General Part

Emil Andersson* and Nicolas Olsson Yaouzis

What Can Historicising Rawls Achieve?

https://doi.org/10.1515/auk-2024-2020

Abstract: This essay explores the implications of historicising John Rawls's theory of justice. While historical research on Rawls and his social context has provided valuable insights, some scholars argue that historicising carries significant philosophical consequences. This paper critically examines one such argument that contends that historicising Rawls's theory demonstrates its contextual nature, undermines its diagnostic powers, and leads to its complete dissolution. We offer a reconstruction of this argument and show that it fails. Further, while we argue that this argument fails, we go on to suggest that historical evidence may contribute to a defensive argument against appeals to expert opinion. By examining the appropriate relationship between historical research and philosophy, this essay contributes to the evaluation of the historicist critique and offers insights into the broader role of historical research in philosophical discourse.

Keywords: John Rawls; liberal egalitarianism; liberalism; historicism; social conditions; nihilism

1 Introduction

In recent years, there has been a growing interest in studying John Rawls's theory of justice and its influence on political philosophy from a historical perspective. This approach involves placing the theory within its social and political context and examining the factors that influenced Rawls's thinking and the development of his theory. The opening of the Rawls archive in 2004 has provided new insights into these subjects and new accounts of the theory's origins (Smith 2021; Smith, Bejan, and Zimmermann 2021).

***Corresponding author: Emil Andersson**, Department of Philosophy, Uppsala University, Uppsala, Sweden, E-mail: emil.andersson@filosofi.uu.se
Nicolas Olsson Yaouzis, Department of Philosophy, Stockholm University, Stockholm, Sweden, E-mail: niklas.olsson-yaouzis@philosophy.su.se

Open Access. © 2024 the author(s), published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.

There is no doubt that this work is of immense value from a history of ideas perspective. However, some scholars have gone further by arguing that this kind of historicising has far-reaching philosophical implications. In this paper, we consider one such argument put forth by Katrina Forrester (2019, 2022). By historicising Rawls's theory and demonstrating its contextual nature as "a product of its time" (Forrester 2019, 279), Forrester aims to establish not only the existence of "elements to be discarded" (Forrester 2019, 278) within the theory but also that it lacks "diagnostic powers", potentially resulting in "the dissolution of liberal egalitarianism altogether" (Forrester 2022, 19).

Unfortunately, and despite the attention her work has received, it remains unclear how the negative philosophical conclusions about Rawls's theory of justice are supposed to follow from historicising the theory. In this paper we aim to rationally reconstruct the historicist argument and provide different interpretations of the central concepts. The negative upshot of the reconstruction is that Forrester's historicist argument fails because its conclusion cannot be reached without accepting dubious philosophical assumptions.

However, everything is not doom and gloom for the historicist and we end by suggesting two philosophical uses for historical evidence. First, uncovering a philosopher's inspiration through archival findings may generate a new hypothesis about a theory's implications and presuppositions. Second, if one has presented such a hypothesis, historical evidence may contribute to a defensive argument to undermine appeals to expert opinion that seem to count against this suggested hypothesis. However, as our negative arguments show, in none of these cases can the historical findings alone establish the hypothesis.

2 Forrester's Argument

The argument we focus on in this paper is presented by Forrester in her book *In the Shadow of Justice* (Forrester 2019) and further developed in her subsequent paper 'Liberalism and Social Theory after John Rawls' (Forrester 2022). While her critique targets various normative theories interchangeably labelled as 'Rawlsianism', 'neo-Rawlsianism', and 'liberal egalitarianism', the primary focus remains on the view presented by John Rawls in *A Theory of Justice* (Rawls 1971) (hereafter referred to as '*Theory*'). It thus makes good sense, for the sake of precision and clarity, to substitute these various labels for 'Rawls's theory of justice'.

In a summary of the argument Forrester writes that she will

> develop a historicist critique of liberal egalitarianism, arguing that it (i.e. Rawls's theory of justice) provides a vision of social reality that is intimately connected to the historical and

> ideological constellation that I call postwar liberalism, and as such cannot account for social reality since the neoliberal revolutions of the late twentieth century. (Forrester 2022, 1)

To support her claims, Forrester employs two types of historical evidence. First, she utilises social and economic historical evidence to demonstrate that the social conditions of the 1950s, referred to as 'postwar liberalism', have been replaced by other social conditions, referred to as 'neoliberalism', which persist into the 2020s. That there have been extensive changes in social and political conditions during this time period is surely undeniable, and we will accept this part of Forrester's view as accurate.

Second, and of central importance to our evaluation, Forrester draws on archival evidence to establish 'an intimate connection' between postwar liberalism and Rawls's theory of justice. Whereas the first type of historical evidence is extracted from different accounts of social change, the second type of evidence is derived from Rawls's correspondence, notebooks, and lecture notes, shedding light on how his thinking and the development of his theory were influenced by the social and political conditions of his time. Ultimately, Forrester contends that due to the intimate connection between his theory of justice and postwar liberalism, and considering that postwar liberalism fails to accurately describe the social conditions of the 2020s, the theory 'cannot account for social reality' in the present.

Stated in this way, Forrester's argument is difficult to understand. The notions of a theory of justice being 'intimately connected' to some set of social conditions, and it failing to 'account for social reality', are far from clear. And although they are central to her argument, Forrester never explains what she has in mind. Throughout this paper we shall consider several different interpretations of these notions, in order to explore whether a forceful historicist argument can be uncovered.

When considering these different interpretations, it is essential to keep in mind what kind of conclusion that Forrester aims to establish. Though it is unclear what she means by the charge that Rawls's theory of justice 'cannot account for social reality', it is abundantly clear that she views this as a fatal shortcoming. Rather than viewing it as a problem that requires revising some minor parts of the theory, she views it as a problem that may lead to its complete dissolution.

With this in mind, let us quickly mention one possible interpretation of 'intimately connected', which it might be good to get out of the way. Forrester claims to provide a historicist criticism, which involves showing how the theory is a product of its time. This might be taken to suggest that the theory of justice is intimately connected to the social conditions of the 1950s in the sense that these conditions exercised an influence over Rawls when he formulated it. In other words, had Rawls found himself in a different political and social context, read other books, met other people, and been exposed to other ideas, his theory might have taken a different

form. On this *causal* reading of 'intimately connected', social and political events are causally connected to the mental events involved in formulating the theory of justice. If this is the correct understanding of 'intimately connected', then archival research plays a significant role in uncovering evidence, such as letters, notebooks, and early lecture notes that demonstrate the influence of the social conditions of the 1950s on Rawls's thinking.

But while this understanding of 'intimately connected' in one sense goes well with the historicist project, it could not be used to arrive at the kind of criticism of Rawls's theory of justice that Forrester aims to provide. Even if the social conditions of the 1950s were necessary for Rawls to develop his theory of justice, it is difficult to see how that fact could pose a serious problem for the theory as such. The merits of the theory must be determined by its qualities, not by its historical origins. Therefore, the causal interpretation cannot be used to establish the kind of conclusion that Forrester aims for.

We will have reason to return to this point later. But first, we shall proceed to consider some more interesting interpretations of Forrester's claims.

3 Reconstructing and Evaluating the Argument

We shall start out by offering a reconstruction of Forrester's argument. By explicitly stating the implicit premise that links her two claims with the conclusion, the argument can be stated as follows:

1. There is an intimate connection between Rawls's theory of justice and the social conditions of the 1950s.
2. By the 2020s, the social conditions of the 1950s had been replaced by other social conditions.
3. A normative theory that is intimately connected with one set of social conditions, S, cannot account for social reality in a society where these social conditions, S, have been replaced by other social conditions, S'.
4. Therefore, Rawls's theory of justice cannot account for social reality in the 2020s.

What makes it especially difficult to understand in what sense Rawls's theory cannot account for social reality is that it is a normative theory. After all, accounting for social reality sounds like a descriptive project – i.e., the project of offering a comprehensive description of the social and political conditions that obtain in some particular society, at some specific point in time. But the core of Rawls's theory of justice is normative, and consists of the following two principles of justice:

> *First principle*: Each person has an equal right to a fully adequate scheme of equal basic liberties which is compatible with a similar scheme of liberties for all.
>
> *Second principle*: Social and economic inequalities are to satisfy two conditions. First, they must be attached to offices and positions open to all under conditions of fair equality of opportunity; and second, they must be to the greatest benefit of the least advantaged members of society. (Rawls 2005, 291; see also 1971, 302)

These principles of justice are clearly not attempts at describing the social and political conditions of some particular society. Rather, they are normative principles intended to provide an account of what any society must be like in order to be just. Since they are not supposed to contribute to a purely descriptive project, it is difficult to see how they could fail to account for social reality in that sense.

3.1 'Cannot Account For' as Failing to Offer Adequate Political Recommendations

Forrester must thus have some other understanding of adequately accounting for social reality in mind. In order to see what it might be, let us return to the notion of a theory of justice being intimately connected to some set of social conditions. On an *analytic* reading of the connection, the relata are propositions rather than political and mental events. One way to establish such a connection is to view the theory of justice as consisting, not only of the two principles of justice, but also of some set of empirical claims about the social conditions of society.

This raises the further question of the role of these empirical claims, and how exactly they relate to the normative principles. On one possible account of this relation, the role of the empirical claims is to guide the application of the normative principles. After all, normative principles need to be combined with empirical claims in order to provide particular political recommendations. Recommendations, that is, concerning what kind of institutions and policies that ought to be adopted in society. For example, Rawls's two principles of justice cannot by themselves tell us how to organise the healthcare system in the U.S. in the 2020s. In order to derive a particular political recommendation about its organisation, empirical claims about what organisation best satisfies these principles in the U.S. in the 2020s are needed. The relationship can be illustrated in the following way:

> *General normative claim*: We ought to adopt the healthcare system that best satisfies the principles of justice.
>
> *Particular empirical claims about the social conditions*: In the U.S. in the 2020s, healthcare system X best satisfies the principles of justice.

Particular political recommendation: Thus, in the U.S. in the 2020s, healthcare system X should be adopted.

Consequently, a political philosopher who persists in using obsolete empirical claims from the postwar period may end up offering inadequate political recommendations in the 2020s. If the social conditions of the 2020s are sufficiently different from the conditions of the 1950s, it will simply not be the case that the recommended health care system satisfies the principles of justice.

This account of how normative principles relate to empirical claims about social conditions suggests how we can understand the idea that the theory cannot account for social reality. To say that a political theory cannot account for social reality in the 2020s is simply to say that the theory cannot provide adequate political recommendations in the 2020s. This fits quite well with Forrester's claim that "egalitarian liberals do not recognize that the very dynamics of capitalist societies they find objectionable may not be remediable by the institutional fixes they advocate" (Forrester 2022, 17, see also 18). This criticism seems not to be directed against the normative convictions of Rawlsians, but rather against the "institutional fixes" that they falsely believe that their theory recommends. They have simply failed, Forrester appears to suggest, to appreciate the fundamental changes in social conditions since the 1950s. And this results in them offering political recommendations that are in fact in conflict with their own normative principles.

However, this interpretation does not align well with Forrester's stated goal of initiating a discussion about the proper "relationship of normative theory to the study of social conditions within liberal political philosophy" (Forrester 2022, 2–3). Even if it were the case that Rawlsians are profoundly mistaken about what social conditions presently obtain, that would not raise new or interesting questions about the relationship between normative theory and the study of social conditions. After all, it is obvious that any set of normative principles must be combined with an adequate account of the social conditions we find ourselves in if we are to effectively satisfy their requirements. Thus, rather than providing reasons to rethink the relationship between normative theory and the study of social conditions, it merely calls for Rawlsians to update their understanding of current social conditions. This is no challenge to Rawls's principles of justice themselves and provides no reasons to believe that the complete dissolution of the theory is called for. Consequently, this interpretation renders the argument toothless.

3.2 'Cannot Account For' as Failing to be Applicable

If Forrester's argument is to deliver the intended kind of criticism of Rawls's theory, we need another account of how his principles of justice relate to empirical claims

about social conditions. It must be the case that the account of social reality presupposed by Rawls's theory of justice is not merely used to guide the application of the normative principles of the theory; instead, the principles themselves must somehow depend on this mistaken account. If this were the case it would be impossible to consistently affirm both the theory of justice and an accurate description of the social conditions of the 2020s. Since an adequate theory of justice should be compatible with our best understanding of our social conditions, so this reasoning would go, the normative component of Rawls's theory must be abandoned or drastically revised. This line of thought appears to align with Forrester's ambition of providing a criticism that calls for a complete dissolution of the theory.

One way of making sense of this kind of relation between normative principles and social conditions is by way of the idea of *applicability conditions*. As an example of this idea, consider that one of the main assumptions in Rawls's *Theory* is that we find ourselves in the circumstances of justice. These circumstances include moderate scarcity of resources. That is, there is not such an abundance of resources that cooperation becomes superfluous, nor is scarcity so severe that cooperation inevitably breaks down (Rawls 1971, 127). Hence, it can be argued that the theory in *Theory* presupposes the social condition of moderate scarcity in the sense that the principles of justice are only applicable in situations where this condition obtains.

On this reading of 'intimately connected,' the third premise becomes plausible: if a normative theory presupposes certain social conditions S that are incompatible with social conditions S', then the theory cannot account for a social reality characterised by S', in the sense that its principles are not applicable under conditions S'. If, for instance, Rawls's theory of justice presupposes the social condition of moderate scarcity, it becomes logically impossible to consistently affirm that the theory should be used to evaluate some particular society while denying that this society is characterised by moderate scarcity.

However, the fact that its principles are inapplicable under certain conditions does not show that Rawls's theory of justice is false. A theory may be correct although its applicability is limited to circumstances where conditions such as moderate scarcity holds. Nonetheless, showing that the two principles of justice are only applicable if the social conditions of the 1950s obtain would be a serious criticism of the view. Rawls's theory is surely intended to apply to our present conditions as well, and to provide us with an account of what our societies ought to be like in order to be just. Hence, it would be a significant problem if it turned out that while not false, the two principles are completely irrelevant for questions of justice in our present circumstances. In order to make the theory relevant, it would perhaps have to be revised to such an extent that the result could be described as its complete dissolution.

For this argument to work, however, we need an account of the applicability conditions of the theory, such that the two principles are inapplicable in the social conditions of the 2020s. It has to be shown that the theory presupposes something false; something other than moderate scarcity, as that obviously still holds. One possible account of such applicability conditions that might do the trick is one in terms of feasibility. That is, in order to be applicable, Rawls's political theory must be feasible.

Forrester claims that we want our political theories 'to be realistic', and then goes on to argue that Rawlsians are insufficiently realistic since they "have come erroneously to believe that the institutions their visions of justice defend and promote can satisfy their own commitments to justice and equality" (Forrester 2022, 17–18). This is, as we have already explained above, a complaint that in no way challenges the two principles of justice, which make up the normative core of Rawls's theory. But this appeal to realism could perhaps be understood as an appeal to a political version of the Kantian demand that 'ought implies can'. According to this understanding of realism, political theories ought to be action guiding, to be action guiding they must be feasible, and to be feasible they must offer recommendations that somehow could be accepted under current political, social and economic circumstances.

It could be suggested that an egalitarian view like Rawls's was feasible in the optimistic political climate of the 1950s, which was "characterized by a new consensus" (Forrester 2019, 277), and its "exceptional, unprecedented economic growth" (Forrester 2019, 180). In the 2020s, where economic growth has been stifled, inequality soared, and political partisanship rules supreme, Rawls's political recommendations are no longer realistically realisable, and therefore his theory is "unable to deal with the current political situation" (Forrester 2019, 277). Thus, it could be argued that Rawls's theory of justice is no longer feasible, and therefore not applicable.

There are several problems with this line of reasoning. One is that applicability conditions of this kind cannot be derived from Rawls's theory. While he was indeed concerned with feasibility, it was feasibility of a very different kind. According to Rawls, a theory of justice must be feasible in the sense of being "realistically utopian", where "the limits of the possible are not given by the actual" (Rawls 2001, 4–5). On this view, the recommendations of Rawls's theory of justice are utopian in the sense that they may be highly unlikely to be realisable in our present circumstances and given the current political climate. But these recommendations are still realistic, in the sense that they are possible under favourable conditions and given the fundamental and enduring facts about human societies. On this account of feasibility, the principles of justice are no less applicable today than they were in

the 1950s. Hence, it could not be reasonably claimed that it somehow follows from Rawls's theory itself that it is not applicable in the 2020s.

Another, and more serious, problem is that it appears highly implausible to hold that principles of justice are only applicable in circumstances where the political climate and public opinion make their recommendations feasible. If we have some set of principles of justice that apply to us, then it seems strange to suggest that we could free ourselves from their demands simply by making the social conditions such that it becomes unrealistic to expect justice to be realised. If the principles truly apply, then we cannot escape their demands by making our society less just, or by making people not believe in the principles. Though some feasibility constraints may make sense, this one does not.

There is, regardless of what one thinks about the plausibility of this suggestion, a more fundamental problem with it. But before explaining that problem, we shall consider yet another possible interpretation of Forrester's argument.

3.3 'Cannot Account For' as Failing to be Justified

There is yet another way of understanding the idea that Rawls's principles of justice depend on an outdated and hence mistaken account of social conditions. Forrester claims that "the possibility of consensus was assumed at a very deep level" (Forrester 2022, 9) in Rawls's theory; that it presupposed "that deep down, social life rested on the possibility of consensus and ethical agreement" (Forrester 2019, xx). Though she does not specify what 'deep level' she has in mind, this could be understood as referring to the underlying justification of the principles of justice. That is, to the original position.

The primary justification of Rawls's two principles of justice is that they would be selected by the parties in the hypothetical original position. Consensus features in a fundamental way in this thought experiment, as the resulting principles are justified by being agreed to by all the parties (Rawls 1971, 17–21). Since the justification of the two principles of justice depends on a certain kind of agreement, it could plausibly be claimed that Rawls's theory presupposes the possibility of consensus 'at a very deep level'.

Further, one could reason that the rationale for building agreement into the original position in this way is drawn from a certain account of social conditions. In the optimistic circumstances of the 1950s, consensus on principles of justice seemed a genuine possibility. The principles of justice are therefore intimately connected to the social conditions of the 1950s in the sense that their justification depends on empirical assumptions regarding the possibility of widespread agreement that seemed plausible at that time.

It could be further argued that what appeared possible in the 1950s is very different from what appears possible given the social conditions of the 2020s. The political and social developments leading up to our present day have shown us that the faith in the possibility of consensus was misguided. Some of the assumptions that underlie the original position are thus simply false. Rawls's theory of justice cannot, on this view, account for social reality in the 2020s in the sense that the justification of the two principles rely on assumptions that we now know to be mistaken. Hence, Rawls's justification of his two principles of justice fails.

This line of reasoning is unpersuasive since it fundamentally misconstrues why the original position incorporates agreement. Nowhere in *Theory* (or in Rawls's later writings) can we find the claim that the role of agreement in the original position depends on the possibility of consensus in the sense described above. The claim that the correct principles of justice are those that would be agreed to in the hypothetical original position is not premised on the assumption that such agreement can be realised, but rather on moral assumptions. The idea is that if certain principles would be chosen, then "it will be true that whenever social institutions satisfy these principles those engaged in them can say to one another that they are cooperating on terms to which they would agree if they were free and equal persons whose relations with respect to one another were fair" (Rawls 1971, 13). The original position models citizens as free and equal, and the principles they could agree to as such, under conditions that are fair, are the correct principles of justice. The rationale for caring about what would be agreed to in the original position is thus not that agreement is actually possible, but rather that it shows us what principles of justice that follow from certain moral assumptions – in particular, the account of citizens as being free and equal.

But an even more fundamental problem, for this suggestion as well as the one in Section 3.2 above, emerges once we remind ourselves of what kind of argument it is that Forrester claims to provide. The argument is supposed to be a historicist one; it depends on archival evidence and is intended to reveal the intimate connection between Rawls's theory of justice and the social conditions of the 1950s. However, the type of analytical connection between propositions required for the argument to work simply cannot be established through such evidence.

To further elaborate this point, let us distinguish between two different ways of construing the versions of Forrester's argument considered in this and the previous section. On the first construal, it is claimed that Rawls's theory incorporates certain philosophical claims, which establishes an analytical connection between social conditions and the two principles of justice. As mentioned earlier, archival evidence provides insights into causal relations between events, such as social or political events influencing Rawls's development of *Theory*. However, no number of archival insights into Rawls's influences, inspirations, or mental states can establish

that his theory presupposes some set of propositions. It makes no difference if, say, the optimistic political and social environment of the 1950s played a causal role in shaping how Rawls thought about the original position. What matters is the content of the theory he came up with, not what made him think in some way or another. Hence, establishing a claim about the analytical relation between social conditions and the principles of justice requires philosophical engagement with the content of Rawls's theory. The kind of philosophical analysis that we find, for instance, in G. A. Cohen's critical analysis of how the Rawlsian methodology makes principles of justice dependent on certain empirical facts (Cohen 2008).

On a second understanding of these arguments, the claim is not about what is already present in Rawls's theory. Instead, it is claimed that we ought to think about applicability conditions and the justification of principles of justice in certain ways. It could, for instance, be argued that we have good reasons to reject the account of applicability conditions that we find in Rawls's theory. It is a philosophical mistake, according to this line of argument, to think that principles of justice are applicable in social conditions where they fail to provide political recommendations that are realistically feasible.

Though we have already suggested that applicability conditions of this kind are philosophically implausible, one could certainly try to argue otherwise. But that would, once again, require doing philosophy. No amount of historical evidence could possibly establish philosophical claims of this kind. This route thus involves going way beyond the historicist project of exposing the social and political factors that influenced the development of certain philosophical theories.

4 Alternate Uses of Historical Evidence

Our critical reconstruction and analysis of Forrester's argument has demonstrated the limits of the historicist project. But while the kind of historical evidence that Forrester employs cannot be used to establish the critical philosophical conclusions that she aims for, it is clear that such evidence can be of use in other ways. We shall therefore end by briefly considering two uses of this kind.

One fairly obvious use for historical evidence is that it may provide reasons to reevaluate our previous philosophical analyses. For instance, discovering a note in the Rawls archives stating that his theory presupposes a set of propositions describing the social conditions of the 1950s would prompt a fresh examination of it. Similarly, in the unlikely event that we were to uncover a letter revealing that Rawls was a high-ranking member of the KKK, that would motivate an examination of whether his theory presupposes any racist propositions. Lastly, encountering an argument in the lecture notes that aims to establish an intimate connection between the

theory and specific social conditions would call for philosophical analysis to assess the soundness of the argument. However, in all these cases, the archival evidence plays merely an indirect role, providing reasons to engage in philosophical analysis.

This indirect role for archival evidence might be sufficient on its own to motivate the historicist project, and there are interesting questions to be asked regarding the relative weight we should assign different forms of evidence. But we shall now put this role for historical evidence aside. We will instead proceed to show that recent debates in metaethics suggest another, much less obvious, use for it. This may be of interest to those who, as Forrester evidently does, find certain kinds of critical theory appealing.[1]

An alternative, dialectic approach to historical evidence can be learned from moral nihilists who suggest that evolutionary evidence can be used as part of a defensive argument against moral realists who appeal to widely held intuitions. For example, Jonas Olson (2014, 146) concedes that although evolutionary debunking arguments cannot by themselves be used to establish moral nihilism, they may still serve a dialectic purpose. He appeals to metaphysical intuitions to support moral nihilism but admits that it is unlikely that it will convince moral realists who may appeal to widely held intuitions that there are moral facts. There will thus be a clash of intuitions, and a resulting standstill. It is at this point that evolutionary evidence can serve a dialectic purpose. Olson suggests that evolutionary debunking arguments may be used to argue against the realist by severing the evidential link between the widely held intuitions and the alleged moral facts. Hence, the debunking arguments can break the standstill not by targeting the content of the intuitions but rather their evidential value.

To further explain this strategy, and simultaneously show how it could be used by the historicist, assume the following situation. The historicist engages in philosophical analysis and presents a philosophical argument indicating that Rawls's theory of justice presupposes a set of propositions describing the social conditions of the 1950s. However, since philosophical arguments typically involve contested premises, acceptance of the argument relies on intuitions and prior philosophical commitments. This may lead to a standstill between the historicist and the Rawlsians, where their respective philosophical stances prevent them from reaching consensus.

To break the standstill, the Rawlsian may appeal to less contested evidence. For instance, she may argue that if the theory genuinely presupposed the social conditions of the 1950s, Rawls would have acknowledged it. However, Rawls explicitly stated that the theory was non-historical and not limited to a specific society

1 For more on how the reasoning below connects to ideas in critical theory and Marxism, see the related argument developed in Olsson Yaouzis 2018.

at a particular time in history (Rawls 1971, 139; 2005, 273). Moreover, other political philosophers who have thoroughly examined the presuppositions and entailments of the theory share this view. The Rawlsian may therefore conclude that the historicist argument should be rejected because the best explanation of this consensus is that Rawls's theory of justice does not presuppose the social conditions of the 1950s.

At this juncture, the historicist can leverage the archival evidence to counter the Rawlsian's appeal to expert opinion. Although the archival evidence cannot establish what the theory presupposes or entails, it may be used to undermine the Rawlsian's contention that Rawls's and the other experts' statements accurately track the theory's presuppositions.

There are three types of evidence that could do so. Firstly, we might discover that the claims are *insincere* and do not reflect the experts' beliefs but instead align with, say, the wishes of the CIA. Secondly, we could find that the claims are *sincere* but that the beliefs are *unreliable* due to factors such as brainwashing by the KGB. In both cases, the connection between the experts' claim and its truth would be severed, weakening the Rawlsian's appeal to experts' claims.

Alternatively, and more realistically, the historian may uncover information in the archives that show that while the experts' claims are sincere and their beliefs reliably formed, they represent a minority position or a special interest. For example, archival evidence may reveal mechanisms of the kind outlined by Charles W. Mills (2005, 172). Mechanisms, that is, that ensure that reliable and sincere feminists, Marxists, and critical theorists who do not share Rawls's beliefs about the presuppositions of his theory tend not to become tenured and influential professors in political philosophy. While the archives might uncover evidence of a conscious conspiracy upholding these mechanisms, they could also reveal that the mechanisms are byproducts of other behaviour. For instance, the archival research might show that (1) high-ranking philosophy departments were dominated by white, middle-to-upper class men, (2) dissenting graduate students in political philosophy tended to be either non-white, working class, or non-male, and (3) non-white, working class, and non-male graduate students left political philosophy due to their minority status. Although these findings would not sever the connection between an individual expert's claims and the truth of those claims, they would sever the connection between the consensus claim and its truth. In other words, the archival evidence would suggest that the best explanation for the consensus is not that it is correct, but rather that all dissenters have left the profession due to their minority status.

For our purposes, there is no need to assess how likely it is that the kind of evidence described above will be discovered. Finding such evidence is, after all, not the task of philosophers. We only suggest that drawing on Olson's recent work on moral nihilism reveals an additional role for archival evidence, which adds to its more obvious indirect significance. While it still cannot independently yield philosophical conclusions, it can play a role within a defensive argument.

5 Conclusions

In conclusion, we have argued that Forrester's historicist criticism of Rawls's theory of justice fails. Although there could perhaps be a philosophically argued case for thinking that the theory presupposes a faulty account of social conditions, such a case would have to rest on philosophical analysis rather than archival evidence. No amount of historicising will show that the complete dissolution of Rawls's theory is called for, since only philosophical analysis can establish such a claim.

However, we have also suggested that historical evidence may serve other purposes. In addition to providing reasons for philosophical re-evaluation of theories and ideas, it may indirectly contribute as part of a defensive argument against Rawlsians who appeal to expert opinions. This is not a belittlement of the work conducted by researchers studying the origins and development of philosophical theories, but rather a reminder of the limitations inherent in such research.

Acknowledgments: We would like to thank Jens Johansson and Jonas Olson for helpful comments on an earlier draft of this paper. Thanks also to the editors of *Analyse & Kritik* for their valuable objections and suggestions. Emil Andersson gratefully acknowledges the generous financial support from the Swedish Research Council (grant 2020-06502). Nicolas Olsson Yaouzis similarly acknowledges the generous funding from the Swedish Research Council (grant 2019-02828).

References

Cohen, G. A. 2008. *Rescuing Justice and Equality*. Cambridge: Harvard University Press.

Forrester, Katrina. 2019. *In the Shadow of Justice: Postwar Liberalism and the Remaking of Political Philosophy*. Princeton: Princeton University Press.

Forrester, Katrina. 2022. "Liberalism and Social Theory After John Rawls." *Analyse & Kritik* 44 (1): 1—22.

Mills, Charles W. 2005. "'Ideal Theory' as Ideology." *Hypatia* 20 (3): 165—84.

Olson, Jonas. 2014. *Moral Error Theory: History, Critique, Defence*. Oxford: Oxford University Press.

Olsson Yaouzis, Nicolas. 2018. "'That is Just What They Want You to Believe': A Modest Defence of Marxist Paranoia." *European Journal of Philosophy*: 827—39. https://doi.org/10.1111/ejop.12335.

Rawls, John. 1971. *A Theory of Justice*. Cambridge: Harvard University Press.

Rawls, John. 2001. *Justice as Fairness: A Restatement*, edited by Erin Kelly. Cambridge: Harvard University Press.

Rawls, John. 2005. *Political Liberalism (Expanded Edition)*. New York: Columbia University Press.

Smith, Sophie. 2021. "Historicizing Rawls." *Modern Intellectual History* 18 (4): 906—39.

Smith, Sophie, Teresa M. Bejan, and Annette Zimmermann. 2021. "The Historical Rawls: Introduction." *Modern Intellectual History* 18 (4): 899—905.