SELF-REFERENTIAL ARGUMENTS IN PHILOSOPHY

Elke BRENDEL
Universität Mainz

Summary
The paper discusses the strengths and weaknesses of arguments of proper self-reference, arguments of self-application and arguments of iterative application. A formalization of the underlying logical structure of these arguments helps to identify the implicit premises on which these arguments rest. If the premises are plausible, the conclusions reached by these arguments must be taken seriously. In particular, all the types of argument discussed, when sound, show that certain theories that purport to be universally applicable are not tenable. The argumentative power of such arguments then depends on how devastating it is for the theories in question to give up their claim of universal applicability.

1. Introduction

Self-referential arguments are highly regarded in philosophical disputes. They are often considered to be particularly profound and strongly persuasive. Self-referential arguments are typically destructive in nature. They are often employed in order to show that a philosophical theory leads to a contradiction—or at least to seriously counterintuitive results—with respect to certain self-referential propositions the theory allows one to construct. In logic and formal semantics, there are paradoxes yielded by self-referential arguments—like Russell’s paradox and the famous Liar paradox—that have had an enormous impact on our views about fundamental concepts in set theory and on our view about the concept of truth, respectively. But also in other philosophical disciplines, self-referential arguments are often used to undermine some of the most basic assumptions of a theory. In spite of their ubiquity in philosophical reasoning, self-referential arguments are rarely the topic of metaphilosophical examination. Most textbooks on informal logic or critical think-
ing do not even address the strengths and weaknesses of these kinds of arguments.¹

In the following, I will argue that there is not just one single type of self-referential argument, but rather that we have to distinguish between at least two main types of self-referential argument: arguments of proper self-reference and arguments of self-application. I will also argue that arguments of self-application need to be divided into three subversions: arguments of propositional self-application, arguments of predicative self-application, and arguments of individual self-application. I will furthermore analyze another type of argument which I refer to as arguments of iterative application. Although these latter arguments are not strictly speaking self-referential arguments, they share some important logical features with arguments of self-application.

After a formalization of the underlying logical structures of all these arguments, I will examine their argumentative power. In particular, I will identify the implicit premises on which these arguments rest and discuss the intuitive plausibility of these premises. If the premises are plausible, the conclusions reached by such an argument must be taken seriously. In particular, all the types of arguments discussed, when sound, show that certain theories that purport to be universally applicable are not tenable. The argumentative power of such arguments then depends on how devastating it is for the theories in question to give up their claim of universal applicability.

2. Arguments of proper self-reference

Arguments of proper self-reference employ a sentence \( A \) which is said to be equivalent to a sentence that attributes or denies a certain property to the very same sentence \( A \). It is then shown that this self-referential structure generates an inconsistency or reveals counterintuitive results of a given theory.² The structure of arguments of proper self-reference can be put

1. Holm Tøtens' recent book *Philosophisches Argumentieren*—see Tøtens 2005—is an exception here. He examines only very briefly certain kinds of self-referential arguments and dismisses them as unpersuasive. According to Tøtens, despite their “mass production” in philosophy, convincing self-referential arguments are extremely rare. In contrast to Tøtens, I will make a more positive appraisal of the argumentative strength of self-referential arguments in this paper.

2. The term “theory” is used throughout this paper in a rather lax way. It does not only apply to full-fledged philosophical systems, but also to less elaborated accounts or to single hypotheses.