Words on Kripke's Puzzle

Maciej Tarnowski, Maciej Głowacki

m.tarnowski3@student.uw.edu.pl

m.glowacki6@uw.edu.pl

Abstract

In this paper we present a solution to Saul Kripke's Puzzle About Belief (1979) based on Kaplan's metaphysical picture of words (1990, 2011). Although it is widely accepted that providing such a solution was one of the main incentives for the development of Kaplan's theory, it was never presented by Kaplan in a systematic manner and was regarded by many as unsatisfactory. We agree with these critiques, and develop an extension of Kaplan's theory by introducing the notion of "conservative" word uses, as well as arguing for the restriction of Kripke's Disquotational Principle only to such uses. This restriction allows us to solve Kripke's Puzzle by arguing that the assent of the puzzle's protagonist does not allow disquotation and ascription of corresponding inconsistent belief, as it contains the non-conservative use of a proper name. At the end of the paper, we defend two of Kaplan's theses that we find essential for our argument to work: that individuation of word tokens involves taking into account the intention of the speaker, and that different uniquely referring proper names should be regarded as different homonymous words.

Keywords: metaphysics of words, Kripke's Puzzle, belief, David Kaplan, intentionalism.

1. Introduction

David Kaplan's paper "Words" (1990) started the ongoing philosophical discussion about the metaphysics of words and other linguistic entities. It posed questions fundamental to the debate and explicitly formulated an original metaphysical theory of words. In the paper and his other works preceding "Words", such as *Demonstratives* (1989a) and *Afterthoughts* (1989b), Kaplan emphasized that his most important motivation for introducing the theory of metaphysics of words was to deal with the identity puzzles, like Frege's Puzzle (Frege 1879/1967, 1892/1980) or Kripke's Puzzle About Belief (Kripke, 1979).

These puzzles pose severe difficulties for any direct reference semantics of proper names. As Kaplan describes his project of resolving the tension between the puzzling cases and the direct reference theory:

> After arguing for years, unconvincingly, that semantic value (properly understood) is not affected by substitution, I hit upon a brilliant, new, and completely successful, strategy: argue, instead, that semantic value *is* affected by substitution. But I also argue that, contrary to my own long-standing misapprehension, this result is not contrary to direct reference theory.
>
> (Kaplan 1990, pp. 94-95 [emphasis in the original text])

According to Kaplan, an adequate solution to these problems should take into account that the medium, through which the semantic value is expressed, is crucial for understanding the difference in the cognitive value of co-referential proper names (see Kaplan, 1989b, pp. 598-599). In this case, the medium is the proper name itself, and hence, a theory of words is needed.

Being aware of the motivation behind Kaplan's metaphysics of words, it is striking that he did not state the explicit solution for either of the puzzles. He left the reader with several vague remarks about the identity puzzles in

"Words" and a short paragraph about Kripke's Puzzle at the end of his "Words on Words" (2011, p. 529). The absence of explicit solutions to the puzzles is even more surprising when one realizes that constructing a solution based on Kaplan's metaphysical view about words is not a trivial task. Sylvian Bromberger (2011, p. 489) confessed that "in spite of sincere efforts, [he] was unable to reconstruct Kaplan's solution." Other critiques suggested that the solution is either incomplete, flawed, or cannot be accommodated within Kaplan's theory (see e.g., Hawthorne & Lepore 2011, pp. 27-29; McCulloch, 1991, p. 77-78).

We believe that a solution to identity puzzles can be given within Kaplan's account of the metaphysics of words (albeit, a slightly modified one). In this paper, we fill the gap of Kaplan's theory and provide a Kaplanian solution to Kripke's puzzling case of Paderewski. For the sake of the argument, we will adopt Kaplan's metaphysical view of words, and adjust it only when necessary. Although we believe that a Kaplanian solution to Frege's Puzzle can also be provided, we decided to restrict ourselves and focus solely on Kripke's Puzzle. Constructing a solution to Frege's Puzzle is a matter for another essay.

The plan for the paper is as follows: In the first part, we recall Kripke's Puzzle about Belief (1.1), and introduce the debate about the metaphysics of words (1.2). In the second part, we present Kaplan's intentional account (2.1), and we reconstruct the solution to Kripke's Puzzle suggested by him and point toward its shortcomings (2.2). In the third part, we propose a modification of Kaplan's solution based on the notion of the conservative use of a word (3.1-3.2). In the final parts of the paper, we discuss some objections to the proposed solution and provide answers to them (4).

1.1 Kripke's Puzzle About Belief

Is it possible for an ideally rational agent to acquire a contradictory belief? Intuitively, it seems that if the agent is rational and has a consistent set of beliefs and accepts all and only their logical consequences, they cannot be

caught having such a belief. Saul Kripke challenged this view in his famous paper "A Puzzle about belief" (1979).

The puzzle goes as follows: Imagine Peter, who is an ideally rational agent in the sense stipulated above, namely an agent with a consistent set of beliefs enclosed in relationships of logical consequence. Let us also assume the following Disquotational Principle:

> (DP) If a normal speaker of the language L, on reflection, sincerely assents to ⌜ p ⌝, then he/she believes that p, where p is a sentence of L without indexical expressions.

Peter has heard about the successful politician Paderewski—the third Polish Prime Minister after the First World War. He is aware of his important role in meeting with President Woodrow Wilson, and in obtaining the explicit inclusion of independent Poland in President Wilson's peace terms in 1918. Peter is also a great admirer of the famous pianist and composer named Paderewski. He has listened to records of Paderewski's opera *Manru* many times, and enjoyed it a great deal. Among the variety of Peter's beliefs about the world, there is one that no successful politician can be a great artist. In accordance with his belief, he would sincerely assent to the English sentence:

(1) Paderewski is not a great musician

having Paderewski-the-politician in mind. However, he would also assent to the sentence:

(2) Paderewski is a great musician

having Paderewski-the-composer-of-*Manru* in mind.

Based on the DP we can infer that Peter believes that Paderewski is a great musician, and that he believes that Paderewski is not a great musician.

Since Peter is an ideally rational agent who accepts logical consequences of his beliefs, he believes that Paderewski is, and is not, a great musician. Therefore, despite being ideally rational, Peter has a contradictory belief. To state it even more boldly, since one person cannot have different properties than themselves, Peter believes that Paderewski is a different person than Paderewski.[1] It seems that we have an ideally rational agent, who has come to the conclusion that there exists an object A for which the principle of identity of the form "A=A" does not hold. This is an unacceptable consequence.[2]

Kripke's Puzzle described above is well known to philosophers of language, and it has received a decent amount of attention over the last decades. Since the publication of Kripke's paper, many philosophical solutions to the puzzle have been proposed by, among others, Ruth Barcan Marcus (1981), Andrzej Zabłudowski (1986), Nathan Salmon (1986, 1995), and Scott Soames (2003). Most of them require some kind of modification of the Disquotational Principle. Despite their sophistication, none is widely accepted. We believe that an adequate and intuitive solution can be given, by following Kaplan's insight about the importance of the medium for cognitive value.

1.2 The Metaphysics of Words

It is obvious that even to be able to state the puzzle, not to mention its solution, we have to know what it means for a sentence or a proposition to be of the form "A ≠ A".[3] It is also trivial to say that the sentence "A ≠ A" contradicts the identity principle only if on both sides of the non-identity sign there are two

---

1 It requires internalization from him of an additional principle, namely the so-called Leibniz Principle, according to which two objects are identical if, and only if, they have the same properties (or at least the weaker implication of the identity to the sameness of properties, known as the principle of the indiscernibility of identicals).

2 It is the second version of the puzzle given in Kripke's paper. The other version, concerning the bilingual Pierre who seems to have inconsistent beliefs about London and Londres, is more complicated and needs an additional Principle of Translation. Although we believe that our solution suits this version as well, we focus on the one-language version of the puzzle, for the sake of simplicity.

3 To avoid problems and controversies linked to the concept of a proposition, we will focus here solely on sentences.

tokens of the same name. Otherwise, we may just be tricked by the appearance of the sentence, as in "David (Beckham) ≠ David (Kaplan)".[4] This sentence, although it may be mistaken for a sentence of the form "A ≠ A," is perfectly meaningful and even necessarily true. We claim that one cannot resolve Kripke's Puzzle without first answering the question about a principle of individuation of word-tokens. To ascribe to Peter a belief of the form "A ≠ A," we have to know whether the name *Paderewski* in (1) and the name *Paderewski* in (2) are tokens of the same name-type. It is impossible to know the lexical form of a sentence without knowing which words were used to build it.

Such considerations are part of a broader philosophical enterprise concerning the metaphysics of language and linguistic entities such as words, sentences, phonemes, and so on. In this paper, we are especially interested in the metaphysics of words (and in the metaphysics of names in particular). The interest in the ontological status of words began with David Kaplan's "Words" (1990).[5] This paper started the ongoing philosophical discussion on the metaphysics of linguistic entities, and engaged such prominent figures in philosophy as Sylvain Bromberger (2011), Herman Cappelen (1999), Jerrold Katz (2000), Linda Wetzel (2009), and many others. The dispute concentrated mainly on the following two questions:

(A) When are two tokens instances of the same word-type?

(B) What are word-types?

The first question is crucial in addressing the problem raised by Kripke's Puzzle About Belief. Given an identity statement, like "Hesperus is Phosphorus," or "Paderewski is Paderewski," we may ask about its lexical form. Is it of the form "A = B" or "A = A"? To answer this seemingly trivial question we need to know whether the name "Hesperus" is different from

---

4 We will come back to such examples later on when discussing Kaplan's distinction between generic and common currency names.

5 Although Kaplan's paper started the most recent discussion on metaphysics of words, it is worth remembering that the subject was within philosophical interests much earlier. See for instance, MacIver, (1937) and Russell (1950, pp. 23-30).

"Phosphorus" or whether *Paderewski* on the right is a token of the same name as *Paderewski* on the right of the identity sign. To be more precise: We seek the answer to the question of whether these two word-tokens are of the same word-type.[6] The most common answers to these questions among philosophers are Platonic ones. According to them, word-types are abstract entities. Word-tokens are spatio-temporal objects that are instances of abstract word-types. Two word-tokens are tokens of the same type if they share the same essential properties of the shape of this word-type. It is a view widely assumed among analytic philosophers like Ramsey (1923),[7] MacIver (1937), and Quine (1982). We label such a view the orthographic conception of words.

The orthographic view is commonsensical and gives us a reasonable condition of word-token individuation. Despite that, it faces several difficult problems. One group of problems with the orthographic conception is strictly related to its Platonic ontology—problems of the existence of abstract objects and the problem of the nature of instantiation relation[8]. The other group is related to the questions that concern words specifically. Firstly, there are tokens of the same word-type "schedule" that have different phonetic forms: It is pronounced /ˈʃɛdjuːl/ and /ˈskɛdʒuːl/ in the United Kingdom and the United States, respectively. One word-type can also have tokens with different standard spellings like British "colour" and American "color." Moreover, there

---

6   The distinction between types and tokens, due to Charles Sanders Pierce, is not necessary for stating the problem, but we decided to retain it despite some controversies it is linked to (see e.g., Wetzel [1993] and Kaplan [2011]). We believe that the type/token language may be used consistently without any Platonic ontological commitments.

7 Ramsey writes in his *Critical Notice of Wittgenstein's Tractatus logico-philosophicus*: "[the] 'propositional sign' has type-token ambiguity; the tokens (like those of any sign) are grouped into types by physical similarity (and by conventions associating certain noises with certain shapes) just as are the instances of a word" (Ramsey, 1923, pp. 468-469).

8 One could ask, for example, how an abstract object (like the word: „esperanto") could be *created* (e.g. by Ludwik Zamenhof, the inventor of Esperanto language), if the abstract objects are cannot have any direct causal relationship with the material world. This problem is adressed by Irmak (2018), but in a fashion which significantly distances itself from the ortographic conception.

is a great variety of idiosyncratic differences in the spelling and pronunciation of each English word (Kaplan, 1990, pp. 98-99).[9] From the above examples, it is obvious that orthographic similarity cannot be a necessary condition for two tokens to be tokens of the same word-type. It also cannot be a sufficient one, because of the existence of the homonymy phenomenon. For instance, in English, there are two different words "bank"—the one meaning financial institution and the other meaning a riverbank. It is even more visible in the case of the three different words with the orthographic form "mean," as they are differently classified by grammaticians: one is a noun, one is an adjective, and one is a verb.

It should be noted here that the failure of the orthographic view as a theory of word individuation offers a chance to seek a solution to Kripke's Puzzle. For, according to this theory, Peter's belief that Paderewski is a different person than Paderewski should be expressed by a sentence in the form "A ≠ A," because both tokens of the name "Paderewski" share the same orthographic properties. We believe that the prevalence of the orthographic conception may be suspected of obscuring the solution to the puzzle. In the next part of the paper, we present an alternative theory of the metaphysics of words that offers foundations for a Kaplanian solution to Kripke's puzzle.

Before we go any further, we should clarify our usage of the term "token." It should be noted that the term "token of a word" is used very vaguely in most of the philosophical literature. One meaning of the term is a produced token—a physical entity. In this reading, tokens of a word are objects or events like ink marks on the paper, a sequence of vowels, or a smoke signal. In the other reading, a token of a word is a specific use of the produced object. These notions very often overlap, for instance in the case of a normal conversation, but they may split in some cases. For example, I may use the sentence "Do not disturb" produced by someone else on a hotel sign. In the paper, we use this term as the latter, more fine-grained, meaning. We understand "token of a

---

9 This is only a small sample of problems with the orthographic conception. For a deep and interesting analysis and critique of this stance, see. eg., Wetzel (2009, p. 53-72), Kaplan (1990, 2011), Hawthorne and Lepore (2011), and Irmak (2018).

word" as a specific use of the physical entity (i.e. of the token in the first sense). To denote specific tokens in this sense we use *italics*, while for word-types we use quotation marks.[10]

2. Kaplan's solution to Kripke's Puzzle

2.1 Kaplan's intentionalism

The Platonic view, despite its popularity, seems to be untenable. It is hard to imagine that there are any properties shared by all and only tokens of a given word. David Kaplan in "Words" continues this line of criticism and claims that between two tokens of the same word "the difference in phonographology, the difference in sound or shape or spelling, can be just about as great as you would like it to be" (Kaplan, 1990, p. 101). He criticized the condition of shape similarity of tokens as both unnecessary and insufficient for the individuation of word-tokens, and proposed an alternative, intentional theory of the metaphysics of words.

To back up this controversial claim, Kaplan offers the following scenario, called the Repetition Game (Kaplan, 1990, pp. 102-103). The game is played by two players—let us call them Spoiler and Duplicator. Duplicator's task is to repeat a name uttered by Spoiler. The name is known to both players and Duplicator is given a certain amount of money for each successful repetition. For example, when Spoiler utters the name *Bertrand*, Duplicator should repeat *Bertrand*, *Rudolph – Rudolph*, *Alonzo – Alonzo*, and so on.

In the light of the criticism of the orthographic conception, presented in the previous part of the paper, we should be aware that the "orthographic" differences between Spoiler's and Duplicator's utterances may be quite significant, due to physiological differences between players, their different accents, or their level of sloppiness in speech. However, at least according to

---

10 Later in the paper, we also speak of lexical representations of words, which we denote using a `distinct font`.

Kaplan, if we are only sure that Duplicator intends to repeat the name by their utterance, it should be obvious that they succeed in doing so. It seems then, that there are no other criteria for success, except the intention which is their reason for making the utterance. Kaplan claims that if only Duplicator sincerely wants to repeat the word uttered by Spoiler, there is no reason to doubt that they succeeded, no matter how their utterance differs from Spoiler's or any "standard" articulation of the name.

From this observation, Kaplan derives a conclusion that the only criterion of word individuation is the speaker's intention to utter a specific word. According to him, the speaker's intention is both a necessary and a sufficient condition for using a token of a certain word-type. No matter how deviant Spoiler's articulation of the name "Bertrand," it would nevertheless be a token of the name "Bertrand," if only they had a proper intention of articulating this very name. The necessary and sufficient condition for two tokens to be tokens of the same word is that they were produced with an intention of being used as the same word.

The notion of intentional repetition gives us a good sense of the continuity of word articulations. According to Kaplan, this notion is crucial for understanding how different entities, like a properly accented utterance of the word "cabbage" and an inscription of this word made by a trembling hand, can still be tokens of the same word-type. He refers to his theory as a "stage-continuant model."[11] In this theory, word-types are naturalistic objects composed of their tokens—spatiotemporal objects or events (like inscriptions, Morse code, or utterances). They preserve their identity because every new token of a word has to be produced with the intention of repetition of one of the previous tokens—to utter the same word as one of them.

As we have seen, according to Kaplan's intentional conception of words, tokens of different words can have the same shape. It is a very common

---

11 This name is also due to the reservations Kaplan has about the type/token distinction usage in the context of the metaphysics of words (Kaplan 1990, p. 98). Although we find these reservations legitimate, we chose to retain this terminology. See also ft. 7.

phenomenon when it comes to proper names. To show this, let us introduce one more device of Kaplan's metaphysics: his distinction between generic names and common currency names. Generic names do not have any particular referents. These are names that can be found in books like *How To Name a Baby?* and are sources of particular, individual common currency names, which only refer to the individual named by it. According to Kaplan's view, a common currency name "Donald" referring to Donald Trump, and a common currency name "Donald" referring to Donald Davidson are different words as the intentions of its use in particular situations and the structures of its uses (which constitute the word as such) differs from one another. Two different common currency names derived from one generic name are a clear example of two different words having the same orthographic form. The token of the name "Donald" referring to the 45th President of the United States may be indistinguishable from the name of the author of "Mental Events". But, according to Kaplan's theory, they are different words, because they have different histories of use and different circumstances of introduction.

This distinction will be crucial in solving the mysterious case of Paderewski. For, as we have noticed, even to state the puzzle, we have to know whether in Peter's statement "Paderewski ≠ Paderewski" the name *Paderewski* on the right side is a token of the same type as the name-token *Paderewski* on the left side. As we have seen, the mere orthographic similarity is both unnecessary and insufficient for that.

2.2 Kaplan's solution

It is commonly recognized that Kaplan's account of the metaphysics of words was mostly prompted by the hope of using it to combat identity puzzles such as Kripke's: "perhaps syntax, in some vague sense, was a key to the [identity]

puzzles that I had been unable to solve"[12] (Kaplan, 1990, p. 93). On how his theory may be exactly applied to Kripke's (or Frege's) Puzzle, Kaplan is unclear (Bromberger, 2011, p. 489). It was only pointed out almost twenty years later in "Words on Words", that "Kripke's Peter had made a 'linguistic error' and thus had not satisfied the requirement for disquotation" (Kaplan 2011, p. 505, ft. 3), but again how Kaplanian intentional criteria impact disquotation in the Paderewski–Paderewski case is not made clear.

Let us see where the straightforward modification of the Disquotational Principle along Kaplan's lines may take us (this modification is of course insubstantial as it only explains Kaplan's understanding of *assent*):

> (DP+) If a normal English speaker, on reflection, sincerely assents to ⌜ p ⌝ (which means that ⌜ p ⌝ is produced with an intention to utter ⌜ p ⌝ and is understood as such by the speaker), then he/she believes that p, where p is an English sentence without indexical expressions.

Does DP+ take us any further towards solving Kripke's Puzzle? On the surface, it seems that DP+, along with Kaplan's distinction between generic and common-currency names, should lead us to believe that we should not ascribe to Peter two contradictory beliefs based on his assent to (1) and (2), as he utters these sentences with an intention to say ⌜ p ⌝ and ⌜ ~q ⌝, not ⌜ p ⌝ and ⌜ ~p ⌝. In turn, we should either ascribe to Peter the beliefs that Paderewski_musician had a musical talent, while Paderewski_politician had not (as suggested by Larson & Ludlow, 1993, pp. 318-319), or do not ascribe to him any beliefs about Paderewski (at least based on DP) since he intends to utter two distinct words: "Paderewski_politician" and "Paderewski_musician."

However, both of these solutions presuppose the existence of two distinct common-currency names, which is not the case in Peter's example (this had

---

12 Kaplan in his article uses the notion of „syntax" in a peculiar way: „When I use the word "syntax" or "syntactical" (...) or "logical syntax" it is usually this *syntactical-lexical form* notion that I have in mind" (Kaplan, 1990, p. 94, ft. 4).

been pointed out by some of Kaplan's critics: see McCulloch [1991, p. 77] and Hawthorne & Lepore [2011, pp. 27-29). Peter's intention, as envisaged by Kaplan, is necessary and sufficient for word utterance, but not word creation—and there are no two distinct names for Paderewski-the-politician and Paderewski-the-composer which could be performed distinctly by Peter. Moreover, with Kaplan's liberal identity-criteria for word-token production, it seems that he needs to agree that Peter actually produced two tokens of the same common-currency name. If the only criteria we impose is an intention to repeat the same word we acquired, then it is obvious that our protagonist *repeats the same word "Paderewski"* in (1) and (2) although acquired twice in different cognitive circumstances. Tokens of "Paderewski" in (1) and (2) belong to the same causal-historical chain and are, therefore, stages of the same common-currency name. Therefore, either Kaplan is compelled to hold that somehow there are actually two different common-currency names just by virtue of someone thinking that, or his metaphysical analysis fails to solve the puzzle.

We think, however, that Kaplan's solution points in a promising direction. Kaplan's identity criteria for word-tokens and the impact they have on DP need to be seriously refined, though. What we need is a philosophically viable justification for the modification of DP, which would prevent us from agreeing that Peter assents to (1) and (2) by virtue of his attitude toward names. In the next section we present this kind of justification, which is a modified version of Kaplan's identity criteria and DP, and apply it to Kripke's Puzzle.

3. A modified Kaplanian solution

3.1 Conservative uses of words

Let us start with the following intuition: Under no circumstances, should DP allow the inference of belief from assent to an utterance that involves a linguistic mistake of the agent in question. If someone utters: *My 5-year-old*

daughter has a very rich dictionary for her age (mistaking "dictionary" and "vocabulary"[13]), even if they reflected long before uttering these words, it would be inappropriate to infer that they have a very strange belief about their daughter and some book. Whether this mistake has its source in one's lexical performance ("*You meant 'vocabulary,' 'Yes, of course, my tongue slipped!'*") or competence ("*No, I meant 'dictionary.' 'She knows so many words!'*") should be irrelevant.[14] The problem lies in that our propositional attitude ascriptions are made in a public language, while the utterance we infer it from is performed in one's idiolect; puzzling cases involve roughly the circumstances in which these two do not overlap. We could, of course, restrict ourselves to simply ascribing beliefs in the idiolect of an interpreted agent (as does e.g., Zabłudowski [1986] or, in their own way, Larson and Ludlow [1993]), but this does not seem to take us in the direction we want DP to work. By stating "the speaker believes *that p*," we aim at describing the truth-conditions of the speaker's belief in public language. Suggesting that "the speaker believes that p in idiolect *L*" leaves the question of what their beliefs are about (and how this information might be used in folk-psychological explanation and prediction of the speaker's behavior) unanswered. What we need is a device that lets us troubleshoot these differences; a device that involves, if Kaplan is at least partly right, distinguishing which of the expressions uttered are tokens of words conforming to their public pattern of use, and which are peculiar malapropisms (which probably should not be counted as public word-tokens) that should not inform our theory of belief.

That being said, let us consider the thought experiment which may provide some level of precision to this intuition: namely an extension of Kaplan's Repetition Game. As is well known, native Japanese speakers, if not acquainted with English as youngsters, have difficulties with distinguishing and accurately producing English rhotic /r/ and lateral /l/ phonemes (in

---

13 Not an uncommon mistake among English as a Second Language speakers – and certainly a struggle for one of the authors.

14 These reservations may be already present in Kripke's demand that the speaker in question is "normal." If one wants to read Kripke this way, one should treat further considerations as an explication of the linguistic competence condition.

Japanese these are not distinguished – instead a liquid /r/ phoneme is present; Goto [1971]). Suppose that two different Japanese students, Akira and Botan, have come to an American university (let's call it Harvey Yard) for a summer exchange to work as native-speaking lectors of Japanese. During the exchange, they come to know some English as well. Akira learns that his position name "lector" could easily be confused with the word "rector," which at Harvey Yard refers to the university's chancellor. Botan is not so lucky; hearing the word "rector" in conversation, he takes it to be the same word as "lector," even though he infers that in different contexts it may mean different things. In short, through the history of their interactions, Akira possessed two lexical representations, homonymous from his viewpoint (we might refer to them as (r|l)ecto(r|l)ᴿ and (r|l)ecto(r|l)ᴸ), while Botan has one, which he considers as polysemous ((r|l)ecto(r|l)₍ᴿ|ᴸ₎).

Is there a difference in how we should apply DP to Akira and Botan? We believe that there is. Even if Akira mispronounces the words by uttering /*I am a rector at Harvey Yard*/, we should interpret his words in accordance with his intention. If he wanted to use the word he represents as (r|l)ecto(r|l)ᴸ, then the phonetic shape should be, as Kaplan rightly predicts, irrelevant to the lexical interpretation of his utterance. Akira's utterance is simply a case of mispronunciation, or a peculiar pronunciation, as British /'ʃɛdjuːl/ is for Americans. Botan's utterance /*I am a lector at Harvey Yard*/, on the other hand, should not be regarded as possessing a determinate lexical form. Botan wanted to use the inexistent, ambiguous word represented by him as (r|l)ecto(r|l)₍ᴿ|ᴸ₎—even if the pronunciation accidentally makes it easier to make sense of his words; in Botan's mouth and mind /lector/ it is not the pronunciation of either "lector" or "rector," but a non-existent polysemous word. It does not mean that we cannot understand the meaning Botan tries to convey by some pragmatic means or other folk-psychological heuristics, but it is surely not DP that allows us to ascribe the relevant belief to him.

We draw two conclusions from this scenario. The first one stands in contrast to Kaplan's view: It is not the intention to repeat a word acquired in

the past which matters in the successful use of word-token, but the intention to use the very word in question. Kaplan's original theory cannot capture the difference between the cases of Akira and Botan, as they both, in a sense, repeat the word they acquired. Although they conceptualize this differently, Kaplan's theory has no devices to describe these processes, since both take place in the mysterious "black box." The second conclusion we draw is that the intention to use some word must involve its appropriate, public-rule conforming lexical representation to count as successful. Since Kaplan presents his views on these matters as "liberal" (1990, p. 99), we might label our position as "conservative" and define the following property of word uses:

(Conservativeness) The use U of a token of the word-type "W" is *conservative* if U was performed with an intention that conforms to public standards of individuation of "W."

Such a definition allows us to call a speaker's use of the word non-conservative if the use of this word in the speaker's idiolect differs in a significant way from the public language in which we wish to interpret them. Diverging from Kaplan even more,[15] we think of conservative uses as those produced with the means of lexical representation in the speaker's mind that may be adequately one-to-one paired with a word-type from the public lexicon—the set of all publicly available and used word-types in the given language. This is the case of Akira (since $(r|l)ecto(r|l)_R \rightarrow$ "rector" and $(r|l)ecto(r|l)_L \rightarrow$ "lector"), but not Botan (because $(r|l)ecto(r|l)_{(R|L)}$ does map onto both public words).

We do not want to be too specific yet about the relevant standards of individuation for word-types in the public lexicon—we believe that for the present purpose of modifying DP "public standards of individuation of *W*"

---

15 We specifically differ from Kaplan in acknowledging the role of mental representation in word-token production and abandoning behavioristic „black box"-talk. A more charitable reading of Kaplan perhaps would be to treat lexical representations as his "mysterious *intra*personal stages" of words (Kaplan, 1990, p. 98), but apart from these rather obscure remarks we see no specific evidence for such interpretation in Kaplan's texts.

may be based on Kaplan's stage-continuant model as well as on some other theories of word-type individuation. What we mean here are not any standards of *appropriate performance* as envisioned by Hawthorne and Lepore, for instance, in their principle of Tolerance (2011, p. 17), but merely a standard of individuation which allows us to ensure that the person intends to use the words actually present in the public lexicon, however they are to be individuated.[16]

The appropriate modification of DP which allows us to infer an appropriate conclusion for the cases of Akira and Botan is simply:

(CDP) If a normal English speaker, on reflection, sincerely assents to ⌜ p ⌝, *provided that words featuring in ⌜ p ⌝ are used conservatively and/or understood as such by the speaker*, then they believe that p, where *p* is an English sentence without indexical expressions.

3.2 The Solution

Let us now apply the conservative formulation of DP to the puzzling case of Kripke's Peter. This case works in a reverse way to the Akira/Botan scenario: While Botan mistakenly takes two words to be one,[17] Peter takes one word

---

16 Note that Hawthorne and Lepore's Tolerance actually does not help us to get Akira and Botan's cases; if "[p]erformance p is of a word w only if it meets relevant performance standards" (Hawthorne & Lepore, 2011, p. 17), then Akira's /rector/ should not be counted as a successful performance of the word "lector", while Botan's /lector/ should, despite the only difference in how these standards were met between the two is blind chance of getting the pronunciation right.

17 Interestingly, an even closer analogy between the cases could be drawn if we imagine that Botan takes two proper names, not common nouns, to have the same referent. Paolo Bonardi (2021) proposes such a scenario, which involves a rational speaker believing that the names of Cicero—the Roman orator, and Cicero—the German II World War spy, to be names of the same person and assents to the statement (S) „Cicero is an orator and a spy." We believe that our proposal works here, as well: As this speaker intends to utter the name "Cicero" which maps onto both common-currency names in the public language, we should abstain from disquoting (S).

"Paderewski" acquired in different epistemic contexts to be two. Similarly, as we did with Akira and Botan, we may postulate that Peter possesses two distinct lexical representations: Paderewski$_P$ and Paderewski$_M$. Whenever Peter intends to state something about Paderewski, he does so by employing one of these two. Therefore, if Peter wants to say:

(1) Paderewski is not a great musician.

he clearly intends to utter the name he represents by Paderewski$_P$. When by considering (1), (2), and being aware of the indiscernibility of identicals he concludes that:

(3) Paderewski ≠ Paderewski.

Peter interprets two tokens of the same common-currency name "Paderewski" in (3) as being of different types, through pairing these occurrences with the two distinct lexical representations. This leads us straightforwardly to the conclusion that Peter's uses of "Paderewski" in (1), (2), and (3) are not conservative, and blocks the ascription of contradictory beliefs by means of CDP. Such a solution does not require, contrary to Kaplan's, the existence of two *actual* distinct common currency names, merely supposing that Peter uses two different lexical representations allows us to deal with the puzzle. In fact, it relies on an assumption that there are no two such distinct names in the public lexicon.

A much-needed caveat is that our formulation of CDP limits the scope of views in the metaphysics of words which enables the solution of the puzzle. As we noted before, our definition of conservativeness in itself is pretty agnostic with respect to standards of individuation of the word-type it mentions, and describes a philosophically interesting property of word uses. Analysis of the Akira/Botan case introduced above aims to prove that there are uses that are conservative despite being phonologically atypical (Akira's /rector/), and uses which are not conservative despite conforming to "community standards of

pronunciation" (Botan's */lector/*). Such a conclusion should be acceptable to anyone who believes that there is no distinct set of intrinsic properties of word-tokens that determines their type, and that the speaker's intentions do play a (necessary or sufficient) role in word-token individuation.

We also claim that Kaplan's account of word-token individuation purely based on repetition fails to distinguish between the two cases, and should be refurnished to properly acknowledge the role of lexical representation in order to do so. Our restriction of DP to only conservative utterances (or acts of assent preceded by conservative lexical interpretation) stems from the intuition that in cases when one's idiolect diverges from public rules of language (becomes too "liberal"), we should abstain from ascribing to this person the belief in public language on that basis. But CDP prevents us from ascribing the belief to Peter only if we believe in a possibility of two distinct uniquely referring homonymous proper names, or at least the possibility of the speaker lexically representing proper names in such a way. We think that Kaplan's distinction between generic and common currency names explicates this possibility well—but there are other views around which to secure this kind of fine-grained distinction (see e.g., García-Carpintero, 2018). We believe, however, that this feature should be treated more like an argument for such theories rather than a limitation of our solution as it provides a straightforward response to Kripke's Puzzle.

4. Objections

Kaplan's intentional metaphysics of words gained much criticism after the publication of "Words"; some of it, as discussed earlier, we believe to be fully accurate. As we defend a version of Kaplan's stance, we should try to defend it where both we and Kaplan are in agreement. We will therefore defend two elements of Kaplan's theory: the necessity thesis, according to which having an appropriate intention to utter the word "W" is necessary to produce a token of

the word "W," and the thesis on homonymity of various common-currency names.[18]

The most prominent critique of the necessity thesis was put forward by Herman Cappelen in his reply (1999) to Kaplan's paper. In the paper, he presents two counterarguments against the necessity thesis. The first one (Cappelen, 1999, p. 95-96) is based on the apparent possibility of reusing different physical entities (sometimes unintentionally produced) as word-tokens because of their shape; if I find a sign on the street with "Can you spare a quarter?" written on it, I may use it to ask for money despite the fact that I do not know it's causal–intentional history. This is, however, as we noted before, easily explained away as the confusion between the meaning of the token as a physical object and its specific use. In the same sense we could say that the printed title of this paper is also a token of an expression "Dunkirk sozzle swopper" (since one easily could rearrange the letters with a pair of scissors), however we prefer to think of it as (at best) a potential token of "Dunkirk sozzle swopper" realized only when someone actually uses it with an intention of inscribing such an expression.

Cappelen's main argument against the necessity thesis, which is much more powerful in our opinion, may be presented as follows. It is obvious that in order to communicate, people need to be able to recognize words. To understand the utterance *Donald is a Republican* I have to be able to recognize each of the words: *Donald*, *is*, *a*, *Republican,* and to which type the given tokens belong. It is a *conditio sine qua non* for successful communication. If the necessity thesis is true, and word-tokens are individuated through intentions with which they are produced, then people need to have access to the

18 In particular we do not think we are committed to defending Kaplan's view that having an appropriate intention is sufficient for word-token production, as well as his view that it is the intention to repeat some other token (we explicitly denounced this thesis earlier). Although we are sympathetic to Kaplan's wide tolerance and we certainly agree that the variety of idiosyncratic differences is not easily covered by some strict canon envisaged by the orthographic conception, our solution to Kripke's Puzzle works as well if we, for instance, limit the scope of acceptable difference from the standard pronunciation or inscription as Hawthorne and Lepore do with their principle of Tolerance (2011, p. 17).

intentions of their interlocutors in order to communicate. But they clearly do not have such access to other people's minds. Therefore, if we do not want to accept the rather controversial claim that communication is impossible, we have to abandon the necessity thesis.[19]

We believe that this argument may be accused of mixing epistemic conditions of identification and metaphysical conditions of identity. Of course, from the epistemic point of view, the interpreter indeed has access only to the physical properties of a word-token. Hence, they can identify the word only by the physical properties of its token, the context of the articulation, and so on, but they do not have access to the intention of the producer. By any means, only guessing the intention of the speaker is a fallacious method of identifying words. Nevertheless, these are epistemological considerations. The intentional account claims that the metaphysical principle of word individuation has to take into account the utterer's intentions. This does not mean that there is a well-defined procedure of gaining access to the metaphysical fact concerning the identity of the word.[20] However, we may describe our recognition of the speech of others as a way of investigating their intentions—the vocal shape allows us to determine what the other person intended to say, but is not itself decisive in the matter, as noted by Kaplan in *Demonstratives*:

> When I utter a particular vocable, for example, the one characteristic of the first person pronoun of English, you must decide what word I have spoken or indeed, if I have spoken any word at all (it may have been a cry of anguish). In associating a word with my utterance you take account of a variety of features of the context of utterance that help to determine what I have said but that need not be any part of what I have said. My egotism, my intonation, my demeanor may all support the hypothesis that it was the first person pronoun in English. But these

---

19 This line of critique is quite common in discussions about different forms of intentionalism in the philosophy of language. For a discussion see, for instance, Gauker (2008) and Åkerman (2009).

20 A similar and more elaborate critique of Cappellen's argument is offered in Alward (2005, pp. 176-179).

aspects of personality, fluency and mood are no part of any semantic theory of the first person pronoun.

(Kaplan 1989a, p. 559).

The sound or shape of the word may be therefore conceptualized as an epistemic aid for the interpreter to recognize the speaker's lexical intention, that is what word they intend to use. As Bertrand Russell (1950, p. 24-25) noticed, we might be more prepared to count someone's performance as a performance of the word if we know that someone belongs to another linguistic community and has a specific accent (a token of the shape /*dock*/ may be a token of "dock" in the mouth of an English person, but a token of "dog" in the mouth of a German). This speaker-to-speaker differences in lexical interpretation points toward the fact that while the adherence to general rules of pronunciation may be necessary in counting certain performances as performances of the word, the understanding of lexical intention is also crucial, and needs to be recognized as a part of word-token classification theory.

Another important issue is the controversy surrounding Kaplan's proposal of treating uniquely referential proper names—common-currency names—as different *words* differentiating them from generic names. Some have argued that such treatment unnecessarily "multiplies meanings beyond necessity" (Rami, 2014, p. 124) in the sense that it postulates functioning linguistic ambiguity between common-currency names and generic names, as well as between homonymous common-currency names. Other (e.g., Hawthorne & Lepore 2011, p. 21) critics have pointed out that Kaplan's account is contrary to common-sense intuitions regarding the fact that individuals may be said to share their name (as in "David Kaplan and David Beckham share a first name"). According to proponents of such treatments, proper names are better regarded as indexicals or metalinguistic predicates, which unify all these uses of names and treat them as uses of one word.

In our view, a proponent of Kaplanian distinction may provide two responses to such doubts. First, one may point out various data suggesting that

people actually sometimes seem to distinguish names in a manner predicted by common currency conception. Here is one example provided by Mons Nyquist:

> A primary school teacher asks all the pupils in the class to write their names on the blackboard as a writing exercise. There are two pupils named "Paul" in the class. All of the pupils except for one of the Pauls write their names on the blackboard. The teacher now asks, "Is everyone's name now up on the blackboard?" The students all say "Yes, teacher!. The teacher abruptly responds: "But there are only fourteen names on the blackboard, and there are fifteen of you!?" to which the pupil named Paul who has not written his name says "Yes, my name is up there, the other Paul wrote our name up, so I didn't have to."[21] (Nyquist unpublished, p. 159-160).

It seems intuitive that, although there is a sense in which pupil's utterance was true, there is also a way in which it is "false, and at best a silly joke: in this sense, it is just not the case that a (token of) his name is on the blackboard, even though a token of the Generic name 'Paul' clearly is."[22] If there is a sense in which Paul's utterance is intuitively false, then it cannot be true that in the ordinary language there exists no differentiation between common currency names. Therefore, the perception of Kaplan's distinction as superficial is not grounded in the way speakers of ordinary language use names.

The second thing one may notice is that such distinction may be justified by different linguistic competences matching the uses present in linguistic practice. Consider that a person who believes that a certain name "N" is uniquely referring to some individual *n*, and is not acquainted with anyone else who may be said to bear this name still seems to be perfectly capable of using this name as a tool for reference. A person who believes that names are

---

21 The quote is from Nyquist's unpublished PhD thesis, pages 159-160.
22 From Nyquist's unpublished PhD thesis, page 160.

unique and uniquely referential (that they play a role comparable to social security number), should be still regarded as a competent user of this name, even though they are unaware that with an identically sounding name, other people may refer to different individuals. This difference in competence seems to be crucial when we distinguish between homonymy and polysemy[23]—two homonymous words differing from one polysemous word in that in the case of polysemy one's competence to use the word should cover the competence of its use in all of its potential meanings. Someone who understands how the word "lunch" functions should not have a problem with understanding both the sentence "lunch was delicious" and "lunch took forever," although in the first case "lunch" refers to the dish served and in the second it refers to the dining process. This is not true of homophonic words—a person who understands the sentence "Our coach had limited leg-space" (where "coach" refers to the bus) clearly is not anyhow equipped to understand the sentence "the Yankees changed their coach three times this season" (where "coach" refers to the sports instructor).

This observation shows that proper names in their different uniquely referring uses should rather be regarded as different homophonic words—and coded as different lexemes—than as cases of different potential meaning of one word. Similarly, the competence to understand the statement "David Kaplan and David Beckham share their first name" requires further linguistic competence—the recognition of names as abstract linguistic entities (which Willy Van Langendock [2008, p. 17-20] describes as "proprial lemmas"). If these competences are distinct—the competence to refer to someone with "N," to refer to another object with "N," and to recognize that with "N" one might refer to more than one object—then we take this to be a strong argument to regard different uniquely referring names and names as linguistic entities as

---

23 The difference between polysemy and homonymy is usually introduced on the basis of either genetic considerations of word's shape (whether the similarity is coincidental or tied to identical root) or semantic considerations (Lehrer, 1974). The existence of this difference is however a subject of debate among linguists (see Panman [1982]).

different homonymous words (different common currency names and generic names).

5. Conclusion

In this article, we investigated how metaphysical considerations regarding individuation conditions and the intrinsic nature of name types may help us to solve Kripke's Puzzle in a way which is consistent with direct reference theory. We pointed out that Kripke's Puzzle is deeply connected with the issue of lexical interpretation—the question of what grounds the sentence's lexical form being "A = A" rather than "A = B." We have shown that the commonly shared assumption that it is the orthographic shape of names "A" and "B" which determines this lexical form is wrong due to considerations regarding the variety of speakers' performance. An alternative picture, presented first by David Kaplan (1990), which focusses on the speaker's intention of producing the tokens of "A" or "B" is then suggested as a viable alternative. Following Kaplan, we introduced the distinction between common-currency names and generic names. Further in the paper, we argued that his theory should be modified (in order to escape arguments posed by McCulloch [1991] and Hawthorne and Lepore [2011]) to provide a satisfying solution to Kripke's Puzzle. We argued that such a satisfying solution would be to limit conditions of appropriate uses of words only to conservative uses of words, which we defined as uses performed with an intention which conforms with public standards of word-type individuation. This in turn leads to restricting the Disquotational Principle assumed by Kripke to include only utterances composed of such conservative uses. In the last section, we defended two points of Kaplan's theory we regard essential, and which we include in our solution—the necessity of intentional condition in individuation of word-tokens and the thesis that uniquely referring proper names are different homonymous words—against critique posed by Herman Cappellen (1999), Hawthorne and Lepore (2011) and Dolf Rami (2014).

The overall conclusion of this paper is that the intentionalist account of word individuation and identity can be modified to provide a solution for Kripke's Puzzle, despite concerns expressed by Hawthorne and Lepore, and McCulloch and others. We show independent reasons to believe that only the uses which we label as "conservative" should be regarded as correct uses of word-tokens. In this regard, we depart from Kaplan's original criteria of correct word-token usage, which is based on the notion of "intentional repetition." We think that the picture which includes lexical representations and takes into account the causal history and accuracy of these representations with respect to the public conditions of word-type individuation are more precise and escape many problems haunting Kaplan's position.

This conclusion should be regarded in a wider context as an abductive argument for intentionalist theories of word metaphysics which acknowledge the role of words' causal history and origin, and which enables the distinction of common-currency and generic names. We are not particularly attached to Kaplan's original view, and believe that the proposed modifications may be successfully incorporated in other theories which satisfy these conditions, such as Alward's (2005) or Irmak's (2018); these authors are more keen on acknowledging the social nature of words' performance conditions as opposed to Kaplan's individualistic approach (Alward, 2005, p. 184). These theories definitely have their flaws; for instance, it is argued that individuating words *via* their causal history and origin does not provide enough precise conditions that would allow us to distinguish words borrowed from different languages from their foreign ancestors (see Miller 2020a, p. 877), or that it necessitates words' history while the fact that certain words *might have* been spelled or used differently seems intuitive (Miller, 2020b, p. 6). However, we do not see these flaws as definitive for intentionalist enterprise in the metaphysics of words. Moreover, the fact that, when applied to proper names, it provides us (as opposed to the case of theories which take intrinsic properties of word-tokens to be decisive of their type) with an elegant and intuitive solution of identity puzzles should give us a sufficient reason to develop them.

References

Alward, P. (2005). Between the lines of age: Reflections on the metaphysics of words. *Pacific Philosophical Quarterly*, *86*(2): 172-187.

Åkerman J. (2009). A plea for pragmatics. *Synthese*, *170*, 155–167.

Barcan Marcus, R. (1981). A proposed solution to a puzzle about belief. *Midwest Studies in Philosophy*, *1981*(1), 501-510.

Bonardi, P. (2021). The "Cicero"/"Cicero" Puzzling Case. *Theoria*, *87*(3), 628-642.

Bromberger, S. (2011). What are words? *The Journal of Philosophy*, *2011*(11), 486-503.

Cappelen, H. (1999). Intentions in words. *Nous*, 1999(1), 92-102.

Frege, G. (1879/1967), *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle a. S.: Louis Nebert; translated as *Concept Script, a formal language of pure thought modelled upon that of arithmetic,* by S. Bauer-Mengelberg, in J. van Heijenoort (ed.), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*. Harvard University Press, 1967

Frege, G. (1892/1980) "Über Sinn und Bedeutung", in *Zeitschrift für Philosophie und philosophische Kritik*, 100: 25–50; translated as "On Sense and Reference' by M. Black in Geach and Black (eds. And trans.), (1980), *Translations from the Philosophical Writings of Gottlob Frege.* Blackwell, 3rd edition, pp. 56–78.

García-Carpintero, M. (2018). The Mill-Frege theory of proper names. *Mind*, *127*(508), 1107-1168.

Gauker, C. (2008). Zero tolerance for pragmatics. *Synthese*, *165*(3), 359-371.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuropsychologia*, *9*(3): 317–323.

Hawthorne, J., & Lepore, E. (2011). On words. *The Journal of Philosophy*, *2011*(11), 1-39.

Irmak, N. (2018). Ontology of words. *Erkenntnis*, *84*, 1139–1158.

Katz, J.J. (2000), *Realistic rationalism*. MIT Press.

Kaplan, D. (1989a), Demonstratives, in J. Almog, J. Perry, & H. Wettstein, *Themes from Kaplan*. Oxford University Press, pp. 481-563.

Kaplan D. (1989b). Afterthoughts, in: J Almog, J. Perry, & H. Wettstein, *Themes from Kaplan*. University Press, pp. 564-614.

Kaplan, D. (1990). Words, *Proceedings of the Aristotelian Society*, 1990(64), 93-119.

Kaplan, D. (2011). Words on Words, *The Journal of Philosophy,* 2011(11), 504-529.

Kripke, S. (1979). A Puzzle about belief. [In:] A. Margalit (red.), *Meaning and Use*, Dordrecht, 239-283.

Larson, R. K., & Ludlow P. (1993). Interpreted logical forms. *Synthese*, *95*(3): 305-355.

Lehrer, A. (1974). *Semantic fields and lexical structure*. North-Holland.

MacIver, A. M. (1937). Token, type and meaning. *Analysis*, *4*(4), 58-64.

McCulloch, G. (1991). Making sense of words. *Analysis*, *51*(2), 73-79.

Miller, J. T. M. (2020a). On the individuation of words. *Inquiry*, *63*(8), 875-884.

Miller, J. T. M. (2020b). The ontology of words: Realism, nominalism, and eliminativism. *Philosophy Compass*, *15*(7), e12691.

Quine, W. V. O. (1982). *Mathematical logic*. Harvard University Press.

Panman, O. (1982). Homonymy and polysemy. *Lingua, 58*(1-2), 105–136.

Rami, D. (2014). The use-conditional indexical conception of proper names. *Philosophical Studies*, *168*(1), 119-150.

Ramsey, F. (1923). Critical notice of Wittgenstein's Tractatus logico-philosophicus. *Mind*, 1923, 465-478.

Salmon, N. (1986). *Frege's Puzzle.* MIT Press.

Salmon, N. (1995). Being of two minds: Belief with doubt. *Noûs, 29*(1), 1-20.

Soames, S. (2003). *Philosophical analysis in the twentieth century. Volume 2. The Age of Meaning*. Princeton University Press.

Van Langendonck, W. (2008). *Theory and typology of proper names*. Mouton de Gruyter.

Wetzel, L. (1993). What are occurrences of expressions? *Journal of Philosophical Logic*, 1993(22), 215-219.

Wetzel L. (2009), *Types and tokens: On abstract objects*. MIT Press.

Zabłudowski A. (1986). On Marcus' solution to Kripke's puzzle and a few related issues*. Pacific Philosophical Quarterly*, 67, 279-296.