Studia Semiotyczne—English Supplement, t. XXXI (2020), s. 91–126
ISSN 2451-2745
DOI: 10.26333/stsen.xxxi.04
Article

Maciej Tarnowski *

IS HAVING CONTRADICTORY BELIEFS POSSIBLE? DISCUSSION AND CRITIQUE OF ARGUMENTS FOR THE PSYCHOLOGICAL PRINCIPLE OF NON-CONTRADICTION[1]

Summary: The aim of this paper is to present and analyze arguments provided for the Psychological Principle of Non-Contradiction which states that one cannot have, or cannot be described as having, contradictory beliefs. By differentiating two possible interpretations of PNC, descriptive and normative, and examining arguments (ontological and methodological) provided for each of them separately I point out the flaws in reasoning in these arguments and difficulties with aligning PNC with the empirical data provided by research done in cognitive and clinical psychology. I claim that PNC cannot be derived from any metaphysical stance regarding the mental phenomena and that having contradictory beliefs should be regarded as possible. Furthermore, I argue that interpreting a subject as having contradictory beliefs, and therefore abandoning PNC, can be more effective in explaining the phenomena of contradictory beliefs and irrational behaviour than solutions consistent with the PNC.

* University of Warsaw, Faculty of Philosophy. E-mail: m.tarnowski3@student.uw.edu.pl. ORCID: 0000-0003-3824-4134.

[1] I would like to thank Joanna Komorowska-Mach and Tadeusz Ciecierski for the helpful and illuminating remarks made during the writing of this paper, and to the anonymous reviewers for pointing out some issues that were not included in the previous version of this paper.

KEYWORDS: principle of non-contradiction, rationality, principle of charity, intentional stance, belief ascription, delusion, folk psychology.

The logical principle of non-contradiction, codified in the classic propositional calculus as $\neg(p \wedge \neg p)$, is one of the most stable, basic and obviously true laws that apply in every deduction. By rejecting it, in the vast majority of logical systems we are obliged to follow the so-called principle of explosion (*ex contradictione quodlibet*), which states: from contradiction anything follows / can be proven. Then in our inference system it is possible to prove any claim, which makes this system useless from a practical point of view and makes it impossible to distinguish true and false statements. Both classic and many commonly used non-classical logic systems recognize the logical principle of non-contradiction.

But does the principle of non-contradiction also cover our beliefs? Many philosophers, despite its initial counter-intuitive nature, have given this question positive say. The purpose of this article is to thoroughly analyze the arguments supporting this thesis, here referred to as the Psychological Principle of Non-Contradiction, presenting an extensive critique of this view and argue, that its common acceptance is unjustified.

The first philosopher arguing for the Psychological Principle of Non-Contradiction was Aristotle. In Book IV of *Metaphysics* we find the following passage: "For it is impossible for any one to believe the same thing to be and not to be, as some think Heraclitus says. For what a man says, he does not necessarily believe" (Met., 1005b).[2] Łukasiewicz (1987) considered this formulation to be separate from the other formulations of the principle of non-contradiction, which consist of the Logical and Ontological Principle (in later studies also called a "metalogical" principle), calling it the Psychological Principle of Non-Contradiction. In his view, the Aristotelian thesis should be formulated as: "Two acts of believing which cor-

[2] Citations from Aristotle's works in this paper come from ("The Internet Classics Archive", n.d.). Translation of *Metaphysics* by W. D. Ross, *On Interpretation* by E.M. Edghill.

respond to two contradictory propositions cannot obtain in the same consciousness" (Łukasiewicz, 1987, p. 13).[3]

I reformulate this principle for the purposes of this article as follows:

(PPNC) It is impossible for one agent to hold two mutually contradictory beliefs at the same time.

This reformulation aims to eliminate ambiguities in Łukasiewicz's formulation. Łukasiewicz accepts that beliefs, being psychic entities, can't be contradictory—that's why he uses the notion of a proposition. Currently, however, talking about contradictory beliefs has entered everyday philosophical use, and simultaneously many problems regarding the ontological and logical status of propositions arose. This approach seems clearer to me. Also, because of the change in commonly used terminology, I exchange the concept of consciousness for a concept of an agent. The statement "it is impossible", used instead of the original "cannot", more clearly than in the original indicates the modal nature of this thesis.

There are, despite its counterintuitive nature, important and recurring arguments offered in support of PPNC in philosophical discussions. A special case of such a stance is the position of Classical Interpretationism (Quine, 1960; Davidson, 1974) and the theory derived from it, Daniel Dennett's Intentional Stance (1978; 1981a). What will become clear further in this analysis, is the method of justifying PPNC depends on the theoretical context in which we view the concept of "belief" present in the wording included here. In particular, it will be whether we will adopt realism or instrumentalism about beliefs.

In a realistic reading PPNC remains descriptive. It may be regarded as any other sentence about mutually exclusive phenomena: "it is impossible for it to rain and not to rain at the same time", "It is impossible that there will be night and day at the same time", etc. Adopting such a thesis supports a realistic approach to the set of beliefs as a "map by which we steer" (Ramsey, 1931); the model of the world from which we derive justification for our actions, by its very nature, cannot be contradictory according to this thesis. PPNC understood as a *descriptive* statement will be from now on marked as PPNC-D.

---

[3] Translation of citations from Łukasiewicz's work is based on the partial translation of Łukasiewicz's treatise by Vernon Wedin (Łukasiewicz, 1971).

One can also read PPNC (recognizing the instrumental interpretation of "belief') in a normative way, as a result of accepted methodology, not metaphysics. According to the instrumentalist interpretation, the condition for recognition of some agent as having beliefs is the usefulness of its description in such categories. Therefore, PPNC can be defended in an alternative way: two contradictory beliefs cannot exist at the same time in one mind, because such a description of it is forbidden by the theory within which the concept of belief is defined. Therefore, whether or not one can attribute contradictory beliefs to the subject depends not as much on facts about the mind (as realists would postulate) but on whether or not such an attribution shows its usefulness and whether it is allowed by folk psychology (or another true theory of belief ascription).

I refer to the interpretation described above as PPNC in its n o r m a t i v e reading. One can state it in a simple form like this:

> (PPNC-N) An agent cannot be ascribed two contradictory beliefs at the same time.

In this article, I want to reconstruct two argumentation strategies used in justifying PPNC. I define the first of them as ontological argumentation and the other as methodological—they serve respectively to justify the principle in its descriptive and normative formulation. I will briefly discuss their history and its versions put forward by different authors. Then I will criticize these arguments, pointing out the flaws in provided reasoning and citing conflicting evidence from the fields of cognitive and clinical psychology.

Before going on to analyze these arguments, we must clearly note that the two interpretations of PPNC are not mutually exclusive. A realist who believes that folk psychology apparatus is accurate can accept them both: argue that having contradictory beliefs is both metaphysically impossible and impossible to assign to a subject from the point of view of folk psychology. Views on the "methodology" of folk psychology (whether it can be treated akin to a scientific theory, and whether it forbids assigning contradictory beliefs) are only conventionally, not logically, related to one's stance in the metaphysics of mind or philosophy of science. Hence it seems that sometimes the statements supporting both interpretations have the status of a silent premise. To distinguish between these two strategies and ways understanding PPNC while analyzing the correctness of this principle is therefore essential for presenting the debate clearly.

1. The Ontological Argument From the Mutual Exclusion of Properties

The first author arguing for PPNC is its creator, Aristotle. In Book IV of Metaphysics, he devotes a lot of space to it, arguing for adopting the principle of non-contradiction by at least few separate arguments—proving that the principle of non-contradiction should be treated as one of the basic laws of thought (Łukasiewicz, 1987). Łukasiewicz and Gottlieb distinguish chapters 4, 5 and 6 of Met. IV as containing the argument for adopting the principle of non-contradiction in its psychological version (in the wording provided in the previous section). Fragments associated with this line of argument also appear in *On Interpretation* and *Posterior Analytics* (Łukasiewicz, 1987; Gottlieb, 2007). Aristotle tries to show that PPNC may be proven on the basis of ontological and logical formulations (Łukasiewicz, 1987). He writes:

> [I]f it is impossible that contrary attributes should belong at the same time to the same subject (the usual qualifications must be presupposed in this premiss too), and if an opinion which contradicts another is contrary to it, obviously it is impossible for the same man at the same time to believe the same thing to be and not to be; for if a man were mistaken on this point he would have contrary opinions at the same time. (Met. IV, 1005b)

As Łukasiewicz notes, it is problematic in this fragment that Aristotle equates the notions of being contrary and contradictory: the former belongs to pairs of properties (attributes), the latter—to propositions and beliefs. Referring to the fragments of *On Interpretation* (*On Interpretation*, 14, 23, 27–39), he indicates that Aristotle, in order to avoid this difficulty, treats beliefs as properties of the mind—then mutually contradictory beliefs correspond to contrary properties. Thus, if one man believed the same thing to be and not to be—he would hold two contradictory propositions—he would have contradictory beliefs, and therefore could be attributed contrary properties, which would contradict the ontological formulation of the principle of non-contradiction:[4]

[4] Łukasiewicz in his reconstruction of Artistotle's argument holds that Stagirite bases his proof of PPNC on the Logical Principle, which concerns propositions. He writes: "on the basis of the logical principle of contradiction, it is impossible that incompatible characteristics hold of the same object at the same time"

(OPNC) "To no object can the same characteristic belong and not belong at the same time". (Łukasiewicz, 1987, p. 10)

We can try to reconstruct Aristotle's argument as follows:

(1) (OPNC) It is impossible for one object to both possess and do not possess some property.
(2) Beliefs expressing contradictory propositions are contrary properties.
(3) Contrary properties are mutually exclusive.
(4) Beliefs expressing contradictory propositions are mutually exclusive (from [2]—[3]).
(5) It is not possible for a single entity to have mutually exclusive properties (from [1]).
(6) (PPNC) It is impossible for one agent to hold two beliefs expressing contradictory propositions at the same time (from [4]—[5]).

Significantly, from the point of view of the analysis of Aristotle's ontological argument, will be his acceptance of premises (2) and (3). The author clearly emphasizes that this argument depends on the accepting the truth of OPNC, which he considers to be the basic ontological principle: thus questioning this premise does not overtly attack the soundness of Stagirite's reasoning.

Premise (2) is justified, according to Łukasiewicz, as follows: since Aristotle interprets beliefs as properties of mind, it is necessary to equate two concepts, contradiction (with regard to propositions and beliefs expressing them) and contrariness of properties, to justify adopting this premise. Aristotle writes in *On Interpretation*:

> But if, in thought, it is not the judgement which pronounces a contrary fact that is the contrary of another, then one affirmation will not find its contrary in another, but rather in the corresponding denial. (*On Interpretation,* 14, 24b)

(Łukasiewicz, 1987, p. 24). Since the Logical Principle of Non-Contradiction in Łukasiewicz's reconstruction justifies such thesis, and therefore the equivalence of the Logical and Ontological Principle, I assume that a reconstruction provided here, based on OPNC is also valid.

The contradiction that appears between the propositions expressed in beliefs will, in Aristotle's terms, correspond to the contrariness of properties. It is difficult to consider what the contrariness is actually in this argument: Łukasiewicz writes that "contrary beliefs are those that are answered by an affirmative and negative judgment about the same subject, e.g. 'Callias is just'—'Callias is not just'" (Łukasiewicz, 1987, p. 21).

Łukasiewicz subjects the above reasoning in a similar reconstruction to thorough criticism. His attention is focused on two issues: equating the concepts of contradiction and the contrariness of properties in relation to beliefs, and the unjustified mixing of logical and psychological concepts in premise (2). In this fragment, I will reconstruct Łukasiewicz's criticism, analyze it and draw conclusions regarding the status of Aristotle's argument for PPNC.

In analyzing the justification of premise (2), he assumes both the assumption that beliefs can be treated as properties and that properties can be contrary to each other. But for what is Aristotle's contrariness of properties or characteristics—and consequently the contrariness of belief? Łukasiewicz finds a partial answer to this question in fragments of *On Interpretation*:

> We must therefore consider which true judgement is the contrary of the false, that which forms the denial of the false judgement or that which affirms the contrary fact. [...] Now that which is good is both good and not bad. The first quality is part of its essence, the second accidental; for it is by accident that it is not bad. But if that true judgement is most really true, which concerns the subject's intrinsic nature, then that false judgement likewise is most really false, which concerns its intrinsic nature. [...] Thus the judgement which denies the true judgement is more really false than that which positively asserts the presence of the contrary quality. (*On Interpretation,* 14, 24b)

According to Łukasiewicz, there is an unsound transition from the domain of logic to the domain of psychology, especially visible in another fragment from *On Interpretation*, in which Aristotle states that "the judgement that that which is good is bad is composite. For presumably the man who forms that judgement must at the same time understand that that which is good is not good" (*On Interpretation*, 14, 23b). Łukasiewicz points out that a similar relationship (of logical consequence) occurs between propositions, but not necessarily between beliefs. For if we regard beliefs as properties, we cannot attribute either truth or falsehood

to them—those attributes belong then only to propositions or sentences. Talking about the truth or falsehood of beliefs makes sense as long as it refers to their linguistic representation or the proposition they express. This is a problem for Aristotle's reasoning: it is impossible to simply translate the truth or falsehood of the belief that p in the above sense into any characteristic of the property of mind, such as believing that p. Since properties are neither true nor false, they cannot be contradictory either.

Thus, even if we treat beliefs as properties of the mind, we cannot show that the contradiction of beliefs treated as propositional attitudes entails the contrariness of beliefs interpreted as properties, so PPNC cannot be deduced from OPNC. Therefore, Aristotle's argument in favor of PPNC should be rejected.

Another view that derives PPNC from the characteristics of beliefs as properties is dispositionalism, which considers beliefs to be dispositions to display certain behaviors. A representative analysis for this trend is that carried out by Ruth Barcan Marcus. In her analysis of the concept of belief, she reduces it to the disposition "to act as if S, the actual or non-actual state of affairs, obtains" (Barcan Marcus, 1990, p. 241). What would it mean to act as if the law of non-contradiction would not apply? Barcan Marcus seems to follow the views of Wittgenstein from the *Tractatus*[5] regarding the cognitive status of tautology and contradiction. You cannot act, for example, as if it was raining and not raining at the same time, because there are no conditions (a possible world) in which a similar (impossible) state of affairs could be the cause of your behavior. Since we cannot characterize the disposition to act as if p and not-p was true, it is impossible to have two contradictory beliefs.

Without entering the ontological discussion of the status of beliefs, I believe that no form of dispositionalism logically entails PPNC. For the contradiction of beliefs does not translate (for the same reasons as in Aristotle's case) directly into the mutual exclusion of two dispositions. Any behavior that is the basis for ascription of a belief must be behavior that positively indicates a possession of such belief. Believing that not-p can-

[5] "4.461 Propositions show what they say: tautologies and contradictions show that they say nothing. A tautology has no truth-conditions since it is unconditionally true: and a contradiction is true on no condition. [...] (For example, I know nothing about the weather when I know that it is either raining or not raining)" (Wittgenstein, 2020).

not be just simply n o t behaving as if p was the case—then not believing that p would be tantamount to believing that not-p. If, on the other hand, there are patterns of behavior suitable for believing that p and believing that not-p, the consequence that forbids us to ascribe the belief that p and not-p, in the absence of evidence of the agent's rejection of any of the beliefs, is dogmatic. At least intuitively, there are also ways in which a pattern of behavior can be explained by being convinced of some impossible state of affairs, as Wittgenstein notes in *Remarks on the Foundations of Mathematics*:[6]

> I feel a temptation to say: one can't b e l i e v e that 13 × 13 = 196, one can only a c c e p t this number mechanically from somebody else. But why should I not say I believe it? For is believing it a mysterious act with as it were an underground connexion with the correct calculation? At any rate I can say: "I believe it", and act accordingly. (Wittgenstein, 1998, I–106)

1.1 The Ontological Argument From the Function of Mind

However, there is also a version of the ontological argument which, although rarely stated explicitly, seems to have been silently adopted by many contemporary philosophers arguing for PPNC. I will try to refer to it in the hope that it will clear the methodological points brought up further in the paper—even if the argument in the following version is not adopted as such by any philosopher.

This argument, although significantly different from the one described earlier, belongs to the ontological argumentation in the distinction used here, because it tries to derive PPNC in its descriptive version: that it is impossible for an agent to simultaneously have contradictory beliefs. However, it concerns a much narrower group of cases. According to this line of argument, an agent cannot hold two beliefs of which he knows (or thinks) to be contradictory.

---

[6] Barcan Marcus interprets this passage as an introduction of a distinction between "believing" and "claiming to believe" impossibilities (Barcan Marcus, 1990, p. 253). However, accepting a dispositionalist account of belief, this distinction is pretty dogmatic (if we, as Barcan does, assume "claiming to believe" as a form of behavior positively indicating possession of belief). Also, even if Wittgenstein ever maintained such a distinction, he clearly abandons it further in the text (see the remarks I–106 to I–119).

In Wilfrid Hodges's *Logic*, we can find the following formulation, probably the closest to the thesis discussed here:

> It is simply impossible to believe, fully and without reservation, two things which you know are inconsistent with each other. It seems we are obliged to believe only what we think is consistent without having any real choice in this matter. (Hodges, 1977, p. 15)

A similar passage may be found in Quine's and Ullian's *The Web of Belief*:

> [O]ne can't believe a thing if one sees that it is impossible. [...] We saw it as the very reason for taking thought, for sifting evidence and revising one's system of beliefs. When conflicts arise, creating impossible combinations, we cannot rest with them; we have to resolve them. (Quine, Ullian, 1978, p. 37)

Such claims require a certain assumption about the purpose of the system of belief formation—namely, that forming true beliefs is its proper function. As Ruth Barcan Marcus rightly points out, commenting on Hodges' remarks: "Why focus on contradiction? Is it possible to believe that p when you know that p is false?" (Barcan Marcus, 1990, p. 145). One can therefore interpret Hodges' and Quine's thesis that we are "obliged to believe only what we think is consistent" as follows. Assume that a natural inclination and the purpose of the human cognitive system is to have true beliefs: if you are given information that counters certain belief or directly contradicts it—be it empirical evidence or the result of deductive inference—you are forced to reject a belief that turns out to be false. Only if we will assume that the purpose of our cognitive system is to maintain true beliefs, we can consider that it would have some kind of incentive to get rid of those false beliefs. Because the contradiction is an obvious sign of falsehood, one cannot hold contradictory beliefs while being aware that they are such.

1.2 Criticism of the Argument From Function of Mind

The claim that maximizing the amount of true beliefs is a systemic function of our mind appears to be intuitive, however it has rarely been directly defended. The only significant attempt to do that is an argument referring to the principles of natural selection (given by e.g. Fodor and

Dennett)—however, it seems that it results from the insufficient consideration of competing evolutionary strategies.[7] But if we were to accept it, even without justification, it is still possible to question whether this claim can play any role in justifying PPNC.

First, it seems that by arguing for PPNC this way we fall into a vicious circle. The principle of eliminating false beliefs, which we take as a premise in our reasoning may be expressed like this:

(PE) If an agent A has a belief that p, and learns that p is false, A gets rid of the belief that p.

PE therefore means that with acquiring the justified, true belief that p is false an agent has to get rid of the previously held belief that p. From that it immediately follows that an agent cannot simultaneously hold a belief that p and that p is false. The latter belief does not seem to differ significantly from the belief that not-p—if so, then we have already established the PPNC among the premises.

Another noteworthy assumption in this argument is that the recognition of self-contradiction in some belief (of the form p and not-p) is an obvious evidence of its falsehood. One should wonder what exactly counts as an obvious falsehood of contradictory sentences and beliefs. Not every self-contradictory sentence is obviously contradictory: it can be proved by, for example the history of mathematics where this happened more than once in the mathematician community to accept fake proof of the claims that later turned out false. Most of us would probably consider as self-contradictory (and therefore false) certain counter tautologies of propositional calculus: propositions of the form p and not-p or not-not-p and not-p, however, at first glance, we won't say so about the proof of the theorem of algebraic geometry "proven" by Francesco Severi in 1934 (and its falsehood which was proven 34 years later), not to mention troubling inconsistencies in the naive set theory. So where does an obvious self-contradiction of a sentence begin? Doesn't recognizing a sentence as self-contradictory therefore not the same to our understanding as recognizing it is false (which, again, simply assumes PPNC)? Are dialetheists, such as

[7] One may find a complex critique of such an argument (advanced e.g. by Daniel Dennett in his [1978]) in Stich's (1985) and Lewis & Cooper's (1979).

Graham Priest (who believes in true contradictions), wrong in asserting certain sentences or misinterpreting our concept of contradiction?

I am not going to answer these questions here, but rather point out that accepting this seemingly innocent argument requires a precise (and highly debatable) answer to each of them. Rather, it seems that considering self-contradiction as obvious evidence of falsehood, one must act and infer in accordance with the PE, and therefore this premise also cannot rightly justify PPNC. So it is possible that the Hodges-Quine condition concerning awareness of the contradiction (as we seem to understand it) that an agent must possess, is already assumed to be acting in accordance with PPNC and therefore cannot help us in justifying it.[8]

1.3 The Aposteriority Problem

As has been shown above, the indicated attempts to prove PPNC in its descriptive reading fail due to an unjustified mixing of logical and psychological concepts or the tacit adoption of conclusions along the premises. I would also like to draw attention to a more general argument, which in a similar form was directed by Łukasiewicz against the PPNC itself in its Aristotelian formulation (Łukasiewicz, 1987, pp. 30–34).

---

[8] Another, although similarly interesting in this context group of cases are the cases of contradictory beliefs which mutual inconsistency cannot be recognized by the agent even if we assume (s)he is ideally rational. Those might be e.g. Kripke's Puzzle (Kripke, 1979) or Richard's Problem (Richard, 1983): in both cases, generally speaking, we have to do with beliefs acquired in isolated epistemic or linguistic contexts, which are about the same object, given in those two contexts in different ways. Then, as Kripke points out, "no amount of pure logic or semantic introspection suffices for [an agent] to discover his error" (Kripke, 1979, p. 451): figuring out the internal contradiction by the agent may be done only by acquiring new belief on the basis of empirical evidence (e.g. that "London = Londres" in Kripke's case or that "you (the person I'm talking to by the phone] = he [the person I see on the street]" in Richard's case). If an argument which uses those cases as evidence that one may possess contradictory beliefs is sound (which I do not want to get on in this paper), then not only (as I have tried to show above) it is possible, that some agents internal inconsistency may not be an obvious evidence of the falsity of their beliefs, but also that there are some pairs of beliefs the inconsistency of which cannot be recognized through logical analysis—then even the assumption of strong procedural rationality of an agent does not suffice to prove the weakened version of PPNC-D.

As Lewis Carroll (1995) famously noted, an attempt to justify inference by referring to the axioms themselves (that is, justifying logical inference using purely logical tools) leads to an infinite regress. In addition to the axioms—recognizing certain sentences or formulas as true—we also need to adopt a rule of inference. As Penelope Maddy puts it:

> There [...] would be [no such problem] if I stipulated the truth of all the axioms of ZFC, but when we try to stipulate the truth of logic itself, we find our explicit conventions must be general, and then that these general conventions are without their intended force unless logic is already available to oversee the derivation of particular logical truths from the generalities. (Maddy, 2012, p. 496)

Whether or not a subject reasons in accordance with the principles of logic cannot therefore be determined by reference to any general laws of belief formation. Even if we (hypothetically) discovered in the human mind a representation of the law of non-contradiction, in order to justify PPNC in this way, it must be assumed that the agent thinks logically, applying the general law to its individual cases. We would have to do likewise with PE or other psychological belief formation laws. Thus, no general psychological principle has sufficient strength to prove PPNC-D either.

Therefore, if it is impossible to derive PPNC-D from the general laws of belief formation, the only possible form of justifying this principle is to interpret it as a well-proven empirical hypothesis. This can be viewed as the aposteriority problem: PPNC in its descriptive version cannot be justified *a priori*, but only as a result of empirical research.

Such an approach to the matter seems quite problematic in a philosophical discussion—even if PPNC-D was a well-confirmed hypothesis, many philosophers would find it undesirable to grant an empirical status to a principle that was initially described as one of the basic principles governing human thought. The assumed 100% compliance of the studied cases with PPNC would not prove that it is (in principle) impossible to have contradictory beliefs.

However, if this were the current state of psychology research, it could convincingly justify PPNC-D or at least make it sufficiently plausible. I would like to conclude my deliberations on the descriptive reading of PPNC by challenging its interpretation as a positively verified empirical hypothesis.

As long as we do not question the conclusiveness of the results of these studies, as philosophers who interpret PPNC in a normative way will try, we will have no reason to consider PPNC-D as a well-confirmed claim of psychology. I will provide two groups of examples: well-known research on cognitive heuristics, which shows how often agents unknowingly adopt contradictory beliefs, and clinical cases, which are radical examples of irrational and self-contradictory beliefs.

The first set of examples of interest to us that may undermine the truth of PPNC as an empirical hypothesis are studies from the so-called "heuristics and biases" research programme, which have been conducted by cognitive psychologists and cognitive scientists since the 1960s. These studies try to show that the majority of people (even up to 87% of respondents [Tversky, Kahneman, 1983]) use simple heuristics rather than rules of logical and probabilistic inference in their everyday thinking[9]—and that the two strategies often come into conflict with each other. I will briefly present two studies showing two popular inference fallacies: conjunction and disjunction fallacies, which seem to be the closest related to logical inference and as such may prove that the agent holds obviously contradictory beliefs.

The conjunction fallacy may be illustrated by the classical study of Tversky and Kahneman (1983).[10] A group of 93 respondents was given the following task:

> Suppose Bjorn Borg [a famous Swedish tennis player] reaches the Wimbledon finals in 1981. Please rank in order the following outcomes from most to least likely.
>
> A. Borg will win the match (1.7)
>
> B. Borg will lose the first set (2.7)
>
> C. Borg will lose the first set but win the match (2.2)

[9] The beliefs about the probability of some events happening are mostly dismissed as atypical or unimportant class of beliefs; one may although notice it's commonness in the everyday use of such phrases as: "Under condition, that...", "We need to be prepared for...", "It's likely to happen that...", "It's very unlikely that...", etc.

[10] A classic example from this study is, of course, "Linda, the feminist bank teller"; however, due to its wide coverage in the literature and the methodological concerns it has raised, I decided to use a different experiment illustrating the same effect.

> D. Borg will win the first set but lose the match (3.5). (Tversky, Kahneman, 1983, p. 302)

The numbers in parentheses represent the average rank given to this opportunity among the other four. A large part of the respondents considered that C is more probable than B, which is impossible from the point of view of probability calculus (C is the intersection of two events, one of which is represented in B—it cannot therefore be more probable). The researchers explained this phenomenon by the existence of so-called representativeness heuristic—the subjects, knowing Borg's reputation as a great tennis player, immediately considered any sentence that predicted his victory to be highly probable. Similar studies (Bar-Hillel, Neter, 1993) also concerned reasoning in a situation when we are dealing with the union of two events, expressed as a disjunction—due to the use of the representativeness heuristic, the respondents often also considered that one of the events is more likely than its union with another event.

It is worth emphasizing that the vast majority of respondents in both studies cannot be described as unaware of the basic laws of probability. In the first of these studies, it was even checked in a separate study whether the fallacy was not caused by the common interpretation of the conjunction as an implication (Tversky, Kahneman, 1983, p. 302); it seems that most of the people who made a mistake in one or the other study are also familiar with the rules of probability calculus and can apply them in some cases (this is confirmed, inter alia, by studies using a different, statistical approach to the "Borg problem" [Fiedler, 1988]).

Another, much more direct example of agents having conflicting beliefs are cases of patients with clinical delusions.[11] The presence of similar disorders—resistant to counterexamples, not following the norms of rational inference of beliefs—may indicate that the empirical hypothesis of human rationality may be thoroughly false: those are not cases of minor or explainable deviations from rationality as in research on heuristics, but very serious impairments of the ability to think logically and evaluate given evidence. However, does it also contain direct cases of contradiction?

[11] A similar interpretation is presented in the paper by Tadeusz Ciecierski (2017) whom I thank for bringing my attention to this group of cases.

The most direct example would be certain cases of Cotard's delusion,[12] consisting in the patient having the belief that (s)he has no internal organs, is dead, immortal, currently in hell or does not exist. A patient examined by Ryan McKay and Lisa Cipolotti, LU (2007, p. 353), when asked how she knew she was dead and whether she had ever seen a dead person, replied that she had seen her grandmother's body after her death and knew she was dead because her eyes were closed and she was not moving. An example of a similar contradiction can be seen in the case of the patient described in the study by Nishio and Mori:

> He said to his doctor (Y. N.), "I guess I am dead. I'd like to ask for your opinion". Later, his conviction about death became firmer. He said, "My death certificate has been registered. You are walking with a dead man", and "I am dead. I will receive a death certificate for me from my doctor and have to bring it to the city office early next week".
>
> His discussion of his demise was not associated with a depressed mood or feelings of fear. When his doctor asked him whether a dead man could speak, he understood that his words defied logic, but he could not change his thinking. (Nishio, Mori, 2012, pp. 217–218)

There is little doubt that we can attribute contradictory beliefs to these two patients: for example (1) that they speak, (2) that they are dead, and (3) that the dead cannot speak. In the second case, we even seem to deal with the recognition of this contradiction by the patient himself, combined with the inability to renounce clearly incompatible beliefs.

I am not saying that the cases mentioned above cannot be disputed. One may try to argue that people with Cotard's syndrome really mean something else by "death", and the cases of incorrect inference in the research of Tversky and Kahneman are not examples of contradictory beliefs. These statements, however, belong to the argument for a normative reading of PPNC, as they provide clear indications on how to interpret the given results. Moreover, it will be the obligation of every philosopher supporting PPNC-N to show that the interpretation of the research results given here is wrong—later in this paper I will explain why such stances do not meet this requirement. However, if we interpret PPNC

[12] A good overview of other delusionary cases is presented in (Breen et al., 2000), and their philosophical implications are robustly discussed in (Bortolotti, 2010).

only as an empirical hypothesis, we are forced to consider it as at least dubious in the light of the examples given above.

2. Methodological Arguments for PPNC

In this section I will focus on the methodological arguments in favor of adopting PPNC. According to the distinction introduced above, it will be an argument supporting the principle in its normative reading:

> (PPNC-N) An agent cannot be ascribed two contradictory beliefs at the same time.

Adoption of such a thesis, as outlined above, is set in a different philosophical context than the adoption of PPNC-D. First of all, it is most often associated with instrumentalism or at least agnosticism with respect to the ontological status of beliefs. The essence of the arguments presented below is to recognize "belief" as a theoretical concept of folk or scientific psychology and to show that the adoption of PPNC-N is necessary precisely from the point of view of theories allowing the possibility of belief ascription. Such a position seems to more or less presuppose interpretationism with regard to beliefs—and I am not going to question that assumption in here.

A useful distinction—before going on to discussing the arguments properly—is of that between individual and scientific belief ascriptions. I borrow the terms from Richard Dub (2015) who uses them to distinguish two levels of argumentation for the rationality assumption put forward by Daniel Dennett. However, I consider this distinction to be much more basic and crucial for the disclosure of specific assumptions and goals that individual theories and arguments set for themselves.

Individual ascription is, in short, a belief ascription that every competent language user familiar with the notion of "belief" makes in everyday situations. They accompany the most common uses of phrases such as: "He / she believes that...", "I think he thinks...", "He thinks that...". These ascriptions are quite frugal in terms of the data used: we make them under time pressure, often without having much knowledge about someone else's life, behavior, habits, etc. They are formulated somewhat automati-

cally and are not subject to advanced process of reflection.[13] Maintaining the PPNC-N with regard to individual ascriptions would mean that our everyday use of folk psychology requires assuming the consistency of the beliefs of the agent to which we ascribe beliefs. Traditionally, the justification for such a claim will be based on describing the ordinary use of language and showing that it results from the "method" and nature of the notions of folk psychology.

A completely different context for using the above-mentioned linguistic constructs is to make a scientific ascription. "[Individual and scientific ascriptions] are distinguished by who it is that does the ascription: the first is employed by individuals in real-world situations, and the second is employed by scientists and philosophers in the development of theories" (Dub, 2015, p. 98). The arguments for PPNC-N with regard to scientific ascriptions will focus not on how the concept of belief is, but how it should be used when terms derived from folk psychology are adopted in scientific psychology. Even if the everyday use allows us to break PPNC-N, it cannot be allowed to do so when it comes to "adult" belief theory—some theorists seem to say. Arguments following a similar line will try to prove PPNC-N by referring to the methodological foundations of psychology, anthropology or linguistics. I will try to show that even with the assumptions made by the authors, the adoption of PPNC-N in terms of both individual and scientific assignments is at least problematic.

2.1 The Argument From Daniel Dennett's Intentional Stance

The main supporter of PPNC-N with respect to individual ascriptions is Daniel Dennett from the period of developing the Intentional Stance theory and in later works.[14] Dennett assumes that the concept of belief is part of a broader strategy of predicting and describing the behavior of other cognitive systems, which he calls "Intentional Stance". It consists in assuming that the described subject is rational and ascribing him/her the

[13] Dennett himself describes them in his response to Dub as "the time-pressured quick-and-dirty attributions of folk psychology" (Dennett, 2015, p. 206).

[14] "In *Content and Consciousness*, Dennett is clear that his concern is mental ascription of the second [scientific] type. [...] The ground shifted somewhat when Dennett developed the Intentional Stance. The Intentional Stance became a piece of individual ascription: interpretation was now spoken as something that we all naturally do" (Dub, 2015, p. 98).

beliefs, desires and intentions explaining his/her action and their consequences in accordance with the accepted canon of rationality ("assign those beliefs that an agent should have") and further predicting his/her actions as consistent with assigned beliefs.[15] According to Dennett, intentional stance is part of our daily practice: something we do when we use the term "belief' in everyday language to interpret the behavior of others. Hence, the main emphasis in Dennett's argument is on individual ascriptions (his argument for the adoption of PPNC-N in scientific psychology is in line with the remarks made by Quine and Davidson, whose views I will discuss later).

It is difficult to say at first whether Dennett considers PPNC inviolable. Certainly, there is an important fragment in which he considers the inclusion of cases of the interpretation of irrational behavior into the principles of intentional stance:

> What rationale could we have, however, for fixing some set between the extremes and calling it the set for belief (for S, for earthlings, or for ten-year-old girls)? This is another way of asking whether we could replace Hintikka's normative theory of belief with an empirical theory of belief, and, if so, what evidence we would use. "Actually", one is tempted to say, "people do believe contradictions on occasion, as their utterances demonstrate; so any adequate logic of belief or analysis of the concept of belief must accommodate this fact". But any attempt to legitimize human fallibility in a theory of belief by fixing a permissible level of error would be like adding one more rule to chess: an Official Tolerance Rule to the effect that any game of chess containing no more than k moves that are illegal relative to the other rules of the game is a legal game of chess. (Dennett, 1978, p. 21)

However, do we seek an explanation following the ideal of rationality—or do we refrain from judgment—in cases such as delusions or mental

[15] Dennett differentiates the intentional stance from other strategies of description: "design stance" (which refers to the function an object is designed to perform) and a "physical stance" (which refers to physical properties and the laws of physics. Those stances are differentiated by the complexity and accuracy of its predictions: an operation of an alarm clock may be described and predicted by a physical model, by referring to its designed function (e.g. ringing at the exact time it was set) or by ascribing it a set of beliefs and desires (e.g. a desire to wake us up at specific time and a belief concerning times of day; see Dennett, 1981a).

illness, or in everyday cases of actions suggesting a deviation from the normative pattern? It seems that although we can begin our process of interpretation by referring to the model of a fully rational agent, with time we abandon this assumption, adapting our model to accommodate new evidence. In a discussion with Stephen Stich (1981), who made similar allegations, Dennett (1981) argued that an explanation for such cases is only available through a lower-level stance.

In Dennett's interpretation, the description of irrationality in the language of intentional stance is impossible: expressions such as "it slipped my mind", "I made a mistake", etc. are made from a stance explaining my behavior through the malfunction of one of the functions (memory, vision, etc.) of the subject. Examples of irrationality, therefore, can only be explained as performance errors, not competence errors.[16, 17] So it is not that in the above-mentioned situations I am obliged to explain it (using intentional stance) by referring to the extensive rationalization of my behavior or the rejection of the possibility of interpretation. I am simply referring to the error at a lower level of explanation, as when the alarm clock failure (i.e. acting against the predictions of the "design stance") is blamed on a wiped gear or battery discharge (i.e. events described by a "physical stance").

Providing a wide range of counterexamples is therefore not sufficient to challenge Dennett's argument. There are two key theses in it. First, that the pattern on which we construct a particular concept of an intentional system is an ideal agent that rationally formulates beliefs (avoiding contradictions, among other things) and acts in accordance with them. Second, when a given intentional system works against expectations, we are obliged to explain its behavior by referring to the error at the functional ("design"), not the intentional level. People who otherwise make a mistake in using the concepts of folk psychology. These theses are both descriptive (this is how we use the concept of belief) and normative (in the case of the inconsistency of predictions with effects, we should prefer

[16] This distinction is, of course, borrowed from Noam Chomsky—it is used in this context e.g. by Stich (1985).

[17] A similar line of argument—treating all the irrationality cases as "performance errors"—is supported by authors seeking justification of validity of logic in psychology (Cohen, 1981; Sober, 1978). A critique of their stances is included in (Stich, 1985; Thagard & Nisbett; 1983).

an explanation at the functional level). Below I will try to elaborate on the critique of this line of argument offered by Stich (1981; 1985), first referring to the descriptive side of Dennett's premises, and then to the normative side.

Stich (1981), arguing with Dennett, makes the accusation that his concept, regardless of the adopted interpretation, does not allow to explain the simplest cases of deviations from rationality. He divides Dennett's analyzes and suggestions into "hard" and "soft" lines, accusing him of inconsistency in his arguments. According to Stich, the "hard line" includes the assumption of rationality of an agent, the consequence of which is the adoption of the PPNC-N. The "soft line", on the other hand, consists in viewing this assumption and approach to PPNC-N (which can be deduced from some fragments of Dennett's writings) only as a necessary condition for starting the interpretation process; these conditions do not apply to us later, after acquiring more knowledge of the agent's behavior. Stich finds a hard line impossible to defend. It is not just the intuitive absurdity of the idea that anyone who knows the basics of classical propositional calculus also believes the infinite number of tautologies. The hard line strategy fails to describe the most common cases of irrationality:

> When a neighborhood boy gives me the wrong change from my purchase at his lemonade stand, I do not assume that he believes quarters are only worth 23 cents, nor that he wants to cheat me out of the 2 cents I am due. My first assumption is that he is not yet very good at doing sums in his head. (Stich, 1981, p. 50)

On the other hand, the "soft line" suffers, in Stich's view, from another drawback—if we accept it, it is difficult to understand why the image of "ideal rationality" would be the starting point for our theory and why we do not use the modified concept of rationality based on how usually our inference process is carried out, for example based on research of cognitive heuristics. The "soft line" thus leads to the recognition that the "ideal" we consider to be the first model in the process of interpretation differs from Dennett's understanding of rationality.

As I outlined above, Dennett's theory, however, escapes the simple division into "soft" and "hard line". The author himself writes: "These distinct lines are Stich's inventions, born of his frustrations in the attempt to make sense of my expression of my view which is both hard and soft—that is to say, flexible" (Dennett, 1981b, p. 73). Dennett sees cases similar to the one cited above as only possible through the lens of the "design

stance". After all, the explanation that the boy is "not yet very good at doing sums in his head" seems to come from this level of description. With such an interpretation, Dennett is able to retain the full power of the "hard line" while explaining its hypothetical ineffectiveness.

However, it is difficult not to notice some problems with this formulation of Dennett's position. Do we always prefer an explanation in terms of "design stance"? And do these explanations really result from our aversion to breaking PPNC? In my opinion, the answer to both of these questions is no.

First of all, the design stance is not always available to us—our common intuitions about it often seem ambiguous. Interesting material for consideration is the research conducted by Wason (1969), in which the impact of explanation and the pointing to contradictions on the improvement in solving the Wason selection task was examined (Wason, 1968). The subjects who failed the task of selecting cards, not following the rules of elementary logic (in this case *modus tollens*),[18] the researcher tried to present the subject as having made a mistake in their reasoning so as to convince them to change the previous answer. First, it was made sure that the subject understood the question well and knew that the given rule, which (s)he was asked to check, could also be false. The experimenter began by asking, "If there were a [stimulus mentioned first by the subject] on the other side, could you say anything about the truth or falsity of the sentence?" (Wason, 1969, p. 474). And when increasingly persuasive, but still hypothetical considerations failed, in which the respondents remained self-contradictory when declaring answers (they maintained that the conjunction of the premise and the negation of the conclusion did not falsify the implication), the researchers asked the respondents to reveal the cards. If the subject was still unable to choose the appropriate card, the experimenter would directly inform him/her that (s)he was wrong and asked him/her to think about his/her answer. The

[18] Original research (Wason, 1968) consisted in showing the subject four two-sided cards with two letters and two numbers (e.g. "D", "3", "B", "7"), where on one side of the card was the letter and on the other—a number, with a task of selecting the cards which should be turned over to find out whether a certain implication is true (e.g. "If there is a D on one side of the card, then on the other there is 3"). Only a relatively small group of subjects was able to select the cards appropriately (select the cards "D" and "7"). In the further research the content of cards and a formulation of implication has varied.

study showed that 12% of the respondents were unable to change their minds at any stage of the considerations.

At the moment when a given subject does not want to admit that (s)he made a mistake despite the best efforts of the experimenter, are we still able to recognize it as a "performance error" and use the "design stance"? It seems that it is not—the illogicality here is not only a matter of a temporary disturbance of inference competence, because despite long attempts this mistake cannot be corrected. Although it should not be ruled out that often these errors can (and are) corrected, and our language allows us to "rationalize" them in the manner given by Dennett, however,

> What is really remarkable about these and other experiments in which everything was done to encourage the subjects to gain insight is not the improvements in performance so much as the numbers of subjects who never, no matter what was done to them, selected [the wrong answer]. (Manktelow, 1981, p. 259)

The same is true when we turn again to the examples of people suffering from delusions. The method of "Socratic discussion", according to which by demonstrating the contradiction hidden in the words of a patient, one can persuade him/her to change his/her mind and thus heal, is also often ineffective (Bortolotti, 2010, pp. 86–96). An example could be a case of the patient from the above-cited study, who, after recovering and leaving the mental hospital, continued saying, "Now I am alive. But I was once dead at that time" and "I saw Kim Jong-Il in the hospital where I stayed" (Nishio & Mori, 2012, p. 218). At the same time, it is not absurd or inconsistent with the ordinary use of language in the light of the above data to say that Nishio's patient is convinced that he was once dead, or that the respondents in the Wason test are convinced that the card containing the premise and denial of the conclusion does not falsify the rule stating its implication. This means that, in a situation where we only obtain a little more data about the subject, describing the contradiction of beliefs as part of an intentional stance is perfectly possible and preferable to Dennett's alternative: using a design stance or refraining from describing it in any terms.

Another descriptive element of Dennett's proposal is the recognition that the use of design stance stems from our reluctance to describe others or ourselves as irrational or having contradictory beliefs. However, this claim needs to be substantiated: to prove that these ways of speaking or

linguistic constructs derived from functional strategy are "rationalizations", we must show that it is precisely rationality and consistency that we care about when we use them. This, however, is not the case. One can find many other justifications for this use of language, not having much to do with rationality.[19] However, it does not even seem necessary. As already mentioned, we ascribe errors on the "design stance"-level automatically—it is our first assumption, and not a rationalization that comes to the fore when the possibilities of a consistent explanation of our behavior are exhausted. Importantly, therefore, in Stich's argument, Dennett is wrong in explaining the course of our interpretation of the situation, and not in the conclusion to which his theory obliges him.

It is therefore necessary to carry out the critique to the end and turn to the normative aspect of Dennett's stance. By adopting an interpretationist stance, we must further consider why it is the rationality and consistency that should be the ideal that we follow in individual ascriptions. If we take into account the above-mentioned studies by Wason, Kahneman and Tversky, Bar-Hillel or cases of delusions, it should surprise us how much the intentional stance deviates from actual human behavior in its predictions—how many cases such a theory excludes. If we believe that the intentional stance should allow us to predict someone else's behavior in the best possible way, we must assume that, at least statistically, the most useful description of our behavior is its description in terms of a rationally acting and belief-forming agent. This, however, as the examples above show, is at least far from certain: the human system of inference and belief formulation simply does not seem to follow these standards. A famous example can be the gambler's fallacy—an incorrect inference according to which the probability of an event decreases if the event has happened frequently before (e.g. that the probability of an eagle falling in a toss of a reliable coin is less than ½ if it has previously fallen twenty times). Committing this error is relatively intuitive for most respondents and common among them (Tversky & Kahneman, 1971), they often use a similar principle in predicting facts that depend on probability,

[19] To stipulate such explanations one may discern between consistency and ordinarily understood cohesion: there is nothing contradictory or illogical in many of our actions we tend to explain in a similar way (e.g. slips of the tongue or "socially awkward" or unwanted behavior).

alternating with the contradictory "hot hand fallacy", [20] according to which the probability an event increases when it is repeated enough times (Konold et al., 1993). It is not important here, as in the situations mentioned earlier, whether people are able to recognize such behavior as wrong, but that they often act in accordance with these wrong principles. Thus, if an intentional strategy were to depend on a model that most often produces correct predictions, it should not assume that the subject is procedurally rational, but rather that it forms its beliefs based on certain heuristics consistent with the gambler's and "hot hand" fallacies—a useful "intentional stance" should allow for contradictory beliefs.

The indicated problems with Dennett's theory can be generalized to all stances treating the concept of belief as a concept of folk psychology, which postulate PPNC-N as an element of the practice of individual belief ascriptions. For if there are indeed cases of individual ascriptions that favor the ascription of contradictory beliefs instead of describing a given behavior as a "mere deviation" from the PPNC-N, the thesis about its universal validity is empirically false. However, even if we turn a blind eye to these cases or deny the intuitiveness of such individual ascriptions, there is a much more serious problem for each of these theories. Since in so many cases people, even superficially, tend to act in accordance with the procedurally irrational rules allowing for the inference of mutually contradictory information, the rules governing "time-pressured, quick and dirty ascriptions of folk psychology" should contain these rules and not a rigid canon of procedural rationality.

Maintaining the PPNC-N with respect to individual ascriptions and recognizing it as a methodological principle of folk psychology in its everyday use is therefore unjustified, and the PPNC-N itself presented in such a context is probably false. Therefore, it is necessary to refer to the arguments that defend PPNC-N with regard to scientific ascriptions.

2.2 Consistency and Meaning

Donald Davidson, one of the main supporters of the PPNC-N among contemporary philosophers, shares with Daniel Dennett a set of intuitions about the origin and conditions of the correct use of the terms of folk

[20] This fallacy was first discovered and described in the famous study concerning perception of free throws by basketball fans (Gilovich et al., 1985).

psychology. Davidson's theory, however, clearly refers to interpretation theory as a scientific theory that allows us to produce a "unified theory of meaning and action" inspired by the preference-based belief ascription models proposed by Frank Ramsey in decision theory (Ramsey, 1926; Davidson, 1980). The ascriptions that Davidson talks about will therefore be scientific ones, resulting from appropriate theoretical reflection, explaining to us in the most truthful way verbal and non-verbal human behavior.

Both theories are inspired by the observations of Willard Quine, of whom Davidson and Dennett were students,[21] especially by the thesis of indeterminacy of translation. While Quine's main focus has been on translation and the notion of linguistic meaning, many of his remarks also apply to belief ascription and correspond to the views of his successors. In a famous passage from *Word and Object*, Quine argues that every translation must follow the basic laws of logic:

> That fair translation preserves logical laws is implicit in practice even where, to speak paradoxically, no foreign language is involved. Thus when to our querying of an English sentence an English speaker answers "Yes and no", we assume that the queried sentence is meant differently in the affirmation and negation; this rather than that he would be so silly as to affirm and deny the same thing. Again, when someone espouses a logic whose laws are ostensibly contrary to our own, we are ready to speculate that he is just giving some familiar old vocables ("and", "or", "not", "all", etc.) new meanings. (Quine, 1960, p. 59)

According to Quine, we are obliged to interpret the statements made by others in such a way that will be in accordance with the laws of logic—including the principle of non-contradiction. This thesis can also be presented in the following way: the subject's acceptance of mutually contradictory sentences proves that our translation of a language or idiolect of a given subject is wrong rather than (s)he possesses such beliefs. Logical connectives are functionally defined (by a truth table) and it is impossible by definition to understand a conjunction or negation as we under-

[21] An extensive analysis of similarities and influences between their views may be found in (Dub, 2015, pp. 94–98).

stand them in logic and at the same time recognize the proposition of the form p and not-p.[22]

A similar motivation seems to stand behind Davidson's Principle of Charity. According to it, in order to start the interpretation process at all, it should be assumed that as many beliefs as possible of a given subject are true, and that this subject does not have overtly false beliefs—e.g., logically contradictory ones. Where Quine is looking for a translation, that is, to use its terminology, a stimulus synonymy between sentences of two languages, Davidson tries to find the equivalence at the level of the truth conditions of sentences of both languages—and in order to talk about the knowledge of truth conditions by language users, we must assign certain beliefs to them. In some readings of Davidson's thought, it is often believed that the Principle of Charity consists of two separate principles: the principle that as many beliefs and sentences as possible expressed by the interpreted subject should be true, and the principle that the statements and beliefs of the subject should agree with the canon of rationality (Joseph, 2004, pp. 62–64). These two principles, however, seem to have a common origin: rationality is understood in them as a principle of action aimed at preventing the maintenance of overtly false beliefs, including those that are internally contradictory, and thus maximizing the number of true beliefs.

The above reasoning leads Davidson to the adoption of the following principle as one of the main methodological laws in the process of interpreting others language or idiolect:

> (PC) If an agent asserts or utters mutually contradictory sentences according to the current interpretation of his/her language or idiolect, then interpret his/her statements as non-contradictory in the language or idiolect of this agent.

Since one of the main methodological recommendations made by Davidson is to treat all statements as honest and true for the interpreted subject—and thus entailing that (s)he believes their content—PC can be considered a consequence of adopting PPNC-N.

---

[22] A similar argument against the notion of "paraconsistent logic" (as changing the subject rather than logic) may be found in (Slater, 1995).

So is PC a good and universally valid methodological principle? The argument most often presented for the affirmative answer takes the form of a slippery slope. According to it, once we suspend the validity of the PC, then we are forced to adopt a different rule for the interpretation of the subject, which, while remaining in accordance with the modified inference rules, will result in beliefs and statements "as queer as one pleases" (Quine, 1960, p. 58). However, a similar reasoning cannot be accepted as a justification for the universal application of PC and PPNC-N. The partition between translations and interpretations in accordance with the laws of classical logic and those in accordance with different laws of inference is not complete, as it does not include different degrees of agreement; interpretations can also vary by subject and may not necessarily cover the entire community and language.

The analysis of two different interpretations: compatible and incompatible with the PC, can be carried out on the example of the heated debate on the ontological status of delusions. Their general characteristics have already been outlined above. Many authors, following the suggestion of Dennett and Davidson, have denied giving delusions the status of beliefs, explaining them as imaginations of which agents mistakenly believe to be beliefs (Currie, 2000), or as cases of distinct propositional attitudes referred to as "in-betweenish or grey-area-cases of belief" (Schwitzgebel, 2010) or "bimaginations" (Egan, 2008). These solutions, although compatible with PC (not imposing "responsibility for the given word" on the subject), do not seem to be scientifically useful, but rather constitute a trick needed due to the failure of the more fundamental hypothesis, according to which delusionary patients differ from the standard in their understanding the concepts of "being dead" or "identity". One of the most extensive discussions of various solutions formulated in this spirit is the paper by John Campbell (2001). The basic intuition of Campbell and others seems to be summarized in the following passage:

> Indeed, the patient may say that she is dead even though she realizes that no one else would accept this claim. The trouble is, how can the patients really be said to be holding on to knowledge of the meaning of their remarks when they are using words in such a deviant way? (Campbell, 2001, p. 91)

Campbell recognizes two possible strategies for describing delusions that are compatible with PC: he labels the first as empiricist and the second as rationalist. The empiricist strategy tries to explain the patient's behavior as resulting from data that (s)he begins to receive at some point

(e.g. as a result of damage to the centers responsible for facial recognition in the brain, see [Ellis, Young, 1990]). One can then try to explain the patient's behavior as rational in a broad sense—if a close person, although identical in appearance, ceases to be associated with a subjective feeling of familiarity, the patient comes to the conclusion that the close person has been replaced by an impostor (Capgras delusion) or that the patient himself is dead (Cotard's delusion), which is the reason for the lack of emotions related to a perception identical to the previous one.[23] As Campbell himself notes, this is not a satisfactory "rationalization" strategy for delusions: there are people with similar neurological problems who do not draw similar conclusions. There is also nothing in our experience that could be a possible rationale for accepting the claim of self-non-existence—the neurological disorders listed, at least in the case of Cotard's syndrome, may thus be correlates, but rather not causes of delusions.

The rationalist strategy seems to follow the indications of PC more directly, interpreting the behavior of delusional patients as a result of adopting different framework propositions (Campbell, 2001, p. 96). The concept of framework propositions is borrowed from Wittgenstein's *On Certainty*, who describes them as irresistibly certain propositions which create a frame within which the process of inference and evaluation of the truth or falsehood of other sentences is made. All justification must take place within such a framework (Wittgenstein, 1969).

Although Wittgenstein's concept has never been refined in detail, we may assume, following Campbell, that the beliefs specific to a particular delusion: "I am dead", "This [now seen] woman is not that [once seen] woman" could constitute framework propositions for patients in the proposed sense. Going further, it can be concluded that some patients use some "deviant logic"[24] that allows them to align data from the world with the content of their framework propositions. According to it, e.g. Leibniz's Law of Indiscernibility of Identical functions as a law allowing the identity of objects that have several different properties, which allows us to justify the view according to which e.g. the patient is identical to the

[23] A further discussion on such explanations may be found in the work of McKay and Cipolotti (2007, pp. 351–352).

[24] After Quine, by "deviant logic" I mean here a system of inference in which terms such as "negation", "identity" or "conjunction" have different properties than their counterparts in classical logic (although the rules of inference stay the same).

Virgin Mary (Evnine, 1989, pp. 7–8). However, here we are not dealing with a real violation of the principles of logic and contradiction—such a patient simply understands in a different way (in a different theoretical context or frame) concepts such as "negation", "identity" or "being dead".

There are two problems with the proposal cited above—I will start with a less fundamental one, and then show a more general methodological problem with PC visible in these examples. First, as Bayne and Pacherie (2005) and Bortolotti (2010) note, not all delusions can be defined as framework propositions, because in many cases patients feel the content of their delusions improbable; not all delusions present us with a similar dilemma of application of PC (e.g., persecutory delusions), although we could also describe them as instances of introducing further framework beliefs into a belief system. In many other domains, the reasoning of delusional people remains invariably correct, and they also seem to recognize the gap between laws of logic and their own words (as in the study by Nishio and Mori [2012] cited above).

Second, the fact that the patients *use* the notions of "being dead" or "negation" differently than logicians or doctors do not imply that they *understand* it differently or that it *means* something else to them. Why should we assume that they understand them the same as we do? As I have pointed out, patients, when not asked about the content of their delusions, seem to reason in a classical way, and not according to any "deviant logic". In this context, PC would require postulating that the patients possess an inference system explaining the changes in the rules of inference depending on its content. However, let us recall for a moment the question about the probability of certain events and facts in the research on heuristics. There are multiple contextualizations[25] that significantly increase the number of correct answers among the respondents—and in fact researchers describe the respondents as using different methods of inference. But does it mean that for the respondents the word "probability" means something different in one context than in another? No—the most effective way of explanation is to say that they have contradictory intuitions regarding the interpretation of specific situations and, what is more, those are not conscious intuitions, and the change of reasoning is not volitionary.

[25] See, e.g. a paper by Fiedler (1988) mentioned above.

A much more fertile hypothesis is to recognize that in all these cases we are dealing with objectively false sentences also in the idiolect of an agent, which, however, (s)he considers to be true. The problem for patients with Cotard's delusions, or with the respondents in the research of Kahneman, Tversky and Wason, is not a purely linguistic problem. When using PC, we keep asking the question: how can someone rationally be convinced of such a preposterous thing, and each subsequent answer to it seems to be sensitive to the counterexamples provided by forthcoming empirical data. Thus, this leads to the high instability of hypotheses, which ought to be avoided in empirical sciences. Moreover, it seems strange to prefer the hypothesis of such a bizarre and similarly irrational way of using language over the hypothesis of having a bizarre and irrational belief system. It is more scientifically useful to assume that a person believes in absurdity and try to find out: what may be the cause[26] independent of the agent for the emergence of such a strange belief formation system? This question is empirically decidable on the basis of the assumption adopted and constitutes a step towards a fertile scientific explanation.

I do not want to delve into the extent to which acting in a manner consistent with the PC and PPNC-N (note that it is not to say: in accordance to PC and PPNC-N) is necessary in interpreting the language and behavior of entire communities. If we are to believe in some anthropological evidence (e.g., Rudiak-Gould, 2010; Thagard & Nisbett, 1983, pp. 253–255), assigning entire communities the possession of certain contradictory beliefs, as well as the content of delusions considered both contradictory and true by those who support them, is not problematic and may lead to interesting conclusions.[27] Empirical data, however, are too scarce and the methodology of anthropological research too heterogeneous to draw decisive conclusions.

---

[26] To reverse a Davidsonian slogan: I mean here the causes which are *not* reasons.

[27] Rudiak-Gould (2010) attributes the natives living in the Marshall Islands contradictory beliefs regarding their past, which is simultaneously portrayed as idyll destroyed by the coming of colonizers and the state of war and barbarity brought to an end by the Christian morality brought by the colonizers as well. The author explains it by grounding these beliefs by the natives in two different identities: national or communal (Marshallian) and religious (Christian). The natives have seen the inconsistencies in their visions of past; however, they could not abandon any of them.

3. Summary

In this article, I tried to analyze in detail the argumentation presented in favor of adopting the Psychological Principle of Non-Contradiction. I have singled out two different interpretations: descriptive and normative, which correspond roughly to the realistic and instrumentalistic approach to folk psychology. I examined arguments proposed for a descriptive reading of PPNC, both those based on the interpretation of beliefs as properties and on assumptions about the systemic functions of the human mind, which would be to uphold true beliefs. I have shown that both of these arguments are insufficient for the adoption of the PPNC. Then I pointed out a more general argument showing why PPNC in a descriptive reading can only be interpreted as an empirical hypothesis, and cited studies in cognitive and clinical psychology that allow us to regard it as implausible. Later, I singled out two main lines of argument for PPNC in a normative reading: the argument from Daniel Dennnett's intentional stance and the argument of Donald Davidson and Willard Quine which states, that scientific belief ascription requires the assumption of mutual consistency of beliefs. Using the example of the debate on the interpretation of delusional cases in clinical psychology, I showed why the methodological strategy offered by Davidson and Quine leads to high instability of the initial hypotheses and therefore is no more scientifically useful than assigning contradictory beliefs to the patient.

In the light of the above arguments, it should be concluded that the Psychological Principle of Non-Contradiction in the formulation adopted here does not find a satisfactory ontological or methodological justification. As the Principle of Non-Contradiction seems to be one of the key elements of procedural rationality, it is therefore also doubtful that assigning beliefs and other intentional states requires assuming the rationality of the interpreted agent. The question remains whether the criticism offered here requires a reformulation of the assumptions of interpretationism that most strongly posits a similar assumption, or the adoption of a different model of ascribing beliefs. I believe that the construction of such a model is possible and will allow for a more adequate explanation of the empirical data showing that rationality is a much less common human trait than some philosophers suggest.

REFERENCES

Bar-Hillel, M., Neter, E. (1993). How Alike Is It Versus How Likely Is It: A Disjunction Fallacy in Probability Judgments. *Journal of Personality and Social Psychology*, *65*(6), 1119–1131.

Bortolotti, L. (2010). *Delusions and Other Irrational Beliefs*. Oxford: Oxford University Press.

Breen, N., Caine, D., Coltheart, M., Hendy, J., Roberts, C. (2000). Towards an Understanding of Delusions of Misidentification: Four Case Studies. *Mind & Language*, *15*(1), 74–110.

Campbell, J. (2001). Rationality, Meaning and the Analysis of Delusion. *Philosophy, Psychiatry, & Psychology*, *8*(2–3), 89–100.

Carroll, L. (1995). What the Tortoise Said to Achilles. *Mind*, *104*(416), 691–693.

Ciecierski, T. (2017). Attitudes and Normativity. *Axiomathes*, *27*(3), 265–283.

Cohen, L. J. (1981). Can Human Irrationality Be Experimentally Demonstrated? *The Behavioral and Brain Sciences*, *4*(3), 317–370.

Currie, G. (2000). Imagination, Delusion and Hallucinations. *Mind & Language*, *15*(1), 168–183.

Davidson, D. (1974). Belief and the Basis of Meaning. *Synthese*, 27(3–4), 309–323.

Davidson, D. (1980). Toward a Unified Theory of Meaning and Action. *Grazer Philosophische Studien*, *11*(1), 1–12.

Dennett, D. (1978). *Brainstorms*. Cambridge: MIT Press.

Dennett, D. (1987). True Believers: The Intentional Strategy and Why It Works. In D. Dennett (Ed.), *The Intentional Stance* (pp. 13–36). Cambridge: MIT Press.

Dennett, D. (1981). Making Sense of Ourselves. *Philosophical Topics, 12*(1), 63–81.

Dennett, D. (2015). Not Just a Fine Trip Down Memory Lane: Comments on the Essays on Content and Consciousness. In C. Muñoz-Suárez, F. De Brigard (Eds.), *Content and Consciousness Revisited* (pp. 199–218). New York: Springer.

Dub, R. (2015). The Rationality Assumption. In C. Muñoz-Suárez, F. De Brigard (Eds.), *Content and Consciousness Revisited* (pp. 93–111). New York: Springer.

Egan, A. (2008). Imagination, Delusion, and Self-Deception. In T. Bayne, J. Fernandez (Eds.), *Delusion and Self-Deception: Affective and Moti-*

vational Influences on Belief Formation (Macquarie Monographs in Cognitive Science) (p. 263–280). London: Psychology Press.

Ellis, H. D., Young, A. W. (1990). Accounting for Delusional Misidentifications. *British Journal of Psychiatry*, *157*(2), 239–248.

Evnine, S. J. (1989). Understanding Madness? *Ratio*, *2*(1), 1–18.

Fiedler, K. (1988). The Dependence of the Conjunction Fallacy on Subtle Linguistic Factors. *Psychological Research*, *50*(2), 123–129.

Gottlieb, P. (2007). Aristotle on Non-Contradiction. *The Stanford Encyclopedia of Philosophy*. Retrieved from: https://plato.stanford.edu/archives/sum2015/entries/aristotle-noncontradiction/

Hodges, W. (1977). *Logic*. Harmondsworth: Penguin Books.

Joseph, M. (2004). *Donald Davidson*. Chesham: Acumen.

Kahneman, D., Tversky, A. (1971). Belief in the Law of Small Numbers. *Psychological Bulletin*, *76*(2), 105–110.

Kahneman, D., Tversky, A. (1983). Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review*, *90*(4), 293–315.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., Lipson, A. (1993). Inconsistencies in Students' Reasoning about Probability. *Journal for Research in Mathematics Education*, *24*(5), 392–414.

Kripke, S. (1979). A Puzzle About Belief. In A. Margalit (Ed.), *Meaning and Use* (p. 239–283). Dordrecht: D. Reidel.

Lewis, H. A., Cooper, D. (1979). The Argument from Evolution. *Proceedings of the Aristotelian Society*, *53*, 207–237.

Łukasiewicz, J. (1971). On the Principle of Contradiction in Aristotle. *The Review of Metaphysics*, *24*(3), 485–509.

Łukasiewicz, J. (1987). *O zasadzie sprzeczności u Arystotelesa*. Warsaw: PWN.

Maddy, P. (2012). The Philosophy of Logic. *The Bulletin of Symbolic Logic*, *18*(4), 481–504.

Manktelow, K. I. (1981). Recent Developments in Research on Wason's Selection Task. *Current Psychological Reviews*, *1*(3), 257–268.

Marcus, R. B. (1990). Some Revisionary Proposals About Belief and Believing. In R. B. Marcus (Ed.), *Modalities: Philosophical Essays* (pp. 143–162). New York: Oxford University Press.

McKay, R., Cipolotti, L. (2007). Attributional Style in a Case of Cotard Delusion. *Consciousness and Cognition*, *16*(2), 349–359.

Nishio, Y., Mori, E. (2012). Delusions of Death in a Patient with Right Hemisphere Infraction. *Cognitive and Behavioural Neurology*, *25*(4), 216–223.

Pacherie, E., Bayne, T. (2005). In Defence of the Doxastic Conception of Delusions. *Mind & Language*, *20*(2), 163–188.

Quine, W. V. O. (1960), Word and Object. Cambridge M.A.: MIT Press.

Quine, W. V. O., Ullian, J. S. (1978). *The Web of Belief*. New York: Random House.

Ramsey, F. (1926). Truth and Probability. In R. B. Braithwaite (Ed.), *The Foundations of Mathematics and Other Logical Essays* (pp. 156–198). New York: Harcourt Brace.

Ramsey, F. (1931). *General Propositions and Causality*. In D. H. Mellor (Ed.), *Philosophical Papers* (pp. 145–163), Cambridge: Cambridge University Press.

Richard, M. (1983). Direct Reference and Ascriptions of Belief. *Journal of Philosophical Logic*, *12*(4), 425–452.

Rudiak-Gould, P. (2010). Being Marshallese and Christian: A Case of Multiple Identities and Contradictory Beliefs. *Culture and Religion: An Interdisciplinary Journal*, *11*(1), 69–87.

Schwitzgebel, E. (2010). Acting Contrary to Our Professed Beliefs or The Gulf Between Occurrent Judgment and Dispositional Belief. *Pacific Philosophical Quarterly*, *91*(4), 531–553.

Slater, B. H. (1995). Paraconsistent Logics? *Journal of Philosophical Logic*, *24*(4), 451–454.

Sober, E. (1978). Psychologism. *Journal for the Theory of Social Behaviour*, *8*(2), 165–191.

Stich, S. (1981). Dennett on Intentional Systems. *Philosophical Topics*, *12*(1), 39–62.

Stich, S. (1985). Could Man Be Irrational Animal? *Synthese*, *64*(1), 115–135.

Thagard, P., Nisbett, R. (1983). Rationality and Charity. *Philosophy of Science*, *50*, 250–267.

The Internet Classics Archive. (n.d.). Retrieved from: http://classics.mit.edu/index.html

Wason, P. C. (1968). Reasoning About a Rule. *The Quarterly Journal of Experimental Psychology*, *20*(3), 273–281.

Wason, P. C. (1969). Regression in Reasoning? *British Journal of Psychology*, *60*(4), 471–480.

Wittgenstein, L. (2020) *Tractatus Logico-Philosophicus* (Side-By-Side-By-Side Edition). Retrieved from: http://people.umass.edu/klement/tl

Wittgenstein, L. (1998). *Remarks on the Foundations of Mathematics*. Oxford: Blackwell.

Wittgenstein, L. (1969). *On Certainty*. Oxford: Blackwell.

Originally published as "Czy posiadanie sprzecznych przekonań jest możliwe? Omówienie i krytyka argumentów za psychologiczną zasadą niesprzeczności". *Studia Semiotyczne*, *33*(2), 323–353, DOI: 10.26333/sts.xxxiii2.11. Translated by Maciej Tarnowski.