

## replacement.1 Extrinsic Considerations about Replacement

sth:replacement:extrinsic:  
sec

We start by considering an *extrinsic* attempt to justify Replacement. Boolos suggests one, as follows.

[...] the reason for adopting the axioms of replacement is quite simple: they have many desirable consequences and (apparently) no undesirable ones. In addition to theorems about the iterative conception, the consequences include a satisfactory if not ideal theory of infinite numbers, and a highly desirable result that justifies inductive definitions on well-founded relations. (Boolos, 1971, 229)

The gist of Boolos’s idea is that we should justify Replacement by its fruits. And the specific fruits he mentions are the things we have discussed in the past few chapters. Replacement allowed us to prove that the von Neumann ordinals were excellent surrogates for the idea of a well-ordering type (this is our “satisfactory if not ideal theory of infinite numbers”). Replacement also allowed us to define the  $V_\alpha$ s, establish the notion of rank, and prove  $\in$ -Induction (this amounts to our “theorems about the iterative conception”). Finally, Replacement allows us to prove the Transfinite Recursion Theorem (this is the “inductive definitions on well-founded relations”).

These are, indeed, desirable consequences. But do these desirable consequences suffice to *justify* Replacement? *No*. Or at least, not straightforwardly.

Here is a simple problem. Whilst we have stated some desirable consequences of Replacement, we could have obtained many of them via other means. This is not as well known as it ought to be. But the brief point is this. Building on work by Montague, Scott, and Derrick, Potter (2004) presents an elegant theory of sets. This is sometimes called **SP**, for “Scott–Potter”, and we will stick with that name. Now, in its vanilla form, **SP** is strictly weaker than **ZF**, and does not deliver Replacement. Indeed,  $V_{\omega+\omega}$  is an intuitive model of Potter’s theory, just as it was of **Z**. However, **SP** is a bit stronger than **Z**. Indeed, it is sufficiently strong to deliver: a perfectly satisfactory theory of ordinals; results which stratify the hierarchy into well-ordered stages; a proof of  $\in$ -Induction; and a *version* of Transfinite Recursion. In short: although Boolos didn’t know this, all of the desirable consequences which he mentions could have been arrived at *without* Replacement.

(Given all of this, why did we follow the conventional route, of teaching you **ZF**, rather than **SP**? There are three reasons for this. First: Potter’s approach is rather nonstandard, and we wanted to equip you for reading more standard discussions of set theory. Second: when it comes to dealing with foundations, **SP** may be more philosophically satisfying than **ZF**, but it is harder to work with at first. So, frankly, you will only be in a position to appreciate **SP** *after* you’ve studied **ZF**. Third: when you are ready to appreciate **SP**, you can simply read Potter 2004.)

Of course, since **SP** is weaker than **ZF**, there are results which **ZF** proves which **SP** leaves open. So one could try to justify Replacement on extrinsic

grounds by pointing to one of these results. But, once you know how to use **SP**, it is quite hard to find many examples of things that are (a) settled by Replacement but not otherwise, and (b) are intuitively true. (For more on this, see [Potter 2004](#), §13.2.)

The bottom line is this. To provide a compelling extrinsic justification for Replacement, one would need to find a result which *cannot* be achieved without Replacement. And that's not an easy enterprise.

Let's consider a further problem which arises for any attempt to offer a purely extrinsic justification for Replacement. (This problem is perhaps more fundamental than the first.) Boolos does not just point out that Replacement has many desirable consequences. He also states that Replacement has "(apparently) no undesirable" consequences. But this paranthetical caveat, "apparently," is surely absolutely crucial.

Recall how we ended up here: Naïve Comprehension ran into inconsistency, and we responded to this inconsistency by embracing the cumulative-iterative conception of set. This conception comes equipped with a story which, we hope, assures us of its consistency. But if we cannot justify Replacement from within that story, then we have (as yet) no reason to believe that **ZF** is consistent. Or rather: we have no reason to believe that **ZF** is consistent, apart from the (perhaps merely contingent) fact that no one has discovered a contradiction *yet*. In exactly that sense, Boolos's comment seems to come down to this: "(apparently) **ZF** is consistent". We should demand greater reassurance of consistency than this.

This issue will affect any *purely* extrinsic attempt to justify Replacement, i.e., any justification which is couched solely in terms of the (known) consequences of **ZF**. As such, we will want to look for an *intrinsic* justification of Replacement, i.e., a justification which suggests that the story which we told about sets somehow "already" commits us to Replacement.

## Photo Credits

## Bibliography

- Boolos, George. 1971. The iterative conception of set. *The Journal of Philosophy* 68(8): 215–31.
- Potter, Michael. 2004. *Set Theory and its Philosophy*. Oxford: Oxford University Press.