief in the antecedent may (i.) increase, (ii.) decrease, or (iii.) remain the same.

I will first quickly examine whether the mentioned approaches to learning from conditionals get the intuitively expected results.

The Sundowners example

According to (EXP), after learning the conditional, i.e. setting the conditional probability $Pr(\neg sundowners|rain)\approx 1$, the explanatory status of the antecedent does not change, and Adams' conditioning is applied. Applying (AC) always results in unchanged probability of the antecedent, so (EXP) gets the intuitively expected results (Douven 2012).

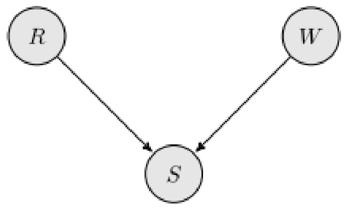


Figure 1: The Bayesian network for the sundowners example (Hartmann and Rad)[13]

(INFOMIN+CAU) also leads to expected results, although in a slightly more complex way. Hartmann and Rad, for instance, propose the following Bayesian network (Figure 1), which in their words “properly represents the causal representation [of the variables]”. They then make sure that all the events are probabilistically dependent with regards to the network, adapt probabilities to the story from the example, and minimise the Kullback-Leibler divergence to get the correct results.

One of the objections I hold against their analysis of the Sundowners example is that in their collider

[13] *R* denotes rain, *W* the wedding party and *S* the sundowners at the hotel.

model (Figure 1), the variable *W* is independent of *R*, although Marian is told from the hotel staff that the wedding taking place inside the hotel depends on the rain. To be fair, Hartmann and Rad actually define *W* as a binary variable with two values, W: "There is a wedding party" and ¬W: "There is no wedding party" Thus defined *W* is obviously independent of *R*.

But is the correct result still achieved if we define *W* as a variable with two values, W: "There is a wedding party inside the hotel" and ¬W: "There is no wedding party inside the hotel" and add the dependence of *W* on *R* (Figure 2)?

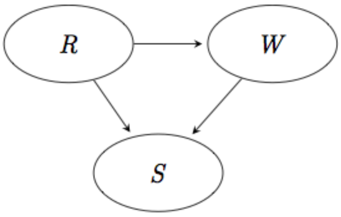


Figure 2: Alternative Bayesian network for the sundowners example

It turns out that if the Bayesian network is represented in this way, and if the Kullback-Leibler divergence is minimized, Sarah's posterior probability of tomorrow's rain decreases after learning the conditional "If it rains tomorrow, we can't have the sundowners at the hotel."[14] This does not refute (INFOMIN+CAU), but rather shows that more detailed criteria on how to properly represent situations is needed.

The Judy Benjamin Problem

(EXP) proceeds similarly in the Judy Benjamin Problem, although the probability of the conditional is not close to 1, but .75. The reason for a similar procedure is that learning the conditional "If you are in the Red Army area (R), the odds are 3:1 that you are in their Headquarters Area (RH)," i.e. setting $Pr(RH|R)=0.75$, does not change the explanatory status of the antecedent. This means that Adams' conditioning is applied and the posterior probability of being in the red area remains the same, .5.

The results of (INFOMIN+CAU) are similar, although the procedure is, again, more complex. According to Hartmann and Rad, there is no direct causal connection between being in the Red Army area and being in their Headquarters Area. The only possibility is, therefore, that both events have a common cause *X*. It is not necessary to determine what the common cause is, as it does not have any influence on their procedure, which after minimising the Kullback-Leibler divergence, again it yields the expected results.

The Ski Vacation and The Driving Test examples

The Ski Vacation (SV) and The Driving Test (DT) examples intuitively require the agents to adjust their degrees of belief in the antecedent in two opposite ways: either to increase it (SV) or decrease it (DT). Both examples, however, turn out to be deeply related on a structural level.

According to (EXP), learning the conditional makes the explanatory status of the antecedent (Sue passed the exam) go up in (SV), while it goes down (exactly the opposite) in (DT). Posterior degrees of belief in the antecedent in the two examples thus adjust in two opposite directions: it increases in (SV) and decreases in (DT).

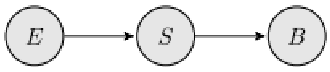


Figure 3: The Bayesian network for the ski vacation example

Structural similarity in the two examples is also noted in (INFOMIN+CAU). Hartmann and Rad thus propose structurally identical models: the event from the consequent depends on the antecedent, and the additional observed event further depends on the consequent. They get the expected results after minimising the Kullback-Leibler divergence. While the proposed network for (SV) (Figure 3) seems obviously appropriate, the proposed network for (DT), on the other hand, does not. It is not hard to imagine that spading the garden prevents the garden party, especially considering it temporally precedes the garden party. Such representation of the situation would then result in a collider network (like in Figure 1) and, likely, in wrong results. An examination of the consequences of such a Bayesian network is omitted because of its complexity.

The Deadly Robber Example

The Deadly Robber Example (DR) was originally provided as an example of epistemic entrenchment, which was not further developed into a general approach to belief updating on conditionals. It is, however, not hard to envision how (EXP)

[14] Proof omitted because of its complexity and length.

would solve (DR) and get the correct results. After learning that "If Henry stole the watch, he also shot the owner," the explanatory status of the antecedent goes down, and as a result, Kate finds it less likely that Henry was the culprit.

It is not clear what the appropriate representation of the problem would be in (INFOMIN+CAU), although it seems reasonable that both clauses would have a common cause. The problem would then be similar to the Judy Benjamin Problem, although it is intuitively not acceptable that the degree of belief in the antecedent remains unchanged.

As I have shown, both (EXP) and (INFOMIN+CAU) provide intuitively expected results in all the examples. But look at what happens in the modified driving test example:

(6) The Modified Driving Test Example

Betty knows that Kevin, the son of her neighbours, was to take a driving test yesterday. She knows he has been illegally driving since he was 12 and <u>thinks it is very likely that he passed the test</u>. She notices his parents have just started spading their garden. Sue's mother, who is friends with Kevin's parents, later tells her: "If Kevin passed the driving test, his parents will throw him a garden party." Sue has two possibilities, depending on the strength of her belief that Kevin passed the test: she could either decrease her degree of belief in the antecedent (like in the original example), or keep her belief in the antecedent unchanged and increase her degree of belief in the consequent, i.e. come to think that the garden party will take place despite spading.

According to both (EXP) and (INFOMIN+CAU) it is, however, only possible that Sue decreases her degree of belief in the antecedent. Problems of this nature motivated my proposed approach, which seems to overcome them.

5. Learning from conditionals revised

As is clear from the previous cases, when one learns a conditional and actually learns something new, then she keeps, increases, or decreases her prior degree of belief in the antecedent. But the antecedent only provides half of the story. In the case that the probability of the antecedent increases or decreases, the probability of the consequent remains unchanged. In the case that the probability of the antecedent remains the same, the probability of the consequent increases or decreases. There are therefore not just 3 possibilities (ant. unchanged, increased, or decreased), but 4.

After learning "If *P*, then *Q*," the degrees of belief in the antecedent and the consequent may adjust in the following 4 ways:

Learning "If *m*, then *n*."	The degree of belief in the antecedent *m*	The degree of belief in the consequent *n*
a)	does not change: $Pr'(m)= Pr(m)$	increases: $Pr'(n)> Pr(n)$
b)	does not change: $Pr'(m)= Pr(m)$	decreases: $Pr'(n)< Pr(n)$
c)	increases: $Pr'(m)> Pr(m)$	does not change: $Pr'(n)= Pr(n)$
d)	decreases: $Pr'(m)< Pr(m)$	does not change: $Pr'(n)= Pr(n)$

Table 2: Possible cases after learning from conditionals

As is clear from Table 2, strength of belief in one of the clauses always remains unchanged (antecedent in lines a) and b); consequent in c) and d)). The clause which retains the same strength of belief is epistemically more entrenched. My aim is to provide guidelines that would enable a determination of which of the clauses is more entrenched.

5.1 Epistemic entrenchment

But what is epistemic entrenchment (EE) after all? The term originates from AGM belief revision theory, although it can also be applied to partial beliefs. Douven and Romeijn (2011), for instance, claim that Kate from the Deadly Robber Example finds it easier to give up her partial belief in Henry's theft because her (partial) belief that he is not a killer is more entrenched.

I claim that one of the clauses is always more entrenched than the other, i.e. degree of belief in this clause remains unchanged. But why and when is one clause more entrenched than the other? I will try to answer this in the next section.

5.2 What can we learn from the examples?

When one is told something in the conditional form, she does not just (i) *explicitly* learn that the consequent will occur in the event of the antecedent, but (ii) (in some cases) also *implicitly* learns something about other events.

The Sundowners Example serves as a good illustration of implicit learning from conditionals: when Marian tells Sarah that they cannot have sundowners if it rains tomorrow, Sarah implicitly infers that the inside area of the hotel will be occupied. Otherwise Marian's conditional would not make sense, as it is clear from the story that she initially thinks they could have their sundowners inside in the case of rain. This observation then plays an important role in the actual learning from the conditional.

Another observation that can be taken from the examples is that learning from conditionals depends on how coherent the newly learned conditional is with not just previous (partial) beliefs in its clauses, but also with other beliefs that support them.

For instance, Betty from the driving test example has strong *support* for her belief that the garden party is unlikely (she observed the spading of the garden). Her belief about Kevin's (non)success with the test is, on the other hand, not supported at all, because she has no knowledge about his driving skills.

Finally, which belief is more entrenched depends on its strength. Strength of belief does not coincide with the degree of belief, although it is determined by it. I define strength in a simple and straightforward manner: beliefs of degree .5 (it is as likely as not) are very weak. Strength of beliefs with degrees lower or higher than .5, on the other hand, symmetrically increases.

This, again, can easily be illustrated with the driving test example. Betty's prior degree of belief in Kevin's success was exactly .5 (she thought it was as likely as not that he passed the test), and thus had no strength, while her belief that it is unlikely a garden party will take place was very strong, and hence more entrenched.

5.3. The method

These three observations are also at the core of the presently proposed model of learning from conditionals.

Step 1: When one learns "If *M*, then *N*," she explicitly learns that *N* follows from *M*, i.e. *N* can be inferred from *M*.

Step 2: Check if the conditional conveyed any implicit information. The simplest way to do so is to answer the following question from the point of view of the learning agent: Could the consequent follow from the antecedent alone? If not, what other conditions would also need to be fulfilled? These implicit additional conditions are thus also learned.

Step 3: How (if at all) are beliefs in the antecedent and the consequent supported? Further, if the supporting beliefs of the consequent are in conflict with the antecedent, its supporting beliefs, or additional implicitly learned conditions, then this casts doubt on them (in the context of the conditional), and their strength is reduced to zero.

If the implicitly learned additional conditions are, however, in conflict with the consequent or its supporting beliefs, this similarly casts doubt on them and reduces their strength to zero.

Step 4: Determine the degree of belief in both clauses with regards to support and then determine their strength, with the following simple entrenchment function:[15]

$$S(m) = \begin{cases} -2Pr(m) + 1, Pr(m) \in [0, .5) \\ 2Pr(m) - 1, Pr(m) \in [.5, 1] \end{cases}$$

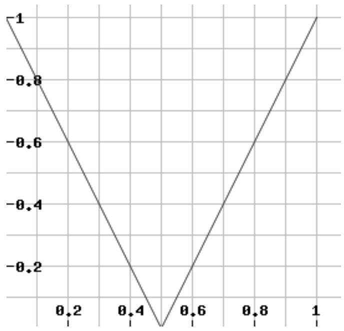


Figure 4: Graph of the simple entrenchment function

Step 5: Compare the strength of the antecedent and the consequent. This leaves three possibilities:

i.) $S(ant) > S(cons)$

ii.) $S(ant) < S(cons)$

iii.) $S(ant) = S(cons)$

In (i) the antecedent is stronger (hence, more entrenched), and thus remains unchanged after learning the conditional, in (ii) the consequent is more entrenched, and in the rare event of (iii) the antecedent is more entrenched (*).

Step 6: I now know which belief remains unchanged. Whether the other increases or decreases depends on their prior degrees of belief:

If $S(ant) > S(cons)$,

then if $Pr(ant) > .5$,

$Pr(ant) \geq Pr'(cons) > Pr(cons)$

Else, i.e. $Pr(ant) < .5$,

$Pr(ant) \leq Pr'(cons) < Pr(cons)$

[15] The range of strengths, i.e. interval [0,1], was motivated by probabilities, which are spread out in the same range. The function could be more generally defined as $S(m) = \begin{cases} -x\,Pr(m) + \frac{x}{2}, Pr(m) \in [0, .5), x \in \mathbb{R} \\ x\,Pr(m) - \frac{x}{2}, Pr(m) \in [.5, 1], x \in \mathbb{R} \end{cases}$

If $S(ant)<S(cons)$,

then if $Pr(cons)>.5$,

$$Pr(cons)\geq Pr'(ant)>Pr(ant)$$

Else, i.e. $Pr(cons)<.5$,

$$Pr(cons)\leq Pr'(ant)<Pr(ant)$$

In the case in which beliefs in both the antecedent and the consequent are of the same strength (*), the antecedent is entrenched, as the agent explicitly learns that the consequent depends on the antecedent.

5.4. The Examples Revisited

Let me now demonstrate how the proposed method works "in action." I will show that it leads to intuitively expected results not just in the examples from the literature, but also in modified examples where other approaches fail, e.g. the modified driving test example.

The Sundowners Example

What does Sarah learn when her sister tells her "If it rains tomorrow (*R*), we can't have sundowners at the hotel. ($\neg Su$)"?

She first explicitly learns that ($\neg Su$) can be inferred from (*R*). She then considers whether (*R*) entails ($\neg Su$). It does not: Sarah thinks they could still have sundowners inside the hotel if it rains. Sarah thus abductively infers that the inside area of hotel will not be available.

Further, her low degree of belief in ($\neg Su$) is supported by her background knowledge that they can have the sundowners either outside or inside the hotel *and* that both areas are very rarely occupied. But Sarah also implicitly learned that the inside area of the hotel will not be available if it rains (or at least one area otherwise). This is in conflict with her support for low degree of belief in ($\neg Su$): she now thinks they can only have the sundowners outside. Strength of her belief in ($\neg Su$) is reduced to 0 and her $Pr(\neg Su) = .5$

As $S(R)\geq S(\neg Su) = 0$, the posterior degree of belief in (*R*) remains unchanged. If she thinks rain is unlikely (although not impossible), her posterior $Pr'(\neg Su)$ lies in this range:

$$Pr(R)=Pr'(R)\leq Pr'(\neg Su)<Pr(\neg Su)=.5$$

If, however, Sarah thinks it's likely that it will rain, she increases $Pr'(\neg Su)$:

$$Pr(\neg Su)<Pr'(\neg Su)\leq Pr(R)= Pr'(R).$$

The results are intuitively correct.

The Judy Benjamin example

When Judy learns that "If you are in Red Army area, the odds are 3:1 that you are in their Headquarters Area," she does not learn anything implicitly. Her beliefs for being either in the Red Army area or being in the Headquarters Area (or Red Headquarters area, given that she is in the Red Army area) have no support, and hence no strength: Judy and her platoon are lost and think it's as likely as not that they are in any quadrant of the equally divided area (Figure 5).

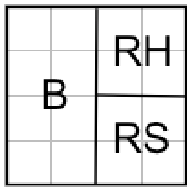


Figure 5: Judy Benjamin's initial mental map

As both clauses have the same strength, the antecedent is more entrenched. If the conditional information was provided with odds 1:0, i.e. "If you are in the Red Army area (*RH* v *RS*), then you are in their Headquarters Area (*RH*)," she would increase $Pr(RH)=.25$ to $Pr'(RH)=.5$. As the conditional is only partial (with odds 3:1 or, equivalently, probability of .75), this needs to be taken into account, so: $Pr'(RH)=.5*.75=.375$. As (*R*) is entrenched, $Pr'(R)=Pr(R)=.5$. This is also what one would expect.

It must be noted, though, that this may not be the only possibility. As Lukits (2014) pointed out, intuitions are not always correct. Judy Benjamin has more possibilities, depending on which of her prior beliefs is entrenched.

She starts with prior beliefs $Pr(R)=Pr(B)=.5$ and $Pr(RH)=Pr(RS)=.25$.

If she entrenches her belief in (*R*), then the above results are correct. But she could also entrench her belief in (*RH*) and still update according to the conditional:

$$Pr(RH)=3/12,\ Pr(RS)=1/12,\ Pr(B)=8/12.$$

In this case, her degree of belief in the antecedent would decrease from .5 to 1/3.

She could also entrench her belief in (*RS*):

$$Pr(RS)=1/4,\ Pr(RH)=3/4,\ Pr(B)=0.$$

In this case the probability of the antecedent (*R*) would increase to 1.

The odds for being in (*RH*) are 3:1 in comparison to (*RS*) in all of the above possibilities, and it is not impossible to imagine cases when this could be true.[16] For example, the last possibility could make sense if the blue area is fully occupied by Red Army, and thus does not exist any more. It is not clear why the radio operator would then use a conditional, although it is a possibility.

All of the alternative possibilities, however, depend on supporting beliefs which are not available to Judy, but rather to the radio operator. Judy thus correctly entrenches (*R*). But if she had different supporting beliefs, her belief updating wouldn't proceed in the same way. My proposed account is thus resistant to Lukits (2014) counterexamples.

Ski vacation example

When Tom tells Harry "If Sue passed the exam (*E*), her father will take her skiing. (*Sk*)," Harry doesn't learn anything implicitly, as (*Sk*) could follow from (*E*). His degree of belief in the success of her exam is low, which is supported by her past results (known to Harry), her testimony, or by other sources. It is important to note that it is not in conflict with Harry's support for partial belief in (*Sk*).

Which belief is stronger—(*Sk*) or (*E*)—depends on how "extreme" they are. Does Harry think it is more likely that Sue's father will take her skiing because she was buying a skiing outfit, or that it is more *un*likely that she passed the exam? Both options seem possible:

1.) If Harry thinks Sue had almost no chance of passing the exam, and thinks that it is somewhat likely that she was just wishfully buying her skiing outfit, then $S(E)>S(Sk)$ and as $Pr(E)<.5$:

$$Pr(E)=Pr'(E)\leq Pr'(Sk)<Pr(Sk)$$

2.) If Harry initially thinks it's very likely that Sue's father will take her skiing since she was buying the skiing outfit, and thinks it's only somewhat likely that she didn't pass the exam, then $S(E)<S(Sk)$ and as $Pr(Sk)>.5$:

$$Pr(E)<Pr'(E)\leq Pr'(Sk)=Pr(Sk)$$

The results of our method are thus in line with intuitions, but also cover the special case, which is not covered either by (EXP) or (INFOMIN+CAU).

Driving test example

When Betty learns "If Kevin passed his driving test (*T*), his parents will throw him a garden party (*G*)," she doesn't learn any implicit information—passing the driving test would suffice for Kevin's parents to throw him a garden party.

[16] For further discussion refer to Lukits (2014).

Betty initially doesn't have any support to believe he actually passed the test, so her $Pr(T)=.5$ and $S(T)=0$. She has strong support to think a garden party is unlikely (she noticed the spading of the garden had just begun), so her $Pr(G)<.5$, and thus $S(T)<S(G)$. As $Pr(G)<.5$:

$$Pr(G)=Pr'(G)\leq Pr'(T)<Pr(T) \qquad \text{(DT-1)}$$

As we expected, she learns that it is unlikely that Kevin passed the test.

If Betty, on the other hand, has strong support for (*T*), and hence $Pr(T)>.5$, as in the modified example, then everything would depend on which belief is stronger, i.e. whether (i) $Pr(T)$ is closer to 0 or (ii) $Pr(G)$ to 1.

In case (i) $S(T)>S(G)$ and as $Pr(T)<.5$:

$$Pr(G)<Pr'(G)\leq Pr(T)=Pr'(T)$$

Betty now comes to think a garden party is likely despite the spading.

In case (ii), the results are the same as in (DT-1) above. Both results seem to be in line with intuitions.

The Deadly Robber Example

I conclude the demonstration of the proposed method with the deadly robber example. When Kate learns "If Henry stole the watch (*W*), he also killed the owner (*K*)," she implicitly learns that the culprit was a single person (*SP*). Otherwise she couldn't infer (*K*) from (*W*).

Kate thinks (*W*) is likely. This is supported by her belief that Henry had financial troubles and that the thief was not necessarily also the killer. The additional condition, (*SP*), is in conflict with her supporting belief for (*W*), which reduces the strength of (*W*) to 0. Her belief in (*K*) is very low, which is supported by her knowledge of her friend Henry and his character. Either way, $S(K)>S(W)=0$.

As $Pr(K)<.5$:

$$Pr(K)=Pr'(K)\leq Pr'(W)<Pr'(W)$$

The result is again in line with intuitions. This example is interesting, because the implicitly learned information undermines the strength of the antecedent. Exactly the reverse was the case in The Sundowners Example.

6. Conclusion

My proposed approach is far from being the final answer to the problem of learning from conditionals. After all, I limited myself to simple conditionals, and it is not clear what would happen in the case of nested conditionals. This needs to be considered in the future.

Another aspect that deserves more attention and is usually overlooked in the discussion of updating on conditionals is their social nature. One can learn something from her perceptions, reasoning, introspection, or testimony.[17] However, all of the conditionals from the examples were learned through testimony, and as such all the examples also touch on questions of social epistemology. For instance, why should the recipient of the conditional trust the sender?

Does this reveal something about the epistemological nature of conditionals? Do I learn anything if I come to know about conditionals from my own reasoning? One could explain this phenomenon (away) by claiming that all the conditionals from the examples were testimonial because otherwise the stories would be, mildly said, modernistic streams of thoughts, or simply boring. This seems a bit rash. I will leave this question open—as it is not of direct relevance—and simply emphasize that (based on the examples) learning from conditionals seems to be essentially a product of testimony.

Another important question that was excluded is how learning from counterfactuals proceeds. After all, counterfactuals are important in philosophy of science and science in general. For instance, causal relations are often described with counterfactuals: "If *A* had not occurred, *B* would still occur." One learns that B is causally independent of A, but the procedure is obviously much different from our proposed approach. After all, neither *A* nor *B* are more likely after we get to know this counterfactual.

It would also be insightful to see how people psychologically learn from conditionals, especially in cases where multiple outcomes seem possible (for instance, in the "normal" and modified driving test example and in the ski vacation example). This would allow for a check of whether the empirical data supports this method.

It is thus clear that my proposal is still in the early phase and has a lot of space for growth in different directions, but still provides (intuitively) correct results not just in the cases already addressed by other approaches, but also in some cases where other accounts do not provide the expected results.

References

Bennett J (2003) *A Philosophical Guide to Conditionals*. Oxford: Clarendon Press.

Dancygier B, Sweetser E (2005) *Mental Spaces in Grammar: Conditional Constructions*. Cambridge: Cambridge University Press.

Douven I (2008) The evidential support theory of conditionals. *Synthese* 164(1): 19–44

Douven I (2012) Learning conditional information. *Mind and Language* 27(3): 239–263

Douven I (2013) The epistemology of conditionals. In *Oxford Studies in Epistemology* eds T S Gendler, J Hawthorne. Oxford: Oxford University Press.

Douven I, Romeijn J (2011) A new resolution of the Judy Benjamin problem. *Mind* 120: 637–670

Douven I, Dietz R (2011) A Puzzle About Stalnaker's Hypothesis. *Topoi* 30:31–37

Gibbard A (1981) Two recent theories of conditionals. In eds W L Harper, R Stalnaker, G Pearce, 211–247. Dordrecht: Ifs. Reidel.

Hájek A (2012) The fall of "Adams' thesis"?. *Journal of Logic, Language and Information* 21(2): 145–161

Krzyżanowska K (2015) Between if and then. Dissertation, University of Groningen.

Lewis D (1976) Probabilities of conditionals and conditional probabilities. *The Philosophical Review* 85(3)

Lukits, S (2015) Maximum Entropy and Probability Kinematics Constrained by Conditionals. *Entropy* (17): 1690–1700

Sanford D H (1989) If P, then Q. New York: Routledge.

Van Fraassen, B C (1981) A Problem for Relative Information Minimizers in Probability Kinematics. *The British Journal for the Philosophy of Science* 32(4): 375–379

[17] Memory is omitted as it merely retains knowledge. Although there are counterarguments to this, the epistemological nature of memory does not interest us for present purposes.

BALKAN JOURNAL OF PHILOSOPHY. Vol. 7, Issue 2, 2015

BETWEEN EXPLANATION AND UNDERSTANDING. THE OVERVIEW OF AN INVESTIGATION INTO THE FUNCTION OF CONTRADICTION IN THE EXPLANATORY MODELS OF THE HUMAN SCIENCES

Dragoş Popescu
Institute of Philosophy and Psychology,
Romanian Academy, Romania
e-mail: dragosgp@yahoo.com

Abstract:
The present paper argues that although the problem of contradiction is particularly important within the domain of human sciences, a close examination of its function within the main methodological options has not yet been undertaken. Therefore, the aim of this paper is to provide an overview of a possible investigation into the function of contradiction in explanatory models of human sciences.

Key words: explanatory model, causal/teleological explanation, hermeneutic circle, communicative action, contradiction.

Introduction

The starting point of this paper is the observation that where the human sciences are concerned, the old dispute regarding method persists. Its origins can be traced back to the 19th century, and, in certain respects, back to Kant. Are the sciences of the spirit (*Geisteswissenschaften*) autonomous from a methodological point of view, or is there a unique method which is viable both for the sciences of nature (*Naturwissenschaften*) and those of spirit?

While this paper subjects both the arguments in favour of a methodological monism and those proclaiming the autonomy of the sciences of the spirit to a critical examination, the purpose of this project is not to support one or the other of the existing methodological options, but to emphasize the implications of these options for the function of contradiction in both monistic and dualistic theories. Although the problem of contradiction is particularly important within the domain of the human sciences, a close examination regarding its function within the main methodological options has not yet been undertaken. The research has mainly focused on either highlighting the narrow significance of contradiction in the case of the option supporting the uniqueness of the scientific method, or the unacceptable consequences of an uncontrollable widening of its significance in the dualistic case.

The hypothesis supporting this project is that a contextualization of the problem of contradiction could transform the monism–dualism dispute into a constructive dialogue. Such a tendency can already be noticed. The last few decades' research on the issue of contradiction within both logic and philosophy of language offers strong conceptual and technical support for a reassessment of the reciprocal exclusion which has characterized the relation of the monistic and dualistic options.

Although the approach to examining this issue taken here is philosophical, the results of research in this direction go beyond philosophical territory. The problem of contradiction has wide implications for both the domain of human sciences and its applications. Since terminological oscillations and imprecise uses of related concepts can often be encountered within the applicative attempts in the domain, a conceptual clarification and an integrated model of the functions of contradiction would result in an increased rigour of applicative approaches.

Monism, dualism, and pluralism

Where the human sciences are concerned, the present time has been widely characterized as the epoch of methodological pluralism *par excellence*. Still, the diagnosis might be premature: once subjected to a close analysis, this characterization so far mostly expresses intentions and not realities. This is because at the root of the numerous methodological innovations (which often proved themselves to be fertile) within both the theoretical reconstructions yielded by the natural sciences and those construed within the sciences of spirit, the classical dilemma still persists: explanation or understanding (cf. Hollis, 2000, p. 22). An authentic methodological pluralism would have transformed the disjunction into a conjunction.

Historically and systematically speaking, painstaking research has been done with regard to explanation and understanding. Attempts to reduce and hierarchize, or—on the contrary—to deepen the opposition between explanation and understanding are numerous. We can gather them under three fundamental types of methodological options:

a) The primacy of explanation over understanding (which sometimes goes as far as to dissolve the latter within the former) is the position specific to methodological monism, with origins going back to 19th century logical positivism;

b) Setting up understanding as the fundamental human attitude which precedes the birth of explanatory sciences (not just in a temporal sense), which is a position specific to a methodological dualism whose main purpose is the preservation of the autonomy of the sciences of the spirit in the context of the rapid development of natural sciences (which may be regarded as a confirmation of their methods).

c) Constituting explanation and understanding by starting with the distinction between technical and practical interests, which is a position that attempts to avoid both the difficulties of monism and those of methodological dualism by employing conceptual offerings of Hegelian dialectics (before the methodological schism of the second half of the 19th century).

The first option, which subordinates understanding to explanation in differing degrees, has its origin within 19th century positivism (Auguste Comte, John Stuart Mill). From this very early stage, methodological monism argued in favor of unity of method in the context of a diversity objects under scientific investigation. It also assumed that the evaluator's methodological standard for the development of all sciences is the one given by the natural sciences, especially physics, and it gave the explanation that individual cases must be subordinate to general laws of nature (causal explanation) (von Wright, 1995, p. 27).

The second phase of methodological monism—initiated by the great developments within logic in the beginning of the 20th century—is the one still in place. It contains the Covering Law Model, or the "subsumption-theoretic model of explanation of action" (von Wright, 1995, p. 33).

The second option, which affirms the fundamental and irreducible character of understanding, may be seen as a reaction against the positivist attempt to exclude all the disciplines not fulfilling the methodological standards imposed by positivism from the corpus of sciences. A special part is played by Wilhelm Dilthey. "Erleben" or "Erlebnis" as complex experience, "Erlebnisausdruck" as "Erlebnis" (objectivation), and "Verstehen" as a reduction of the manifestations of life to their origins are separated from the explanation as the reduction of a process to natural law.

Starting with Dilthey, the tradition methodologically distinguishes between explanation and understanding as conditions of conceiving experimental science (Riedel, 1989, p. 39). Passing through Heidegger's contribution, this tradition reaches up to Hans-Georg Gadamer, for whom the autonomy of the sciences is the result of their free choice, expressed in the methodological delimitations constituted by the object (Hufnagel, 1981, p. 74).

Last, but not least, the third option often evokes Hegelian dialectics, though it retains some distance from some of the fundamental Hegelian standpoints. It has its origins in the "critical theory" of the Frankfurt School, and has been reshaped by Jürgen Habermas. Apart from rejecting the final identity between subjectivity and objectivity and adopting materialism, the critical theory argues that the understanding of contemporary society is not a mathematical science of nature—which is a perpetual logos—but the critical theory of a given society (cf. Marga, 1988, p. 190). As a consequence, Habermas puts forth a concept of understanding which presupposes the consent of all members of a linguistic community regarding the rightness of an assertion according to a normative foundation they recognize (Marga, 1988, p. 272).

In the postmodern context, all the above-mentioned fundamental methodological options have entered an acute process of dialogue, or to put it otherwise, a process of hybridism. The theories of science and hermeneutics, logic and the methodologies of human sciences are all profoundly involved in this process. One can still notice, though, that within any theory regarding the relation between explanation and understanding, apart from assuming or rejecting some marginal issues on the matter, one of the three methodological options can still be recognized as foundational.

In order to simplify this investigation, I will lay aside their different origins in disciplines like theory of science, hermeneutics, logic, etc., and instead focus on gathering theoretical reconstructions characterized by the same methodological option within an explanatory model.

Explanatory Models in Human Sciences: General Coordinates of the Investigation

I understand an "explanatory model within human sciences" to be more than a theory which embraces one of the above fundamental methodological options. An explanatory model not only supports—from a theoretical point of view—the primacy of explanation over understanding, establishes understanding as a fundamental attitude, or envisages these two concepts as utterly independent, but also manages to reconstruct the entire domain. An explanatory model manages to define the theoretical status of both natural and human science by making use of:

(1.) specific concepts, which

(2.) thoroughly describe the domain of every group of sciences,

(3.) by justifying their claims with regard to knowledge (based on a theoretical perspective on knowledge and truth), and

(4.) at the same time, defining their limits.

Thus, an explanatory model has conceptual resources, argumentative procedures, and methods of investigation through which

(5.) it can establish a dialogue on a set of positions which other explanatory models do not share.

Hence, as a consequence, though numerous divergences regarding the relation between explanation and understanding may exist among different explanatory models, an argument within one explanatory model cannot be solely aimed at invalidating argumentation pertaining to a different explanatory model. To accomplish this aim, a comparative analysis of the conceptual resources, argumentative procedures, and methods of investigation would first be necessary; in other words, this would presuppose the initiation of a process of semantic homogenization between the respective models.

One can easily identify numerous attempts at semantic homogenization both within the tradition of methodological monism and within the defining of explanation and understanding by starting from various human interests. Though rather sparse, these attempts are not absent in the case of methodological dualism either.

Thus, we are dealing with three explanatory models corresponding to the fundamental methodological options. The first one inherits the monist option; it can be denoted as the analytical-empirical model of the human sciences. The second one, which affirms methodological dualism, can be defined as the ontological model of the human sciences, since in its case we can recognize ontological reconstructions which were very influential in the 20th century. Finally, the third model is the inter-subjective one. Through its close-to-Hegelianism branch, this model has absorbed concerns regarding the phenomena of labor and domination, which the inter-subjective model of human sciences shares—or disputes—with Marxism.

a) The analytical-empirical model of the human sciences represents the tradition built around methodological monism. It finds its classical theoretical exposition in Georg Henrik von Wright's work *Explanation and Understanding*. In this work, we will analyse the five above-mentioned traits of explanatory models.

(1.) Among the conceptual resources specific to the analytical-empirical model of human sciences, the notions of cause and causal explanation, intentionality and teleological explanation are the most important.

The traditional causal explanation has been reconsidered within the analytical-empirical model. The Covering Law Model (the theory of explanation as subsumed to laws)—confronted with the traditional problem of causation—has been amended and redefined by von Wright. It defines the notion of total state (or possible world). The state-description is "the conjunction (the order of conjuncts being irrelevant) of sentences and/or their negations, describing the states which are the 'atoms' or 'element' of such a possible world" (von Wright, 1995, p. 62). The successive state-descriptions of a world composed of several elementary states compose the histories of the possible world. A fragment of the history of the possible world is a system (von Wright, 1995, p. 67). Within systems, scientific experiments are carried out (von Wright, 1995, p. 68). The model put forth by von Wright distinguishes between causal analysis and the causal explanation. The former discovers the relations of conditionality within the system, while the latter—confronted with the individual appearance of a generic phenomenon—searches for the system in which generic phenomenon can be correlated to its individual appearance (von Wright, 1995, p. 73). Four types of causal explanation are analysed by the author; two of them (of a predictive nature) answer to questions such as "Why necessary?" and "How possible?" (von Wright, 1995, p. 78), while the other two (which can be utilized for retrodictions) show the manner in which something was or became possible. These types of causal explanation are classified as quasi-teleological.

The teleological explanation, as opposed to the quasi-teleological causal explanation, is, of course, oriented towards the future. Still, an authentic teleological explanation is not limited to this definition, but presupposes an intentionalist understanding of behaviour: "The explanandum of a teleological explanation is an action, that of a casual explanation an intentionalistically noninterpreted item of behaviour, i.e. some bodily movement or state." (von Wright, 1995, p. 135).

(2.) Specifically regarding human sciences, the analytical-empirical model applies a relativist rationalism which, in contradistinction to absolute rationalism, does not attribute an aim to history or social processes as a whole, but "views actions in the light of set purposes and cognitive attitudes of agents" (von Wright, 1995, p. 172).

(3.) As a consequence, any claims regarding the particularities of historical or social knowledge are rejected. Knowledge of historical and social phenomena is identical to knowledge of the phenomena studied by natural sciences; still, historical and social phenomena are different from natural phenomena, hence the logical systems approaching them must undergo several adjustments.

(4.) The analytical-empirical model implies certain limits regarding claims to explanation in historical research. Among them is the impossibility of a complete understanding of the past; this impossibility is a consequence of the fact that the historical process has not come to an end. A second limit is derived from the continuous attribution of new significances to past events; "the re-evaluation of the past" is not, however, an expression of the researcher's subjectivity, but a constitutive trait of the model.

(5.) The teleological explanation, as present in von Wright, was the object of several investigations referring to traditional standpoints regarding the significance of action. Research in this direction is not based on analytic philosophy; this makes possible a thorough evaluation of the analytical-empirical standpoint from several different perspectives (cf. Riedel, 1989, p. 162–187).

b) The ontological model of the human sciences. Methodological dualism stands at the core of this model, where reconstructions such as Heidegger's or Gadamer's are representative. For identifying the fundamental notions of the model, I will investigate Gadamer's contribution in *Truth and Method*.

(1.) The hermeneutic circle plays a central part among the concepts used by this model. It describes a phenomenon which the ontological explanatory model keeps as fundamental for the constitution of human sciences: understanding presupposes anticipation and simultaneously makes possible a project of identifying the pre-structure of understanding, which can no longer be offered by traditional historical-philological sciences under the pressure of exact sciences (Hufnagel, 1981, p. 54). The hermeneutical circle is not vicious, and does not represent a denial of the laws of logic. Also, the ontological explanatory model cannot be exempt from hermeneutic experience, which stands at the core of the theoretical reconstruction in *Truth and Method*. In order to delimit hermeneutic experience, Gadamer redefines the notion of judgement, charging it with negative value: "«Vorurteil» heisst also durchaus nicht falsches Urteil, sondern in seinem Begriff liegt, dass es positive und negative gewertet warden kann" (Gadamer, 1975, p. 255). The disrepute of prejudice is correlated with the modern affirmation of scientific knowledge, philosophically fathered by Descartes. "Die grundsätzliche Diskreditierung aller Vorurteile, die das Erfahrungspathos der neuen Naturwissenschaft mit der Auklärung verbindet, wird in der historischen Aufklärung universal und radikal" (Gadamer, 1975, p. 260). But prejudices are conditions of the possibility of understanding. Legitimate prejudices, able to rehabilitate authority and tradition, can dissolve abstract oppositions between tradition and history, or between history and knowledge (Gadamer, 1974, p. 267).

(2.) The human sciences and social life gain their 'form' through understanding (Verstehen) and comprehension (Verständigung). Understanding occurs within communication (cf. Marga, 1988, p 221): "Die Sprache ist der Mitte, in der sich die Verständigung der Partner und das Einverständnis über die Sache vollzieht" (Gadamer, 1944, p. 361).

(3.) Hermeneutic experience stands at the core of a knowledge which is different from the one specific to natural sciences, a knowledge in which authority and tradition play a central part. Still, hermeneutic experience is not competing with the one specific to natural sciences.

(4.) Gadamer clearly formulates the limits of the ontological explanatory model: "*Der Grundbezug von Sprache und Welt bedeutet daher nicht, dass die Welt Gegenstand der Sprache werde*" (Gadamer, 1974, p. 426). The language-related character (Sprachlichkeit) of human experience of the world does not encompass within it the world's objectification (Vergegenständlichung).

(5.) The basic notions of the ontological model do not exclude or subsume deduction or other demonstrative procedures of the classical model of explanation specific to the analytical-empirical explanatory model. Still, they gain certain particularities which deserve careful examination. They have casued several scholars to put forward a "model of comprehension by presumption" (Riedel, 1989, p. 50–52).

c) The inter-subjective model of human sciences. Habermas' contribution is, in this regard, the most representative contemporary example. Departing from "critical theory," Habermas accomplishes an important theoretical extension. Through the "theory of communicative competence," the intersubjective model defines its relations with the analytical-empirical and ontological models of the human sciences.

(1.) Within the inter-subjective model, universal pragmatics, communicative action, and discourse stay at the core of the identification of the structuring role of language, which is investigated from the point of view of its role in mediating social life and according to the functioning of the social system (Marga, 2006, p. 206). Neither the natural sciences with their causal explanation model, nor the classical theories regarding sciences of spirit constitute a proper point of departure for investigating the social system. The system put forth by this explanatory model modifies the relation between both a theory and its object, and between theory and experience. The social system is neither considered an organism (more than the sum of its parts), nor a class (which could be determined by assembling all the elements) (Habermas, 1983, p. 89).

As opposed to empirical pragmatics—which examines behaviour by distinguishing between contexts of different speech situations—universal pragmatics assumes the task of reconstructing the system of rules of speech (Habermas, 1983, p. 191) by explaining the actions the participants in a dialogue have previously assumed (Habermas, 1983, p. 193).

Communicative action and discourse can be considered stages in the process of understanding. Communicative action presupposes more that verbal expression. In the case of discourse, understanding only takes place through the use of linguistic locutions (Habermas, 1983, p. 201).

(2.) The perimeter of research in the inter-subjective model of human sciences is drawn by Habermas through five theses: 1. The achievements of the transcendental subject are grounded in the natural history of the human species; 2. Knowledge is an instrument of self-preservation to the same extent to which it transcends simple self-preservation; 3. The interests leading to knowledge are formed within the environments of work, language, and domination; 4. As far as the power of self-reflection it is concerned, knowledge and interest are identical; 5. The unity of knowledge and interest constitutes a dialectic which—from the historical traces of what has been repressed—reconstructs what has been repressed (Habermas, 1983, p. 134-137). The solidarity between knowledge and interest reveals itself as a consequence of a critique which dissipates “the objectivist illusion” dominated by science.

(3.) The consensual theory of truth formulates the condition of truth as everybody’s potential agreement (Habermas, 1983, p. 209). It manages to avoid the difficulties in defining reality implied by the ontological theories of truth. It asks for a symmetrical division of chances to use and choose the speech acts (Habermas, 1983, p. 220).

(4.) The inter-subjective model of the human sciences is mostly confronted with the problem created by the demand for the abandonment of the conventional postulate of value-neutrality.

(5.) The inter-subjective model is, among the presented models, the most preoccupied with establishing a dialogue, both with the monist and with the dualist tradition. It employs the results of analytical philosophy of language (the contributions of J. L. Austin and J. Searle are partially integrated into Habermas’ theory of communicative competence), and also recognizes itself as indebted to Gadamerian hermeneutics (Marga, 2006, p. 63–70).

Contradiction and the Explanatory Models. The three above-mentioned explanatory models allow particular investigations regarding contradiction. First, one should notice that contradiction has a different function within each model.

In the analytical-empirical model—which uses the structure of 20th century logic—contradiction is understood as logical inconsistency. It is the result of an imprecise definition of concepts, an incorrect demonstrative path, etc; it never expresses fundamental traits of the analysed phenomena. The aim of the analytical-empirical model is the non-contradictory reconstruction of the human sciences, and the general attitude towards the possibility of reaching this purpose is optimistic. Contradictions are to be identified and removed through formalization and the logical analysis of language.

The attitude towards contradiction in the ontological model is ambivalent. On one hand, the hermeneutic circle constrains the representatives of this model to accept the fact that, in certain circumstances, contradiction does not express a logical inconsistency as such, but a modification of the content of concepts due to hermeneutic experience. On the other hand, though, the hermeneutic experience must itself be recuperated in a non-contradictory theoretical form: hermeneutic investigation does not presuppose the assumption of antinomies within the theoretical discourse, but abides by classical argumentation procedures. Contradictions must be identified—not in order to be solved, but rather with the aim of determining the outcome of the hermeneutic experience.

The inter-subjective model is closer to the Hegelian account regarding contradiction. Here contradiction is, so to say, unavoidable. Once the consensus of the participants in the dialogue is realized, the contradiction is solved.

The points of departure for this analysis of different accounts regarding contradiction in the above-mentioned explanatory models could be:

a. Explanation and contradiction in the analytical-empirical model of the human sciences. I would mainly focus on the theory of action debate between causalists (who argue that intention can establish a causal connection whose terms are independent) and intentionalists (who see the relation between intention and behaviour as a practical syllogism). This burning issue could be also examined from an ontological and inter-subjective perspective. Also, an examination of the problem of double negation within the analytical-empirical model—as well as of its ontological and inter-subjective counterparts—is required.

b. Contradiction and historicity in the ontological model of the human sciences. In a 1938-1940 work, Heidegger wrote: "Widerspruch ist nicht Widerlegung, d.h. nicht Vorbringen und Begründen entgegengesetzter Aussagen über Gegenständliches, sondern Ergründung einer anfänglichen Grundstellung in der Wahrheit des Seyns und Inständigkeit in ihr" (Heidegger, 1998, p. 15). The understanding of contradiction as rejection has discredited the notion of prejudice, which the ontological model rehabilitates. Thus, a comparative analysis of the ontological and analytical-empirical model is necessary in this regard.

c. Contradiction and communicative action in the inter-subjective model of the human sciences. According to Habermas, linguistically mediated communication has a two-fold structure: inter-subjectivity (the speaker and the listener speak to each other) and the objects of speech (things, events, emotions, persons) they communicate about (Habermas, 1983, p. 193). The object of this investigation would be the correspondence between the inter-subjective sphere and the ontological model on one hand, and the sphere of the objects of speech and the analytical-empirical model on the other.

Final remarks

The main result of an investigation on contradiction could be a precise definition of this notion and its function within the analytical-empirical, ontological, and inter-subjective explanatory models of the human sciences. One key question in this particular context is how to determine the compatibility/incompatibility relations between the definitions and functions of contradictions within the three explanatory models. Also of great importance is the redefinition of the main concepts of the classical models (causal and teleological explanation, hermeneutic circle, communicative action) on the basis of the relations between the different definitions and functions of contradiction.

References

Gadamer H-G (1975) Wahrheit und Metode, Grundzüge einer philosophischen Hermeneutik, 4. Auflage, J. C. B. Mohr (Paul Siebeck), Tübingen.

Habermas J (1983) *Cunoaştere şi comunicare, prefaţă şi îngrijirea versiunii în limba română: Andrei Marga*, traducere din limba germană de Andrei Marga, Walter Roth, Iosif Wolf. Bucureşti: Editura Politică.

Haberma J (2000) *Conştiinţă morală şi acţiune comunicativă*, traducere:Gilbert. Lepădatu, Bucureşti: Editura All.

Heidegger M (1998) Gesamtausgabe, Band 69, Die Geschichte des Seyns, Vittorio Klostermann, Frankfurt am Main.

Hollis M (1994) *The Philosophy of Social Science. An Introduction*. Cambridge: Cambridge University Press [Romanian edition: *Introducere în filosofia ştiinţelor sociale*, traducere din limba engleză de Carmen Dumitrescu, verificarea traducerii Dragan Stoianovici, Bucureşti: Editura Trei, 2001].

Hufnagel E (1976) *Einführung in die Hermeneutik*, Verlag W. Kohlhammer GmbH [Romanian edition: *Introducere în hermeneutică*, în româneşte de Thomas Kleininger, Bucureşti: Editura Univers, 1981].

Marga A (1988) *Introducere în filosofia contemporană*. Bucureşti: Editura Ştiinţifică şi Enciclopedică.

Marga A (2006) *Filosofia lui Habermas*. Iaşi: Editura Polirom.

Riedel M (1978) *Verstahen oder Erklären? Zur Theorie und Geschichte der hermeneutischen Wissenschaften*. Stuttgart: Klett-Cotta [Romanian edition: *Comprehensiune sau explicare? Despre teoria şi istoria ştiinţelor hermeneutice*, traducere din limba germană şi prefaţă de Andrei Marga, Cluj-Napoca: Editura Dacia, 1989].

von Wright G H (1971) *Explanation and Understanding*. Cornell University Press [Romanian edition: *Explicaţie şi înţelegere*, notă introductivă de Mircea Flonta, traducere de M. D. Vasile, Bucureşti: Editura Humanitas, 1995].

BALKAN JOURNAL OF PHILOSOPHY. Vol. 7, Issue 2, 2015

INTENTIONALISM AS A THEORY OF SELF-DECEPTION

Ivan Cerovac
University of Trieste, Department of Philosophy,
Trieste, Italy
e-mail: ivan.cerovac@phd.units.it

Abstract:

Is self-deception something that just happens to us, or is it an intentional action of an agent? This paper discusses intentionalism, a theory claiming that self-deception is intentional behavior that aims to produce a belief that the agent does not share. The agent is motivated by his belief that *p* (e.g. he is bald) and his desire that *not-p* (e.g. not to be bald), and if self-deceiving is successful, the agent will end up believing *not-p*. Opponents of intentionalism raise two different objections: it seems that self-deceiver should then simultaneously hold two incompatible beliefs (namely, that *p* and *not-p*), as well as simultaneously intend the deception and be unaware of it. This paper reviews possible answers to anti-intentionalist objections (temporal partitioning, psychological partitioning, and the attentional strategy account) and offers guidelines to strengthen intentionalist claims.

Key words: Self-deception, Intentionalism, Anti-Intentionalism, Mind partitioning, Attentional strategy.

I'm not fat. I'm big-boned.
– Eric Cartman

Eric Cartman, a fictional character from *South Park* series, believes that he is not fat, despite the overwhelming available evidence to the contrary (all his friends and teachers think he is fat, and his weight greatly exceeds the one recommended for his age and height). Moreover, Eric has come with what he holds to be a reasonable explanation for their mistake—he is not fat, but big-boned. We should also consider the following: (i) his epistemic standards are not defective in any relevant way—given the described evidence he should be able to conclude that he is fat, and (ii) he has a strong desire not to be fat. Is Eric self-deceiving himself (or being self-deceived)?[1] To answer this question, we must first define self-deception. But how are we to go about it?

Alfred Mele distinguished two main approaches to the analysis of self-deception: the lexical and the empirical (or example-based) approach. While the former—usually associated with traditionalists—"[s]tarts by establishing a definition of deception from the interpersonal case, and uses it to deduce the meaning of self-deception"[2] independent of the usage of expression by ordinary people, the latter—usually associated with deflationists—"[s]tarts by scrutinizing representative examples of self-deception and attempts to identify their essential common features."[3] Both positions generally agree that self-deception is "the acquisition and maintenance of a belief (or the avowal of that belief) (i) in the face of strong evidence to the contrary (ii) motivated by desires or emotions favoring the acquisition and retention of that belief."[4] It seems that Eric is indeed deceiving himself (or being self-deceived), since he is believing something against his better judgment because of his strong desire.

But why does he end up believing that ~*p* despite the fact that he has both adequate cognitive capabilities and relevant evidence to conclude that *p*? Certainly there has to be some reason why Eric

---

[1] Ariela Lazar (1999) distinguishes these two formulations and connects them with two main theories of self-deception: "deceiving oneself" with intentionalism and "being self-deceived" with anti-intentionalism.

[2] Lynch, K. (2009) Prospects for an intentionalist theory of self-deception. *Abstracta* 5 (2) p. 126–138

[3] Mele, A. (2001) *Self-Deception Unmasked*. Princeton: Princeton University Press

[4] Deweese-Boyd, I. (2006) Self-Deception. *Stanford Encyclopedia of Philosophy.* [Online] Available from: http://plato.stanford.edu/entries/self-deception/ [Accessed 27/11/2010]

ends up falsely believing that he is not fat. First of all, we must perceive that the phrase "Eric has a reason to believe that he is not fat" has more than one meaning: "[a] reason of the first sort is evaluative: it provides a motive for acting in such a way as to promote having a belief, and a reason of the second kind is cognitive: it consists in evidence one has for the truth of a proposition."[5] In many examples of self-deception, including the former one, irrationality arises when someone has only an evaluative (motivating) reason for a belief that *p*, and he acquires that belief despite the fact that he has no cognitive (normative) reason to believe it. The problem with self-deception emerges when a person starts having thoughts that emphasize the obtaining of cognitive (normative) reasons for a belief because of desires or motivating reasons for holding it. Self-deception is the process (or the state) that leads from a desire or a motivating reason to a reason that the self-deceiver regards as normative for a certain belief. It usually consists of intense epistemic activity that attempts to justify the false belief that ~*p*—since initially the self-deceiver has better (epistemic) reasons to believe that *p*, he has to thoroughly reconsider the evidence he has (find new evidence) in order to end up believing that ~*p*. There are many strategies one can use to manipulate the belief-formation process to favor a certain belief, such as selective evidence gathering and selective focusing of attention. I will not discuss these strategies in detail; we should rather inquire into how one applies them. Is Eric intentionally gathering evidence selectively, or is this cognitive bias something that happens to him? Is Eric the author or the victim of self-deception?

Contemporary discussion of self-deception distinguishes two central positions—intentionalism and anti-intentionalism. Intentionalists compare self-deception with interpersonal deception, intentional behavior that aims to produce a belief that the agent does not share. Unlike interpersonal deception, where A intentionally gets B to believe that *p* while A knows that ~*p*, in self-deception A is simultaneously the deceiver and the deceived—he intentionally gets himself to believe that *p* while also knowing or believing that ~*p*. Anti-intentionalists, on the other hand, consider self-deception to be a form of motivationally biased belief, not at all structurally similar to intentional interpersonal deception. They hold that our false beliefs are not the result of an intentional process, but unintentionally motivated by a desire or emotion.

This paper is divided into three distinct sections. The first provides a summary of main arguments that support intentionalism as a theory of self-deception. Following Davidson (1985) and Bermudez (2000), I first defend intentionalism through inference to the best explanation using the lexical approach to the analysis of self-deception, and then, noticing the paradigmatically selective nature of self-deception, argue that only intentionalist approaches can explain why self-deception took place in that particular situation. In the second section I consider the most common objections to intentionalist approach—the "static" and "dynamic" (or "strategic") paradoxes. Since there is more than one way to answer these objections, I briefly sketch possible intentionalist accounts that avoid these paradoxes: Bermudez's (2000) "temporal partitioning," Davidson's (1982) "psychological partitioning," and finally the "attentional strategy." In the final section, I shall briefly summarize the main explanatory virtues of intentionalism and provide some additional arguments that support this claim.

1. The advantages of the intentionalist approach

Considering self-deception analogous to interpersonal deception enables us to distinguish it from a mere error, since in self-deception acquisition and maintenance of a false belief is intentional. Self-deception can be easily shown by distinguishing between belief—a non-intentional but causally produced mental state that one cannot be held responsible for—and acceptance—an intentional process based on reasons and subject to one's control. Self-deception would thus be a process in which a person acquires and maintains a false belief (i) despite many reasons available to her, and (ii) tries to accept false belief by means of filtering the information available in various ways.[6] The intentionalist approach gives us a clear distinction between self-deception and other sorts of motivated believing, like wishful thinking and weakness of warrant.[7] Davidson thus distinguishes wishful thinking—the

[5] Davidson, D. (1985) Deception and Division. In: LePore, E. and McLaughlin, B. (1985) eds. *Actions and Events*. Oxford: Basil Blackwell

[6] Cohen, J. L. (1989) Belief and Acceptance. *Mind*, New Series, Vol. 98, No. 391.

[7] For a detailed distinction between self-deception and other sorts of motivated believing, see Davidson's paper "Paradoxes of Irrationality."

sequence in which the desire that *p* produces the belief that *p*—and self-deception, which requires "the agent to do something with the aim of changing his own views."[8] The self-deception must not only be self-induced, it also must be intentional.

> It is not self-deception simply to do something intentionally with the consequence that one is self-deceived, for then a person would be self-deceived if he read and believed a false report in a newspaper. The self-deceiver must intend the 'deception'.[9]

Davidson explains how self-deception occurs.

> A has evidence on the basis of which he believes that *p* is more apt to be true than its negation; the thought that *p*, or the thought that he ought rationally to believe that *p* motivates A to act in such a way as to cause himself to believe the negation of *p*. […] All that self-deception demands of the action is that the motive originates in a belief that *p* is true (or that the evidence makes it more likely to be true than not), and that the action be done with the intention of producing a belief in the negation of *p*. Finally, and it is especially this that makes self-deception a problem, the state that motivates self-deception and the state it produces coexist; in the strongest case, the belief that *p* not only causes a belief in the negation that *p*, but also sustains it.

Although most philosophers who discuss the self-deception problem here include Davidson's example of Carlos and his driving license test, I believe anther example, also made by Davidson, is more appropriate: the bald man example. It displays not only how a desire, along with an appropriate belief, can induce self-deception and lead to an agent believing something that is false, but also how this process can continue indefinitely into the future, with the belief that *p* constantly sustaining the belief that ~*p*. D believes that he is bald and has good reason to believe that he is bald. Since he does not want to be bald (and does not want to believe that he is bald) he chooses views and lighting that create a hirsute appearance. Now, with his new hairstyle, D starts to believe that he is not all that bald. His belief that he is bald, along with his desire not to be bald, have led to selective evidence gathering, and as a result D ends up believing that he is not bald. However, he keeps choosing views and lighting that create a hirsute appearance. Why does he keep choosing them if he now truly and completely believes that he is not bald? The answer seems rather obvious: though D now believes that he is not bald, he still has a belief that he is, and it is exactly this belief—along with his desire not to be bald—that motivates him to keep choosing certain views and lighting. It seems that anti-intentionalists cannot explain self-deception in this example by the usage of desires and emotions alone; it is certainly correct that our desires play a great role in self-deception, but it also seems that they alone are not enough. Suppose D had the desire not to be bald from his childhood, for a very long period of time. 10 years ago, when he did not have good reasons to believe that he is bald (probably because he wasn't) he still had the desire not to be bald, but it did not motivate him to choose views and lighting that create a hirsute appearance.

Most intentionalists claim that D's self-deception is the result of his practical reasoning. If asked why D intentionally forms a belief that is undermined by the weight of the evidence, they suggest that the irrational belief is acquired in order to attain a goal that is frustrated by the presence of the rational belief. Thus Davidson introduces the Carlos example in order to elaborate how a belief (or the perception that one has sufficient reasons for a belief) can sustain (cause) a contrary belief. Davidson writes

> So he has a perfectly natural motive for believing he will not fail the test, that is, he has a reason for making it a case that he is a person who believes he will (probably) pass the test. His practical reasoning is straightforward. Other things being equal, it is better to avoid pain; believing he will fail the exam is painful; therefore (other things being equal) it is better to avoid believing he will fail the exam.[10]

Self-deception is also paradigmatically selective. "Any explanation of a given instance of self-deception will need to explain why motivational bias occurred in *that* particular situation."[11] According to

[8] Davidson, D. (1985) Deception and Division. In: LePore, E. and McLaughlin, B. (1985) eds. *Actions and Events*. Oxford: Basil Blackwell
[9] Ibid.
[10] Ibid.
[11] Bermudez, J. L. (2000) "Self-deception, Intentions and Contradictory Beliefs." *Analysis* 60: 309–319.

the anti-intentionalist account, a strong desire and a low acceptance/high rejection threshold for the hypothesis that *p* are enough to induce self-deception and make a subject believe that *p* (no intention from the subject is required). But there are also many cases in which we have both the desire and a low acceptance/high rejection threshold, but we don't end up believing that *p*. Why does a person end up self-deceiving in one example, and not in other, while she had an equally strong desire and low acceptance/high rejection threshold for both? Bermudez (2000) concludes that "it (also) requires an intention on the part of the self-deceiver to bring it about that he acquires the belief that *p*." Since anti-intentionalists cannot solve the selectivity problem, this solution offered by intentionalists proves to be a strong argument for their position.

Since (unlike intentionalism) anti-intentionalism is often argued for indirectly through attacking incoherencies in the intentionalist account, in the next section we shall review these objections and possible solutions.

2. Common objections to the intentionalist approach

In this section we shall discuss the most common objections to intentionalist approach—the "static" and "dynamic" (or "strategic") paradoxes. Both paradoxes arise from the analogy of self-deception and interpersonal deception. Davidson (1985) claims that "the only intentions a liar must have are these: (1) he must intend to represent himself as believing what he does not (for example, and typically, by asserting what he does not believe), and (2) he must intend to keep his intention hidden from his hearer."[12] In other words, (1) the deceiver must believe the negation of the belief he wants to induce in the hearer, and (2) he must intend to deceive the hearer, and the hearer must not know about the deceiver's intentions. Since in self-deception both the deceiver and the hearer are the same person, this means that (1) the self-deceiver must believe both the belief he wants to induce in himself and the negation of that belief (he must believe both *p* and ~*p*), and that (2) he must intend to deceive himself and simultaneously be ignorant about his own intentions.

The first requirement leads to the "static" paradox—it seems impossible to consciously believe both *p* and ~*p* at the same time. The self-deceiver must (as a deceiver) believe that ~*p*, while at the same time (as a hearer) believe that *p*. This, however, seems to be an impossible state of mind, because it violates the principle of non-contradiction.

The second requirement leads to the "dynamic" paradox—it seems impossible to simultaneously intend the deception and remain unaware of our own intentions. Thus, even if intentionalists somehow manage to avoid the "static" paradox by asserting that the self-deceiver gets a change of mind (believes that *p*, but after reconsideration of the evidence ends up believing that ~*p,* so he never simultaneously believes both *p* and ~*p*), they still have to face the fact that the intention to deceive ourselves renders the process self-defeating.

There are several intentionalist accounts that avoid these paradoxes, and we shall briefly review three main responses.

Temporal Partitioning

One of the possible intentionalist strategies is to admit that it is impossible to simultaneously believe both *p* and ~*p*, but claim that this does not imply that self-deception is impossible, since it is often extended over time. Thus a self-deceiver can believe that *p* in t_1 and that ~*p* in t_2 without ever simultaneously believing both *p* and ~*p* and falling into the 'static' paradox. Though the "dynamic" paradox cannot be avoided so easily, Bermudez claims that "one can simply lose touch with an intention while one is in a process of implementing it, particularly when that implementation is a long drawn out process."[13] And even if one still remembers his own intention to deceive oneself, he might still be so overwhelmed with new evidence he had acquired through self-deception that he no longer cares about his former intention.

Psychological (Mind) Partitioning

Another possible intentionalist strategy is to directly confront the 'static' paradox and claim that it *is* possible to simultaneously believe both *p* and ~*p* without ever making the conjunction ($p \wedge {\sim}p$). Davidson thus claims that the mind can be partitioned into quasi- independent structures, and that there can

[12] Davidson, D. (1985) Deception and Division. In: LePore, E. and McLaughlin, B. (1985) eds. *Actions and Events*. Oxford: Basil Blackwell.

[13] Bermudez, J. L. (2000) "Self-deception, Intentions and Contradictory Beliefs." *Analysis* 60: 309–319.

be non-logical causal relations between them.[14] We can have two contrary beliefs, but placed in different "brain areas"; this is Davidson's solution to the "static" paradox. The "dynamic" paradox proves to be a serious problem for the intentionalists; it seems impossible to deceive oneself since it would require doing something with the intention that the intender does not recognize this very intention. However, Davidson claims that there is a distinction between self-deception and lying to oneself: while the latter entails the existence of a self-defeating intention (and is thus struck by the "dynamic" paradox), the former "pits intention and desire against belief, and belief against belief."[15] Therefore the subject of self-deception tries to find other reasons to justify his selective acquisition of the available information; he is not lying to himself that he is bald (because he acquires reasons to believe that he is not bald after he has chosen views and lighting that create a hirsute appearance), and he chooses views and lighting because of some other selectively chosen evidence (e.g. someone once told him that he looks better that way). The deceptive subsystem is thus hidden from consciousness in a complex network of motivated and interconnected reasons, and the "dynamic" paradox is avoided.

The Attentional Strategy

Yet another possible intentionalist strategy includes intentional manipulation of our attention, as presented by Lynch (2009). "A person who believes or at least suspects that *p*, but who wishes to believe that *not-p*, by turning his attention away from the unwelcome considerations supportive of *p*, and by attending to welcome considerations supportive of the contrary *not-p*, may end up losing the (conscious) belief that *p* and acquiring the belief that *not-p*."[16] Consider a student who has good reasons to believe that he should study for his final exam, but does not want to do so. If he tries to convince himself that he will probably pass the exam without studying he faces the "dynamic" paradox; however, if he turns his attention to something else, like playing football or watching movies, he successfully avoids it. When he is playing football, he is both successfully self-deceived and unaware of the deception he has just made. "Although his distraction seeking can be done intentionally, this does not imply that agent is intentionally not thinking about the unwelcome evidence when she succeeds in distracting herself from thinking about it."[17]

3. Conclusion

This paper emphasizes two main arguments for the endorsement of intentionalism: its attitude towards the nature of self-deception offers us a clear and intuitive explanation of that phenomena and supports our intuitive idea that the self-deceiver should be responsible for his self-deception, and the inclusion of intention offers a solution to the problem of the paradigmatically selective nature of self-deception. In the second section we have considered possible answers to the "static" and "dynamic" paradoxes. Although some of these answers may raise objections, the discussion is still active and there is no final conclusion. Another suggestion is presented by Bach (1981), who claims that the self-deceiver does not intend the deception, but nonetheless acts intentionally. "Although the self-deceiver does what he does intentionally, he does not do it under the description of 'deceiving myself' or anything of that sort. Rather, he is motivated to avoid the thought that *p* but is unaware of (or denies the impact of) this motivation and of his uncharacteristic violation of his own rational standards."[18] This suggestion successfully avoids the two paradoxes and is somewhat similar to Davidson's position, especially in respect to unconscious beliefs and intentions. I believe that psychology has yet to provide its main contribution to this discussion, and that strategies involving unconscious beliefs and desires have significant potential.

References

Bach K (1981) An Analysis of Self-Deception, *Philosophy and Phenomenological Research*. 41(3): 351–370

Bermudez J L (2000) elf-deception, Intentions and Contradictory Beliefs *Analysis* 60: 309–319.

[14] Davidson, D. (1982) "Paradoxes of Irrationality." In: Wollheim, R. and Hopkins. J. (1982), eds., *Philosophical Essays on Freud*. Cambridge: Cambridge University Press.

[15] Davidson, D. (1985) Deception and Division. In: LePore, E. and McLaughlin, B. (1985) eds. *Actions and Events*. Oxford: Basil Blackwell.

[16] Lynch, K. (2009) Prospects for an intentionalist theory of self-deception. *Abstracta* 5 (2) p. 126–138.

[17] Ibid.

[18] Bach, K. (1981) An Analysis of Self-Deception, *Philosophy and Phenomenological Research*, Vol. 41, No. 3 pp. 351–370.

Cohen J L (1989) Belief and Acceptance. *Mind*, New Series, 98(391).

Davidson D (1982) Paradoxes of Irrationality. In *Philosophical Essays on Freud*, eds R. Wollheim, J. Hopkins. Cambridge: Cambridge University Press

Davidson D (1985) Deception and Division. In *Actions and Events*, *eds* E. LePore, B. McLaughlin. Oxford: Basil Blackwell.

Deweese-Boyd I (2006) *Self-Deception*. Stanford Encyclopedia of Philosophy. [Online] Available from: http://plato.stanford.edu/entries/self-deception/ [Accessed 27/11/2010]

Lazar A (1997) Self-deception and the Desire to Believe. *The Behavioural and Brain Sciences,* 20: 119–120.

Lazar A (1999) Deceiving Oneself or Self-Deceived? On the Formation of belief 'Under the Influence', *Mind*, 108: 285–290.

Lynch K (2009) Prospects for an intentionalist theory of self-deception. *Abstracta* 5 (2): 126–138.

Mele A (2001) *Self-Deception Unmasked*. Princeton: Princeton University Press.

Pedrini P (2006) Self-Deception: What is it to Blame After All?. *Annali del Dipartimento di Filosofia 2005.*

BALKAN JOURNAL OF PHILOSOPHY. Vol. 7, Issue 2, 2015

HOW CAN AN ARTEFACT BECOME AN ARTWORK IN THE 20TH CENTURY?

Mihaela Pop
University of Bucharest,
Bucharest, Romania
e-mail: pop.mihaela.a@gmail.com

Abstract:
The purpose of this paper is to identify those characteristics that transform a common artefact into an artwork, especially in our postmodern contemporary world. My analysis is based on aesthetic principles which are used in artistic theories and debates.

Key words: artworld, artefact, aesthetic appreciation, artistic transfiguration.

When we want to artistically appreciate an object (artefact), the main question is whether the producer (the artist) intended to make it as a work of art, or as an object useful in everyday life.

This question helps us especially in the case of ancient civilizations such as Greece, where we find a lot of artefacts, such as a large amount of pottery that is artistically decorated. Should we consider them to works of art, or common artefacts?

The answer is not one specific to Aristotelian binary logics, which impose a clear option between two contradictory choices. We have to admit that each piece of pottery is beautiful and has real artistic value, but this value should not be the fundamental one in appreciating the artefact. Anyone could agree on the beauty of the decorations and the chromatic combinations that embellish these artefacts, just as anyone could also recognize some mythological Greek gods and heroes represented on these pieces of pottery. However, a piece of funerary pottery was created to fulfill certain conditions required by religious rituals. For instance, it had to have a certain form due to the content it had to contain. Therefore it is evident that aesthetic value is not the fundamental reason for the creation of these pieces. The Greek craftsman did not want mainly to make a work of art, but rather an artefact that was useful for a specific purpose, which in this case is a religious one.

Ancient Greek pottery exhibited in the National Museum of Archeology in Athens

This is also clear for Byzantine icons and for all Orthodox icons.

Christian Orthodox Holy images exhibited in the Byzantine and Christian Museum, Athens

We already know that Byzantine painters of Orthodox icons had to undertake a religious ritual and practice ascetic behavior before being allowed to paint holy images. Dionysius of Fourna, a religious painter in the beginning of the 18th century, wrote a manual on how to paint holy images. Religious meditation and ascetic behavior were fundamental. The painter should live a certain period within a monastery, have a confessor, and think deeply about the holy person he was going to represent in his painting. He considered himself only an instrument of the divinity and the Holy Spirit, since the image that he was going to paint was an image of the Invisible, of the divine itself. There was an entire dogmatic theory developed by John of Damascus, Theodor of Stoudion, and Nikephoros—the patriarch of Constantinople during the 8th and 9th centuries—that established the role of the holy images within Byzantine dogmas, dogmas which have been obeyed since the Second Council of Nicaea (787). The painter was considered only an instrument of a divine manifestation, and the holy icon was thought of as an object of the religious cult. It was considered an intermediary, a bridge between the absolute and invisible divine world and the mundane one. Its role was to guide the believer towards the transcendent realm, to make him focus on his religious condition. Their Christology made possible, from a dogmatic perspective, the representation of the divine, as Christ had a double condition: divine and human at the same time. What was to be suggested in the holy image was also part of the dogma: the icon was not the divinity itself, but only an intermediary. The Aristotelian concept of homonymy was used for this purpose: both the divine person Christ and the holy image "Jesus Christ" had the same name but they differed in substance, and therefore their difference was an ontological one. This marks the great distinction between the holy images and what they represent, and also between holy images and pagan idols.

From an artistic point of view, painting a holy image presupposed a specific painting canon that the painter should obey rigorously. Personal creativity was not encouraged. This is a reason for the fact that holy images are not signed by their painters. We can thus remark that the religious value of these artifacts is primordial, and takes priority over their aesthetic value. And the examples are nearly innumerable…

What happens in cases like the Sistine Chapel or the *Last Judgment*, painted by Michelangelo,

or **Stanza della Segnatura** in the Vatican, painted by Raphael?

Raphael – *Parnasus*

Raphael – *School of Athens*

In the case of these paintings, though they decorate the most religious places of the Catholic world, their religious value tends to be overlooked in the face of their artistic value. But this attitude is not completely new: these paintings were appreciated artistically even when they were painted in the 16th century, and this artistic appreciation grew more and more evident through the following cultural periods of the European history of art since the Renaissance. The producer (the artist) becomes more and more aware of his special status, and tended to manifest himself as a special person in his society. Art exceeded the domain of craftsmanship, and the production of artefacts became a distinct realm.

Thus, even when the artist paints a religious subject, he chooses his own technique and the way that he is going to represent the subject artistically. His individual contribution becomes fundamental and marks his product, which is now a work of art, and which remains unique, unrepeatable, and irreplaceable. The artwork contains particular materials, displays the artist's personal vision, and arises from a specific cultural background, all of which ensures the uniqueness of the artwork.

*

What happens during the 20th century? The Futurists claimed that all classical works of art in museums should be destroyed and libraries should be set on fire.[1] They dreamt of a day when a new art would completely replace the classical one. They claimed the right to the status of artistic value for what was usually considered to belong to everyday life, “the commonplace.”[2] The commonplace—usual artefacts—could be a snow shovel, a wheel of a bicycle, even a urinal vessel. Marcel Duchamp exhibited such common objects as artwork, and called them *readymades.*

Marcel Duchamp – *Ready mades*

These are artefacts present in our daily life; they are made to be useful, to help people to perform daily tasks and duties. They are certainly all cultural products if we consider *culture* in its wider meaning of including any product of human activity.

---

[1] Filippo Tomasso Marinetti, *The Manifesto of the Futurism* (Romanian translation), Art Publishing House, Bucharest, 2009, pp. 75-76. It was published in Paris, in *Le Figaro*, on February 20th, 1909, and translated and published in Romania during the same day in the journal *Democrația (Democracy*), led by Mihai Drăgănescu in Craiova, a Romanian town.

[2] We use this term taking into consideration Arthus Danto's work *The Transfiguration of the Commonplace, (La transfiguration du banal)* – French translation, Ed. Du Seuil, Paris, 1989.

Therefore, what could be the difference between a useful artefact and a work of art? What are we to make of Duchamp's works, or Andy Warhol's *Brillo boxes?* The scandal generated in the artistic world is evident: where is the artist's contribution if the object is already made? In the case of Warhol's *Brillo Boxes*, there are two objects which look alike, but belong to different worlds.

Andy Warhol – *Brillo Box*

They are similar to a perceptive level, but are ontologically different (an artefact and a painting). The artwork is and remains distinct from all the other objects. While the artefacts restrict themselves to being simple objects that have a utility, the work of art supposes a chain of intentions which the artist incorporates in the process of creativity. The artwork has the power to represent something; it is a symbol which has meaning and artistic value. It could therefore be characterized by intentionality and the force of representation. These qualities suggest that the artwork exists, manifests itself not only at an ontic level of pure material existence, but also at an ontological level, through a process of interpretation of that representation (image). The artwork is therefore a semiotic structure which helps us to identify it. As Arthur Danto remarked, any artwork is a metaphor manifesting at two levels: a) the level of the artistic material, and b) the formal and formative level. This means it is not a mere artefact; it represents an artistic idea, and so its form also becomes meaningful, rather than simply transparent.

Usually art critics consider Hegel the philosopher who announced *the death of the art.*[3] The Avant-garde of the 20th century was perceived as the real application of this cultural prediction. In fact, the Avant-garde helped art to enter a new era. While classical and modern art and even some of the Avant-gardes (Cubism, Abstractionism) searched for the philosophical essence of art, postmodern art seems to respond in a strange way to Hegel's theory. Hegel considered art a preliminary step of human spirituality in its way to philosophical knowledge of the self or high spirituality. According to Danto,[4] this step has already been taken by contemporary art, which seems to reflect more deeply on itself and in a more philosophical way. It is closer to philosophical way of thinking, and thus seems to be reviving Hegelian thought, as Danto points out.

The Futurist and Dada movements tried to reveal some different ways of understanding the new conditions of living. The world of the 20th century is no more that of Titian's *Venus*. It is a dynamic world full of many changes, including an increase in the mass production of artefacts which suggest the benefits of consumerism and modern life. Thus, even artefacts such as a bottle drier or a bike-wheel should be considered interesting and capable of articulating meanings, symbols, perceptions, and intellectual or practical experiences. Objects which had previously belonged solely to the commonplace now seem as able to become sources of inspiration for an artist as any myth, heroic gesture, or divine narration. In fact, the ready mades (the industrialized artefacts) deserve our deep attention, as they are part of our daily human universe. And this change of priorities has been evident since Duchamp and Andy Warhol through the process of *transfiguration of the commonplace*, as Danto calls it.

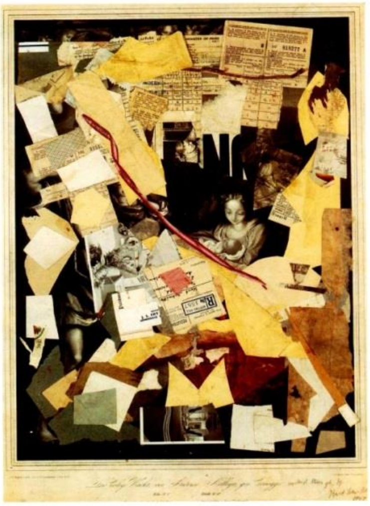
Andy Warhol – *The sickle and the hammer*

[3] Especially in his *Lectures on Aesthetics*.

[4] See for instance his article "The End of Art: A Philosophical Defense," in the volume *History and Theory: Danto and his Critics*, vol. 37, no.4, 1998, pp 127–143; www.jstore.org/2006, Feb.9.

We can see how the two artefacts (the sickle and the hammer) are represented in these artworks. Warhol very carefully drew even the name of the industrial company that produced the two objects, suggesting the idea of an artistic involvement in what usually belongs to the commonplace. This convincing presence of the artefacts does not, however, diminish the artistic and meaningful content of Warhol's creativity. If we take into consideration that these paintings were made during the Cold War, their meaning is rich, and a highly political message is more evident.

Not only may certain common objects become works of art in contemporary artistic movements, but also what is considered waste or garbage, dust or pure naturalistic elements such as minerals, land, rocks, and above all, the human body itself.

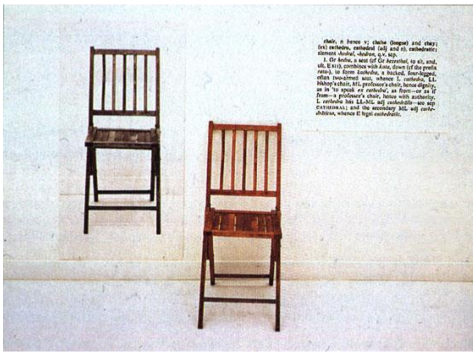


K. Schwitters – *Merz*

J. Tinguely – *Homage to New York*

Here are two examples of the artistic use of what industrial civilization generates in huge quantities—garbage and waste materials. Kurt Schwitters was among the first artists who used pieces of garbage in his paintings, and Jean Tinguely assembled a huge number of waste parts from machines in order to create his works of art.

Joseph Kosuth – *Five words in green*



J. Kosuth – *One and three chairs*

Joseph Kosuth, an artist of the conceptualistic movement (which belongs to post-modern artistic thought), claims that art should eliminate any aesthetic, emotional, and sensitive elements. Only the artistic side should be present, and it should have an intellective content. He uses definitions of words which name various artefacts, such as a chair, or a neon light which itself is composed of words. In his opinion, art should replace philosophy during this period when philosophy itself seems to have lost its identity.[5] He is also known for his collection of works *Art as Idea as Idea*, which are simple definitions of certain meaningful words shown on black frames.

Joseph Beuys was an artist of the 1960s and 1970s who belonged to the Fluxus movement. He created installations and performances and participated with his own body in his works. He also employed animals, as in his famous performance *Coyote*, in which he lived for three days with a coyote in an art gallery. In another performance he planted 7,000 oak trees. He considered the elements

[5] See for instance J. Kosuth's article "Art after Philosophy," http://www.ubu.com/papers/kosuth_philosophy.html

of nature very inspiring. In his installation *Honey Pump*, he used honey and fat as symbolic energizing substances necessary in any human activity, as the two substances are capable of stimulating the metabolism of the human body.[6]

J. Beuys – *Coyote*

J. Beuys – *Honey pump*

J. Beuys – *Oaks*

In this way, contemporary art seems to prove what John Dewey had already said: art should not be separated from the experience of real life.[7] The experience of art is one of the human experiences; it belongs to what is called *Lebenswelt*, the live world. This continuity between art and life should be re-established. This is the new philosophy of postmodern art. In this way, any artefact can become an artwork, as Danto also pointed out. It is not the aesthetic qualities which are called to determine artistic value, but rather the intentionality included in each experience is to be considered artistic.

George Dickie pointed out another characteristic: "an artwork is an artefact to which a society or a social group gives the status of an object susceptible to artistic appreciation."[8] This means that art has a historical dimension; it is in strict connection to the cultural period to which it belongs. Art is now considered an image of reality, of life itself, but also a transfigured image, since it was created by an individual—the artist—whose intention and thought are built into the work of art itself through the entire process of artistic creativity.

We have to remember that the artistic value of the African art was discovered at a certain historical and cultural moment in the European history of art. In 1907, Picasso discovered the aesthetics of African artefacts while visiting an exhibition of African civilization in Paris. This was possible because European artistic thought had reached at that moment (the beginning of the 20th century) a level at which African artistic values could be perceived aesthetically, could attain a new meaning. At the same time, we have to remember that Hegel already mentioned that an artwork should be an expression of the spiritual level, it should have a wide meaning; it cannot belong only to the material level.

A common artefact usually belongs to a series or a chain of objects created to be useful. The nail, for example, was not invented alone; it belongs to a series which includes also a hammer, a wall (or a board), a saw, etc. All come together. Therefore, to understand an artefact means to understand the system of artefacts and gestures to which it belongs, and their role in a specific society. An artwork, on the contrary, does not belong to a series: it is self-

[6] An interesting talk about the use of natural elements and substances in art can be found in a text edited by Volker Harlan, *What is art? Workshop with Joseph Beuys*, (Romanian translation), Publishing House Idea Design, Cluj, 2003.

[7] John Dewey, *Art as Experience,* Perigee Books, Putnam's Sons, New York, 1980, especially chapter 1, "The live creature," pp. 3-19.

[8] George Dickie, "Defining Art," *American Philosophical Quarterly*, 6/1969, University of Illinois Press, pp.253-256.

sufficient, and its purpose is in itself. It could be considered an embodied individual spirit. Danto said that any object can become art through a process of transfiguration. This transfiguration is achieved through imagination and spirituality. If we take into consideration a general definition of artwork—*an object or a gesture that has artistic content expressed in various artistic manners, and which is capable of stimulating the spectator's imagination and aesthetic satisfaction*—it becomes evident that even a simple bottle drier can "speak" to the spectator's imagination about symbolism, about adaptation to necessities, and about ingenuity in creating objects in compliance with human purposes. It speaks also about the human being as a creator of values.

Here at the end of my paper, I would like to show how artefacts could become parts of an artistic event in contemporary performance art. I chose two events performed by the Serbian artist Marina Abramowicz: *Rhythm 0* and *Lips of Thomas*. In my opinion the artefacts used in these performances—though very familiar—reveal profound meanings and questions which are unveiled through this artistic manner of expression.

Marina Abramowicz – *Rhythm 0*

Marina Abramowicz – *Rhythm 5*

Marina Abramowicz – *Lips of Thomas*

In *Rythm 0,* seventy-two objects were exposed (lipstick, scissors, a knife, a pistol, a bullet, etc.).The artist herself was present and stood still for six hours, allowing the spectators to do anything they wanted with those objects to her body. Finally someone took the pistol and put it into her mouth. The performance stopped. It was an exploration of human violence against another human being.[9]

In *Rhythm 5* the artist used another kind of violence against herself. She entered a large star full of burning candles, which emitted a lot of hot monoxide, which is toxic. After a while the artist fainted, and the spectators intervened and extracted her body.

Lips of Thomas is another violent performance using the artist's body as the object of violence. Marina Abramowicz comes and sits nude on a chair in front of a table, upon which she finds bread and wine. She drinks and eats. Then she flagellates herself and draws a star on her abdomen with a blade. With her skin full of injuries and bleeding wounds, she lies down on an ice cross and remains there without moving until the audience grows anxious enough to pull her away from the ice cross themselves. The allusions to the Christian Eucharistic ritual and to Christ's Passions are evident.

We could thus remark that contemporary art has accepted any artefact produced by human beings in what Danto called the artworld. Furthermore, contemporary art also explores the human body as an artistic object, an object which is also the subject of the art. As we live within our bodies, our condition of being in this world is a bodily-being-in-the-world, as Maurice Merleau-Ponty proved in his *Phenomenology of Perception*. Thus contemporary art has already succeeded in using artefacts as artistic material. This could be another topic of analysis.

References

Marinetti F. T. (2009) *The Manifesto of the Futurism* (Romanian translation), Bucharest: Art Publishing House, 75–76.

Danto A (1989) *The Transfiguration of the Commonplace, (La transfiguration du banal)* – French translation, Paris: Ed. Du Seuil.

"The End of Art: A Philosophical Defense," in the volume *History and Theory: Danto and his Critics*, vol. 37, no.4, 1998, pp 127–143; www.jstore.org/2006, Feb.9.

Kosuth J (1969) "Art after Philosophy," http://www.ubu.com/papers/kosuth_philosophy.html

Harlan V ed. (2003) *What is art? Workshop with Joseph Beuys*, (Romanian translation) Cluj: Publishing House Idea Design.

Dewey J (1980), The live creature. In *Art as Experience,* 3–19. New York: Perigee Books, Putnam's Sons.

Dickie G (1969) Defining Art. *American Philosophical Quarterly*, 6: 253–256.

Goldberg R L (2001), *Performance art. From Futurism to the Present.* London: Thames and Hudson.

Rosse C (2006) The Paradoxical Bodies of Contemporary Art. In *A Companion to Contemporary Art since 1945, ed.* Amelia Jones, 378–396. Oxford: Blackwell Comp.

Fischer-Lichte E (2008) *The Transformative Power of Performance, A New Aesthetics*, London: Routledge.

Davies D (2004), *Art as Performance.* London: Blackwell Publishing House.

[9] See Rose Lee Goldberg, *Performance art. From Futurism to the Present*, Thames and Hudson, London, 2001; *A Companion to Contemporary Art since 1945,* edited by Amelia Jones, Blackwell Comp, Oxford, 2006, especially Christine Rosse, "The Paradoxical Bodies of Contemporary Art," pp.378–396; Erika Fischer-Lichte, *The Transformative Power of Performance, A New Aesthetics*, Routledge, London, 2008; and Donald Davies, *Art as Performance*, Blackwell Publishing House, London, 2004.

BALKAN JOURNAL OF PHILOSOPHY. Vol. 7, Issue 2, 2015

THE SECOND ENLIGHTENMENT AS AN AESTHETIC ENLIGHTENMENT AND ITS RELEVANCE

Fan Meijun*, Wang Zhihe*
*Professor of Humanities, Harbin Institute of Technology,
Director of the China Project of the Center for Process Studies
**Director of the Center for Constructive Postmodern Studies, Harbin Institute of Technology,
Director of the Institute for Postmodern Development of China
e-mail: claremontwang@yahoo.com

Abstract:
The Second Enlightenment is a deep reflection and an immanent transcendence of the first Enlightenment (17th and 18th centuries). Although the great achievements of the first Enlightenment cannot be denied, its limits are increasingly being exposed. Among the many limitations of the Enlightenment, the suppression of beauty in general and natural beauty in particular is one of its main drawbacks, caused by its blind worship of reason and the domination of a modern mechanistic worldview. The suppression of beauty and natural beauty has produced destructive consequences which are responsible for the ecological, social, and spiritual crisis facing us today. In order to prevent ecological catastrophes and create an ecological civilization, a second Enlightenment whose core concept is beauty is needed. The Second Enlightenment challenges us to rethink the value and importance of beauty and natural beauty, and regards beauty as the aim of the universe, as well as of ecological civilization. In this sense, the Second Enlightenment can be seen as an aesthetic Enlightenment which not only helps us to resist consumerism, guard spiritual dignity, nurture our souls, find a sense of belonging, and lead a poetic life, but also helps Chinese society to overcome a defiant attitude toward rural civilization and farmers, and eventually to remove the Great Wall between city and country.

Key words: Second Enlightenment; Aesthetic Enlightenment; Ecological Civilization, Ecological Aesthetics.

I. The Suppression of Beauty in the First Enlightenment

Beauty in general and natural beauty in particular have been neglected by most of the leading Enlightenment thinkers. Although there was resistance from Romanticism—represented by Jean-Jacques Rousseau and others—the suppression of beauty and natural beauty was an overwhelming theme of the first Enlightenment (17th and 18th centuries). In Sandra Lubarsky’s words, “For the past two hundred years, it must be admitted that we in the West have abandoned beauty. We have allowed it to remain in the background of almost every human endeavor. And in abandoning it, we have made way for a singular value system—economism.”[1]

Lubarsky is not alone. According to Steven Shaviro—the author of *Without Criteria: Kant, Whitehead, Deleuze, and Aesthetics*—even during the twentieth century “most aesthetic theorists and innovative artists tend to disparage the very idea of beauty.”[2] Beauty was either regarded as merely subjective, or it was *only* in the eye of the beholder. Accordingly, aesthetics has become “a poor stepsister, a despised and neglected subject of theoretical discussion in the general field of philosophy, or beauty is only skin deep.”[3] In the words of J. Baird Callicott, “Natural aesthetics is in even worse straits, since what little aesthetic theory there is has centered almost entirely on art: painting, sculpture, architecture, drama, literature, dance,

---

[1] Sandra B. Lubarsky, “The Importance of Beauty.” Conference Paper for International Conference on Moral Education from Process Perspective, November 2010, Claremont, CA.
[2] Steven Shaviro, *Without Criteria: Kant, Whitehead, Deleuze, and Aesthetics*, Cambridge, MA: MIT Press, 2009, p.153.
[3] Sandra B. Lubarsky, “The Importance of Beauty.” Conference Paper for International Conference on Moral Education from Process Perspective, November 2010, Claremont, CA.

music, and more recently, cinema."[4]

While China's situation is quite different from the West, the repression of beauty and natural beauty is similar due to the deep influence of the European Enlightenment, during which Chinese traditional culture suffered a "devastating blow."[5] China's first Enlightenment began with the May 4th movement in 1919. It advocated Democracy and Science, which was called Master Democracy and Master Science. At that time, people in China believed that Science could solve every problem in the universe, including the meaning of life. Like its European teachers, China's first leading Enlightenment thinkers also held a nihilistic attitude toward tradition and the past. In Europe, the past was viewed as the "dark ages." In China, tradition and the past were treated as trash which should be totally and completely abandoned. "Down with Confucianism" was the most famous slogan at that time. Such a radical rupture caused us to sever the intrinsic link to our own tradition, effectively cutting us off from a lot of excellent spiritual resources, such as respect for heaven and awe of the Dao, the concept of harmony in difference, and a deep appreciation of nature and natural beauty.

A story can illustrate this point to some extent. Professor Shen Congwen was one of the most important "native soil" writers in modern Chinese literature. He was famous for describing the natural beauty in his remote hometown, Xiangxi, Hunan Province. But when he applied for an associate professor position at China National Southwestern Associated University (where he taught), he received strong objections from many professors who had studied overseas because of his identity as a "country fellow" and his enthusiastic praise of the natural beauty of the country. One of these professors—Liu Wendian—showed his contempt for Shen with the following words: "Other professors like Chen Yingque deserved 400 (silver dollars), myself 40; Shen Congwen only deserves 4, but even then, I would rather not give him 40 cents."[6]

It is obvious that beauty and natural beauty have been banished in the modern world, both in the West and in China. Why have beauty and natural beauty been forsaken in this way? Some critics seek the reasons inside aesthetics itself. Some ascribe it to the "narrowness of aesthetics," others to a "change of aesthetic taste."[7] Arthur Danto ascribes it to "kalliphobia"—fear of beauty.[8] Although these explanations have great value and can help us to understand the whole picture of the repression of beauty, they fail to explain the root causes of why beauty and natural beauty have been neglected and suppressed in the modern world. The discussion is still about the results rather than the causes. In our opinion, it is the Enlightenment (called the first Enlightenment in this work)—especially its anthropocentric stance and imperialistic attitude toward nature, its nihilistic attitude toward tradition, its blind worship of reason, and its underlying mechanistic worldview—that condemns beauty and natural beauty to a forsaken fate.

The domination of modern or Western reason and the deportation of beauty were synchronous. "The modern era—from the Renaissance to the Enlightenment to socialism, from positivism to scientism, from the Industrial Revolution to the information revolution—has been characterized by rational, structural thinking,"[8] the space for beauty has become smaller and smaller. "Beauty has been sidelined, accused of being merely subjective."[9] Although some scientists reflect enthusiastically on the beauty they experience in their work, with some speaking of "elegance" as an intuitive measure of the veracity of their discoveries, the scientific method itself rules out the explicit search for beauty as part of the quest for knowledge. The result is that in the Enlightenment garden, "pure reason" ruled all the land. It left little space for nature and feeling, not to mention beauty and natural beauty. Anything related to nature was relegated to "the sphere of irrelevance and even crudeness."[10] The poet John

---

4 J. Baird Callicott, "Land Aesthetics," Christopher Chapple, ed. *Ecological prospects: scientific, religious, and aesthetic perspectives.* New York: SUNY Press, 1994, p169.

5 Bihui Xu, "On the Rebirth of Traditional Chinese Aesthetics." *Beauty & Times* 7 (2002):4-10.

6 Yue Nan, "The One Century Life of Chinese Culture Master Liu Wendian." *Beijing Evening News*, May 7, 2014.

7 Ronald Hepburn, " Contemporary Aesthetics and the Neglect of Natural Beauty" in Allen Carlson and Arnold Berleant, *The Aesthetics of Natural Environments*. Broadview Press, 2004. p.45. Arthur C. Danto, "Kalliphobia in Contemporary Art." *Art Journal*. Vol. 63, No. 2 (Summer, 2004): pp. 24–35.

8 Patrick *Slattery, Curriculum Development* in the *Post-modern Era*. New York: *Garland* Pub, 1995, p. 24

9 Sandra B. Lubarsky, "The Importance of Beauty." Conference Paper for International Conference on Moral Education from Process Perspective, November 2010, Claremont, CA.

10 Charlene Spretnak, *The Resurgence of the Real*, New York: Routledge, 1999, p.143.

Keats's indicts mechanistic philosophy in his *Lamia,* in which he writes that "all charms fly at the mere touch of cold philosophy."

Since the assumption that "the ultimate units of nature are vacuous actualities" has been "so widespread in modern thought,"[11] almost all leading Enlightenment thinkers believed that reality is composed mostly of lifeless materials. To picture the world as a machine is to conceive of an entirely new world order. This new mechanistic worldview in turn endorses the Enlightenment's imperialistic attitude toward nature based on anthropocentrism, and treats the natural world merely as a resource for human purposes. The mechanistic worldview ruled beauty out in principle. "As beauty is reduced to something figuratively in the 'eye [really in the mind] of the beholder,' so all values are literally figments of mentality, in this worldview."[12] It is clear that as long as the imperialistic attitude toward nature reigns, it is impossible to take a stance in favor of beauty and natural beauty.

The consequence of suppressing beauty and natural beauty is very destructive, and to use Alfred North Whitehead's language, is even "a disastrous error" which modern civilization has made. It not only has caused the exploitation of nature and the subsequent ecological crisis, but also led to the downplaying of human feelings, and caused us to become a "homeless" and rootless people. In James Hillman's opinion, it is the absence of beauty that results in the ugliness that prevails in our modern society. For him, almost all of the causes of major social, political, and economic issues of our time can "be found in the repression of beauty."[13] In other words, the Enlightenment's repression of beauty should be seen as responsible for the ecological, social and, spiritual crisis facing us today. In turn, the modern crisis has resulted in the crisis of modern aesthetics or "the defeat of modern aesthetics."[14] In order to stay away from ecological catastrophes, create an ecological civilization, and allow for a resurgence of beauty and aesthetics, a Second Enlightenment whose core concept is beauty is needed.

II. The Second Enlightenment is an Aesthetic Enlightenment

The Second Enlightenment is a deep reflection and an immanent transcendence of the first Enlightenment (the 17th and 18th century), which stressed faith in reason and science. The Second Enlightenment can be called "constructive postmodernism," since it includes but also transcends the greatest achievements of the first or modern Enlightenment, including its emphasis on human rights, which "made possible enormous advances towards social justice."[15] Also, the Second Enlightenment is a coming together of certain excellent aspects of Western and Eastern culture. If the first Enlightenment was a solo played by Western culture, then the Second Enlightenment is a symphony played by both Western and non-Western culture. Such an Enlightenment creates space for valuable traditions of the West, such as respect for science and innovation, individual worth, and encouraging citizens to participate in policy making.[16] Meanwhile, traditional Chinese values such as harmony, *Ren*, and organic consciousness would be revitalized in order to help heal the ills of the modern age.[17] Among these values, dynamic harmony and beauty are what the Second Enlightenment thinkers particularly pursue. They prioritize harmony and beauty as ideals for human life.

It is worth mentioning that the Second Enlightenment thinkers have no intention of either replacing Western culture with Eastern culture or vice-versa. What they are interested in is the complementarity and mutual flourishing of the two cultures. In the words of Daiyun Yue, a contemporary pioneer of the Second Enlightenment in China, the goal of the Second Enlightenment is to "solve the common problems bothering both by

---

11 David Ray Griffin, *Reenchantment without Supernaturalism*, Ithaca and London: Cornell University Press, 2001, p.97.

12 Frederick Ferre, *Being and Value*. State University of New York Press, 1996, p. 11.

13 James Hillman, "The Practice of Beauty," in *Uncontrollable Beauty: Toward a New Aesthetics*, Bill Beckley and David Shapiro, eds. New York: Allworth Press, 1998, p.265.

14 *Masahiro Hamashita*: "*Paradox of Eco- Aesthetics*: Defeat of Modernism, Revival of the Pre-modern and Prospect of Post-modernity." Qingben Li, ed. *Ecological Aesthetics Abroad: A Reader*. Changchun: Changchun Press, 2010, p.241.

15 Mary Midgley, *Animal and Why They Matter*. University of Georgia Press, 1998, p. 51.

16 Jay McDaniel, "*Mei's Invitation*: A Gentle Asceticism for Chinese and Americans." *Cross Currents*, vol. 57, Winter 2008.

17 Meijun Fan, *Contemporary Interpretation of Traditional Chinese Aesthetics.* Beijing: Peking University Press, 2006, p. 17.

mutual complementarities and reaching a consensus."[18]

What are the defining features of the Second Enlightenment? In our book, *Second Enlightenment*, we outlined seven defining features of the Second Enlightenment: 1) Beyond anthropocentrism to ecological awareness; 2) Beyond Western-centeredness to complementary awareness of China and the West; 3) Challenging homogenous thinking and appreciating the beauty of diversity; 4) From abstract freedom to deep freedom with responsibility; 5) Disregarding a single form of democracy and moving toward a democracy; 6) Challenging hegemonic science and moving toward a science with compassion; 7) From pure reason to aesthetic wisdom.[19]

Among the seven features, the core features are deep appreciation of beauty and harmony. In this sense, we would like to call the Second Enlightenment "an aesthetical Enlightenment." The Second Enlightenment is deeply aesthetical in three respects: 1) It regards beauty as its core and fundamental idea; 2) It espouses an aesthetic wisdom; 3) It deeply appreciates harmony, and stresses the importance of harmony.

First of all, the Second Enlightenment regards beauty as its core and fundamental idea. In contrast to the first Enlightenment, which undervalued beauty and treated beauty and natural beauty as something irrelevant, the value of beauty becomes significant. The Second Enlightenment thinkers challenge us to rethink the value and importance of beauty and natural beauty. Accordingly, they challenge us to "revalue the importance of aesthetics."[20]

For Second Enlightenment thinkers,[21] beauty is universal. It "touches every place."[22] For Whitehead, the well known pioneer of the Second Enlightenment, "beauty is a grand fact in the universe."[23] Beauty is not only the aim of every actual entity, but is also the aim of the whole universe. "The teleology of the Universe is directed to the production of Beauty."[24] This means that the dynamic process of the universe is not aimless; instead, "every process of becoming aims at the achievement of beauty."[25] Frederic Ferre coined the term *kalogenesis* to describe the generation of beauty. According to his interpretation, every actual entity is intrinsically kalogenic: "Every pulse of actualizing energy represents in itself an act of kalogenesis."[26] This means that in the process of becoming actual, they "give rise to beauty."[27] The whole universe consists of kalogenic entities, and the combinations of these entities are just "the by-product of beauty."[28] In this sense, scholars such as Brian Henning argue that "Whitehead's Metaphysics is an aesthetic because it attempts to give a consistent, adequate, coherent, and applicable account of process, a process that is an achievement of beauty."[29] To Whitehead, beauty is "wider" and "more fundamental than truth." Truth would sink to triviality without beauty. "Truth matters because of Beauty."[30] "Beauty is the internal conformation of the various items of experience with each other, for the production of maximum effectiveness. Beauty thus concerns the interrelations of various components of reality, and also the interrelations of various components of appearance, and also the relations of Appearance to Reality."[31] For Charles Hartshorne, "in a generalized sense" aesthetics has to do with "what makes experiences good in themselves."[32] For Sandra Lubarsky, beauty is not something trivial that should only be left to beauty salons and fashion magazines, or only exist in the

[18] Daiyun Yue, "How Does Chinese Culture Face the World?" *Chinese Culture Newspaper*, August 7, 2012.
[19] Zhihe Wang, Meijun Fan, *Second Enlightenment*. Beijing: Peking University Press, 2011, pp.24-38.
[20] *Masahiro Hamashita*: "*Paradox of Eco-Aesthetics*: Defeat of Modernism, Revival of the Pre-modern and Prospect of Post-modernity." Qingben Li, ed. *Ecological Aesthtics Abroad: A Reader*. Changchun: Changchun Press, 2010, p.244.
[21] Here we use the term "Second Enlightenment Thinkers" to refer to those process-relational oriented thinkers worldwide who constructively reflect on the first enlightenment, although not all of them consciously identify themselves as "Second Enlightenment Thinkers."
[22] Stephen David *Ross*, *The Gift of Beauty*, State University of New York Press, *1996*, p.296.
[23] Whitehead, *Modes of Thought*. New York: The Free Press, 1981, p.120.
[24] Whitehead, *Adventures of Ideas*. New York: The Free Press, 1967, p.265.
[25] Brian G. Henning, *The Ethics of Creativity: Beauty, Morality, and Nature in a Processive Cosmos*. University of Pittsburgh Press, 2005, p.100.
[26] Frederick Ferre, *Being and Value*, p.358.
[27] Frederick Ferre, *Being and Value*, p.340.
[28] Frederick Ferre, *Being and Value*, p.358.
[29] Brian G. Henning, *The Ethics of Creativity: Beauty, Morality, and Nature in a Processive Cosmos*. University of Pittsburgh Press 2005, p.100.
[30] Whitehead, *Adventures of Ideas*. New York: The Free Press, 1967, p. 267.
[31] Whitehead, *Adventures of Ideas*. p.265.
[32] Charles Hartshorne, *Creative Synthesis and Philosophic Method*, Open Court Publishing Company, 1970.p. 303.

ivory tower of art or in the malls of commerce; it is "the primary cosmological value," an intrinsic "part of the structure of life."[33] Individual action has value only when it is driven by beauty and oriented toward beauty. Therefore, when we exclude aesthetics from our ordinary experience, we actually limit our way of understanding a very deep and intimate part of our lives.

Secondly, the Second Enlightenment espouses an aesthetic wisdom. This is unlike the first Enlightenment, which worships reason, especially pure reason or instrumental reason. Due to the fatal flaws of pure or instrumental reason and technological rationality, some thinkers have proposed a new concept to replace or transcend them. Ulrich Beck, a noted German sociologist, suggested the notion of "ecological rationality," and Herbert Marcuse advocates the use of "aesthetic rationality." We prefer "aesthetic wisdom," not only because beauty is wider than truth, but also because reason which pictures the world as an organic whole is no longer reason in the traditional sense—it already ascends to wisdom, a wisdom that emphasizes the harmony between parts and whole, community and individual, humankind and nature.

By aesthetic wisdom, we refer to an integrative thinking based on the conception of organic interrelatedness, which aims at harmonizing truth, good, and beauty. In aesthetic wisdom, not only scientific rationality and cognitive reason, but also artistic intuition and religious experience can find room to freely develop. All of these modes of human experience can complement and enrich each other. This means that in aesthetic wisdom not only truth, good, and beauty, but also scientific thinking, rational thinking, sensual thinking, artistic thinking, and religious thinking can find their space to freely develop and to complement and enrich each other.

Appreciating feeling can be regarded as a defining feature of aesthetic wisdom. For Enlightenment rationalism, feeling is something we should get rid of. Therefore, feelings and emotion have been deeply repressed. However, for Second Enlightenment thinkers, emotions are what matter most in our lives. In Charles Birch's words, "Our feelings are the most important aspect of our life. When there is no feeling life has lost its value."[34] To these thinkers, the universe is "the oceans of feelings,"[35] and we are emotional beings. It is "feeling" that "constitutes the core of human growth and social development."[36] They are convinced that wisdom requires not only the preciseness of reason and the breath and depth of thought, but also an emotional dimension that wisdom cannot achieve if it lacks feeling and respect for others. If it lacks empathy and compassion, so does civilization.

Deeply influenced by binary thinking, the reason of the Enlightenment acted on the premise of suppressing feeling, emotion, and sensitivity. It is no big secret that Enlightenment rationalism treated feeling and sensitivity as "merely confused forms or stimulators of the intellect."[37] For Descartes feelings were "corruptions of pure perception."[38] For Kant, as Charlene Spretnak pointed out, "'pure reason' is valuable precisely because it is 'untainted' by emotions, sensate knowledge, social constructions, and non-cognitive awareness."[39] It is apparent that the triumph of reason was at the expense of sensitivity. The triumph of reason and the suppression of feelings have been parallel. That partly explains why mainstream modern philosophers have not had much to say about feelings. On the contrary, "the principal goal of mainstream modern thought in philosophy has been to exile feelings from the field of study."[40] This kind of repression of feelings is undoubtedly harmful. It not only paralyzes our ability to sympathize with others, but also harms the full development of the global community.[41]

In contrast to the Enlightenment rationalism that favors either/or thinking, aesthetic wisdom espouses an integrative thinking based on the concept of organic interrelatedness, which aims at harmonizing truth, good, and beauty. Therefore, it holds an appreciative attitude towards feeling and sensitivity in terms of both/and harmonious thinking.

[33] Sandra B. Lubarsky, "The Importance of Beauty." Conference Paper for International Conference on Moral Education from Process Perspective, November, 2010, Claremont, CA.

[34] Charles Birch, *Feeling*. Sydney: University of New South Wales Press, 1995, p.35.

[35] Whitehead, *Process and Reality*, p. 166.

[36] Joseph Grange. *John Dewey, Confucius, and. Global Philosophy*. State University of New York Press, 2004, p. 81.

[37] Charlene Spretnak, *The Resurgence of the Real*, New York: Routledge, 1999, p. 55.

[38] Charlene Spretnak, *The Resurgence of the Real*, p.137.

[39] Charlene Spretnak, *The Resurgence of the Real*, p.220.

[40] *Charles Birch*, *Feelings*, p. 5.

[41] Wendy Wheeler, *A new modernity*? London,: Lawrence & Wishart Limited, 1999, p. 160.

It believes that the essence of Dao lies in synthesizing and harmonizing seeming opposites. The Second Enlightenment does not reject reason as such; what it rejects is the hegemony of reason. This hegemony represents the triumph of reason over sensibility, of rationality over feeling. For Second Enlightenment thinkers, there is no need to choose sides in the quarrel between reason and feeling. They tend to marry reason to feeling and appreciate both of them. They espouse a dynamic harmony of reason and sensibility.

Starting from such an inclusive, open, and harmonious stance, aesthetic wisdom regards feelings, value, and beauty as complementary partners rather than enemies. Artistic intuition, religious experience, sensitivity, feeling, values, and beauty can complement and enrich scientific rationality and cognitive reason. Like Yin and Yang, which "nurture and help complete each other,"[42] reason and sensitivity are relational polarities. They are intrinsically related to each other, complement each other, and enrich each other. Together, they constitute a dynamic harmony. It is beyond doubt that such an aesthetic wisdom as a combination of Western and Eastern wisdoms is needed by China and the world today if we are to move beyond the shallowness of consumerism into a more meaningful way of living.

Thirdly, the Second Enlightenment deeply appreciates harmony and stresses its importance, since "the beautiful is about harmony and proportion,"[43] and "beauty is nothing other than the 'aesthetic continuum of harmonious interpretation between the many and one' which lies at the base of every moment of our experience."[44] Thus, beauty is inextricably linked to harmony. Harmony is essential to us. As Hartshorne explained, "for Whitehead, the value of truth consists in its contribution to harmony, unity in variety."[45] Harmony can be regarded as "aesthetic teleology,"[46] because by its nature, "beauty involves contrasting, even conflicting, elements but resolves their conflict by making them internal to satisfactory synthesis that retains the elements, adjusted to one another so that they enhance the whole."[47]

The Second Enlightenment's appreciation of harmony is a rejection of first Enlightenment emphasis on competition and struggle. Due to the worship of possessive individualism, upholding struggle and despising harmony has been a main theme of the first Enlightenment. When Thomas Hobbes claimed that man's state of nature was a state of war of all against all, where "man is wolf to man,"[48] he actually laid a theoretical foundation for the worship of struggle which has influenced the modern world for centuries, from Social Darwinism, to Marx's theory of class struggle, to Mao's favorite motto: "Battling the heavens brings unlimited pleasure, battling the earth brings unlimited pleasure, battling human beings brings unlimited pleasure." All of these concepts are products of "struggle thinking," which has reached its peak in the United States. Competition is often conceived of as America's "official state religion."[49] Famous American football coach Vince Lombardi's statement that "winning isn't everything; it's the only thing"[50] reflects the American obsession with contest and struggle. In Alfie Kohn's words, "not only our economic system, education system, but also our family lives are full of fighting. Others are regarded as 'obstacles' to our success."[51]

This sort of thinking helps to explain Samuel P. Huntington's theory of civilization, where conflict prevails. It also explains why the United States ranks first in military expenditure, weapons exports, and international violence. In 2013, according to a report by the Stockholm International Peace Research Institute, the U.S. government accounted for 37 percent of total world military expenditures, putting it far ahead of all other nations (the two closest competitors, China and Russia, accounted for 11 percent and 5 percent

[42] Whitehead, *Process and Reality*, p. 185.

[43] Steven Shaviro, *Without Criteria: Kant, Whitehead, Deleuze, and Aesthetics.* Cambridge, MA: MIT Press, 2009, p.153.

[44] David Ray *Griffin, "*Introduction: *Sacred Interconnections*." In *Sacred Interconnections: Postmodern Spirituality, Political Economy, And Art.* ed. David Ray *Griffin*. Albany: SUNY Press, 1990, p.11.

[45] Charles Hartshorne, *Whitehead's Philosophy*, p.74.

[46] Charles Hartshorne, *Whitehead's Philosophy,* p 82.

[47] Frederick Ferre, *Knowing and Value: Toward a Constructive Postmodern Epistemology.* Albany: State University of New York *Press. 1998*. p.366

[48] Thomas Hobbes, *Leviathan*, ed. Edwin Curley. Indianapolis: Hackett Publishing Company, 1994, p.76.

[49] Paul *Wachtel,* The *Poverty* of Affluence: A psychological portrait of the American way of life. New York: Free Press, 1983, p.284.

[50] Alfie Kohn, No Contest: The Case Against Competition. Boston: Houghton Mifflin, 1992, p.3.

[51] Alfie Kohn, No Contest : The Case Against Competition. Boston: Houghton Mifflin, 1992, p.2.

respectively).[52]

Against the first Enlightenment's worship of fighting—which has caused a great deal of suffering in both human society and the natural world—Second Enlightenment thinkers deeply appreciate and put great emphasis on harmony. In Whitehead's writings, so-called beauty is, in essence, harmony. Beauty is defined by Whitehead as "the mutual adaptation of the several factors in an occasion of experience."[53] "The perfection of Beauty is defined as being the perfection of Harmony."[54]

It is important to note that the harmony the Second Enlightenment thinkers aspire to is not sameness. It is instead a harmony that respects and takes delight in differences. Although Whitehead spoke highly of the ancient Greek theory of "Beauty is harmony" and called it a great discovery, "a landmark in the history of thought,"[55] for him static harmony is "a debased type of harmony,"[56] because it is a homogeneous harmony that does not tolerate difference. Such a harmony is often achieved at the cost of sacrificing individuals.

This kind of harmony is not what Second Enlightenment thinkers pursue. What they appreciate is a dynamic harmony which does not rely on suppressing individuals. It holds a very positive attitude toward diversity in human life, including ethnic, racial, sexual, cultural, and religious differences. It not only pays respect to diversity, but also appreciates and "honors" diversity.[57] It emphasizes the mutual flourishing of individuals and community. It is a comprehensive, inclusive harmony in which individuals own their own unique beauty, but at the same time "lend themselves to the beauty of the whole."[58]

Unlike static harmony, the harmony the Second Enlightenment pursues is a dynamic harmony which is always in process. It is fluid and evolves over time. Therefore, it is open to adventure and appreciates discord. Whitehead once warned that the Gospel of Uniformity is dangerous. To him, "the differences between the nations and races of mankind are required to preserve the conditions under which higher development is possible."[59] "A diversification among human communities is essential for the provision of the incentive and material for the odyssey of the human spirit. Other nations of different habits are not enemies: they are godsends. Men require of their neighbors something sufficiently akin to be understood, and something sufficiently different to provoke attention, and something great enough to command admiration."[60] In *Ideas of Adventure*, he stressed that "progress is founded upon the experience of discordant feelings. The social value of liberty lies in its production of discords." "The value of discord is a tribute to the merits of imperfection."[61] "Even perfection will not bear the tedium of indefinite repetition. Thus, 'adventure is essential.'"[62] This means the search for new perfection is essential, which means the search for a new harmony. No beauty is immortal; moving from harmony to disharmony, from disharmony again to a new harmony is the moving rhythm of the universe. If people are obsessed with such "merely qualitative harmony," civilization will decay.

III. The Relevance of the Second Enlightenment to China and the World

Given the serious situation facing human beings, it is beyond doubt that our time needs such a Second Enlightenment, an aesthetic enlightenment. It could play an instrumental role in leading us out of various modern crises and toward an ecological civilization.

1) First of all, the Second Enlightenment can help advocate an ecological consciousness.

Due to its emphasis on the universality of beauty, the Second Enlightenment challenges us to pay attention to the beauty which is immanent in nature. It makes us realize that nature is not comprised of insentient, "vacuous actualities."[63] Its existence is not merely for us to use. On the contrary, nature has its own value, feeling, dignity, and beauty. Therefore, we human beings should

52 Lawrence S. Wittner, "The United States is No. 1 — But in what? Military Expenditure, Weapons Export and International Violence," Global Research, October 13, 2014, http://www.globalresearch.ca/the-united-states-is-no-1-but-in-what-military-expenditure-weapons-export-and-international-violence/5407683

53 Whitehead, *Adventures of Ideas*. p.252.

54 Whitehead, *Adventures of Ideas*. p.252.

55 Whitehead, *Adventures of Ideas*. p.149

56 Whitehead, *Adventures of Ideas*. p. 264.

57 Charlene Spretnak, "Postmodern Directions." In David Griffin, *Spirituality and Society: Postmodern Visions*, p. 39.

58 Adventure p.264.

59 Whitehead, *Science and Modern World*, Free Press, 1997, pp.206-207.

60 Whitehead, *Science and Modern* World, p. 207.

61 Whitehead, *Adventures of Ideas*. p. 257.

62 Whitehead, *Adventures of Ideas*. p 258.

63 Alfred North Whitehead, *Process and Reality*, New York: Free Press, 1978, p.167.

learn to pay respect to her, appreciate her, and revere her. In doing so, the Second Enlightenment will help us to overcome anthropocentrism and its manifestations, discard our imperialistic attitude toward nature, and advocate an ecological consciousness.

The ecological awareness the Second Enlightenment promotes regards nature as a "subject."[64] It challenges people to realize that we are "a part of the unfolding process, inherently linked with the stars, the winds, the rocks, the soil, the plants, and the animals. They are not a valueless 'resource' that we are to exploit as detached managers for the sake of a GNP."[65] Unlike the popular concept of environmental protection—which still treats nature merely as an object to be protected by humankind—ecological awareness emphasizes that it is not we who are protecting nature, but rather nature that protects us. Nature is not only the provider of our food and clothing and sustenance for our bodies, but also "nurture for our mind."[66] Therefore, we should not only protect her but also love her, respect her, and be in awe of her.

Such an ecological consciousness helps overcome the alienation between humankind and nature, and calls for our responsibility for the fate of the earth. Unlike the first Enlightenment, which is deeply anti-ecological, the Second Enlightenment engenders both ecological and social responsibility. It would play a crucial part in creating an ecological civilization, a new and higher stage in the development of human civilization.

2) Second, the Second Enlightenment can help us resist modern consumerism and encourage people to lead a poetic life with spiritual dignity. The Second Enlightenment's emphasis on the importance of beauty and prioritizing *beauty* as the ideal for human life rather than material possessions can help us to fight back against the increasingly popular consumerist culture in both the West and China, and then encourage people to lead a poetic life with spiritual dignity.

Consumerism equates life's meaning with the possession of material goods and the pursuit of wealth. The philosophy of consumerism teaches people that they are fulfilled or made whole by purchasing more and more consumer goods each year without ever having to say "enough." It tells them to measure their own worth by how attractive they are by consumer-driven standards, by how much money they have, and by how much status they have in terms of the goods they own. It tells them that they are consumers first and citizens second, such that their private good is more important than the public good. It asserts that high paying jobs are more important than their families, and that paid work is the only work worth doing. It says that life is a race toward material success, in which there must be winners and losers. Consumerism has led many people to find more value in life in material things than, for example, in friendship and family and community. Being materially successful becomes more important than being a good parent, neighbor, and friend. People are more interested in new clothes than old friends. The point is, when life is reduced to things, a sense of emptiness will come.[67]

The Second Enlightenment as aesthetic enlightenment challenges people to move beyond the shallowness of consumerism into a poetic life. A poetic life is a life aimed at beauty, valuing and prioritizing beauty. In the eyes of Second Enlightenment thinkers, it is apparent that copying others' lifestyles and catching up with the Joneses is not a poetic life. A poetic being is one who has a deep feeling that we are small but included in a larger whole called Nature, and that we can gain guidance for our lives by being attuned to nature. Therefore she or he has respect for the earth, our homeland, and other forms of life. A poetic life is full of curiosity, imagination, and sympathy. People who lead poetic lives can be called "poetic beings" who have a sense of the sacred, a sense of the spirituality found in harmony between people and harmony with nature. They regard nature as a subject and are fully aware that we human beings are a part of an unfolding process. They have a clear idea that it is nature that protects us. Nature not only provides our food and clothing and nurtures our body, but also nurtures our mind. Therefore, poetic beings not only enjoy and appreciate the beauty of nature, but also *add beauty to the world. Although poetic beings live in a harmonious relationship with nature and the social community, they make every*

[64] Charlene Spretnak, *The Resurgence of the Real*, New York: Routledge, 1999, p.73.

[65] Charlene Spretnak, "Postmodern Directions." In David Griffin, *Spirituality and Society: Postmodern Visions*. Albany: State University of New York, 1988, p. 37.

[66] Meijun Fan, *Contemporary Interpretation of Traditional Chinese Aesthetics*, Peking University Press, 2006, p.18.

[67] Jay McDaniel, "Mei's Invitation." *Crosscurrents* 4, 2008.

effort they can to live out their own lives in their own uniquely beautiful way. A poetic being is also a creative being which sustains a poetic being, because a poetic being relies on unique creativity.

Thus the Second Enlightenment regards creativity as a crucial component of a poetic life. For Second Enlightenment thinkers the universe is creative. This means that the universe unfolds moment by moment in a way that is not entirely reducible to the influence of the past. At any given moment, a human being is prompted not only by impulses from the past, but also by possibilities for the future, which she or he can actualize in the moment itself. For David Griffin, as part of the universe we human beings are "essentially creative beings."[68] That is, to live is "to be creative."[69] This explains why Whitehead felt sick of closed-mindedness and the "decadent habit of mind" in his times. To Whitehead, "the most un-Greek thing that we can do is to copy the Greeks."[70] The Greeks were not copyists, but were adventurous, "eager for novelty."[71] From the point of view of the Second Enlightenment, to be a creative being means making a difference. Therefore a creative being highly values and appreciates diversity. This can be an effective remedy to consumerism.

3) The Second Enlightenment as aesthetic enlightenment can help us overcome the defiant attitude toward rural civilization and farmers, and eventually remove the great Wall between city and country.

While the first Enlightenment promoted industrial civilization, it was at the expense of destroying rural and community life. The Second Enlightenment is dedicated to preserving and reconstructing the countryside, not only economically, but also aesthetically. It aims at building an aesthetic countryside.

Such an aesthetic countryside can not only elevate the quality of the farmers and villagers' lives, but also their pride. It will make them as well as city citizens delighted to live in the country. Also, rural areas will become the aesthetic base for aesthetic education, because country life, as Ralph Waldo Emerson said, is full of "the beautiful"—richness, joys, and wisdom.[72] In the eyes of Wei An, a Chinese eco-author, "wheat is the most beautiful, most elegant, and most touching of crops."[73] Of course, there will be more and more young people who do not want to be mortgage slaves, and would like to go to work and live in the beautiful countryside. The place Heidegger longed for in which people can poetically dwell cannot be the urban jungle of armored concrete, but must be beautiful rural land which is rich and solid. People are more likely can find the sense of belonging they need in the countryside. In a rural setting, one's life has a place, and one's spirit has its trust. In other words, in the countryside, people are more likely to find their "spiritual dwelling place."[74]

If this happens, then the great Wall between city and countryside which has plagued China for many years would break down naturally. This would be a portion of the fruit of the Second Enlightenment as an aesthetic enlightenment. While it is not an easy task, it is a challenge worth pursuing.

References

Birch C (1995). *Feelings*. Sydney: University of New South Wales Press.

Callicott J. B (1994) Land Aesthetics. In *Ecological Prospects: Scientific, Religious, and Aesthetic Perspectives*, ed. Christopher Chapple. New York: SUNY Press.

Danto A C (2004) Kalliphobia in Contemporary Art. *Art Journal* 63(2): 24–35.

Emerson R W(1904) Country Life. *The Atlantic Monthly* 94: 594–604.

Fan M (2006) *Contemporary Interpretation of Traditional Chinese Aesthetics*. Beijing: Peking University Press.

Ferre F (1996) *Being and Value*. Albany: State University of New York Press, 1996.

Ferre F (1998) *Knowing and Value: Toward a Constructive Postmodern Epistemology.* Albany: State University of New York Press.

Grange J (2004) *John Dewey, Confucius, and. Global Philosophy.* Albany: State University of New York Press.

Griffin D R (1990) Introduction: *Sacred*

[68] David Griffin, Spirituality and Society. Albany: State University of New York Press, 1988, p. 149.
[69] Charles Hartshorne, *Whitehead's Philosophy*. Lincoln: University of Nebraska Press, 1967, p.132.
[70] Whitehead, *The Adventures of Ideas*, p, 274.
[71] Ibid.
[72] Ralph Waldo *Emerson* , "Country Life", *The Atlantic* Monthly 94, (1904): p. 594-604.
[73] Wei An, *The Things on Land*. Beijing: China Translation & Publishing Corporation, 1995, p. 13.
[74] Frederick Ferre, *Being and Value,* p. 370

Interconnections. In *Sacred Interconnections: Postmodern Spirituality, Political Economy, And Art.* ed. David Ray Griffin. Albany: SUNY Press.

Griffin D R (1988) *Spirituality and Society*. Albany: State University of New York Press.

Hartshorne C (1970) *Creative Synthesis and Philosophic Method.* Open Court Publishing Company.

Hartshorne C (1967) *Whitehead's Philosophy, Whitehead's Philosophy*. Lincoln: University of Nebraska Press.

Henning B G (2005) *The Ethics of Creativity: Beauty, Morality, and Nature in a Processive Cosmos.* University of Pittsburgh Press.

Hepburn R (2004) Contemporary Aesthetics and the Neglect of Natural Beauty. In *The Aesthetics of Natural Environments*, eds Allen Carlson, Arnold Berleant. Broadview Press.

Hillman J (1998) The Practice of Beauty. In *Uncontrollable Beauty: Toward a New Aesthetics*, eds Bill Beckley, David Shapiro. New York: Allworth Press.

Lubarsky S B (2010) The Importance of Beauty. Conference Paper for International Conference on Moral Education from Process Perspective, Claremont, CA.

Kohn A (1992) *No Contest : The Case Against Competition*. Boston: Houghton Mifflin.

Masahiro H (2010) Paradox of Eco- Aesthetics: Defeat of Modernism, Revival of the Pre-modern and Prospect of Post-modernity. In *Ecological Aesthetics Abroad: A Reader,.* ed. Qingben Li. Changchun: Changchun Press.

McDaniel J (2008) Mei's Invitation: A Gentle Asceticism for Chinese and Americans. *Cross Currents* 57.

Midgley M (1998) *Animal and Why They Matter.* University of Georgia Press.

Ross S D (1996) *The Gift of Beauty*. State University of New York Press.

Shaviro S (2009) *Without Criteria: Kant, Whitehead, Deleuze, and Aesthetics*. Cambridge, MA: MIT Press.

Spretnak C (1988) Postmodern Directions. In *Spirituality and Society: Postmodern Visions* ed. David Griffin. Albany: State University of New York.

Spretnak C (1999) *The Resurgence of the Real*. New York: Routledge.

Wachtel P (1983) *The* Poverty *of Affluence: A psychological portrait of the American way of life.* New York: Free Press.

Wang Z, Meijun F (2011) *Second Enlightenment*. Beijing: Peking University Press.

Wei A (1995) *The Things on Land*. Beijing: China Translation & Publishing Corporation.

Wheeler W (1999) *A new modernity*? London: Lawrence & Wishart Limited.

Whitehead A N (1967) *Adventures of Ideas*. New York: The Free Press.

Whitehead A N (1981) *Modes of Thought*. New York: The Free Press.

Whitehead A N (1978) *Process and Reality*. New York: Free Press.

Whitehead A N (1997) *Science and Modern World.* New York: Free Press.

Yue D (2012) How Does Chinese Culture Face the World? *Chinese Culture Newspaper*, August 7.

BALKAN JOURNAL OF PHILOSOPHY. Vol. 7, Issue 2, 2015

REPUBLICANISM: A NEW POLITICAL PHILOSOPHY?

Cristian-Ion Popa
The Institute of Political Sciences and International Relations
of Romanian Academy, Bucharest, Romania
e-mail: popaioncristian2004@yahoo.com

Abstract:

This article examines some aspects of current republican philosophy, starting from its central idea of human freedom—*freedom as non-domination*—then describes in general terms its political agenda, attempting therein to undertake a critical analysis of its position in relation to the prevalent political philosophies of our time—liberalism and social democracy.

Current republicanism is a school of thought cultivated by philosophers and historians of ideas such as Philip Pettit, Quentin Skinner, Frank Lovett, John Maynor, and others who—drawing on the classical republican tradition from Cicero to James Madison—propose a new political philosophy as a way to establish and legitimate some general, "ecumenical" positions on the major policy issues of contemporary societies.

At its core, the central objective of this political philosophy is to rethink almost all substantial issues of legitimacy and democracy, welfare and social justice, public policy and constitutional design from within the conceptual framework that its central idea of liberty as non-domination provides. Formally, republicans aspire to articulate a universal language, a political *lingua franca* in which all individuals and social groups are able *to have a voice* in the current and ardent political debates, especially in order to express their diverse "complaints" or "grievances." For unless different individuals and groups can find a common language in which to talk about their problems, the complaints of each group are nothing but noise in the ears of the others; in effect, they would have the status, and the (in)significance, of inarticulate mewlings.

The paper will argue that republicanism, consistent with its (overly) "ecumenical" propensity, risks overshadowing its *sui generis* existence by being insufficiently different from Socialism, as well as by its exaggerated insistence on providing a comprehensive alternative to Liberalism.

Key words: republicanism, liberalism, social democracy, non-domination, non-interference, social protection.

Introduction

This paper examines some aspects of current republican philosophy, starting from its central idea of human freedom—*freedom as non-domination*—then describes in general terms its political agenda, attempting therein to undertake a critical analysis of its position in relation to the prevalent political philosophies of our time—liberalism and social democracy.

Current republicanism is a school of thought cultivated by philosophers and historians of ideas such as Philip Pettit, Quentin Skinner, Frank Lovett, John Maynor, and others who—drawing on the classical republican tradition from Cicero to James Madison—propose a new political philosophy able to establish and legitimate some general, "ecumenical" positions on the major policy issues of contemporary societies.

The theoretical background of this philosophy is made up of political ideas that emerged in the ancient Roman Republic, ideas that were subsequently appropriated and developed by political thinkers of medieval and Renaissance Italy, Europe in the seventeenth and eighteenth centuries, and the "Founding Fathers" of America. The emblematic figures of this "classical republican" (or "Neo-Roman") tradition of thought are Niccolò Machiavelli, English republicans John Milton, James Harrington, and Algernon Sidney, Montesquieu, and the founders of the American Republic—Thomas Jefferson and James Madison.

Quid prodest a (new) political philosophy? For according to a widespread mindset—widespread even in some intellectual circles—*normative philosophical ideas* (or "political ideals") are barely relevant (or even futile) in ordinary political activity, which is fundamentally determined by the *interests* of participants, both for elected politicians and citizen-voters. According to this superficial mode of thinking, ideas and interests seem mutually exclusive: if a political action or decision is motivated by interests, it could not have been inspired by an "idea" or even a "political ideal." However, in current electoral democracies elected politicians seek the *support* of their voters to pursue policies, and they can get such support if they describe these policies as *legitimate*—that is, based on a normative, general commitment. Even the secret police of former communist regimes have from their inception continually nourished and cultivated a certain faith in the legitimacy of the political and "constitutional" order that defended them: collective ownership, the "state-party" (an oxymoronic phrase: a part can never be the whole!), the myth of the wise and benevolent Great Leader, the soteriological promise of a general welfare postponed to an uncertain future, etc.

As can be seen, general political ideas—even if they are wrong—play an important role in political life. More or less explicitly, current policy-makers often express such general ideas to *legitimate* public policies, which they promote in order to obtain the *consent* of their fellow citizens. But to reiterate a *locus communis*, they are often speaking in prose without knowing it, for these ideas have already become truly multiple, and rival *languages* or *discourses* for legitimating institutional and constitutional arrangements have already arisen in these contemporary plural societies. The role of political philosophy is precisely to systematically examine these languages of debate and political legitimacy in order to reveal the tacit assumptions on which they are based, to analyze their plural "grammar" and even "semantics," and to determine their internal consistency and their coherence with one another (both for contemporary theories and those of other times and places), with the ultimate goal of grounding the more comprehensive theoretical terms of political debate. As the most prominent current republican philosopher Philip Pettit says, without such a systematic reflection on the language in which it takes place, political dialogue remains a "Babel of dogmatic assertion and counter-assertion. If political philosophers did not exist, we would have to invent them."[1]

Therefore, republican theorists aspire to articulate a universal language, a political *lingua franca* in which all individuals and social groups are able to have a voice in the current and ardent political debates, especially in order to express their diverse "complaints" or "grievances." For unless different individuals and groups can find a common language in which to talk about their problems, the complaints of each group are nothing but noise in the ears of the others; in effect, they would have the status, and the (in)significance, of inarticulate mewlings.[2]

The republican idea of freedom

In the most general philosophical terms, the current republican literature develops three main aspects of the classical idea of *Res publica* (or *Commonwealth*): *free individuals*, who are not subject to any form of "domination" exercised by their fellows; *the free state*, which is designed to prevent its members from dominating one another, but which itself does not dominate them in any way; and *the civism*, which is the commitment of citizens to remain constantly vigilant in maintaining the state in its role of defender and guardian of freedom, thus understood. The central objective of this philosophy is to "to rethink issues of legitimacy and democracy, welfare and justice, public policy and institutional design, from within the framework that these basic ideas provide."[3]

According to the most representative contemporary authors of this school of thought—Philip Pettit, Frank Lovett, Quentin Skinner, and others—republicanism defines freedom in a negative and basic mode as being essentially the absence of *domination* or *dependence* in the relationships of citizens. "This idea of freedom as non-domination has become the crucial unifying theme for those who work within the neorepublican framework…"[4]

What does it mean, however, to say that a person is (or should be) free? According to a widespread view theorized in the seventeenth and eighteenth centuries by authors such as Thomas

[1] Philip Pettit (1997) *Republicanism: A Theory of Freedom and Government*. Oxford University Press, Oxford, UK, p. 4.

[2] *Ibidem*, p. 130.

[3] Frank Lovett and Philip Pettit (2009) Neorepublicanism: A Normative and Institutional Research Program. *Annual Review of Political Science* 12: 2.

[4] *Ibidem*, p. 2.

Hobbes and Jeremy Bentham, a person is free if her actions or "choices," whatever they may be, are not the subject of "interference" by others, i.e. if she is simply left alone to do what she wants. For example, some slaves may have a benevolent master who leaves them in peace, despite the fact that he has the power to treat them harshly or tyrannically. From this perspective, we might say that *the slaves are free in their slavery*, which is quite contradictory, because no matter how gently they are treated, the slaves remain under their master's supervision and control, with the master able to impose his will on their will in any way and at any time that he wishes. Therefore, whether the master is good or bad in any imaginable degree, slavery is "the quintessential state of unfreedom."[5] This is why in the traditional or classic republican view freedom is not (only) the absence of interference, but the lack of a need for obedience to any individual (or group) which acts as a master: *freedom is the absence of domination*. John Locke, one of the great founders of liberalism—which generally associated freedom with non-interference—illustrated this republican conception in an exemplary way: freedom consists not in "a liberty for every one to do what he lists," but rather in being not "subject to the inconstant, uncertain, unknown, arbitrary will of another man."[6]

In a study that greatly influenced the contemporary philosophical debate on this topic, Isaiah Berlin distinguished between two main conceptions of liberty: "negative" and "positive." The first conception typically describes non-interference: I am free, in a negative sense, "to the degree to which no man or body of men interferes with my activity."[7] The second conception describes a sort of "self-mastery" in and through the Society or State, referring—in the writings of J. J. Rousseau, for example—to the active participation of individuals in the formation of a "general will" of society. "We (collectively) are free, on this view, only when our (individual) passions and desires are governed by laws and constraints that we (collectively) have actively chosen for ourselves."[8]

However, "positive" liberty—as Berlin convincingly argued—requires the coercive intervention of the collectivity or the state on the individuals' activities in the name of their interests, collectively defined, and thus *hypostatized* as truly real or authentic. Republicanism does not support "positive" freedom so understood. "The republican view is negative in the formal sense that it characterizes freedom as the absence of something—namely, the absence of domination."[9] From the days of ancient Rome up to the present time, freedom means that you are not under the power of a master—*in potestate domini*—a condition in which you can do certain things or not only according to his will (*cum permissu*).

On the other hand, from the republican perspective interference is not always arbitrary. A republic is (should be) "the empire of laws and not of men," classics said, so that the laws, although restrictive, may not be arbitrary if they are enacted with the consent of people in order to protect themselves from domination—whether it is mutual, internal, or exercised by foreign rulers or people from other societies. In the words of John Locke, "the end of Law is not to abolish or restrain, but to preserve and enlarge Freedom."[10]

Therefore, a person is free if she is not subject to domination, but domination can exist even in the absence of actual interference, and interference can exist without being arbitrary and therefore dominant. In republican sense, however—and beyond this abstract level of analysis—freedom can be exercised only if it is "*institutionalized* in an appropriate manner,"[11] and, we might also add, if it is *generalized* for the benefit of all members of society as equal citizens. In the early Roman Republic, as Machiavelli noted, freedom was not fully accomplished because the *plebeians* were deprived of civil and political rights, and this situation

[5] *Ibidem*, p. 4.
[6] John Locke (1980) *The Second Treatise of Government* (1690), edited by C.B. Macpherson. Hackett Publishing Company, Indianapolis, p. 17.
[7] Isaiah Berlin (1958) *Two Concepts of Liberty*. Oxford University Press, Oxford, p. 7.
[8] Frank Lovett and Philip Pettit, *op. cit*., p. 5.
[9] *Ibidem*, p. 5.
[10] John Locke, *op. cit*., p. 32.
[11] Frank Lovett and Philip Pettit, *op. cit*., p. 6, italics mine.

changed only later, with the extension of the equal citizenship of all.[12]

The republican policy agenda

What public policies typically promote this philosophical republicanism? “If we take *the goal of the state* to be the promotion of freedom as non-domination—if we reappropriate the republican ideal—then we can begin to see the outlines of a commanding political philosophy,”[13] says Philip Pettit in his foundational book. This philosophy, which seems to *reify* the “State”—implicitly presenting it as *sui generis* with a self-reliant existence, beyond and even above the civil and economic society—claims that (almost) all *social demands* should be transformed into public policies undertaken by government. It attempts to provide a general, “ecumenical” justification for all these fragmented, “sectional,” even contradictory demands, based on a critical, *contestatory* conceptual predisposition for democracy: whether or not the government does what its citizens want, for these republican philosophers it is important for people to challenge government policies.

Their philosophical commitment to the ideal of freedom as non-domination contains in itself some specific policy suggestions which are easy to illustrate througha comparison with what republicanism has declared to be a rival political program: contemporary liberalism, particularly the right-of-centre alternative of so-called contemporary libertarianism. According to the latter, the state, through its policies, should be concerned almost exclusively with the promotion of freedom as non-interference, the old political ideal of classical liberalism. By contrast, republicanism—concerned with the elimination (or reduction) of all possible forms of domination—conceives *freedom* together with other values (considered equally important) such as *equality* and *welfare*, so that the public policies inspired and supported by republicans are much closer to left-of-centre liberalism.[14]

In addressing the issues of material and social protection, for example, there is considerable overlap with the theory of “equality of basic capabilities” developed by Amartya Sen,[15] arising from the convergence of general conceptions of human freedom. In this context, however, republicanism declares its merit for cultivating a single supreme value—non-domination—by the manner in which it designs institutional and constitutional arrangements, and in which elaborates its public policies.

From this perspective, the “infrastructure” of non-domination must include an active civil society, a strong economy, a rule of law in which citizens can trust, a culture of general civic trust, and a healthy and sustainable environment.

Furthermore, a republican state should develop public policies in the domains of education, health care, and social insurance, protecting its citizens from possible domination imposed by some private, onerous “engineering”; in the sphere of the “vulnerability” of certain members of society, such as workers, women, and minorities; and in the domains of domestic crime, external threats by other states, exploitative corporations, and terrorist networks.[16]

The main claim of workers’ movements in the nineteenth century—whether they were radical, Marxist-inspired, or more moderate or socially democratic—was that they needed liberation from “exploitation”—or even from the “new slavery”—of industrial wage labor.

According to socialists, the very existence of the capitalist entrepreneurs transforms workers into “slaves,” and not only through the exercise of their power as direct employers. But socialism was nothing else but the historical successor of classical republicanism. “Not only did socialism depend on the republican conception of freedom as non-domination in order to advance the idea of wage-slavery, but also that very formula had its origins in republican circles.”[17]

Along with socialism, the feminist movement has “valorized” this republican conception of

[12] Niccolò Machiavelli (1970) *Discourses on Livy* (1531), edited by L.J. Walker. Penguin Books, London, p. 110.
[13] Philip Pettit, *op. cit*., p. ix, italics mine.
[14] Frank Lovett and Philip Pettit, *op. cit*., p. 8.
[15] Amartya Sen (1980) Equality of What? In: *The Tanner Lecture on Human Values*, vol. I. Cambridge University Press, Cambridge.
[16] Philip Pettit (2010) Republican Law of Peoples. *European Journal of Political Theory* 9: 70–94.
[17] Frank Lovett and Philip Pettit, *op. cit*. p. 10.

freedom, criticizing domestic "white slavery,"[18] the counterpart of "wage slavery" denounced by the socialists.

What solutions do republicans advance for the protection of vulnerable social groups—workers, women, minorities of all kinds—against this domination which is, in their view, ubiquitous? They argue, sometimes redundantly,[19] for the enactment of laws, institutions, and practices that have *already existed* in developed societies for a long time, such as the right for workers to associate in professional syndicates, the right to strike, the right of women to divorce, the creation of public housing for female victims of domestic abuse, etc., adding the right to a "basic income" unconditionally granted to all by the state, including to non-citizens, whether or not they have a vulnerable economic status. "By *universal basic income*," says Belgian philosopher and economist Philippe van Parijs (one of the most ardent supporters of this right), "I mean an income paid by a government, at a uniform level and at regular intervals, to each adult member of society. The grant is paid, and its level is fixed, irrespective of whether the person is rich or poor, lives alone or with others, is willing to work or not. In most versions—certainly in mine—it is granted not only to citizens, but to all permanent residents."[20]

Basic income is, however, different in principle from the "means tested guaranteed minimum income," which was enacted long ago in many societies, including (for example) in Romania, by the "Law no. 416/2001 on minimum guaranteed income,"[21] which sets out the principles for granting this right as follows: "Article 1: (1) Families and single persons, *Romanian citizens*, are entitled to a guaranteed minimum income as a form of social assistance, (2) The guaranteed minimum income is achieved through monthly social aid, as provided by this law; (3) Establishment of minimum guaranteed income is based on the principle of social solidarity."[22] By the "law on social assistance" the categories of beneficiaries of social assistance granted by the state are redefined as follows: "Art 4 (1) *All Romanian citizens living in Romania, as well as foreigners and stateless persons who are domiciled or resident in Romania* are entitled to social assistance under Romanian law and regulations of the European Union and the agreements and treaties of which Romania is part."[23]

As Lovett and Pettit argue, all these institutional forms of social assistance must be inspired and shaped by experience in each country, taking into account their real effects on vulnerable groups and on society as a whole. Republican philosophy can only describe and reflect on some great alternative public policies, such as: should social protection be granted selectively based on verification of means, or to everyone through a basic income provided by the state? Should health care services be freely offered to all, or provided through a universal health insurance system? Should the dominating power of wealth be reduced by progressive taxation and substantial redistribution—a privileged formula during the second half of last century—or by offering some "facilities" and "incentives" so that the rich allocate more resources for certain public purposes?[24]

Anyway, the enactment of a "basic income," freely provided by the state, must satisfy (according to Pettit) two main desiderata: *adequacy* of its levels, found in every society, even intuitively; and *independence*, i.e. granting this right without checking the material resources available of members of society, regardless of their "employment history," and with no added obligation to perform public services.[25]

The "attraction" of republicanism is that its argument in favor of basic income satisfies both of these desiderata better than the alternative political philosophies can. Liberalism, for example—attached as it is to the cause of non-interference in "basic liberties," including the liberty to undertake private

[18] Carole Pateman (1988) *The Sexual Contract*. Stanford University Press, Stanford, p. 123.
[19] Robert E. Goodin (2003) Folie Républicaine. *Annual Review of Political Science* 6: 55.
[20] Philippe Van Parijs (2001) A Basic Income for All. In: *What's Wrong With a Free Lunch?*, edited by J. Cohen and J. Rogers, vol. 1, Beacon Press, Boston.
[21] *Monitorul Oficial al României* (2001), Part I, no. 401, July 20.
[22] *Ibidem*. Italics mine.
[23] http://www.mmuncii.ro/pub/imagemanager/images/file/Legislatie/LEGI/L292-2011.pdf. Accessed December 15, 2014. Italics mine.
[24] Frank Lovett and Philip Pettit, *op. cit.*, p. 11.
[25] Philip Pettit (2007) A Republican Right to Basic Income? *Basic Income Studies. An International Journal of Basic Income Research* 2 (2): 2.

gainful activities (*free enterprise*)—stands in principled contradiction to the "right" to a basic income freely provided by the state. In contrast with liberalism—which interprets freedom as being essentially the absence of interference by others, especially the intrusion of the state in the "private affairs" of individuals—republicanism defines freedom in a maximalist manner, as the absence of any domination (private or public) of citizens. Unless they are provided a basic income, Pettit argues, their rich peers may interfere in their lives in every conceivable way. In "hard times" such as those of an economic crisis in which there are many available employees but few available employers, the former could not get a decent salary, preventing them from "functioning properly in society," not to mention the potential abuses of employers and their ability to arbitrary terminate employment at any time. Of course, employees can seek the protection of unions and use their power to negotiate with employers. "But the most effective of all protections, and one that should complement other measures available, would be one's ability to leave employment and fall back on a basic wage available unconditionally from the state."[26] Similarly, women (and their children) who are without their own income are completely materially dependent on their husbands, who thus subject them to "dominant control," even if they never resort to violence or other abuses. Therefore, a basic income freely granted by the state can provide them an "effective alternative."

In order to make this republican conception of social assistance intelligible, we can remind ourselves here of the great distributive principles imagined in the tradition of Western political thought. Karl Marx, for example, formulated two main distributive principles applicable in the socialist society he imagined. In the (first) *socialist* phase, "to each according to his work," which still respects liberal or "bourgeois" constitutional and legal principles; in the (second) *communist* phase, "from each according to his capacity, to each according to his needs," which exceeds all "narrow limits of the bourgeois right."[27]

As can be seen, the current republican conception—as well as the related social-democratic one that advocates for the granting of incomes completely independent from any useful work done in return—seems to further assume the communist distributive principle envisioned by Marx in the nineteenth century. In this spirit, Pettit argues, the free provision of basic income by the state would make it possible for people to "function properly in society," without having to beg the favor of the rich and powerful (including state-appointed officials) to distribute "social aids" in money or goods.

But here the fundamental theoretical issue is: "why give the basic income right to all, not to only those in need?"[28] Granting this right to all, regardless of available means and independent of any obligation to perform any useful work, would bring common benefits to the whole society by becoming a strong "bastion" against domination, which seems to be ubiquitous. Those *dependent* on this right—as opposed to those who are *independently* wealthy—do not become, Pettit believes, a special and inferior social category: "people who depend on others' goodwill and are easier targets of control and domination."[29] The *general* character of this right would mean the fundamental equality of all citizens in front of the state.

The second fundamental issue: would not a state which provides a basic income exert in this way a direct, quantifiable, not-fictitious domination on certain categories of citizens who are also the real payers (taxpayers) of such income redistribution, thus contradicting one of the republican principles that the state itself should not dominate its citizens? Pettit's answer is less categorical, limited to evoking the *formal* requirement that a basic income is to be established by a majority parliamentary vote or referendum, given the "common good" of reducing global intra-societal domination.

Liberalism, as he admits, even in its left-of-centre version, cannot convincingly support the right to basic income given its conception of freedom as non-interference. "Only republicanism serves well in

[26] *Ibidem*, p. 5.
[27] Karl Marx (1964) Critica Programului de la Gotha (1875). In: *Opere*, vol. 19. Editura politică, București.

[28] Philip Pettit (2007) A Republican Right to Basic Income? *Basic Income Studies. An International Journal of Basic Income Research* 2 (2): 5.
[29] *Ibidem*, p. 6.

the required role"[30] through its comprehensive vision, which assigns the state the perpetual task of eliminating (or at least reducing) all possible forms of domination within the society as a whole.

Conclusions

To conclude this short investigation with a more comprehensive, theoretical evaluation, the current republicanism can ultimately be understood as a post-socialist critique of the contemporary capitalist society.[31] The market economy is usually conceived as the privileged *locus* in which people can effectively exercise their freedom, primarily the "negative" freedom cultivated by liberalism or libertarianism: freedom as non-interference. From the perspective of the republicanism, the declared rival of liberalism, the market economy is seen differently. This conception of freedom as non-domination argues for some specific constraints on production and distribution of wealth within the free market, but avoids the rhetorically extreme positions that " *property is theft*," or conversely that "*taxation is theft*,"[32] so that, as the republicans say, the "market economy" does not turn into a "market society."

From a broad liberal perspective, the market economy primarily facilitates *individual freedom*, the unlimited choices of individuals multiplying and diversifying indefinitely under the pressure of the competition that is inherent in the economic process. But the republican conception argues for *social freedom*, which requires state intervention in the economy and the society, but in a way that is non-arbitrary, and therefore non-abusive and non-dominating. Social freedom, as opposed to individual freedom, is essentially a property of the *chooser*, not of *choice*—this is one of the crucial assertion of the republican argument.[33] The person is, as an individual, free to the extent that his choices are not obstructed in any way. Of course, there are *natural* limits or restrictions on individual choices, such as a genetic bad luck, accidental disability, chronic illness, etc., but this freedom is intrinsic to the choice. As a chooser, however, individual freedom can be obstructed *by others*. One can protect his or her freedom, so understood, by an organizational affiliation, financial resources, and cultural influence, but as republican thinkers emphasize, including a set of protective rules and institutions in the framework of a constitutional democracy can provide to the individual a certain social status (*social standing*), and transform their condition from *servus* to *liber*.

Finally, do republicans believe in a "big government"? Yes and no. Like many socialists, they are inclined to assign great responsibilities to the government, but at the same time, like many liberals they fear of the "tyranny of majority," inherent discretion, and governmental abuses. While they claim a vigorous state intervention will mitigate the kind of private domination (*dominium*) of citizens that is caused by a high degree of inequality in the amount of resources available to them, they also fear and remain "vigilant" about the danger of extending the public domination (*imperium*) of the government itself. "Perhaps government," Pettit says again, "should have a large range of responsibilities assigned to it, and perhaps in that sense it should be big. But government should not have a large range of independent powers and in that sense it should be decidedly small."[34] In other words, they assign to the government *great responsibilities*, but authorize *low powers*, which is quite contradictory. Unlike the liberals, republicans are less reluctant about state intervention and less skeptical about its positive effects, but like socialists, they are more radical when speaking of the "social evils" that the state should remove or alleviate.

References

Berlin I (1958) *Two Concepts of Liberty*. Oxford: Oxford University Press.

Gaus G (2003) Backwards into the Future: Neo-Republicanism as a Post-socialist Critique of Market Society. *Social Philosophy and Politics* 20.

Goodin R E (2003) Folie Républicaine. *Annual Review of Political Science* 6.

[30] *Ibidem*, p. 7.
[31] Gerald Gaus (2003) Backwards into the Future: Neo-Republicanism as a Post-socialist Critique of Market Society. *Social Philosophy and Politics* 20: 59-91.
[32] Philip Pettit (2006) Freedom in the Market. *Politics, Philosophy and Economics* 5: 131.
[33] *Ibidem*, p. 133.
[34] *Ibidem*, p. 135.

Locke J (1980) *The Second Treatise of Government* (1690), ed. C. B. Macpherson. Indianapolis: Hackett Publishing Company,.

Lovett F, Pettit P (2009) Neorepublicanism: A Normative and Institutional Research Program. *Annual Review of Political Science* 12.

Machiavelli N (1970) *Discourses on Livy* (1531), ed. L. J. Walker. London: Penguin Books.

Marx K (1964) Critica Programului de la Gotha. In *Opere*, vol. 19. Bucureşti: Editura politică.

Pateman C (1988) *The Sexual Contract*. Stanford: Stanford University Press.

Pettit P (2010) Republican Law of Peoples. *European Journal of Political Theory* 9.

Pettit P (2007) A Republican Right to Basic Income? *Basic Income Studies. An International Journal of Basic Income Research* 2 (2).

Pettit P (2006) Freedom in the Market. *Politics, Philosophy and Economics* 5.

Pettit P (1997) *Republicanism: A Theory of Freedom and Government*. Oxford, UK: Oxford University Press.

Sen A (1980) Equality of What? In *The Tanner Lecture on Human Values*, vol. I. Cambridge: Cambridge University Press.

Van Parijs P (2001) A Basic Income for All. In *What's Wrong With a Free Lunch?* eds J. Cohen, J. Rogers, vol 1. Boston: Beacon Press.

BALKAN JOURNAL OF PHILOSOPHY. Vol. 7, Issue 2, 2015

THE SPANISH INQUISITION AS A MEANS OF INFLUENCING THE COMMON SENSE OF CITIZENS

Slobodan Nešković, B.A. Danijela Marčeta
Faculty of Economics and Engineering Management in Novi Sad,
University Business Academy Novi Sad, Serbia
e-mails: slobneskovic@gmail.com, danijelamarceta@gmail.com

Abstract:
Manipulation of the human mind is a topic that suggests a number of different discussions, but so far most of these have focused on the subject of crowd manipulation as if it were only a contemporary phenomenon that is carried out through the mass media. However, manipulation is a means of engaging, controlling, or influencing the desires of a crowd in order to direct their actions toward the manipulators' best interests. It is clear that manipulation is an essential part of human nature and dates back to the founding of societies. Religion is one of the mechanisms aimed at unification of societies, and therefore unification of their supreme values, principles, and objectives. Unfortunately, there are extreme examples of the manipulation of societies by means of religion, and the Spanish Inquisition is a case in point. The objective of this paper is to draw conclusions regarding the mechanisms of domination and control over human common sense. It also aims to discover how the Inquisition restricted not only people's actions, but also their minds, and how it employed human minds to achieve not only its own goals, but also the goals of the government with which it was mutually dependent.

Key words: Spanish Inquisition, common sense, the social control, religious extremism.

Introductory considerations

Religion is a window to the human soul; it teaches the human to recognize his own nature, to discover the world. When confronted with problems in life, a human seeks after the purpose and meaning of life by developing an understanding of the universe and its existence. Religion is an answer to why humans are the way they are. Unfortunately, history books abound with examples of the forceful abuse of religion. Religion can be a powerful weapon for controlling the thoughts and common sense of people. This creates further possibilities for taking control of their actions. The Inquisition was established within Christianity and the Roman Catholic Church. When studying the functioning of the Inquisition, it is important to take into consideration the period of its development, the conditions under which it evolved, and the identity of its victims in order to reach the proper conclusions.

First of all, the Inquisition as a tool of social control was created after the religious schism, but it was not an exclusively Spanish, medieval, or Roman Catholic phenomenon, since the tribunal of the Holy Office was created in Sicily and would later spread to other countries in Western Europe (France, Great Britain, Switzerland, etc.). What characterizes the Inquisition in Spain is that it was a tool of social control in the religious absolutist monarchy, and was strongest from the end of the 15th till the end of the 17th century.

The Spanish Inquisition was officially established in 1478 by a decree of King Fernando II and Queen Isabella and remained under their direct control, but it was also under the control of the Pope in Rome. It originally had jurisdiction over Christians only, but after the expulsion of Muslims and Jews from Spain, its jurisdiction was extended to all the king's subjects. It is important to emphasize that the Inquisition was created with the aim of controlling the population of Christianized Jews and Muslims suspected of secretly observing non-Catholic religious practices. Isabella aspired to make everyone Catholic, and she was strongly influenced by her confessor, a Dominican friar named Tomás de Torquemada, who became a notorious inquisitor.

1. The levers of power and the Familiari as the intelligence apparatus of the Inquisition

Considering the question of the real power of the Spanish Inquisition to manipulate human common sense and actions, Peter Leposavić (2006) overcomes many old prejudices, emphasizing that "the presence of the Spanish Inquisition in the entire area and the establishment of permanent courts—distributed regionally according to a certain elaborate but logical system of 'visits'—confirms that it was not about decentralization or regionalization, since there were courts in the centers of each district, but simply a good and deliberate centralization," and also "that the Inquisition was set up as a pyramidal structure of ecclesiastical authority and control from its very outset" (p.85). The Spanish Inquisition was associated with the Vatican, but it was also an instrument of the Spanish state seeking to establish control. Leposavić (2006) came to this conclusion by drawing on a series of historical facts, the most important being that the Inquisition in Spain came out as *Consejo de la Suprema y General Inquisición*. The *Suprema* was primarily a supreme state body, and was only nominally under the control of Rome. In reality, it was the Spanish kings who initiated the establishment of the Inquisition in Spain. However, both the church and the state had a shared interest in the Spanish Inquisition. The connection between the *Suprema* and the top of the Roman Catholic Church is evident from the fact that all the great inquisitors elected by the *Suprema* at the time of the elections were bishops or archbishops, and twelve grand inquisitors held the title of Cardinal.

The *Suprema* had a lot of tasks to perform, the most important being the direction and control of the provincial courts. Leposavić (2006) emphasizes that "the *Suprema* was constantly modifying, completing, and reducing its regulations, conclusions, tasks, and goals, aligning them with the political, economic, and psychological state of the society" (p. 86). Of particular significance were inspections of the courts, in which an inspector's duty was to determine whether the general policy prescribed by the *Suprema* was being implemented in court procedure. It also controlled the work of the *Familiari*, who represented a kind of Inquisition intelligence apparatus, and were of particular importance to the mechanism of instilling fear in the entire society and controlling citizens through intimidation. It is interesting that the *Suprema* had entire files on every court, containing data about court cases, inquisitors, and all other members of the courts.

The first court was established in 1481 in Seville. According to the instructions of Torquemada, each court was composed of two inquisitors, one lawyer, one alguazil, and one clerk. In addition, every court had notaries and subordinate officers, as well as royal officials whose job was to confiscate and manage the property of every person sentenced by the Inquisition. Of course, we should not forget the other assistants whose number varied from court to court. At that time, Spain was divided into 12 provinces provided with an equal number of courts. Most courts had two or three inquisitors, and in rare cases one. The Spanish Inquisition was studied by Bartolome Bennassar (1981), who examined biographies of inquisitors and arrived at the conclusion that over 90% of inquisitors were elected from among highly educated people belonging to the hierarchy of the Roman Catholic Church in Spain.

Provincial courts received a permanent status after 1530, and afterwards—given the frequent inspections by the *Suprema*—they faced the need for an increased number of undercover agents and informants to disclose heretics among the population. Their job was to investigate suspicions and to summon, interrogate, torture, and punish those convicted of heresy. The Spanish Inquisition was very interconnected, and every territorial unit in Spain had its own network of *Commissioners* and *Familiari*.

Commissioners of the Inquisition were a sort of secret police, and a large number of them represented so-called *letrados* subject to a very strict selection. *Letrados* were literate and educated officers in the public service. *Commissioners* often worked in places located far away from the provincial courts, and were charged with the tasks of execution, interrogation, examination, questioning, and recounting oral statements of immobile persons. The Inquisition vested *commissioners* with these powers in order to save money.

Familiari and their role in the activities of the Inquisition have been studied by the famous Spanish historians Henry Kamen, Bartolome Bennassar, and Ricardo García Carsel. These authors agree on the fact that *familiari* were lay people loyal to the Inquisition and were granted certain privileges by the Inquisition for their activities. Stephen Haliczer (1990) examined the issue of the origins of the *familiari* and came to the conclusion that *familiari* were the Inquisitors' armed escorts, accompanying them in their travels from town to town. At first they were simply young monks of the Dominican Third Order who were called *familiari* due to their close attachment to the inquisitors. Emil Lucka (1933) states that *familiari* enjoyed immunity from secular courts: "Fear and trembling of the citizens was instilled by the *familiari*, the hull of suspicious people who, as officers of the Inquisition, could not be brought before any secular court," but he also reveals another fact about them: "they often used the name of Inquisition for their own deeds" (p. 131). *Familiari* used to be elected publicly by a court from among old Christians who had never been indicted by the Inquisition (nor had their ancestors). The number of *familiari* was largest in the period of the most active operation of the courts—from 1490 to 1550. Some records indicate that citizens often complained about the work of the *familiari*, whose number was constantly increasing and whose work was financed by the courts. According to Lucka (1933) "complaints to the secular authorities were constantly occurring; in 1514, the Toledo city authorities complained to the king because of the impudence of these people who disturbed peace in the city. Ferdinand and Cardinal Ximenes issued several decrees to curb their iniquity, including that only the people of good reputation were eligible for the *familiari*; in addition, they had to carry with them an authorization signed by three inquisitors" (p.132). As part of his research on *familiari*, Bennassar (1981) also studied their past. He found that out of a total of 78 *familiari* hired at the end of the first large recruitment in 1544 in Cordoba, there existed data on the former occupations of 68 of them. Most belonged to the lower strata of the society, either artisans or small traders. Emil Lucka (1933) points out that there were prominent people who accessed the famous *familiari* brotherhood Confradia de San Pedro Martyr, and that some of them even used to pay large sums of money to the Inquisition in order to join the *familiari*. Another author to write about inquisitors taking money in exchange for admitting people into the office of *familiari* is Bennassar (1981). He cites, as an example, the case of inquisitor Alonso de Hoces from Seville, who was twice accused of bribery. However, the Inquisition solved this problem by announcing in 1651 that one needed to pay 1,500 ducats in order to become a member of the *familiari*. This paid off the most for people who had just converted to Christianity and were afraid of being suspected of heresy and denounced.

Familiari had other privileges as well, including the right to carry weapons, ride horses, eligibility for military ranks, and exemption from rent and from military service in the case of war. The title of *familiari* became a legacy that everyone sought to preserve within his family and pass on to heirs. *Familiari* possessed prestige and people feared their power, although they lost their privileges and influence in the late 17th century. It is clear that *familiari* represented an important element of social control, a kind of an intelligence machine, and most importantly, all citizens feared them. The *familiari* could be anywhere, and anyone could be in touch with them. The function of the *familiari* was to enable the Inquisition to keep control over the people and to rule using the pedagogy of fear.

2. The course of the Inquisition against heretics

We can distinguish different steps taken by the Inquisition from the start of the process until the final outcome. The worst possible outcome of a trial was burning at the stake. The process consisted of the following elements: a complaint, a period of grace, interrogation procedures, torture, *sambenitos* (penitential clothing), *autos-de-fe*, imprisonment, etc. All procedures were carefully planned, and judges followed instructions from books written specifically for them. One of the first books of this type was the "Directorium Inquisitorum" written by the monk Nicolas Eymerich in 1376 in Avignon. It was the synthesis of a number of documents, and it was used to justify the ecclesiastical structure of the repressive apparatus. Moreover, four Dominican monks had already drafted a similar book between 1244 and 1254. This was a famous manual by Bernard Guy, the "Practica inquisitionis." Later on, inquisitors also used a book called "Malleus Maleficarum."

During the medieval Inquisition period that lasted from the 13^{th} to the 15^{th} century, bishops who acted as judges used to apply the method of *acusatio*, waiting for individual complaints. The inquisition method that involved testing and active verification of whether there had been manifestations of heretical behaviour or a moral transgression was introduced by the modern Saint Office. The Torquemada period brought about the need for more serious evidence in order to level accusations, causing the Inquisition tribunal to act more carefully when leveling a charge of heresy considered a crime against the "divine majesty."

Beatriz Comella (2003) came to the conclusion that the usual stages of judicial procedure were "preparatory period, checks, taking precautions, initiation of the trial, presentation of evidence, use of torture, sentencing, enforcement of the sentence (*auto-de-fe*), mitigating of the judgment" (p. 106).

2.1 The preparatory period

Edicto de gracia was the so-called period of forgiveness, an edict which guaranteed forgiveness to all those who signed up as heretics without public exposure. There used to be many self-incriminations during periods of forgiveness. The punishment implied either merely social consequences, or the penitent was whipped, obliged to give charity on a regular basis, or obliged to attend all religious services. Still, such penitents were not deprived of their property.

Edicto de fe was a proclamation issued in the 16th century, and commanded people to denounce all those who they knew to be heretics—even if said heretics were their own family members—or face excommunication. Bartolomé Bennassar (1981) called this proclamation a "pedagogy of fear."

Edicto de delaciones used to be read on one Sunday during Lent every year, and it obliged believers to denounce those who committed heresy within six days (until the next Sunday mass).

Edicto de los anatemas used to be proclaimed eight days after the *edicto de delaciones* as a threat to all believers who did not denounce another person's heresy.

Edicto emplazador applied to those who fled or were absent, i.e., whoever would not show up in person before inquisitors (investigators) at a defined time in a defined place as their trials began. Such persons were convicted of heresy for being stubborn and failing to fully repent.

2.2 Verification

In its quest for signs of heresy, the Inquisition used to avail itself of various sources: accusatory rumours confirmed by honourable persons, the prosecutor's charges substantiated by evidence, individual denunciation before a notary and two witnesses who swore on the Gospels. Initially, three individual denunciatons were necessary to issue an indictment; later on (in the 15th and 16th centuries) this was not necessary, even though the supreme inquisitor Valdes proclaimed that a case would only be taken into consideration if there were several individual denunciations. Starting in the second half of the 16th century, anonymous individual denunciations were allowed in order to protect the informer from revenge. Moreover, when it came to informers, they had to be absolutely confident in what they were doing, as there was a special consideration that they had to take into account in reporting their suspicions to the Inquisition. If an informer accused another person of heresy, they had to be aware of the punishment foreseen for the accused, because should the allegations turn out to be false, under the law of Talion it would be the informer who received that punishment instead. Alternatively, if informers merely expressed their doubts in the form of information, the informer would run no risk of being punished.

Two witnesses were needed to give statements under oath for the tribunal to issue an arrest warrant for the accused. At the end of the 15th century, an obligation was introduced to take a statement from the accused; within 8 days he had to answer whether he knew what he was accused of before the Inquisitor, two priests, and a notary.

2.3 Precautionary measures

Precautionary measures included the imprisonment of the accused and the confiscation of all his property. It is important to mention that an accused person used to be incarcerated and unable to communicate with the outside world; the whole trial used to be conducted in utmost secrecy both from the public and from the indicted person. Preventive imprisonments of up to two years also occurred while the cases was being supposedly investigated, while the indicted person very often did not even know why they were in custody. Regarding the confiscation of the suspected offender's property, it was inventoried according to specific court rules, and confiscated in order to pay off the debt for the costs of the trial. As a result, families of the accused would frequently lose all of their assets.

2.4 Initiating the trial

First the accused would be interrogated by the inquisitors in the presence of two priests and a notary in accordance with the instructions from the manual. The accused had three chances to confess heresy and, should that happen, he would be punished with imprisonment and property confiscation. Should one fail to admit heresy, the content of the denunciation would be read out to the accused, and he would be entitled to a lawyer who would get a share of the seized property. Also, the defendant could request the withdrawal of one inquisitor.

2.5 Presentation of evidence

The presentation of evidence implied either the confession of the defendant or witness statements.

When interrogating witnesses, the inquisitor was obliged to investigate whether any of them were hostile to the accused. All statements were given under oath and signed by witnesses. Emil Lucka (1933) points out in this respect: "Since the inquisitor was anything but a fair judge, as a prosecutor he wanted to prove the guilt of the accused at all costs; he would appreciate any witness incriminating the accused, while opposing all witnesses who would speak in favor of the defendant" (p. 140). Also, the defendant was entitled to make up a list of all his enemies, but also of all witnesses who would guarantee his innocence. If evidence could not be determined at this stage, the Inquisition would proceed to the phase of torturing the accused.

2.6 Use of torture

Inspired by secular trials, the Church introduced the method of torture in the 13th century. It is important to note that torture was always applied in the presence of a doctor. It is rarely mentioned in the literature that victims were tortured naked or with certain body parts just barely covered, which was supposed to have a particular psychological impact on them. The most common methods of torture included winch, water, and crank. The *garrucha* (winch) was a method of torture by tying the wrists of the accused behind his back and then suspending him from the ceiling with weights tied to his ankles. Tormentors would first lift the accused high into the air and subsequently lower him to make him feel as if he was about to hit the floor. This was the mildest method of torture. The *agua* was a water-based method in which a cloth was placed into the mouth of the accused and water was spilled inside his mouth, provoking a fear of drowning. The defendant would be posed in such a position that his head would hang down slightly below the rest of his body, and the tormentors would spill slightly less than one liter of water. The tormentors used to repeat the torture process 6 to 8 times. During pauses, they would ask the accused whether he would like to confess heresy. Later on, these two primitive methods of torture were replaced by the rack. The *cordeles* (rack) was a wooden frame to which the victim of the torture would be tied with ropes. By a gradual straining of the ropes the accused person was stretched, causing unbearable pain. There were also other inventions, like the red-hot iron, red-hot stones, etc. However, even though the Inquisition's activity reached its peak in the 16th century, it rarely used torture.

2.7 Pronouncement of sentence

An advisory committee composed of priests was responsible for the pronouncement of sentence. It used to deliver a decision after reviewing detailed reports submitted by inquisitors. Regarding the most severe sentence (the death penalty), starting from the period of Torquemada the unanimous consent of all members of the advisory committee was required, and thus a single vote would be sufficient to save the defendant from a death sentence. The defendant was also entitled to "canonical compurgation" (a canonical principle taken from Germanic law), which meant that the accused had a final chance to prove his innocence with the help of witnesses who would swear together with him (seven to thirty witnesses). If the judges were not convinced of the guilt of the accused, this method would be used. If the judges accepted the oaths of witnesses, the accused would get a milder sentence, which would be announced to him privately. Should the accused person be convicted, the decision would be announced before the accused, a notary, a representative of the bishop, and the inquisitors.

2.8 Execution of sentence

If proven guilty, the convicted person would have his sentence determined by the inquisitors according to the degree of severity of the misdeed. We can distinguish five types of penalties for offenses that fall under heresy:

1. Disclaimer: this sentence was reserved either for convicts who had committed a minor offense, or for those who had been accused of more serious offenses but whose guilt could not be proven. The sentence required repentance of the convicted at Sunday Masses.
2. Death sentence: this sentence was reserved for those heretics who would not admit their mistakes. This sentence was always followed by excommunication from the Church and the confiscation of all property.
3. Prison sentence: According to Eymeric’s manual (1376) for inquisitors, persons accused of grave heresy were detained and sentenced to prison. Those sentenced to life imprisonment could be subject to a milder regime in prisons, and their punishment never really lasted a lifetime, because later on—as penitents—they would be sent to monasteries or to their homes to atone for their sins. The sentence of house arrest was

applied since 1488. Moreover, convicted persons would have all their property confiscated, with a certain amount left for a dowry if the convict had a daughter, while sons could not claim any right to the convict's property.

4. Financial fines: Besides property confiscation, prisoners could be assigned cash fines. The *ammenda* was the best-known fine, assigned to the descendants of sentenced persons in order to wash away their shame through financial contributions.
5. Other forms of punishment included whippings, exile, and penitence—punishments designed to induce feelings of shame in convicts. Public repentance implied the penitents' duty to attend every ceremony of *auto de fe*, publicly renouncing their sins. During the ceremony of *auto de fe*, graver offenders whose sins were heinous used to be subject to public whipping, which was supposed to serve as a warning to the people observing this torture. Those condemned to exile used to get sentences ranging from three-and-a-half to fifteen years (lifetime sentences were rare). This punishment was typical for wealthy businessmen, who would thus lose their prestige and reputation and were at the same time squeezed economically, which suited the *Suprema*'s interests. Emil Lucka mentions a sentence that required work on the galleys. Lucka (1933) states that this sentence was introduced by Ferdinand the Catholic because he needed ship rowers.

As regards practices that induced shame in convicts, it is important to mention the *sambenito*. The *sambenito* was a penitential garment that "marked" the penitents in the society and made them subject to derision. In the first decades of the Spanish Inquisition's operation, penitents had to wear *sambenitos* any time they left their home. These were yellow tunics with a red cross of St. Andrew on the chest and back. The *sambenito* was a way to perpetuate punishment. Those who had not repented and were handed over to the secular authorities wore black *sambenitos* with painted flames, which symbolized hell. The *sambenito* was a way of punishing not only heretics, but also their descendants, and was thus called the "inherited guilt."

Those prisoners who were condemned to carry *sambenitos* only for a defined period of time had to return these tunics to the church. Priests would hang these *sambenitos* marked with the penitents' names on a special place in the cathedral. There were cases of *sambenitos* disappearing from the cathedral, but *Suprema*'s inspectors controlled all the cathedrals and placed new *sambenitos* in place of the stolen ones so that the descendants of convicted people could not escape public humiliation. The aim of the *Suprema* was to stigmatize the accused forever, together with their descendants, in order for the church to be protected from them in advance, and to make persons carrying *sambenitos* embody a warning to sinners of their fate, instilling fear in the populace. Petar Leposavić (2006) states that the people of Spain sought the abolishment of every practice of public humiliation through complaints and public protests, outraged at the presence of publicly labelled people walking around in towns wearing *sambenitos*; in their opinion, penitents in these tunics were disgracing entire cities.

A heretical or apostate offense could be punished differently depending on whether the convicted person admitted the offense or denied it. Should he deny the offense, the Inquisition would hand him over to secular authorities, who would condemn the offender to public burning at the stake during the ceremony of *auto de fe*. The *auto de fe* was actually a ceremony aimed at spreading the faith and directing people towards orthodoxy. The nation even looked forward to this feast of faith, and used to complain if there was a decrease in number of burned heretics, while official persons had a duty to attend these ceremonies. Emil Lucka (1933) emphasized that the *auto de fe* "was the highest service to God, the victory of the church over all of its enemies, a genuine picture of the last judgment" (p. 161). If the convicted person repented at the last moment, he would not be burned alive, but hung instead, and thus spared of major torment. There were cases of prisoners somehow managing to escape, and pictures bearing their names burned at the stake instead. In the case of someone being convicted of heresy after his death, the inquisition would burn that person's remains.

Besides these public *autos de fe*, there were also *auto particular* or *autilio* that were held in churches or halls of the Inquisition, when the latter wanted to spare some offenders of major public shame, or simply did not have the money to organize a public *auto de fe* in a city square. In the case of *auto de fe* penalties, families of the convicted were most gravely hit by the forfeit of civil rights. This verdict was called *inabilita*. The penalty was designed mostly for rich Jews with the aim of their exclusion from both public and economic life. However, *inabilita* often included the possibility of *ammenda*, which allowed the cessation of punishment through financial contribution, thereby removing the stamp of shame from descendants. It is important to emphasize that the researchers of Spanish Inquisition came to the conclusion that the *Suprema*'s aim was to weaken all heretics economically.

2.9 Mitigation of sentence and appeal procedure

Appeal and mitigation of sentence was possible only for those sentenced to prison. Good conduct of the defendant was a precondition for the mitigation of punishment and its conversion to spiritual atonement, which implied attending Masses, making pilgrimages, and the like. Appeals were frequent from the time of Torquemada. Complaints filed by respectable persons on behalf of the accused could suspend trials on grounds of insufficient evidence.

3. The Inquisition as a tool of social control and mind control

After analyzing the judicial process, we can conclude that the victims of the Inquisition were numerous. Certain numbers presented by some authors during the 19th and early 20th century are not accurate, since some data is lacking, and certain ideological prejudices that existed at that time led to an exaggeration of the number of victims. Henningsen and Contreras thoroughly researched the number of victims of the Inquisition, and came to the most objective conclusions made so far. They found that at the time of the highest activity of the Inquisition—between 1540 and 1700—there were approximately 50,000 trials. Beatriz Comella (2003) emphasizes their conclusions: “Tribunals in Castilla sentenced 44,674 people, 876 were burned at the stake; Tribunals in Aragon sentenced 25,890 people, 520 were burned at the stake. Accordingly, during the period of two and a half centuries of the highest inquisitorial activity, 1,346 people were executed, accounting for 1.9% of the total number of convicts" (p. 116). The most serious of the crimes for which people were convicted were: Crypto-Judaism (baptised Jews who secretly adhered to the faith of Moses), Mohammedanism (baptised Muslims persisting in their customs), Lutheranism, being Illuminati, blasphemy, making up stories against religion, bigamy, etc.

Contreras and Henningsen also came to the conclusion that the most brutal period of the Inquisition (the period between 1540 and 1595) coincided with the secular reign of King Philip II, which proves that the Inquisition was a tool of the Counter-Reformation.

The relationship of synergy has always been ideal for the church and state. Of course, this relationship implies a clear distinction between spiritual and secular authorities, whereby there is a mutual cooperation for the sake of the progress of the community. For this reason, the Church has always followed politics and social changes and tried to contribute to those changes. Before secularization, religious heresy represented a doubly destructive phenomenon for the society, because a heretic violates both God's law and the societal law. And the people themselves have accepted these laws. Beatriz Comella (2003) argues that Inquisition is nothing else than an institutionally embodied timeless and universal social function—the social control that is performed by absolutely all known human societies in order to defend their special values, an essential condition for the preservation of collective identity. Heresy reached socio-political dimensions because religion and politics intertwined, and therefore heresy actually threatened the teachings upon which the society rested. Theorists of the 16th century equated sin with offense. Heresy was not only an insult against God, but also an insult against the king as a Christian ruler, and also a source of civil discord and other problems.

The objective of the *Suprema* was to control all layers of the society, including respectable ones, through the Inquisition. Moreover, the Inquisition often accused even prominent government officials of heresy, and even those within its own ranks. The Inquisition in Spain condemned prominent humanist Juan de Vergara, doctors Constantine and Kazali, inquisitor Jose Fernandez de Toro, archbishop and primate of Spain De Karnas, archbishop of Toledo and the primate of Spain Bartolome de Carnaza y Miranda, and also other prominent people, including Pablo de Olviede, Luis de Leon, Teresa of Avila, etc. Comella (2003) also states that the Inquisition became increasingly dedicated to purging the church's ranks, not allowing the "disease" imported from Europe (Humanism, Reformation, Enlightenment, rationalism, liberalism, and modernism) to create a stronghold on the soil of the most Catholic country—Spain.

Moreover, public execution of penalties and loss of reputation are clear evidence that heresy was condemned in general. The best examples are *sambenitos* and *autos de fe*. Through the ceremony of *auto de fe*, the Inquisition was attributing huge importance to the public pronouncement of judgment in order to teach people a lesson and create a state of fear in the society. Also, the ceremony had a lot of religious symbols and elements of a real drama, which further affected the general state of human consciousness at that time. However, the social control implemented by the Inquisition was based on law, and people accepted it because they were awed by the Inquisition. The *sambenito* was a kind of reminder of shame to both the victim and others. The custom of wearing *sambenito* persisted until the 19th century.

Another mechanism of social control was the mechanism of secrecy. The Inquisition penetrated deep into the private sphere, and the accused persons often did not know why they were arrested, nor had they had any contact with the outside world. This mechanism of secrecy instilled fear among all individuals in the society, because anyone could be accused of something. Emil Lucka (1933) points out that “the most unique invention and a tremendous power of the Inquisition was the secrecy that surrounded all its operations” (p. 141). Subjects of the Spanish king and Christian believers were deeply surprised to learn that someone could be imprisoned without any evidence and kept far away from the outside world; and on top of that, the accused had no right to know who had accused him, nor what the charges were. This is why in the 16th century there were frequent protests appealing to the authorities to ensure a public and fair trial. In 1518, citizens sent an appeal to King Charles V. The appeal was drafted in the form of an amendment containing 25 clear demands. This was not the only attempt of citizens to influence the work of the Inquisition. However, the authorities did not exert any influence in this area because, in a certain way, the Inquisition worked in their favor. The church itself also tried to influence the Inquisition. Pope Leo X tried to influence the work of the *Suprema*, but his edicts did not produce the expected results. The Pope died, and his successor did not attempt to deal with these issues. So secrecy was the most important mechanism employed by the Inquisition for the establishment of social control, since mystery evoked fear, and fear proved to be the best weapon for controlling human minds and common sense. No one even dared to entertain an opinion different from the church's, let alone publicly express such opinions.

Bartolome Bennassar (1981) states that for three centuries the Inquisition ruled by fear, because orders were carried out through instilling fear. He emphasizes that “the most conscientious inquisitors were trying to achieve a result: fear was meant to remove the biggest obstacles in the way of heresy” (p. 94). The opinion of Francisco Peña on the manual for inquisitors—written by Nicolau Eymerich—has been mentioned above in this paper. It is important to emphasize that, when speaking of "terror against the people," Eymerich refers to the psychological unbearability of terrifying scenes created deliberately to induce fear in believers so as to make them cast aside any shadow of doubt in their faith.

Bennassar (1981) still wonders whether the terror of the Inquisition was more terrible than the terror of secular courts. He analyzed the opinion of Louis Sala Molins, who claimed that the Inquisition had invented a technology of torture that was different from others, and that both state courts and the Inquisition pursued the same goal, i.e. forced confessions. Bennassar (1981) disagrees with these views and claims there is no difference between the torture performed by the Inquisition and that performed by courts worldwide, and that civilian courts tried to find material evidence of offenses, but the Inquisition was doing its best to extract confessions from defendants. Emil Lucka (1933) wrote on this subject that "whoever did not confess and whose heresy could not be proven demonstrated the incapacity of the Inquisition, and could be ruled innocent at the end of the day. By way of consolation, the following principle was already in force back in the time of the medieval Inquisition: it is better to punish a hundred innocent people than to let one criminal escape” (p. 145).

However, we must assess things objectively and ask ourselves how the Inquisition managed to control the common sense of each person. The human mind is practically impossible to control, and the same is true for the ideas that emerge in it. The thoughts and common sense of every individual belong only to him, and if he does not express them aloud or act in accordance with them, the Inquisition could have no suspicions of heresy. Therefore, the

Inquisition found another effective mechanism for establishing control—censorship. It took all necessary measures to prevent the spread of new ideas that were not in line with the church's teachings. Intellectuals in particular found themselves in the spotlight of the Inquisition's attention. Books sold in Spain were under control, and the Inquisition was very strict in terms of censorship. By the 19th century there were no works of scientific or philosophical significance left in Spain, as it was basically cut off from the rest of Europe and its culture. The Renaissance, humanism, and the German Reformation left no trace on the cultural heritage of Spain. The most interesting fact is that the church tried to "hide the Bible" from the nation, even resorting to burning translations of the Bible done by priests. It was only in 1790 that the first Spanish translation of the Bible was published. A list of banned books was issued by the Spanish Inquisition in 1559. The famous historian of the Spanish Inquisition Juan Antonio Llorente (1826) analysed in detail the censorship in Spain during the Inquisition, adding that the activities of the Inquisition regarding censorship weakened over time. Emil Lucka (1933) wrote on the cooperation between the church and state regarding the mission of censorship and the whole process at large: "Before a book could be published, its manuscript had to be handed over to the Inquisition, which thoroughly examined it before issuing a license. The license had to be submitted to the *Consejo Real* in Madrid, which conducted governmental censorship" (p. 206). By performing the function of censorship and controlling shipments of every book that was entering the country, the Inquisition worked in favor of the state, particularly after the outbreak of the French bourgeois revolution.

All of this leads us to the conclusion that the Inquisition introduced a form of terror into the society. Dragan Simeunović (2009) defines terror "as a form of intense and terrifying violence carried out by someone who is dominant, which implies not only possession of power, but also the use of any currently superior position in human relations; but, as a political category, terror violence of a certain political authority is vested in someone" (p. 23). Furthermore, terror as political violence is characterized by an organized, planned, and targeted enforcement, provoking large-scale fear in order to achieve certain goals. Also, the symbolic function of terror is that anyone can become an object of violence, and even someone who is completely innocent has no possibility of remaining protected.

Concluding Remarks

In the end, it is unquestionable that the Inquisition included a kind of extreme religious violence. However, it is ironic that Jesus himself was a victim of intolerance. Also, the first Christians suffered great casualties and experienced violence themselves. Nowadays, it can be claimed that the Inquisition violated fundamental human rights, but at a time when the state was not secular, it was state terror and the Inquisition that were nothing but tools of the state, used to combat heresy that had acquired social-political dimensions and threatened the doctrines on which the society was based. Of course, the Inquisition was an apparatus that was created as a common interest of both the state and the Roman Catholic Church in the Vatican.

Governments sometimes resort to terror when they become less powerful, and this is especially characteristic for closed and totalitarian societies in which ideology or religion have important roles. The aim of the Inquisition was not to save the souls of convicts, but to frighten the rest of the population. Therefore, it can be concluded that the Inquisition operated on the principle of "it is better to win than to convince!" The aim was not to convince the masses of the existence of God, but to dominate them by provoking fear, which is typical of a policy of terror. The Inquisition in Spain manipulated people's common sense by employing the mechanisms of secrets, censures, announcing verdicts, etc. The most important element of the Inquisition for causing fear was the existence of secret police, who acted secretly and enjoyed great privileges in society. *Familiari* worked to obtain information about the perpetrators of heresy, and induced fear in people's minds because anyone could be an informer and provide *familiari* with true or false information. Torture performed by the Inquisition was nothing worse than torture performed by secular authorities. The most terrible fact about inquisitorial process was the mechanism of secrecy, as well as public exposure of convicts with the help of *auto de fe* penalties, *sambenito*, and similar sentences. Also, the Inquisition prevented any intellectual development of people by means of censure, and it exhausted the economic assets of the wealthy, since economic power is the basis of any other kind of power.

The conclusion that can be directly drawn from all this is that the Inquisition in Spain was a state instrument for implementation of certain objectives by employing terror as a kind of political and—in this case—religious violence. It can be concluded that the citizens in Spain at the time did not have freedom of speech, and also it can also be assumed that their freedom of opinion was limited.

References

Bennassar B (1981) *Inquisicion española: poder politico y control social*. [e-book]. Grupo editorial Barcelona: Grijalbo. Available via SCRIBID. http://www.scribd.com/doc/115160832/ Benassar-Bartolome-Inquisicion-Espanola, accessed 20 Oct 2014

Comella B (2003) Španska Inkvizicija. Translated from the Spanish language by Biljana Bukvić. Beograd: Clio.

Haliczer S (1990) *Inquisition and Society in the Kingdom of Valencia, 1478–1834*. [e-book]. Berkeley: University of California Press. Available via Google Books, http://books.google.rs/books?id=crCQmS_85EEC&printsec=frontcover&hl=sr&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false, accessed 12 Oct 2014

Leposavić P (2006) *Inkvizicije: srednjovekovna, španska, rimska*. Beograd: IGAM.

Llorente J A (1827) *The History of the Inquisition of Spain, from the Time of Its Establishment to the Reign of Ferdinand VII.: Composed from the Original Documents of the Archives of the Supreme Council, and from Those of Subordinate Tribunals of the Holy Office*. [e-book]. London: G. B. Whittaker,. Available via GoogleBooks, http://books.google.rs/books?id=ZTUvAAAAYAAJ&printsec=frontcover&hl=sr&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false, accessed 10 Oct 2014

Lucka E (1933) *Torquemada i inkvizicija u Španiji*. Translated from the German language by Stevan Milović. Beograd: Nolit.

Simeunović D (2009) *Terorizam*. Beograd: Pravni fakultet Univerziteta u Beogradu.

BALKAN JOURNAL OF PHILOSOPHY. Vol. 7, Issue 2, 2015

THE DEVELOPMENT OF JOHN LOCKE'S IDEAS ON TOLERATION

Petar Cholakov
Institute for the Study of Societies and Knowledge, Bulgarian Academy of Sciences
Sofia, Bulgaria
e-mail: P_Cholakov@yahoo.com

Abstract:

This work analyzes the problem of the development of John Locke's ideas on toleration, in particular the grounds of separation of church and state. The *first part* examines Locke's arguments regarding the prerogatives of the magistrate towards 'indifferent things' and the religious sphere. I distinguish between three stages in the development of Locke's view on toleration: a suspicion toward the plea for it (the *Two Tracts*); an implicit non-verbalized distinction between church and state, and support for toleration (*An Essay on Toleration*); and toleration as a political right (A Letter Concerning Toleration, the Two Treatises and the later letters). The *second part* focuses on the definition of 'commonwealth' in *A Letter Concerning Toleration*. I outline two fundamental sets of interdependent arguments that Locke uses for the separation of church and state. The *third part* is dedicated to the sphere of the church and the dimensions of the duty of toleration. The relation between Locke's views on toleration and political practice explains the shift between the three stages, and is explored throughout the paper.

Key words: John Locke, toleration, separation of church and state.

The starting point of John Locke's (1632-1704) political writings—devoted to the settlement of the religious conflict which divided England in the seventeenth century—is his interpretation of the prerogatives of church and state. His position on this issue did not remain invariable; it was subjected to several profound permutations during his life. The claim of the First Letter (1685) that toleration—as a means for removing severe tensions between the civil government and religion—should be based on the separation of the two spheres, was itself an outcome of the development of Locke's position on the prerogatives of the magistrate in the matters of religion. During this epoch toleration was perceived simply as *one* of the alternatives to religious wars. Only a few political writers supported it, and Locke himself did not immediately adopt it as a cure of *odium theologicum*.

This work analyses the problem of the development of Locke's views on toleration, in particular the grounds of separation of church and state. The term "grounds," in its usage here, has a double meaning. *First***,** in a more theoretical sense, it denotes the arguments coined by Locke in the evolution of his interpretation of the prerogatives of the magistrate regarding religion and especially *res indifferens* (indifferent things). *Second*, in a more empirical sense, the term describes the need to explore the historic factors, or the influence of the historical context on the development of his position. "What has changed in his arguments?" and "what was the role of the empirical factors?"—these questions establish the essence of what we must examine in order to describe the grounds of the separation between church and state.

The contemporary interpretation of the problem doesn't provide a coherent answer to these questions. For example, there is still a debate on the reasons why Locke changed his "conservative" position on the main political issues of the day—stated in the *Two Tracts*, written in the beginning of 1660's—and became a "champion of toleration."[1] As far as the development of Locke's arguments regarding the two spheres is concerned, although scholars generally agree that his position became "more liberal" than his early political works (Tully 1984 and 1991), others claim that in his last essays, Locke again gave priority to the authority of the magistrate (Cranston 1988). This brief overview shows that the problem is still significant and worth exploring.

In accordance with the two tasks outlined above, the *first part* of this paper will examine Locke's arguments supporting his position on the prerogatives of the magistrate towards indifferent things and the sphere of religion (*Two Tracts on*

[1] For the different interpretations of this issue, see the first part of the paper.

Government, An Essay on Toleration). The *second part* focuses on the definition of "commonwealth" given in *A Letter Concerning Toleration (Epistola de Tolerantia).* I discuss the arguments for this definition in the *Letter* and in the *Two Treatises on Government*, in particular its role as a basis for toleration. The *third part* is dedicated to the sphere of the church and the duty of toleration.

1. Indifferent things and toleration

An Essay on Toleration (1667) marks a turning point in Locke's political philosophy. It was influenced by his close friendship with Lord Ashley (later the Earl of Shaftsbury), and was his first main political work which defended the cause of the Dissenters.[2] The Dissenters were repressed and persecuted according to the norms of the Clarendon Code, adopted by the parliamentary majority of the Anglican-gentry alliance. Their political ends, which the *Essay* supported, were two: *the ceasing of persecution and the promulgation of a legal document on the right of toleration*.[3] Locke begins his essay with an important distinction between the different spheres of "actions and opinions" (Locke 1997: 136) and the corresponding "degrees" of toleration. He distinguishes between: actions and opinions that are irrelevant to state and society ("speculative opinions and divine worship"); action and opinions that are neither good or bad—and, although they concern society, are indifferent to religion (indifferent things); and actions and opinions that "concern society, but are also good or bad in their own nature, and these are moral virtues and vices" (Locke 1997: 136).

Locke is prepared to admit an "absolute and universal right to toleration" only to the first category of actions and opinions. This category denotes a sphere in which the civil magistrate does not possess the right to interfere. In it are included not only some "purely speculative opinions, as the belief of the Trinity," but also the "place, time and manner of worshiping God." It should be emphasized that in the *Two Tracts*, which Locke wrote five years before the *Essay on Toleration*, the empirical facts and details regarding worship are put in the second category—that of indifferent things, which is in the jurisdiction of the magistrate. In the *Essay*, Locke limits the prerogatives of the state in order to justify toleration. As in his later political writings, he bases this limitation on the ends of the commonwealth. On the one hand, he admits that the power the magistrate possesses is vested in him "for no other purpose, but to be made use of for the good, preservation, and peace of men." On the other hand, he refutes the claim that the "monarchy is jure divino" (Locke 1997: 135). Thus, Locke reminds his readers of *Magna Charta* (1215) and the rights guaranteed by it.

In the *Essay*, the "levels" of toleration correspond to a certain understanding of the liberty of conscience of the believers.[4] This liberty is most broadly defined and unrestricted in the first sphere of actions and opinions. In order to fully understand the significance of the *Essay on Toleration*, we have to compare it with the *Two Tracts*. The spirit of the *Tracts* is not only pro-absolutist, it is a claim against and a response to the proposal for toleration based on individual conscience developed by Locke's contemporary Edward Bagshaw (Abrams 1967, Tully 1991). Nonetheless, the two texts also share a common ground.

In the first place, in the *Tracts* Locke develops a view on the civil war and the ends of government that he upholds in his later writings. In the *Two Tracts*, Locke is persuaded that religion is unashamedly used by the "clergy of all sects" for the ultimate political end—power. This struggle for power is the real motivation for the conflict (the same statement is made in the *Letter Concerning Toleration* and in the *Second Treatise*). Political power is used not as it should be—to preserve property—but rather to confiscate and transfer it (Tully 1991: 643). In the *Letter* and in the *Treatises*, Locke develops his view of the preservation of property as the main goal of political power and the ultimate reason for the existence of the state. Thus, property is the most important element of Locke's definition of the term "state." The term "property" includes "life, liberty, and estate," therefore the obligation to preserve it is a basis for toleration.

In the second place, Locke states in the *Tracts* that Christian leaders had inculcated two erroneous beliefs in both princes and the laity: "that there is

[2] During the English religious wars, the term "Dissenters" referred to Baptists, Presbyterians, Quakers, and Independents.

[3] Although it was never proven, up to the last day of Charles II's (1600-1649) life, some believed (or suspected) that he was Catholic, and as such sympathised and supported the cause of his coreligionists and the religious minorities.

[4] Locke describes the different spheres and their relation to toleration in order to formulate a reply to the question of the liberty of conscience.

only one true way to heaven; and that it is a Christian duty to uphold and to spread the true way by force and compulsion and to suppress heresy" (1991: 643). Locke claims that the Christian religion allows each man to worship God in the way he sincerely believes is right, and that Christianity should be spread not by coercion, but by love and persuasion (Locke 1967: 161). These arguments appear almost unchanged in *A Letter Concerning Toleration*. However, on the basis of this analysis, Locke advanced two radically different solutions for the religious conflict. The first, in the *Two Tracts*, is the theory of absolutism and the imposition of religious uniformity; the second, in *A Letter Concerning Toleration*, is a theory of popular sovereignty and religious toleration (Tully 1991: 644). The premises, which have been described above, give rise to different interpretations due to the change in the historical context in which the *Two Tracts* and the *Essay on Toleration* were written.

The *Two Tracts* are Locke's project for the political and religious form of the Restoration Settlement of 1660–1662. Locke casts doubt on the proposal for religious toleration stated by Bagshaw, because as long as the two erroneous beliefs continued to be widely held, a policy of religious toleration would have been used by the religious groups to build up strength, and eventually to instigate another civil war in an attempt to gain power. The unlimited liberty of humans to operate with indifferent things is a danger to the very existence of the commonwealth. Locke perceived the civil magistrate as "*conservator pacis*"[5] (Locke 1967: 162). The magistrate is the "supreme legislative authority of any society," and could impose and determine the use of indifferent things, *including* all matters regarding Divine worship (Locke 1967: 125-126). Nonetheless, toleration could be granted within certain limits by the magistrate. In any case this could not be a toleration based on individual conscience, because that would be a threat to authority.

Analysing the question of why Locke decided not to publish the *First Tract*, Tully (1991: 645) claims that Locke perceived the desire for greater power for the English church to be a *greater danger* to the civil peace than the hope for toleration of the Dissenters. Locke was repelled by the policy of severe Anglican uniformity imposed by the government in the summer of 1661. But religious uniformity supports the strong power of the magistrate and stability, which are seen as complementary even in Locke's early writings.[6] This is why the thesis of Abrams seems *more plausible*. He states that Locke decided not to publish his essay because the imposed uniformity made this publication unnecessary, "not by violating his trust, but by fulfilling his hopes" (Abrams 1967: 14).

Two facts changed Locke's mind regarding toleration and the power of the magistrate. *The first* was that instead of solving the religious conflict, the Clarendon Code only made it more virulent. *The second* reason, already mentioned, was that in 1666 (the year of the Great Fire in London) Locke met and become a close friend of the Dissenters' leader, Anthony Ashley Cooper. These two circumstances resulted in the limitation of the prerogatives of the magistrate and the apology of toleration in the *Essay*. According to it, the civil magistrate does not have the right to control citizens' beliefs. And if the magistrate nevertheless makes such an attempt, then he should know that he will be resisted (Tully 1991: 647). Not uniformity, but toleration is the cure for the society divided by aversion. Furthermore, Locke developed a radically subjective definition of religion according to which if a person sincerely believes that something is necessary and not indifferent, then it is necessary for salvation.

In the *Essay*, Locke places the sphere of the ecclesiastic matters beyond the sovereignty of the magistrate over the indifferent things. Indifferent things include "all practical principles or opinions by which men think themselves obliged to regulate their actions with one another"; they regard how people "may dispose with their estates," "when men may work or rest," etc. (Locke 1997: 140). Things in this sphere have only limited right to toleration. The criterion for interference of the magistrate in this sphere is, as in the *Letter*, the public good. Locke's view on toleration in the *Essay* is more restrictive than in the first *Letter*, in which he postulates the separation of the two spheres and formulates two definitions regarding their ends. This interpretation is supported by Locke's view on several "degrees of imposition that could be used in matters of opinion"

[5] The italics are mine.

[6] Abrams demonstrates (1967: 51-52) that Locke devoted two of his early writings (tailor-made poems) to Cromwell and Charles II. In both works he perceives these men as guarantees of peace against the anarchy of the Civil war that thrived on the grounds of the religious hatred and two erroneous beliefs.

(1997: 141)[7] presented in the *Essay*. The mere fact of the existence of such a topic (later omitted in the first *Letter*) is significant. Although Locke grants considerable toleration towards opinions, he claims that the magistrate may prohibit the publishing of any opinions when “in themselves they tend to the disturbance of government” (1997: 141). It was not until 1673, when Charles II withdrew his support for the Declaration of Indulgence and the Anglican-gentry alliance entered the Parliament, that Shaftsbury and Locke turned against absolutism (Tully 1991: 647). This new political position was mirrored in the separation of the two spheres in the *Letter* and the radical ideas of the *Two Treatises*.

To sum up, we can distinguish *three stages* in the development of Locke’s view on toleration. *The first* is characterised by suspicion towards the plea for toleration. The sovereign’s jurisdiction encompasses indifferent things and people’s beliefs (*Two Tracts*). The main arguments are: that a strong magistrate is needed in order to serve as *conservator pacis*; that the plea for toleration masks the demand of the religious elites for political power, resulting in civil war; and last but not least, the role of the two “erroneous beliefs.” *The second stage* is marked by an implicit non-verbalized distinction between the church and state and support for toleration (*An Essay on Toleration*). Matters of religion are excluded from the sphere of the magistrate. The leading arguments are that: political power is given to the magistrate only for the protection of the public good; believers have liberty of conscience; the oppression of Dissenters will result in resistance. During *the third stage*, toleration becomes a political right. Furthermore, the *restrictions* imposed on Dissenters are minimized, while the right of the magistrate to interfere is reduced (*A Letter Concerning Toleration*, *Two Treatises on Government*, and the "later letters" on toleration). I shall analyse the arguments coined at the third stage in the next parts of the paper.

2. Locke’s definition of "commonwealth"

2.1 The two sets of fundamental arguments

In many respects, *Two Treatises on Government* and *A Letter Concerning Toleration* are complementary. The right of resistance defined in the last chapter of the *Second Treatise* is a guardian of the political right of toleration formulated in the *Letter*. Locke’s *Two Treatises*[8] represent Shaftsbury’s turn to revolution, provoked by the persecution of the Dissenters and the fact that Charles II dissolved three toleration parliaments (Tully 1991: 648). Locke’s definition of "commonwealth" is given in the *Letter*, which Locke wrote while in exile in Holland. He fled from England after the Dissenters’ leaders began to be executed, brought on by the Rye House Plot and the Monmouth Rebellion.

Locke uses the term "commonwealth" instead of "state," because the very meaning of it mirrors his idea of the essence of politics.[9] The commonwealth is a “society of men constituted only for the procuring, preserving, and advancing of their own *civil interests*” (Locke 1824a: 10).[10] Locke defines "civil interests" as “life, liberty, health, and the indolency of body; and the possession of outward things, such as money, lands, houses, furniture, and the like” (1824a: 10). The term "civil interests" is almost identical to Locke’s definition of "property" (which consists of “life, liberty, and estate”) in the *Second Treatise* (Locke 2002: 323).[11]

The core of Locke’s definition of the term "commonwealth" is *the relation between civil interests and individual interests*. The reason that a commonwealth is needed is that individuals possess certain undeniable natural rights that must be protected. Locke remarks that it is the duty of the civil magistrate “by the impartial execution of the equal laws, to secure unto all people in general, and to every one of his subjects in particular, the just possession of these things belonging to this life” (Locke 1824a: 10). This duty is based on natural law (Locke 2002: 271), and is in accordance with the essence of the political power as formulated in the *Second Treatise*. Political power is (2002: 268):

> a right of making laws with penalties of death, and consequently all less penalties, for the regulating and preserving of property, and of employing the force of the community, in the execution of such

[7] Obviously here Locke understands toleration in a broad sense- not only a religious toleration towards beliefs, but towards opinions.

[8] For the dating of the *Treatises*, see Ashcraft 1986. According to Ashcraft, Locke worked on the *Treatises* from 1681-1683.

[9] Both terms derive from Latin. While "state" comes from the Latin word *status*, "commonwealth" is a literary translation of the Latin phrase *res publica*, a term used (for example) by Cicero. In the present essay "commonwealth" and "state" are used interchangeably.

[10] The italics are mine.

[11] For a possible connection between Locke’s ideas on property and resistance and his intellectual predecessors—such as the mediaeval Conciliarists—see Skinner 1998.

laws, and in the defence of the commonwealth from foreign injury; and all this only for the public good.

As far as individual's religion is an "inward persuasion of the mind," it could also be interpreted as property, or at least as having such status. A sufficient condition for the activation of the right of resistance is not only an immediate threat to life, liberty, and estate, but to a person's religion (Locke 2002: 404-405). Perhaps Locke perceived religion as a part of man's liberty. If the magistrate attempts to impose a hostile religion on his people, they have the right to resist, just as, for example, passengers traveling on board a ship have the right to organize a mutiny against their captain if he sails them to Algiers (which in Locke's time was the slave market for Christians captured by the Moorish pirates) (2002: 405). These ideas in the *Two Treatises* represented the fears of many Englishmen that the manifest Catholicism of Charles II and James II (who was a cousin of Louis XIV,[12] the omnipotent king of Catholic France) was a considerable danger to the sovereignty of the English Monarchy. These reasonable fears paved the way for the Glorious Revolution (1688) and the abdication of James II.

The relation between civil interests and individual interests highlights the function of the commonwealth, which distinguishes it from any other human institution. This relation is at its core of a *set of fundamental arguments* for the separation of church and state developed by Locke. If the commonwealth has a specific *raison d'être*, then what are the means that could be used for its realization? As Jeremy Waldron mentions, the right to use *coercion* is a unique political instrument of the commonwealth.[13] Locke assumes that the simple limitation of the jurisdiction of the civil magistrate is not a sufficient condition for toleration. He claims that the *nature* of the means implemented by political power for the ends of the commonwealth makes their use in the salvation of souls *irrational* (Locke 1824a: 10). In order to prove the validity of this thesis, Locke coins *three political arguments for toleration*. But coercion itself is unreliable as a basis for the operation of the commonwealth. According to Locke, the commonwealth was established as a result of a contract, and hence *consent* (Locke 2002: 331). When the political power resorts regularly to the use of excessive force and coercion, as well as damages property, this breaks the contract and leads to the dissolution of government and resistance to tyrants (2002: 406-428). Nevertheless, consent is more important as a basis of the church. Decisions in civil society are based on the principle of majority, instead of on mutual consent.

Hence, *the second set of fundamental arguments* includes the relation between civil interests (including religion as a property) and individual interests, and the use of legitimate coercion. To these we must add the grounds by which the church is defined.[14] The term "set" used here denotes a number of *interdependent* and compelling reasons for the separation of the two spheres. *The first set of fundamental arguments* logically precedes the other, and is constructed as an artefact of Locke's rational interpretation of faith and religion. The discussion in the *Letter begins with arguments* regarding the nature of the "religious" conflict in England (and in whole Europe), as well as the reasons for the civil war. This is a restatement of Locke's conviction that religion is used as a camouflage for greed and aspirations for political power.

Thus—and this is *the second argument in the first set*—Locke elaborates the notion (which appears as an undeveloped idea in the *Two Tracts*) of "the true religion." The true religion is not established "in order to the erecting an external pomp, nor to the obtaining of ecclesiastical dominion, not to the exercising of compulsive force; but to the regulating of mens lives according to the rules of virtue and piety" (Locke 1824a: 6). Therefore, Locke concludes that toleration is the "chief characteristical mark of the true church," and an essential part of the Christian religion (1824a: 5). *The third argument* (or group of arguments) in the first set upholds the vital importance of toleration toward people who differ from us in matters of religion, since it is "so agreeable to the Gospel of Jesus Christ, and to the genuine reason of mankind, that it seems monstrous for men to be so blind, as not to perceive the necessity and advantage of it, in so clear a light" (1824a: 9). Apparently Locke's interpretation of dogmatic questions deviates from the strict canonical rendering. Locke combines the views of the persuaded believer with faith in human reason, which is characteristic of the Modern époque. The first set of fundamental arguments also affirms the irrationality of persecution, although in a different way than the second

[12] Louis XIV abolished the Edict of Nantes, which had proclaimed toleration for Protestants, and prosecuted the French Huguenots.

[13] See Waldron 1993: 42.

[14] See below.

one. This fact illustrates the interdependence between these two sets of arguments.

2.2 Locke’s political and epistemological arguments in defence of toleration

Along with the first fundamental set—which includes not only arguments for separation of church and state, but also arguments in defense of toleration—we could distinguish between two other groups of arguments for toleration: *political* and *epistemological*. It has already been mentioned that from the perspective of the second set of fundamental arguments for separation, Locke formulates the thesis of the incompatibility of the ends and the means of political power with the “care of the Salvation of Men's Souls” (Locke 1824a: 10). The *political arguments* ("considerations") in the *Letter* are coined in support of this claim. The Care of Souls is not the responsibility of the civil magistrate, because:

- it “appears not that God has ever given any such authority to one man over another, as to compel anyone to his religion. Nor can any such power be vested in the magistrate by the consent of the people” (1824a: 10);
- the power of the magistrate “consists only in outward force; but true and saving religion consists in the inward persuasion of the mind” (p. 11);
- even if “the rigour of laws and the force of penalties were capable to convince and change men's minds, yet would not that help at all to the salvation of their souls” (p. 12).

The *epistemological arguments* are a development of the idea of “the true religion,” which, as we already established, is part of the first set of fundamental reasons for the separation of the two institutions. They follow from the principle that while we can have knowledge and certainty based on evident empirical facts, in the matters of religion (for example, that Jesus was crucified in Jerusalem) we can have only faith and persuasion. This principle is implicit in the first *Letter Concerning Toleration* and is articulated in Locke’s *Third Letter* devoted to this problem (Locke 1824c: 422).[15] The epistemological and political arguments are *complementary*. A good example of their interaction is the dictum that true religion is based on “inward persuasion of the mind,” and therefore cannot be changed through coercion. This should not be construed to mean that Locke denied that there is only one true way to Heaven. He was undoubtedly neither a sceptic or an atheist, but a believer.

3. The definition of "church" and the duty of toleration

In *A Letter Concerning Toleration*, Locke states that a "church" is a “voluntary Society of men, joining themselves together of their own accord in order to the public worshiping of God, in such a manner as they judge acceptable to him, and effectual to the salvation of their souls” (Locke 1824a: 13). This formulation contrasts with the definition of "commonwealth." There are three important points on which this contrast is based.

In the first place, in the core of the definition of the church lies the relation between the individual and the salvation of his soul. This relation gives *causa coeundi*—the ultimate end(s) for which people join a certain church (the public worshiping of God and salvation) and the means for its fulfilment. In this context, Lock states that “nothing ought, nor can be transacted in this society, relating to the possession of civil or worldly goods” (p. 16). This is its main difference from the state. It is also important to mention that the church creates its own ecclesiastical laws. This fact emphasizes the separateness of its jurisdiction.

In the second place, the church is a voluntary organisation. Hence its members are not bound together on the basis of coercion, and its institutional activities and its means depend not on the use of force, but on consent, “for force belongs wholly to the Civil Magistrate” (p.16). The means used by the church include: “exhortations, admonitions, and advice” (p.16). Locke’s gloss of this characteristic of the church is again based on rationality rather than theology. He states that “nobody is born a member of any church” (p.13). The coercionless character of the church stems from the fact that coercion is incompatible with its ends.

The third and last point concerns the fact that Locke also gives the members of the church the right to choose the particular kind of worship that they will follow. Therefore Locke accepts the principle of the religious pluralism.

These three features of the definition of the church have to be added to the already formulated "content" of *the second fundamental set of arguments*. This set outlines *differentiae specificae* of state and church.

[15] For the arguments in defence of toleration in Locke’s later Letters see Locke 1824b, 1824c, 1824d.

On the grounds of these two sets of fundamental arguments, Locke formulates "the duty of toleration" (Locke 1824a: 16). This duty has *three dimensions*, which regard the relationships and interaction between: the members of each church; the different churches; and the civil magistrate and religious institutions. According to the *first principle* of this duty, although no church is required to retain a member who violates its laws, the excommunication of such a person must not be a detriment to his body or estate (p.17). This principle follows from the second fundamental set of arguments. The *second principle* postulates that "no private person has any right, in any manner, to prejudice another person in his civil enjoyments, because he is of another church or religion" (p.17). This principle is based on the first fundamental set of arguments. The *third and last principle* requires clergy to confine ecclesiastical authority ecclesiastical authority within the bounds of the church (p.21). Both the first and the second sets of arguments serve as grounds for this principle.

To sum up, as far as the relationship between different religious institutions is concerned, the duty of toleration demands that churches have to be tolerant toward their members and toward each other. The most important part of the duty regards the interactions between the commonwealth and the church. Locke postulates that whatsoever "is lawful in the commonwealth, cannot be prohibited by the magistrate in the church" (p. 34). However, if the civil sovereign sanctions a law which is in accordance with the public good but in a disagreement with a private person's conscience and religious convictions, the violation of such law will justly result in the punishment of the transgressor (p. 43). To *the third dimension of the duty* further belong the cases in which the civil magistrate should deny the right of toleration to certain individuals or groups, who:

- share opinions contrary "to the human society, or to those moral rules, which are necessary to the preservation of the civil society" (Locke 1824a: 45);
- arrogate to themselves "some peculiar prerogative… opposite to the civil right of the community," e. g. the prerogative to dethrone the king if they differ from him in religion (p. 45);
- deliver "themselves up to the protection and service of another prince" (p. 46);
- "deny the being of a God" (p. 47), and hence the existence of natural law.

These four issues are hotly disputed in scientific literature. Their analysis and examination goes beyond the scope of this paper. Still, we must mention that these limitations of toleration follow from the reasons for the very existence of the state. Locke perceives these issues as a threat to property and its foundation—natural law.

On the basis of the above analysis, we can fill the lacunae in the third stage of the development of Locke's views on toleration. The separation of church and state in *A Letter Concerning Toleration* is built upon the first set of fundamental arguments (the nature of the religious conflict in England, the concept of the true religion, and the vital importance of toleration), as well as the second set (*differentiae specificae* of state and church). They mirror the elaboration of Locke's ideas on these issues, and interact with the notions and postulates in the *Second Treatise* (which contains Locke's definitions of "natural law," "political power," "religion" as a property, and "the right to resist"). With the help of these sets of arguments, Locke constructs his apology of toleration. Its essential parts are the political and epistemological arguments, as well as the duty of toleration.

The three stages of development of Locke's political ideas are not only derived from political practice, but also have influenced it. According to Lois Schwoerer, Locke became acquainted with many Whig leaders while he was living in exile in Holland, leaders who later organized the Glorious Revolution, and his ideas very likely influenced them (Schwoerer 1990: 532).[16] Unfortunately, the same does not apply to his ideas on toleration. The Toleration Act of 1689 shows "how far outside reasonable opinion was Locke's call for toleration of anyone who believed in any god and for the end of coercion in religion" (Tully 1991: 651). The Act gave a limited right of toleration only to the orthodox Dissenters. Despite the fact that Locke's views on toleration were not fully accepted by his contemporaries, they remain a valuable and inseparable part of his intellectual heritage, without which it is hard to imagine the shape of modern political thought.

References

Abrams Ph (1967) "Introduction" to *Two Tracts on Government.* Cambridge: Cambridge University Press.

[16] Cf. Ashcraft 1986.

Ashcraft R (1986) *Revolutionary Politics and Locke's Two Treatises on Government*. Princeton University Press.

Cranston M (1988) Locke on Toleration. In *Rivista-Internazionale-di-Filosofia-del-Diritto,* 213–219, AP-JE 88.

Dunn J (1990) *The Political Thought of John Locke: An Historical Account of the Argument of Two Treatises on Government*. Cambridge: Cambridge University Press.

Horton D, Mendus S (eds.) (1991) *John Locke: A Letter Concerning Toleration in Focus*. Routledge: New York.

Locke J (1824a) A Letter Concerning Toleration. In *Works of John Locke*, vol. 5, 12th edition, 1–58, London.

Locke J (1824b) A Second Letter Concerning Toleration. In *Works of John Locke*, vol. 5, 12th edition, 59–138, London.

Locke J (1824c) A Third Letter for Toleration. In *Works of John Locke*, vol. 5, 12th edition, 139–546, London.

Locke J (1824d) A Fourth Letter for Toleration. In *Works of John Locke*, vol. 5, 12th edition, 547–574, London.

Locke J (1967) *Two Tracts on Government*. Cambridge: Cambridge University Press.

Locke J (1997) An Essay on Toleration. In *Locke: Political Essays*, ed. Mark Goldie. Cambridge: Cambridge University Press

Locke J (2002) *Two Treatises on Government*, 12th reprint. Cambridge: Cambridge University Press.

Schwoerer L G (1990) Locke, Lockean Ideas, and the Glorious Revolution. *Journal of the History of Ideas* 51(4): 531–548

Skinner Q (1998) *The Foundations of Modern Political Thought*, vol. II. Cambridge: Cambridge University Press

Tully J (1983) Introduction to A Letter Concerning Toleration. In *A Letter Concerning Toleration,* Hackett Publishing Company.

Tully J (1991) Locke. In *The Cambridge History of Political Thought 1450-1700*, ed. Mark Goldie. Cambridge: Cambridge University Press.

BALKAN JOURNAL OF PHILOSOPHY. Vol. 7, Issue 2, 2015

HOW THEOLOGY LOOKS AT NEUROTHEOLOGY

Ali Akbar Ziaee
Director of Cultural Center in Kuala Lumpur
e-mail: aaziaei@gmail.com

Abstract:
In this paper, I discuss neurotheology and its scope as defined by natural scientists. In my opinion, in denying the paranormal world this science has failed to present criteria to distinguish "genuine" from "non-genuine" religious experiences. However, this science can be used to recognize revelation, inspiration, true dreams, and feelings of no-space, no-time, and annihilation of the human body in mysticism, but we should not attribute genuine experiences to mental disorders, since genuine experiences can be distinguished from non-genuine ones.

Key words: neurotheology, religious experience, revelation, mental disorder.

I. The rise of neurotheology

Sub-divisions of neuroscience include neuropsychology, neurophysiology, and neurophilosophy. Another new discipline has recently been added to this list, and it has provoked extensive discussion among contemporary scientists. This newly established discipline is neurotheology—or in other words, neurology of spirituality. Taken together, these four fields can serve as the cornerstones of neuroscience, and as important tools for decoding and dissecting the human psyche and behavioral characteristics. The general concept of neurotheology is nothing more than the process of relating the phenomenon of spirituality and its resultant spiritual experiences with possible scientific explanations (scientific hypotheses) for such experiences. Those interested in this field believe that there is an evolutionary and neurological basis for all internal mental (subjective) human experiences which derive from religious and spiritual practices [1]. In discussions about theology and religion, the term "neurotheology" is not used properly, focusing as it does upon a very few specific areas of religion (e.g. near-death experiences) while ignoring the rest.

The term "neurotheology" was coined by a man named James Ashbrook, an expert in theology who began studying neuroscience. However, so far the meaning of "neurotheology" has remained obscure and ambiguous; a clear definition of the term has never been provided. The clearest thing that may be said about this science is that it tries to find logical and scientific reasons for religious experiences based on scientific descriptions [2]. The goal of neurotheology is to study and "decode theology based on a neuroscientific perspective." In this way neurotheologians hope to understand the innate desire of human beings for religion and religious myths [3].

II. Neurotheology and its scopes

In the early twentieth century, religious experience was considered to belong properly to the realm of psychology, but gradually lost its importance for psychologists over time. In the twenty-first century, however, religious experience has not only seen a resurgence of interest among psychologists, but has become one of the major issues in neurology.

Today neuroscientists study matters that were once considered to be outside the scope of science. They now discuss not only traditional subjects such as memory, pathology, and motivation, but also personality, prejudice, social attachment, moral judgment, and social emotions such as empathy. Furthermore, empirical research on religion and religious experiences based on evolutionary and scientific views has increased dramatically. What is the role of the brain in the human experience of God? Is religious experience specifically different from other human experiences? Or is it different from other experiences merely in terms of its content, with religious and non-religious experiences being otherwise alike neurologically?

Is it possible for neurotheology to investigate "divine revelation" from a scientific perspective? This newly emerged science examines the nerve origins of religious experiences, and wants to know what is happening in the brain during religious and spiritual experiences. On the other hand, it seeks to discover the relationship between brain disorders and religious experience, and ultimately seeks a solution to eliminate brain disorders.

Here neurology enters into the domain of religion and spirituality, and some researchers claim that they have found the area of the brain that generates God, that the influx of electrical pulses in one area of the brain may cause religious and mystical states. Some other scientists have stated that divine inspiration and revelation have been the result of mental disorders. These scientists have claimed that the prophets were suffering from diseases such as epilepsy, schizophrenia, and psychosis. In their view, holy books purported to be divinely inspired are nothing but words repeated in a poetic style by those who suffered from epilepsy and nervous disorders in their temporal lobes. In this paper, I will demonstrate the following: first, neurotheology is about the religious experience of mankind; second, revelation is not a human religious experience, but possesses a divine origin and is outside the scope of neurotheology. On the other hand, I will also discuss the fact that a religious experience can be divided into genuine and non-genuine types, and I will elaborate the neurotheological perspective on these two issues.

III. Human experience of the world outside

The brain plays a role in all human behaviors and experiences. In whatever we do, think, and feel, and in however we behave, the brain plays a role. In the nineteenth century, psychologists and scientists researched anything and everything that we receive from the world through our senses, especially visual perception. Through the further development of these studies due to scientific advances in optics and acoustics, psychologists discovered how the energy of light and sound impulses is changed into an electric stimulus that transmits this acoustic and visual information to the brain. The twentieth century witnessed further developments regarding the study of other sensory information (e.g., touch, taste, and smell). To get sensory information from the environment, there must be receptors which are able to react to a certain energy, and this energy must be of sufficient strength to stimulate the sensory receptors into generating an electrical impulse (sensory operation). When this electrical message is created, it is transmitted by neurons of the nervous system to specific parts of the brain, and this is where the electrical impulses are interpreted (i.e. the process of perception, or the operation of understanding). As far as we are sensitive to different types of energies in our environment, all of these energies are turned into electrical impulses in our nervous systems. The concept of "representation" in the brain refers to how different things are experienced in our brains—in other words, how an external thing is interpreted by the brain as a model, or as a pattern of electrical stimuli.

Feelings and thoughts are the products of electrical impulses that enter into a particular part of the brain or arise out of it. Movement of the body is also caused by electrical stimulation in a part of the brain associated with movement, and from there it is sent to our muscles. In this way, in whatever we do, think, feel, or experience, we are connected to our brain. Even all our perceptions of ourselves are interpreted in our brains. A strip of the brain's surface tissue is responsible for interpreting various parts of our body. For example, when an electrical impulse from the sensory receptor in the index finger begins and finally and reaches the cortex of the brain which is dedicated to interpreting that finger, we are able to understand the feeling in that part of the body. The phantom limb phenomenon in humans demonstrates how such a sensation and perception of feeling are possible. This phenomenon may occur to anyone who has lost a limb, such as hand. Despite losing that limb, the person can still experience the same sense of the organ, because electrical impulses still reach the part of the brain which is responsible for interpreting that organ [4].

IV. Emotions, beliefs, and intellect

Emotions have been one of the most important issues in psychology for decades. Likewise, in recent years neuroscientists have examined emotions through the use of the modern technologies, including brain imaging techniques. According to research by neuroscientists, emotions play a more important role than was previously thought in our thinking and decision-making. Although Descartes believed in the separation of mind from body, recent studies on emotions have shown that affection is closely related to mental ability and even to personal identity [5]. Such studies suggest how emotions play a role in the quality of our reasoning in decision making. According to the view of traditional rationalists, it is reasoning that leads human beings to make decisions, and emotions do not play a role in this regard (not even a little). Haidt [6] showed that emotions about a decision occur before the reasoning. In fact, we use reasoning to justify a decision that has already been shaped by our emotions.

Other researchers suggest that social emotions (e.g., empathy) are associated with brain structures. Farrow et al [7] used MRIs to discover specific patterns of human brain activity that are linked directly with the capacity for forgiveness and empathy. Parts of brain in the prefrontal cortex (orbital and medial sections) and the superior temporal sculus play a role in social activities and moral emotions such as guilt and shame [8]. The brain circuity involved in both positive and negative emotions follows two underlying systems, together with many other elements. The approach system (positive motivation) provides the groundwork for positive emotions, while the withdrawal system plays a role in negative emotions (such as disgust and fear). These systems interface with the prefrontal cortex (PFC) and the amygdala [9] through the brain's circuity.

The prefrontal cortex has many parts, but two of the most important parts of it include the orbitofrontal and dorsolateral sections. The orbitofrontal part of the cortex is more associated with emotions, while the dorsolateral part of the cortex has historically been seen as closely associated with memory, but recent studies have shown that this sector is also associated with emotions. Although not all studies have been confirmed, researchers have shown that positive emotions in the prefrontal cortex hves links with the left hemisphere of brain, while the right hemisphere has links with the processing of negative emotions [10].

The prefrontal cortex is involved in predicting positive and negative effective consequences, and it is this same part of the brain that was grievously injured in a man named Phineas Gage as the result of an explosion that drove an iron rod through his skull; the rod broke his cheekbone and went through his body, reached the orbitofrontal cortex (anterior cortex), and went out of his skull. While his life was saved, he witnessed changes to his character and instability of his emotions [9], and his ability to effectively design works, participate in social environments, and demonstrate a sense of responsibility to himself and others was impaired [11]. Any injury to the prefrontal cortex (PFC) may lead to behavioral disorders in terms of psychopathology, such as lack of social interactions and knowledge, a diminished sense of guilt for behavior, and faulty feelings associated with reward and punishment [12].

Although the amygdala is not as involved in emotions as the prefrontal cortex is, the role of the amygdala in emotions—especially fear—has been outlined well by Ledoux [13]. We have no idea how much the difference between the two hemispheres influences our emotional processes, or if the amygdala is equally effective in both positive and negative emotions [9], but we know that the amygdala plays an important role in acquiring probable risks, evaluating dangerous situations, and stimulating fear [14, 15]. To explain the primary role of emotions in religion, Watts [16] states that both theologians (such as Friedrich Daniel Ernst Schleiermacher) and psychologists (such as William James) have recognized the fundamental role of emotion in religion. Schleiermacher believes that emotions are the source and foundation of religion. According to this philosopher's point of view, the best way to determine a person's religion is to assess the extent of his emotions. Some researchers [17] believe that cancer patients define their faith in positive rather than negative emotional words. Based on these investigations, religious belief can be known on the basis of emotional terminology. Schleiermacher also believes that religious emotions are fundamentally different from other human emotional experiences.

There is no doubt that religion is associated with emotions, but the question here is: do emotions compose the entirety of a person's experience, or can categories of "wisdom" and "rationality" be identified which cannot be seen as "feelings" or "emotions"? It seems that material empiricists do not recognize a category of "wisdom" that is independent from "emotions," and they define all the functions of the human brain within brain sensory processes and electrical pulse transmission in the brain's nerve system. An alternative to this reductive materialistic view is that human wisdom is abstract and non-material, and that this abstract matter uses material brain cells for information processing, and that the nature and truth of this wisdom is independent from the physical brain, and that it uses the brain systems as a tool for self-development. This abstract wisdom is not only in the realm of feelings and emotions, but also includes abstract concepts and absolute abstract rationality. The idea that non-physical matter can exist is a philosophical issue, and many reasons have been given which support the existence of an abstract essence of material. Existence—both of matter and abstractions of matter—is divided into three parts: the scope of the material and it potentials or capacities; the scope of the abstraction of the material with regard to physical representations such as shape, size, position, etc.; and the scope of the abstraction of the material and its effects on "the world of intellect." Emotions and sensory perceptions need material and brain neurons, but emotions are dependent upon the human abstract soul, and their dependence upon the abstract soul is like the relationship between an ocean and waves. Research by empiricists and materialists is limited to rational and sensory perceptions which have not entered into the realm of abstraction from the material domain (i.e., the world of ideas and the world of intellect). In other words, if there were no matter—brain neurons and their inter-brain interactions—there would be no background upon which to create sensitive perceptions in the soul.

V. Neurotheology and reductionism

Referring to the investigations carried out so far, it was said that the brain plays a role in emotions, and that there is a relationship between cognition and emotions. In addition, the importance of cognition and emotion in religious experience was empirically examined. In the last decade, a new way to understand and explain religious experience arose called neurotheology, which seeks to evaluate the way the human mind or brain works when it is connected with God [18]. The role of the brain in religious experience and the relationship between man and God is one of the most important issues for scientists, theologians, and lay people. A recently published book examined the views of twenty-one neuroscientists, psychologists, philosophers, and theologians [19]. An article in *Newsweek* dated May 7th, 2001 with the title "God and the Brain" was dedicated to this issue. Various books have been published on the subject, and stories in newspapers about the brain and religion are published regularly. Some people are afraid that neurotheology is going to prove the nature of God (or deny his existence), or that science is going to find the place for God in the brain. In fact, scientists who study the brain and religion have appealed to philosophical interpretation of this knowledge—which has been recognized as an independent science—or they have used ontologically reductive views in analyzing their experimental data in a way that has led to this knowledge (e.g., Michael Persinger, 20). It must be said that the existence of God can only be shown through non-empirical methods, and that empirical science cannot find the answer by itself. Empirical studies of religious phenomena can show what happens physiologically to a man when he has a religious or spiritual experience. Neurological analysis of a religious experience is a legitimate topic for discussion, but of course there are limitations to what can be studied scientifically. One of the first scientists to examine a religious experience was Michael Persinger. In the early 1980s, he published a number of articles about religious experience and brain activity [21], and at the end of the decade he published a book about belief in God in terms of neuropsychology [20]. He emphasized the role of the temporal lobe (particularly the amygdala and hippocampus in the limbic system) in typical religious and spiritual experiences. One of the reasons for his emphasis on the role of the temporal lobe is that in such experiences there is a type of electrical instability in the firing patterns of that structure that is not necessarily compatible with the actual sensory data. In fact, a kind illusion and remembrance is combined in these kinds of experience. The temporal lobe is the location of seizure disorders, as well as the vascular abnormalities that are directly associated with both illusions and dreams [21].

Based on Persinger's hypothesis, religious experience is the result of a limited and short-term electrical activity called Temporal Lobe Transients (TLT). So according to Persinger, in extreme cases of seizure disorders some other behavioral symptoms such as crying, facial and speech impairment, and even dementia are also seen. Interpretation and explanation of this phenomenon depends on environmental factors, and on when and where it happens. Persinger believes that we can experience a relationship with God because our temporal lobe has become developed enough to do so. He believes that if the temporal lobe had evolved in a different way, these religious experiences of God would not occur [20, p. 14]. In addition to what was expressed regarding the reductive view, Persinger believes that since the religious experience of God is a biological figment of the human brain, there is a risk that such experiences can be controlled or exploited by others. He also expressed concern for those with seizure disorders in the temporal lobe who die or are killed. Persinger says that there are many stimuli in religious activities (such as religious music, specific movements of the body such as swaying and dancing, frequent or repetitive sounds, and even specific smells of religious centers) which can provoke seizures in the temporal lobe [20]. Unfortunately, he has not provided sufficient evidence to prove such a claim. Although the temporal lobe is the main site of seizure disorders, Persinger failed to provide evidence for the defense of his claim that religious experience is merely a figment of such seizures. Persinger also could not prove his claim that belief in God is a cognitive virus, and a kind of the imagination and illusion [22]. Joseph [23] believes that the structures of the limbic system (such as the amygdala, hippocampus, and the inferior part of the temporal lobe) are the basis of religious and spiritual activities. He links the universality of religion with the limbic system's basic human drives. These brain structures are involved with sexual activity, anger, and irritability. He concludes that is why the Torah's God is strongly related to sexuality, and often displayed anger and demanded fear in the presence of a sacred and the transcendent power [23, p. 118]. Joseph believes that due to activation of the limbic system, many religious people are often hateful, adventurous, and/or criminal individuals. Confirming Joseph's claims on the role of the limbic system, Saver and Rabin [24] argue that the same limbic system of the brain is involved in both the religious and non-religious experiences, the only difference being that religious experiences are inexplicable and profound; this is because the limbic system identifies such experiences as ultra-personal and very important, causing rapture and pleasure.

Against such reductive claims, Joseph believes that science cannot deny the role of spiritual and paranormal worlds in the influence of the limbic system in religious activities. Thus, as long as the limbic system is metaphorically known as the location of the soul and of the person's relationship with God, this "transferor to God," according to Joseph's vision, increases survival expectancy, because it helps man to communicate with reality (here spiritual reality is concerned) in just the same way as other brain structures and networks do [23].

Carol Albright is highly critical of the reductive perspective on religious experiences. She believes that identifying the "God Model" or "Transfer to God" with the limbic system obscures the fact that the brain acts as a whole, and that specific parts of it do not act independently from other parts of the brain. Although some activities are enacted in certain parts of the brain, this does not mean that the brain does not through a process of complex interrelationships (compatible and self-organized) in these areas. It is perhaps more correct to claim that the whole human brain is involved in religious and divine experiences [25]. Because of the different aspects of religious experience, it is expected that the brain is involved in religious experience in different ways and with different parts. Albright and colleagues used Paul McLean's triple model to relate religious beliefs and different conceptions of God with three major parts of the brain called the Reptilian, Mammalian (the limbic part), and Neocortex.

The reptilian brain is responsible for human behavior that is common with reptiles, such as the preservation of life and property, reproduction, eating, and the like (any behavior that preserves life). Such behaviors can provide problems for human beings and therefore must be controlled. Religious activities and religious beliefs teach people the right ways to live, e.g. the proper upbringing of children, sound sexual behavior, and control of violent behavior and law-breaking. In other words, religious teachings try to bring the reptilian brain under control [26]. These experts believe that Christian beliefs such as eternity and the unchanging nature of God are related to the activity of this reptilian brain. The mammalian brain is composed of the limbic system, which adds feelings and emotions to reptilian brain activities, and these feelings and emotions help us distinguish secure and insecure behaviors from each other. The limbic system also plays a role in memory. According to Albright and Ashbrook [26], through memory and emotion we can have a meaningful relationship with God. These two parts of the triple brain are surrounded by a third section called the cerebral cortex, which is associated with our higher human functions, such as language. The information related to other parts of the brain is processed in this section, allowing us to decide or judge [26]. All of these three parts are involved in religious experiences. Spiritual experiences of love and belonging are probably related to the limbic system, while different types of virtuous behaviors are related to the reptilian brain, and understanding and a sense of calling from God is associated with the cerebral cortex (e.g. the frontal lobe). Patrick McNamara believes that religious activities are related to the activation of the frontal lobes and particularly the prefrontal cortex, which acts as an intermediary for executive cognitive functions [27]. The frontal lobe of the brain was the last part to mature in the evolution of human beings, and the development of this portion of the brain is finalized in the last stages of puberty. According to McNamara's view, one of the reasons for the importance of "executive cognitive function" in religious rituals is the ability to relate subjective interpretations to other people or factors (such as God). A man with an advanced theory of mind can anticipate the thoughts, actions, beliefs, and intentions of others. In addition, the frontal lobes directly participate in emotional operations, nonprofit and charity behaviors, compassion, empathy, and self-awareness [27]. These executive cognitive functions are also involved in religious activities. The theory of mind enables one to better understand the concept of God. According to McNamara, if the prefrontal cortex is injured, man’s view about God is changed a hypothesis that has not been proven empirically. According to this hypothesis, any damage to the frontal lobe will harm human participation in the emotional aspects of religious experience, weaken self-discipline and self-control related to religious teachings and beliefs, and undermine religious behaviors such as empathy, compassion, and forgiveness of others, as far as it is relevant to these activites. To be able to reach maturity and have free choices as human beings, we must have healthy and active frontal lobes. McNamara believes cultures that encourage certain activities can contribute to the growth and maturation of the frontal lobe, and religious activities that are common in almost all of the world (such as praying, meditation, and spiritual deliberation) can be a means to accomplish this maturity [28].

The frontal lobes—particularly the prefrontal cortex and the limbic system—are important in the somatic marker hypothesis developed by Damasio. According to this hypothesis, physical (bodily) changes such as nervous and hormonal responses create feedback that marks a certain status as a good thing (to be encouraged) or a bad one (to be avoided). This automatic feedback is also affected by cultural conditions, and serves as a good response to the development of the “theory of mind” and making appropriate decisions, and as a result it has lots of benefits on an evolutionary basis. The somatic marker hypothesis of Damasio can justify McNamara's theory on how religion has been a universal phenomenon. In his view religious beliefs, moral character, and religious practices are all part of the physical marker system, and are compatible with evolution because they provide the tools needed to evaluate the situations and make decisions.

It might be argued that the most famous research on the neural foundations of religious experience belongs to Andrew Newberg and Eugene d’Aquili, who raised the issue to the level the public awareness [29] and published stories on it in general publications such as *Newsweek*. Newberg and d’Aquili conducted experiments on the human brain during religious experiences—such as praying and meditation of Franciscan monks and nuns, as well as the meditation of Buddhists—which indicated that some parts of the brain are activated in during those experiences. Using a brain imaging technique called single photon emission computed tomography (SPECT), one can identify the duration of radioactive tracers and discover the activities of different parts of the brain and its structures. A few minutes after tracer injection into the brain, it is possible to extract the image. Newberg and Eugene d’Aquili injected the tracer at the peak of religious activity, and a few minutes later they could take images of brain activity. Thus, taking an image through this method did not interfere with the religious activities. Activity in the parietal lobe near the top rear, upper part of the brain (posterior superior parietal lobe) during the rise of the religious activity is severely reduced (compared with activity in normal conditions). According to Newberg, the role of this part of the brain is to help us to organize ourselves in the world. This means a strict distinction between us and the objects in the world. During prayer, this part of the brain is relaxed is still, which the Franciscan nuns perceive as a close relationship with God, and which the Buddhists perceieve as a sense of timelessness and infinity [29]. According to Newberg, just when this part of the parietal lobe shows decreased activity, some parts of the frontal lobe and the limbic system show increased activity for better focus on spirituality during praying and meditation. It is worth noting that these results are based on experiments conducted on only a small number of people.

As we can see, all of these scientific studies are on hypothesis level, and have examined the background for creating religious perceptions and emotions. No philosophers deny the connection between the physical aspects of religion and the human nervous system, and it is obvious that if a part of the brain is physically disrupted it can affect a person's religious beliefs and behaviors negatively or in some cases reinforce the religious beliefs of people, but the main problem is that the origin of religion and the relationship of religion with the Creator and His prophets are not the things that empiricists and neuroscientists and even the newly emerged discipline of neuotheology have addressed.

VI. Our perception of god in the brain

What does does it mean to say that through prayer and other spiritual practices a part of our brain is “lit up” or “made active”? Some researchers have ventured far beyond empirical certainty and claimed that God can be found in a part of the brain that they call the “God spot.” This claim limits existence to the body and brain and results in total materialism, which contradicts the intellect and common sense; it is the same as claiming that taking pleasure in sailing is located in the brain's "navigation point." Certainly when I enjoy maritime patrolling some parts of my brain become active (turned on), but is this all that happens there? Some researchers like Persinger [30] believe so. Persinger says that experiencing God is nothing but a provocation or activation of the temporal lobe. In other words, it is a biological product of the human brain [31].

Many researchers have greater precision in expressing their research. Among them are scholars like Newberg and D'Aquili [32], who argue that their research does not say anything about God's existence. Scientific findings do not imply that God is merely the result of the stimulation of a group of brain cells; likewise, these studies do not provide any evidence for or against God's existence. If this research, like other scientific studies, is understood properly, then it can be seen that it does not say anything about philosophical positions or beliefs. The results of this research does not hint at the truth of God's existence. It cannot be denied that religious experience and the knowledge of God must be filtered through our brains (which organs besides the brain could we use to grasp such experiences?). But we use our brains in whatever we do and experience, and this is not limited to religious experiences. We use our brains when we study, and when we listen to music!

A leading figure in the integration of science and psychology is Malcolm Jeeves, who believes that we understand a phenomenon at different levels and in different layers [33]. For example, the mental and the physical are two different layers of a personal unit. In Jeeves' view, these layers are associated with each other. These two perspectives create a dual view, but not a dual essence. This irreducible dualism is not just Jeeves' point of view. In the field of psychology, a man named Gustav Theodor Fechner had a similar view regarding the relationship between mind and body [34]. Also, Clive Staples Lewis [35] discussed looking at a beam of light (objective, a third party experience) and looking through the light (subjective, a first person experience).

In the same way we can look at religious experience from different perspectives. We should not follow those previously discussed and examine a phenomenon only from the physical side. Like all human experiences, a religious experience is non-reductive and irreducible. Of course it is clear that our brain (and body) are changed during a religious experience. Through multiple human perspectives on an issue, we can examine religious experiences from both objective and subjective viewpoints. It could be useful to explore religious and spiritual experiences with such an approach, and the role of cognitive neuroscience and physiology in explaining the experience from a physical perspective could be brought to light.

VII. Definition of revelation from a theological perspective and its presuppositions

Revelation literally means to instill awareness in somebody else, and one can describe it as hidden, fast, and wordness. In other words, revelation refers to prophet's connection with God in a direct or indirect way (e.g. through the angel Gabriel) in order to receive a message or knowledge; or it means God affecting the senses of a prophet in order to communicate with them.

Revelation sometimes takes place directly and sometimes indirectly without the mediation of Gabriel, such as God's speech from behind a veil (e.g. God's talking to Moses through a tree), and it is sometimes instilled by an angel in the heart of God's messengers.

"Revelation" in the Christian and Jewish religions is defined as: disclosing a kind of knowledge through a relationship with God or through intermediaries or non-material (paranormal) links.

The aforementioned definitions reveal that revelation is a mysterious and non-obvious relationship between man and God. And through revelation specific knowledge is given to human beings in order to lead them toward the path that God wishes them to follow. The truth of the phenomena of revelation is that a connection lies between the supernatural and human beings, and this connection is not of a known and material kind. The hiddenness and mysterious nature of revelation are its main characteristics, and the domain of human knowledge includes merely the physical and material aspects of man, and not his intellectual aspect. The fact that there is a world of non-material things beyond the material world which we might refer to as "the world of intellect" or "the world of ideas"—as well as the the fact that there is a self-existent thing that is the main cause of all things—is the presumption of revelation. Theology and philosophy prove the rationality of this stance, which is beyond the scope of the empirical sciences. Therefore neurotheology will never be able to reach the realm of revelation and present a non-fallacious interpretation of it. However, it should be noted that one side of revelation is the human being and the other side of it is the immaterial world, and if we accept that man has an immaterial substance and that God has combined it with the physical world in such a way that it can passed down from person to person, then we can see that the inherent essence of the divine prophets and their bodies will carry the message of God's revelation and will be in a secondary form. The main problem empiricists face regarding the meaning of revelation in Abrahamic religions is that they are not able to perform clinical trials on the Prophets' bodies, and it would be impossible to do scientific studies to reveal the material impact of revelation on their bodies. The only thing that empiricists and neurotheologists can do is examine people undegoing spiritual and religious experiences; they can examine the influence of spirituality and religion on man's material aspects, but not his spiritual and non-material dimension.

VIII. Religious experience and the difference between its genuine and non-genuine forms

The writings of neurotheologists suggest that by "religious experience" they do not mean the whole religion as beliefs, rituals, and rites, but rather religion as "speaking in tongues," visual hallucinations, auditory hallucinations, feelings of oneness and unity with the universe or the supernatural, a sense of no-space or no-time, and any other unusual sensation related to religion. Based on this definition, everything that human beings see, hear, or feel due to stress, psychological disorders, hereditary diseases, or any other faulty reason for a belief or sensation would be a religious experience. It is therefore a naturalistic and materialistic scientist who sets the criteria for "true" and "untrue" experiences, and with a general sentence he can relate all matters of the paranormal world—such as revelation, inspiration, and true dreams—to issues arising from psychiatric disorders such as schizophrenia, paranoia, and epilepsy as a categorical assumption, and then attempt to find the cause of that purely material phenomenon in human life. With no argument, these scientists assume that there is no truth to the supernatural, and if anyone has any reports of such cases, he will certainly be affliected by a psychological disorder. The problem with this assumption is the lack of tests, reasoning, or arguments for spiritualism being unreal. And what are the criteria for judging what is real and what is imaginary, anyway? As has already been mentioned, the topic of revelation is outside the scope of neurotheology. If we assume the existence of the paranormal universe—which has been proven intellectually in philosophy—then how can neurotheology be applied? As we know, all the self-existent, general prophecy and particular prophecy have been proven on the basis of intellect, and even many philosophers have insisted on their self-evidence. If we assume that divine prophets and divine revelation are inherently outside the scope of neurotheology, we must examine whether this new science can be applied in the field of contemporary religious experiences in the modern era.

Based on the definition, and according to the perspective of the believers, religious experience is presented in both genuine and non-genuine types. Genuine religious experience includes authentic revelation, inspiration, true dreams, feelings of annihilation in God, the extracting of the soul out of the human body in the final stages of the spiritual journey, the sense of no-time and no-place due to religious austerity, continuous prayer, recitation, and practicing religious duties. The non-genuine type of religious experience is due to psychological disorders which lead to "speaking in tongues," auditory and visual hallucinations with religious themes, and a sense of no-space and no-time based on non-religious austerity such as meditation and appealing to secular mysticism. Thus, since in the contemporary era the concept of revelation has ended (specifically with regard to religious prophecy), the question is about the area of neurotheological studies regarding the two kinds of genuine and non-genuine forms of experience.

Neurotheology can survey the first type of religious experience. This means that if one follows divinely inspired religious teachings (divine revelation) to communicate with the paranormal world, and then one has an experience such as a sense of hearing the divine, seeing a divine light, or feeling a mystic sense of self-annihilation and so on, a neurological scientist will be able to analyze the material effects of those intellectual aspects on their bodies. He will also be able to examine the effect of blood pressure in the temporal lobe, the neurotransmitters such as dopamine and serotonin in the brain, the extent of electrical impulses in the central nervous system, and the quality of connection of religious emotions with the hippocampus and the amygdala. But all of these are examples of abstract world effects on the material world which can be examined by a natural scientist, so that through, but this scientist cannot rule out these experiences as non-genuine ones deriving from psychological disorders solely based on an examination of the physical condition of the believers. The scene of physical interactions in the believer is like a shadow in which all the real elements are behind the curtain, and the viewers and visitors (neurotheologists) can only see a view of all the truth in terms of the shadows, and then they rule out certainly about what is behind the curtain.

Regarding the second type of religious experience, it should be said that this new knowledge can be a good way to solve people's mental problems in religious communities, although of course we should not neglect the fact that God pays special attention to patients who suffer from a mental or physical illness of any kind, and if we accept that God is sovereign over all human affairs, then we cannot accept with certainty that all of the religious experiences of individuals suffering from psychological disorders are unreal.

IX. Conclusion

The existence of God has been proven based on non-empirical methods, and the empirical study of religious phenomenon can only show what happens physiologically to a human being when he has a religious and spiritual experience. No philosophers have denied the connection between the physical aspects of religion and the human nervous system, and it has become obvious that if a part of the brain becomes physically disrupted, it can affect a person's religious beliefs and behaviors. However, the main problem is that the origin of religion and the relationship of religion with the creator and his prophets are not the problems that empiricists, neuroscientists, and the newly emerged neuorotheologians have addressed. In this article, I argued that neurotheology can only investigate the human materialistic aspects of religion and religious experiences, and cannot provide a report on the spiritual world; divine revelations as divine messages sent to the prophets is out of the scope of this science. Even if we accept that the prophets' bodies were subject to materialistic changes during revelation, this science cannot judge their physical condition, because revelations have come to an end with the end of prophecy, and scientists do not have access to their bodies. This article argued that religious experiences are divided into two types—genuine and non-genuine—and that under the mentioned circumstances, both types can be studied by neurotheology.

References

Albright C R (2000) The "God module" and the complexifying brain. *Zygon* 35: 735–744.

Amaral D G (2002) The primate amygdala and the neurobiology of social behavior: Implications for understanding social anxiety. *Biological Psychiatry* 51: 11–17.

Anderson S W, Bechara A, Damasio H, Tranel D, Damasio A R (1999) Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature Neuroscience* 2: 1032–1037.

Ashbrook J B, Albright C R (1997) *The humanizing brain*. Cleveland, OH: The Pilgrim Press.

Ashbrook J (1984) Neurotheology: the working brain and the work of theology. *Zygon* 19 (3):331–350.

Craig A-S (2012) Neuroscience for the Soul. *The Psychologist* 25 (7): 520–523.

Damasio A R (1994) *Descartes' error*. New York: Avon Books.

Damasio A (1999) *The feeling of what happens*. New York: Harcourt Brace & Co.

d'Aquili E G, Newberg A B (1999) *The mystical mind.* Minneapolis: Fortress Press.

Davidson R J, Jackson D C, Kalin N H (2000) Emotion, plasticity, context, and regulation: Perspectives from affective neuroscience. *Psychological Bulletin* 126: 890–909.

Davidson R J, Irwin W (1999). The functional neuroanatomy of emotion and affective style. *Trends in Cognitive Sciences* 3: 11–21.

Davidson R J (2002) Anxiety and affective style: Role of prefrontal cortex and amygdala. *Biologicat Psychiatry* 51: 68–80.

Farrow T F D, Zheng Y, Wilkinson I D, Spence S A, Deakin J F W, Tarrier N et al. (2001) Investigating the functional anatomy of empathy and forgiveness. *NeuroReport* 12: 2433–2438.

Jeeves M (2004) Mind reading and soul searching in the twenty-first century. In. *What about the soul?*, ed. J. B. Green, 13–30. Nashville: Abingdon Press.

Joseph R (2001) The limbic system and the soul: Evolution and the neuroanatomy of religious experience. *Zygon* 36: 105–136.

Joseph R (2002) *NeuroTheology*. San Jose, CA: University Press.

Haidt J (2001) The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108: 814–834.

Hardy A (1966) *The divine flame*. Oxford, UK: The Religious Experience Research Unit.

Hay D (1998) *The spirit of the child*. London: HarperCollins.

Heidelberger M (2004) *Nature from within: Gustav Theodor Fechner and his psychophysical worldview*. Pittsburgh, PA: University of Pittsburgh Press.

LeDoux J (1996) *The emotional brain*. New York: Touchstone.

Lewis C S (1970) *God in the dock*. Grand Rapids, MI: Eerdmans.

McNamara P (2001) Religion and the frontal lobes. In *Religion in mind, ed.* J. Andresen, 237–256. New York: Cambridge University Press.

McNamara P (2002) The motivational origins of religious practices. *Zygon* 37: 143–160.

Moll J, de Oliveira-Souza R, Eslinger P J, Bramati E, Mourao-Miranda J, Andreiuolo P A, et al. (2002) The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *The Journal of Neuroscience* 22: 2730–2736.

Newberg A, d'Aquili E (2001) *Why God won't go away*. New York: Ballantine Books.

Newberg A B, d'Aquili E G, Rause V (2002) *Why God Won't Go Away: Brain Science and the Biology of Belief*. New York: Ballantine Books.

Persinger M A (1983) Religious and mystical experiences as artifacts of temporal lobe function: A general hypothesis. *Perceptual and Motor Skills* 57: 1255–1262.

Persinger M A (1987) *Neuropsychological bases of God beliefs*. New York: Praeger.

Ramachandran V S, Blakeslee S (1998). *Phantoms in the brain*. New York: Quill.

Ramirez-Johnson J, Fayard C, Garberoglio C, Ramirez C M J (2002) Is faith an emotion? Faith as a meaning-making affective process: An example from breast cancer patients. *American Behavioral Scientist* 45: 1839–1853.

Saver J L, Rabin J (1997) The neural substrates of religious experience. *Journal of Neuropsychiatry and Clinical Neurosciences* 9: 498–510.

Watts E N (1997) Psychological and religious perspectives on emotion. *Zygon* 32: 243–260.

Book reviews

THE BERLIN GROUP. ANTHOLOGY [Nikolay Milkov (ed.), Hamburg: Felix Meiner Verlag, 2015]

Constantin Yanakiev

Sofia University "St. Kl. Ohridski," Sofia, Bulgaria
e-mail: kjanakiev@phls.uni-sofia.bg

Die Berliner Gruppe. Texte zum logischen Empirismus, Nikolay Milkov (ed.), Hamburg: Felix Meiner Verlag, 2015 (lxi+490 p.). Introduction, notes, and index of names by Nikolay Milkov. The introduction also contains a list of selected works by members of the group (p. lv–lx). The pagination of the original editions is marked on the margins. Since the German text of two of K. Grelling's papers[1] published in English could not be retrieved, they are not translated back into German, but are printed in their language of first publication. Another paper by K. Hempel and P. Oppenheim[2] had originally been published in French, but in this case the editor opted to provide a German translation.

In the below review, I will refer to this anthology only by its page numbers and paper titles.

In this volume Nikolay Milkov has collected some important texts of the Berlin Group—

Walter Dubislav, Kurt Grelling, Carl G. Hempel, Alexander Herzberg, Kurt Lewin, Paul Oppenheim, and Hans Reichenbach—written between 1925 and 1937. The Berlin Group is commonly considered a minor associate of the Vienna Circle. This is what I thought of the Berlin Group before reading this book, which is able to change one's image of logical empiricism and inspire second thoughts on the idea of a philosophical school.

You would look in vain in this collection for the characteristic ideas of the Vienna Circle: the verification criterion of meaning, the condemnation of the metaphysics associated with it, the idea that the task of philosophy should be the analysis of scientific language, the dualism of tautologies and synthetic statements (*Realsätze*), the tautological character of mathematics, and unified science.

In fact, the lecture of Hans Reichenbach—the most prominent figure of the group—on *Metaphysics and Natural Science* (1925) states some metaphysical problems that he considers entirely meaningful: the reality of the external world, the freedom of will, and the peculiarity of life in the natural world (p. 330). In his lecture he undertakes to solve the first of them. And this he does by means of the principle of induction, which to his mind is subject neither to logical nor to empirical demonstration, but is nevertheless indispensable for natural science (p. 338). It is also a cornerstone of Reichenbach's chief epistemological efforts—probability (cf. his articles in part III "Probability and Induction").

Kurt Lewin in his *On the Idea and Task of Comparative Theory of Science* (1927) puts forward a "theory of science" (*Wissenschaftslehre*) that is rather inept from a neoposi-

[1] *Philosophy of the Exact Science: its Present Status in Germany*; *Realism and Logic*.
[2] *The Logical Significance of the Concept of Type*.

tivist point of view. Instead of reconstructing the logical syntax of scientific language, it seeks out resemblances between sciences one by one (*Wissenschaftsindividuen*), and then rises above the description of their "phenomenal properties" to conditional-genetic ones (p. 39). And he flatly denounces any preconceived unity of all sciences (p. 17).

When dealing with the subject matter of mathematics, Walter Dubislav doesn't say a word about a tautological character of logic and mathematics.[3] He argues for a formalist (in the sense of David Hilbert's formalism) view of mathematical theories, and this brings him close enough to the Vienna Circle, but does not rank him among the proponents of its doctrines.

The kinship between the logical empiricism of the Vienna Circle and that of the Berlin Group is obvious, but the latter is not an extension of the former. The outstanding features of the Vienna Circle are not represented in the Berlin Group. Moreover, Hans Reichenbach characterized the attitude of the Vienna Circle as doctrinaire radicalism.[4] According to N. Milkov, the perception of the Vienna Circle as the leading group arose only after the publication of its famous Manifesto in 1929. Until then Reichenbach was considered internationally to be the most promising German-speaking representative of empiricism (introduction, p. xxi). N. Milkov argues against the "Neurath-Haller thesis," according to which a German-speaking philosophical scene stamped by German idealism has been changed by the Vienna Circle, which imported the ideas and spirit of English empiricism. A German ground for this kind of philosophy as represented by Jakob Friedrich Fries existed even in the first half of the 19th century at the height of Hegelianism, and after Hegel's death a strong tradition of scientific philosophy developed. It was the source for the formation of both the Berlin Group's and the Vienna Circle's adherents (p. xxii–xxiii). The Berlin Group's member Kurt Grelling collaborated with Leonard Nelson for fifteen years, and it may be that he—under the influence of Fries's school—stimulated Reichenbach to deal with probability (p. xxxvi–xxxvii). Walter Dubislav maintained a relationship with Joseph Petzoldt until the latter's death in 1929 (p. xviii), and the formal organization, the core of which was the Berlin Group—the Society for Scientific Philosophy—was heir to Petzoldt's Society for Empirical Philosophy (p. ix-xi).

So are Schlick, Carnap, or Neurath greater philosophers than Reichenbach, Dubislav, and Hempel, or are they undeservedly praised for merits that are at least as much to be attributed to the Berlin Group? What is the measure of greatness for a philosopher? There are names which are remembered in philosophy, and it may be suggested that what made the Vienna Circle the more remarkable group is precisely its "doctrinaire radicalism." If one looks closely at a philosophical community, one may find several ideas and approaches, partly overlapping and partly at cross-purposes. But in order to conceive it as one whole, one needs to form a single image. Thus the Vienna Circle deserves its place in the history of philosophy because it provided the movement with physiognomy. However, the anthology of N. Milkov allows this closer look that reveals the variety behind the unifying scheme. This may be disconcerting. One may feel that the anthology lacks focus. Certainly a collection of papers from the Vienna Circle would be far easier to focus, and going beyond a convenient scheme is always uncomfortable.

But the abolishment of this scheme makes visible a more general background, upon which Karl Popper's philosophy of science has also fed. Although Popper is best known for falsifiability, I believe that the most fundamental layer of this philosophy is its conception of methodological rules, i.e. of the nature of scientific knowledge. According to Popper methodological rules are not statements of fact. They are decisions.[5] Decisions may be wise or not, fruit-

[3] In *On the Relation of Logic to Mathematics* (1925/26) (p. 101–120), *On the So Called Subject Matter of Mathematics* (1930) (p. 121–150).

[4] *Experience and Prediction*, The University of Chicago Press, Chicago & London, 1961 (originally published in 1938), p. 74. Quoted from the German translation in the introduction of N. Milkov, xiii.

[5] *The Logic of Scientific Discovery*, § 10.

ful or useless, but they are neither true nor false. The important thing for them is to be necessary. Thus a convention about basic statements is necessary, because otherwise "the soaring edifice of science would soon lie in ruins."[6] Clearly Popper's methodological decisions are a conventionalist version of Kant's *a priori* conditions of cognition. Also, J. Fries occupies an eminent place in his considerations.[7] But he may have drawn these ideas indirectly through Reichenbach, who is an important opponent in *The Logic of Scientific Discovery*. E.g. Reichebach attributes to the principle of causality the same status Popper ascribes to methodological decisions: it shares with empirical statements the feature that its opposite is conceivable, yet nevertheless is not an empirical statement. It must be valid, because otherwise knowledge would be impossible (*Philosophical Critique of the Probability Calculus*, 168–169). Popper makes the same point, but applies it to methodological rules in general.

[6] Ibidem, § 29.
[7] Ibidem, § 25.