Philosophia (2023) 51:1445–1448
https://doi.org/10.1007/s11406-022-00599-5

A Reply to Haze's Argument Against Arbitrary Reference

Sofía Meléndez Gutiérrez[1]

Received: 26 July 2022 / Revised: 26 July 2022 / Accepted: 27 November 2022 /
Published online: 12 December 2022

Abstract
This paper is a response to Haze's brief argument for the falsity of the theory that instantial terms refer arbitrarily, proposed by Breckenridge and Magidor in 2012. In this paper, I characterise instantial terms and outline the theory of arbitrary reference; then I reconstruct Haze's argument and contend that it fails in its purpose. Haze's argument is supposed to be a *reductio ad absurdum:* according to Haze, it proves that a contradiction follows from the most basic tenets of the theory of arbitrary reference. I will argue, however, that the contradiction in question follows not from these tenets, but from the surreptitious use that Haze makes of a self-referential expression. I conclude, consequently, that Haze's argument is nothing more than an illustration of the well-known fact that self-referential expressions produce paradoxical results.

Keywords Arbitrary reference · Instantial terms · Instantial reasoning · Paradox · Self-reference

This paper is a response to Haze's (2016) brief argument for the falsity of the theory that instantial terms refer arbitrarily (Breckenridge and Magidor, 2012). In what follows, I will characterise instantial terms and outline the theory of arbitrary reference. I will then reconstruct Haze's argument and contend that it fails in its purpose.

Suppose that a reasoner reaches a conclusion conforming to the following schema: $\ulcorner \exists x \Phi x \urcorner$. She may introduce $\ulcorner \Phi b \urcorner$ by existential instantiation, provided that $\ulcorner b \urcorner$ does not occur earlier in her proof. In the latter formula, $\ulcorner b \urcorner$ functions as an instantial term. Informally, $\ulcorner b \urcorner$ would be introduced to the derivation in question by means of a locution akin to the following: ⌜Let b be a Φ⌝. Pre-theoretically, then, we could say

✉ Sofía Meléndez Gutiérrez
sm2272@cam.ac.uk

[1] University of Cambridge, Cambridge, Cambridgeshire, UK

that b is any one of the Φs, or an arbitrary Φ; but, in reality, the semantics of $\ulcorner b \urcorner$ is far from clear: the individuals satisfying $\ulcorner \Phi x \urcorner$ might be more than one, but at no point is any one of them specified to be the referent of $\ulcorner b \urcorner$. Hence, our reasoner intends $\ulcorner b \urcorner$ to refer to a Φ, but no Φ in particular. Breckenridge and Magidor claim that, as long as there are Φs, $\ulcorner b \urcorner$ *refers arbitrarily* in virtue of these intentions: it refers to a Φ, but *nothing* determines which Φ $\ulcorner b \urcorner$ refers to, and no one can identify any one of the Φs as the referent of $\ulcorner b \urcorner$ (2012, pp. 377–381).

Breckenridge and Magidor believe that their theory is counterintuitive: it conflicts, they maintain, with the commonly held view that, for every referring expression, there is something that determines what it refers to (2012, p. 380). However, the burden of their paper is that this view is simply mistaken (2012, p. 380).[z,11]

In 2016, Haze contended that he had formulated a *reductio ad absurdum* of the theory of arbitrary reference. Let us suppose, Haze says, that there are objects that are not referred to by anything ever. Now, by existential instantiation, let c be one of them. On Breckenridge's and Magidor's theory, this stipulation makes it the case that 'c' refers arbitrarily to one of the objects that are never referred to—and that is absurd. Hence, Haze concludes, the theory of arbitrary reference is false (2016, p. 540).

Haze seems to me to be mistaken: his argument, I believe, is nothing more than an illustration of the well-known fact that self-referential expressions produce paradoxical results. Consider, for instance, the so-called liar sentence: 'This sentence is false'. If it is true, then what it says must be the case—so it is false. If it is false, then what it says happens to be the case and, hence, it is true.

The liar's paradox is notable amongst the paradoxes of self-reference, but it is Berry's paradox that I wish to discuss in order to articulate my response to Haze. Suppose that an expression *specifies* a positive integer n just in case it refers to n directly, or is uniquely satisfied by n. There are finitely many words and, hence, finitely many expressions constituted by fourteen words or fewer. There are, then, finitely many expressions involving fourteen words or fewer such that each one of these expressions specifies a positive integer. Since there are infinitely many positive integers, there are positive integers that cannot be specified by an expression that is fourteen words or fewer. By the well-ordering principle, it follows that there is a least positive integer amongst them: let us call it k. Clearly, k uniquely satisfies the following definite description:

1. The least positive integer not specifiable by any expression of fourteen words or fewer.

If it does, though, then k is, after all, specifiable by an expression that is fourteen words or fewer, for (1) comprises exactly fourteen words. It follows that the least positive integer not specifiable in fourteen words or fewer is specifiable in fourteen words.[2]

[1] See also Kearns and Magidor (2012) for a defence of the view that there are brute semantic facts.

[2] For the original formulation of Berry's paradox, see Russell (1908).

I think that Haze's argument shares an interesting feature with Berry's paradox. (1) invokes a set containing itself—viz. the set of expressions of fourteen words or fewer which specify a positive integer. Haze's argument, too, features a self-referential expression—albeit one of a special kind. A definition is *impredicative* if it refers to, or quantifies over a set containing its *definiendum*. Haze's argument involves what might be called an *impredicative stipulation*: the term '*c*' is to refer to one of the objects that are never referred to by anything. The stipulation whereby '*c*' is introduced, that is to say, invokes a set that includes '*c*' itself—namely, the set containing everything.

Any standard stratification of languages will bar self-reference, and dissolve both Berry's paradox and Haze's offensive against the theory of arbitrary reference.[3] Roughly, in the simplest of these stratifications, it is supposed that (1) is formulated in a language L_1, which can only refer to, and quantify over languages of a lower level. It follows that the expressions that (1) quantifies over belong to one such language—call it 'L_0'. If that is the case, then (1) should be precissified thus:

1*. The positive integer that is not specified by any expression$_0$ that is fourteen words$_0$ or fewer.

Let k be the denotation of (1*). Now, (1*) is exactly fourteen words, but it is not an expression of L_0. Hence, the fact that k is denoted by (1*) does not imply that k does not satisfy (1*). The paradox does not arise.[4]

Similarly, we may suppose that the language that '*c*' belongs to proscribes impredicativity: if '*c*' is stipulated to be a term referring to one of the objects that are never referred to by anything, then 'anything' cannot amount to a universal quantification over the elements of a set containing '*c*' itself. Now the object *c* can be consistently not referred to by *anything*, in this very specific sense, and referred to by '*c*'.

Language stratification has well-known disadvantages as a strategy to solve semantic paradoxes.[5] Nevertheless, my aim is not to propose a superior alternative, but merely to show that, like any one of the standard paradoxes of self-reference, Haze's argument can be dismantled if self-reference is banned for the relevant language; and, consequently, to contend that its paradoxical result is best blamed not on the Breckenridge-Magidor theory, but on self-reference.[6]

At this point, I believe, it is clear that Haze's argument and Berry's paradox are structurally alike; and I think that the function of the theory of arbitrary reference in Haze's argument corresponds to that of the well-ordering principle in Berry's paradox. Both Haze and Berry invoke sets whose members are characterised by not being the referents of any one of the expressions of a certain class. Let us call these sets

[3] French, moreover, noted that Berry's paradox can be solved by noting that it relies on the false assumption that "a given description that identifies a natural number cannot be used again to describe another natural number" (1988, pp. 1220).

[4] Quine (1966, pp. 6–10) advanced a solution of this kind for Berry's paradox.

[5] See Bolander (2017) for a comprehensive outline of the corresponding discussion.

[6] Yablo's paradox is a non-standard semantic paradox whose solution requires not mere stratification, but a well-founded version thereof (see Yablo, 1985 and 1993).

B and H, respectively. By the well-ordering principle, it is possible to single out a member of every subset S of the positive integers by means of a definite description: ⌜S's least element⌝. In Berry's paradox, it is this principle that allows us to single out one of the members of B. Analogously, by the theory of arbitrary reference, it is possible to refer arbitrarily to an individual member of any non-empty set. In Haze's argument, this theory allows us to refer arbitrarily to one of the elements of H. The parallelism is patent. To my mind, then, blaming Haze's paradoxical result on the theory of arbitrary reference, as Haze does, is as confused as blaming the well-ordering principle for the genesis of Berry's paradox.

References

Bolander, T. (2017). Self-Reference. In Edward N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy*

Breckenridge, W., & Magidor, O. (2012). Arbitrary reference. *Philosophical Studies*, *158*(3), 377–400.

French, J. D. (1988). The false assumption underlying berry's paradox. *Journal of Symbolic Logic*, *53*(4), 1220–1223.

Haze, T. (2016). A counterexample to the Breckenridge-Magidor Account of Instantial reasoning. *Journal of Philosophical Research*, *41*, 539–541.

Kearns, S., & Magidor, O. (2012). Semantic Sovereignty. *Philosophy and Phenomenological Research*, *85*(2), 322–350.

Quine, W. V. O. (1966). *The Ways of Paradox and other essays*. Harvard University Press.

Russell, B. (1908). Mathematical Logic as based on the theory of types. *American Journal of Mathematics*, *30*(3), 222–262.

Yablo, S. (1985). Truth and reflection. *Journal of Philosophical Logic*, *14*(3), 297–349.

Yablo, S. (1993). Paradox without self-reference. *Analysis*, *53*(4), 251.