Age, Period, and Cohort Effects in Philosophy Journal Citations

Anon

2025-02-12

There are extremely strong age and period effects in citations in philosophy journals. The age effect is that citations are concentrated on articles published two to five years prior. The period effect is that recent years have seen an explosion in the number of articles published, and the number of citations per articles, so many articles are getting more citations per year than they ever had previously. But cohort effects are trickier to detect. In this note I argue that they exist. There are more citations to articles from eras of more dramatic change in philosophy, such as around 1970 and around 2010. And there are fewer citations to articles from periods of consolidation, especially in the late 1970s through the 1980s.

1 Introduction

Before looking at the data, here are two things I believed about philosophy citations. First, philosophers tend to cite very old papers. We still regularly teach a number of papers over half a century old in introductory classes; e.g., (**?**), Thomson (1971), (**?**), (**?**). These

aren't taught as history papers, but as early entries into the contemporary philosophical debate. And, I thought, that's how we cite. Second, the technological changes of the last quarter century meant that this practice was being slowly reversed. The spread of electronic communication in the late 20th century, and then the rise of archives (e.g., Arxiv, SSRN, PhilPapers) and eventually journals publishing in EarlyView, meant that papers could now be cited even before they were published, and certainly without the delays involved in printing and posting journals around the world.

Both of these thoughts were wrong. Historically, philosophy papers have tended, when they are citing other philosophy papers, to cite very recent ones. But this tendency is diminishing, not increasing, over time. I'll offer much more evidence for these claims as we go along, but to make them plausible, I'll start with two simple graphs.

The data for the graphs come from citation data I downloaded concerning XXX papers published from 1955-2021,[1] in one hundred leading philosophy journals. I focussed on the citations to and from journals in this dataset. So every citation is from one of these 100 journals between 1955 and 2021, and to one of these 100 journals between 1955 and 2021. (The details of the journals, including when they start getting indexed for this dataset, are in **?@sec-methodology**.) In total, that gives us YYY citations.

Say the *age* of a citation is the difference between the publication year of the citing article

[1] I would like to have more recent data, but this is the latest full year of data available through my university's contract with Web of Science. I do have substantial partial data for 2022, and it mostly confirms the trends shown here. But in this case I think it's better to leave off partial data than to try to correct for its incompleteness.

and the cited article. So if an article published in 1998 cites an article published in 1985, that's a 13 year old citation.

In Figure 1 I've plotted the number of citations in the dataset with each possible age. As you can see, it's very heavily tilted towards the left-hand edge. It is true that people still cite (**?**). Indeed, it's one of the most cited papers in the last ten years. But it's just one paper; the bulk of citations are to recently published papers which, if history is any guide, will soon stop collecting citations.

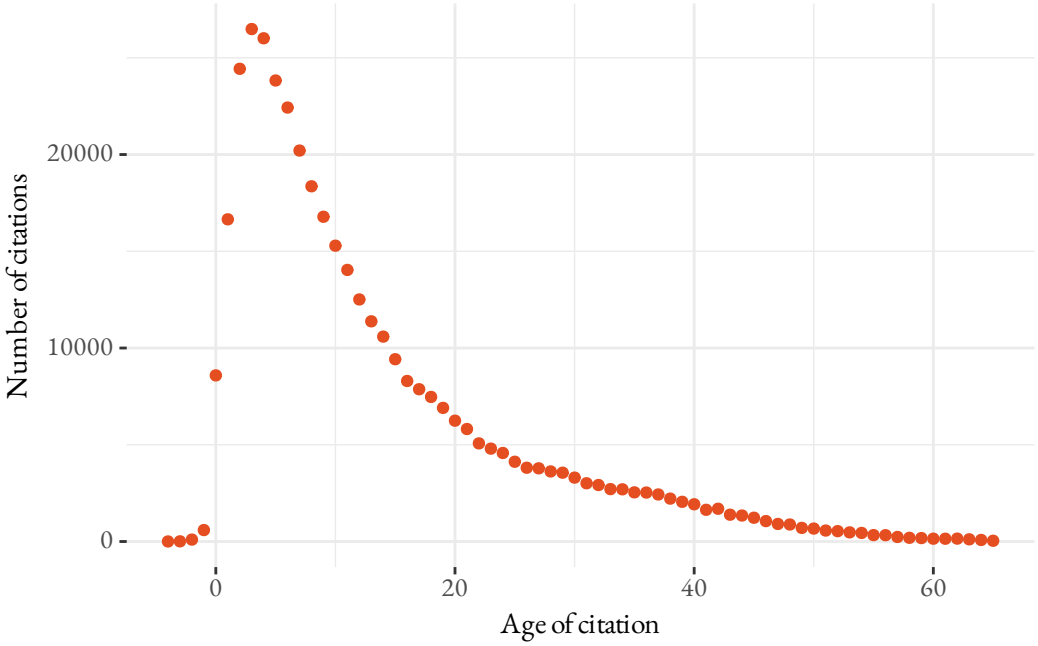
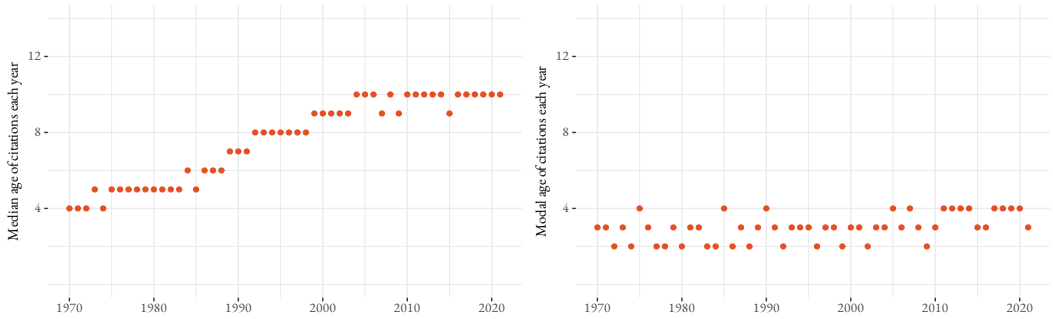


Figure 1: Number of citations with each age.

In Figure 2 I've plotted the median (on the left) and mode (on the right) age of citations in each year from 1980 onwards. Before that the numbers are even lower, but since I'm

only looking at citations to articles published after 1955 (or later if Web of Science started indexing the journal later than that), this is arguably an artifact of how I'm collecting the data. From 1980 onwards, however, there are many older articles that could be, but are not, getting cited. The upwards trends in both graphs look like a real change in citation practices, and not in the direction I antecedently expected.



(a) Median (b) Mode

Figure 2: Summary statistics for outbound citations each year 1970-2021.

There is a third surprise in the data, but it's a little more equivocal, and I'm not sure what to make of it. The 2010s seemed like, and to be honest still seem like, something of a golden age for philosophy. In metaphysics we saw the biggest paradigm shift in many years, away from modality and towards grounding. We saw the growth of important fields of social philosophy, including social epistemology, social metaphysics, and social philosophy of language. Though there were some earlier papers that have become important to the latter two fields (e.g., (**?**), and (**?**)), it would have been a stretch to even call them 'fields' before 2010. Social epistemology was always a bit bigger, and you could point to earlier field defining work by, e.g., Jennifer (**?**) and Adam Elga (2007). But it

grew phenomenally in the 2010s. I'd predicted that would show up in higher citations to work in the 2010s, as these changes were consolidated. The data are a bit messy, and it would be good to have much more data, but this does not look to have happened. There isn't as neat a graph for this, however, and we'll return to this point at the end.

2 Age, Period, and Cohort

To help understand the citation patterns, I'll borrow some terminology that's common in both sociology and medicine. Imagine that we see, in the historical record, some interesting patterns among teenagers in the late 1960s, and we're wondering what could explain the pattern. Two types of pattern spring immediately to mind, along with ways to test them.

First, the behaviour could be explained by the fact the people involved are teenagers. If so, it is an **age effect**. The natural way to test this is to see if similar patterns show up with teenagers at different times.

Second, the behaviour could be explained by the fact that it was the 1960s, and lots of striking things happened in the 1960s. If so, it is a **period effect**. The natural way to test this is to see if the same pattern shows up with non-teenagers in the 1960s.

There is an important third kind of explanation. The people involved are born in the early 1950s, so they are part of the post-war baby boom. Colloquially, they are boomers.

Maybe that could explain the pattern we see. If so, it is a **cohort effect**. The natural way to test this is to see if the same pattern shows up if we look at the same people in other stages of their life.

It's easy to overlook the importance of cohort effects. Sometimes they simply look like age effects. Ghitza, Gelman, and Auerbach (2023) argue that many hypotheses about age effects on voting, e.g., that older people are more naturally conservative, are really just cohort effects. (**?**) argues that understanding the distinctive role the boomers in particular play is crucial for understanding many aspects of modern American life.

There are mathematical reasons that it is hard to tease these effects apart too. Many statistical techniques for separating out influences start to fall apart when there are linear correlations between combinations of variables. In this case there is as tight a correlation as is possible. By definition, cohort plus age equals period. There are some things you can do to get around this problem - see Keyes et al. (2010) for a useful survey of some of the options - but it remains a challenge.

Even conceptually, it is hard to separate out these three effects in cases where there is evidence that the strength of the effects changes over time. As I noted at the start, the natural way to test hypotheses about which effect is strongest involve looking at other times. That works well when the age effects are constant. When they are not (and they might not be here), it is harder.

For most of our story, however, it helps just to have these three effects in mind. Using

them, we can summarise the data reasonably quickly.

- The age effect is that articles get cited most when they are two to five years old, as shown in Figure 1.
- The period effect is that there are many more citations in recent years than in earlier years. This is in part because the number of articles published in these journals has been growing, and in part because the number of citations per article grew substantially over the 2000s and 2010s, and exploded in the 2020s.
- The cohort effect is that articles from the 1970s and 2000s get cited more than you'd expect given these age and period effects, articles from before the late-1960s get barely cited at all, and articles from 1980 through the mid-1990s also get cited considerably less than articles either side of that period. I'll offer some speculations at the end of the paper about the philosophical causes of, and consequences of, these cohort effects.

Elga, Adam. 2007. "Reflection and Disagreement." *Noûs* 41 (3): 478–502. https://doi.org/10.1111/j.1468-0068.2007.00656.x.

Ghitza, Yair, Andrew Gelman, and Jonathan Auerbach. 2023. "The Great Society, Reagan's Revolution, and Generations of Presidential Voting." *American Journal of Political Science* 67 (3): 520–37. https://doi.org/https://doi.org/10.1111/ajps.12713.

Keyes, Katherine M., Rebecca L. Utz, Whitney Robinson, and Guohua Li. 2010. "What Is a Cohort Effect? Comparison of Three Statistical Methods for Modeling Cohort

Effects in Obesity Prevalence in the United States, 1971–2006." *Social Science & Medicine* 70 (7): 1100–1108. https://doi.org/10.1016/j.socscimed.2009.12.018.

Thomson, Judith Jarvis. 1971. "A Defense of Abortion." *Philosophy and Public Affairs* 1 (1): 47–66.