

**Rerum Causae**  
**Volume X, Issue I**  
**2018**

**Foreword:**

We gladly present the tenth edition of *Rerum Causae*, the student journal from the Department of Philosophy, Logic and Scientific Method of the London School of Economics and Political Science. This issue contains nine works by students enrolled in our undergraduate and postgraduate degrees.

This issue accurately resembles the research focus of our department, since the majority of papers published here are either in the field of philosophy of science or in moral and political philosophy. The issue opens with Malvina Ongaro examining John Stuart Mill's proposal of a weighted-voting system and discusses under which conditions this system can be justified. Francesca Bellazi then defends the essentialist view on defining biological species. Martin Vaeth discusses the relationship between internal and external validity and the possibility of a trade-off between them. Silvana Maestro argues against Ludwig Fahrback's position that theory change has ended and that the growth of science has been exponential. We have a break from other topics via a paper in philosophy of mind. Niccolò Aimone Pisano presents a modified version of Searle's Chinese Room thought experiment and argues that this version is better equipped to attack the computational theory of mind. Ragini Dua analyses John Taurek's position in the debate about the duty to save the greater number and, while finding Taurek's argument unsuccessful, concludes that his view challenges our common intuitions about this duty. The issue also includes two papers on an important topic in political philosophy, namely, freedom of movement and migration. Lewis Williams proposes a relational egalitarian defence of the right to migrate. Liam Fields presents a Rawlsian argument for states' right to

deny entry to economic migrants. We close with a familiar problem in philosophy. Julius Schwarzwälder discusses Moore's proof of an external world and finds Moore's argument unsatisfactory, but he does not argue that we should think that the external world does not exist – there are strong pragmatic reasons to believe otherwise

We hope that these essays demonstrate the range of interests of our students and reflect the approach of our department – to do philosophy which is continuous with the sciences and which is socially relevant. Furthermore, we hope that the reader will find works published in this issue insightful and enjoyable.

**The Editorial Team:**

Dmitry Ananyev

Robbie Carver

Emma Jane Curran

Morgan Fairless

Rory Gillis

Rafae Qazi

Laura Schmadl

Sumer Singh

Cheryl Ting

Simon van Teutem

Harry Ward

Shike Zhou

**Acknowledgements:**

We would like to thank the Department of Philosophy, Logic and Scientific Method for its continued support in the publication of this Journal. In particular, we thank Dr Johanna Thoma and Ewan Rodgers for their assistance and advice over this academic year. We also thank the LSESU Philosophy Society for funding this issue.

**Contents:**

**Malvina Ongaro – p.7**

J.S. Mill's plural voting system and the epistemic defence of democracy

**Francesca Bellazzi – p.17**

In sustain of M. Devitt's "Resurrecting Biological Essentialism"

**Martin Vaeth – p.27**

Can there be a trade-off between internal and external validity?

**Lewis Williams – p.35**

Border Control and Relational Egalitarianism

**Ragini Dua – p.44**

How convincing a case does Taurek make that one has no duty to save the greater number?

**Niccolò Aimone Pisano – p.50**

Searle's Chinese Room Reconsidered

**Silvina Maestro – p.60**

Is Fahrback right in thinking theory change is coming to an end?

**Liam Fields – p.71**

Denied Entry: a Rawlsian Defence of the Moral Right of the State to Close its Borders

**Julius Schwarzwälder – p.78**

The Practicality of an External World

## **J. S. Mill' s plural voting system and the epistemic defence of democracy**

By Malvina Ongaro

---

**Abstract** – In this paper I explore the possibility of making a case for plural voting. I start with observing a tension between democratic participation and the epistemic quality of results in collective decisions of a technical nature. The search for epistemic quality may render plural voting appealing, but I argue that this case fails. First, I reconstruct Mill's argument for plural voting, and I show that it rests on the same grounds on which the epistemic defence of democracy is based. Second, I try to show that, if one were to adopt an epistemic perspective, then plural voting should not only be acceptable, but even advocated for. However, I suggest that such an epistemic perspective cannot be used to justify democracy whenever there is an external and independently legitimised epistemic authority. In these cases, either democracy is disposed of, or its legitimacy must be justified on non-epistemic grounds. I conclude exploring a possible non-epistemic argument, and contending that such an argument would be incompatible with plural voting.

---

### • 1. Introduction

On April, 17<sup>th</sup> 2016 Italian citizens were called to express their opinion in a referendum about the opportunity of prolonging gas and oil drilling concessions to extract hydrocarbon within 12 nautical miles of the Italian coast until exhaustion of the sources. The referendum sparked a heated debate on broad environmental values and on free market. However, points like the protection of the environment were defended by both fronts, and campaigners on both sides were persuaded that their position was the one to properly implement them. Thus, people

sharing the same view would end up in opposite fronts. In the end, the turnout was so low (31%) that the result was not validated<sup>1</sup>, and the referendum proved to be only a waste of public resources.

What went wrong? The confusion was partly due to the highly technical nature of the matter at hand<sup>2</sup>: the consequences of either choice were hard to identify for anybody without a proper understanding of engineering and international trade. Therefore, it seems that there are matters on which appealing indiscriminately to the whole population is not the right thing to do – matters that need competence to be settled. Furthermore, there are matters of fact that are either true or false, and on which there seems to be no point in asking the opinion of the majority. If we do not vote on gravity, why should we vote on climate change? And yet, shouldn't a democracy guarantee that all citizens have equal voice on public matters?

Some philosophers have argued that this is not necessarily the case, and that epistemic accuracy needs not be in tension with democracy. Electoral democracy may come in different forms, including systems where influence is unevenly distributed. In his treatises *On Liberty* (1859) and *Considerations on Representative Government* (1861), John Stuart Mill defends an implementation of democracy characterised by a weighted-voting system of *plural voting* (PV henceforth). In such a system, while every citizen (within the scope of suffrage<sup>3</sup>) votes, and thus participates in public decision-making, those with higher competence will have more votes to express, and thus a higher influence.

---

<sup>1</sup> Ministero dell'Interno: [Archivio storico delle elezioni](#).

<sup>2</sup> See Cozzi (2016): “(The) question was too technical to be subject to a referendum, above all because it would hardly been fully understood by those citizens who should have voted”.

<sup>3</sup> Which, for Mill, would include women but still exclude illiterates, bankrupts, and non-tax-payers: see Mill (1861).

A proposal on these lines seems to go against the natural intuition that democracy recognises the equality of all citizens. However, in light of cases like the Italian referendum or climate change denial, there seems to be an appealing case for a system that favours competence. In what follows, I will try to show that the validity of weighted-voting systems depends on the justification adduced for democracy as a legitimate source of political authority. If the case for democracy rests on an epistemic argument that looks at the quality of the decisions it reaches, then weighted voting should not only be valid, but even advocated for. Nevertheless, I will suggest that epistemic arguments cannot justify democracy when there are external epistemic authorities, and that therefore democracy will need some non-epistemic justification at least in those cases. However, I will conclude that, if the legitimacy of democracy rests on non-epistemic grounds, then Mill's case for PV seems to fail.

## • 2. The Epistemic Argument

According to Mill, the value of a form of government is not absolute: it is relative to the consequences it produces within a given society<sup>4</sup>. In his utilitarian view, democracy – as any other system – is not a good thing *per se*, but it is as good as its consequences. Therefore, if we can improve the consequences, then we have a better form of government.

With this criterion in mind, Mill proposes PV as the best democratic system. In his view, PV would improve the consequences of democratic decision-making by (i) avoiding class-legislation, and by (ii) assigning

---

<sup>4</sup> “The ideally best form of government is [...] the one which [...] is attended with the greatest amount of beneficial consequences, immediate and prospective” (Mill (1861), p. 404).

influence proportional to competence and intelligence (Mill 1861).

The need to avoid class-legislation can be seen as a circumstantial claim, that would make PV a temporal solution given the uniform composition of the social classes of the times (as affirmed, e.g., by Brillhante & Roca (2013)). But the call for unequal influence is proposed as a general reason in support of PV. In Mill's words, "that the constitution of a country should declare ignorance to be entitled to as much political power as knowledge" (1861, p. 478) is not only useless, but even hurtful, and this regardless of the composition of society. Let us see on which basis Mill values unequal influence.

Since voting is a tool for decision-making, its main consequence is the outcome of that decision. Therefore, from a utilitarian perspective, a voting system is better than another if it leads to better decisions.

Often, democracy is defended precisely on these grounds. The epistemic defence of democracy (EDD henceforth) claims that democracy is better than other forms of decision-making because it is more reliable in making the right choice. If there is a correct decision, democracy is justified to the extent that it tracks it:

"For epistemic democrats, the aim of democracy is to 'track the truth.' For them, democracy is more desirable than alternative forms of decision-making because, and insofar as, it does that." (List and Goodin (2001), p. 1).

EDD justifies democracy over other forms of government, but there is no reason why a similar standard should not apply also to different voting systems within democracy. Let us call this standard EDD':

EDD': a democratic voting system is more desirable than alternative ones because, and insofar as, it tracks the truth.

It is with something like EDD' in mind that Mill proposes PV. PV is supposed to be a better truth-tracker than other democratic rules, and since he appeals to proportional influence to argue for it, then proportionality of influence must bring some epistemic advantage. So, to complete Mill's argument we need a reason for the claim that assigning greater influence to competence improves the accuracy of decisions. Even though Mill does not provide one himself, Condorcet's (1785) Jury Theorem (CJT henceforth) can be used to support his conclusion.

In CJT, the probability that a group reaches a correct decision increases with the size of the group – as long as every member is sufficiently competent, i.e. has a probability higher than chance of being right. Moreover, the higher the competence, the higher the probability to reach a correct decision. Therefore, if the quality of the final decision is the relevant criterion, then democratic procedures should value competence, and PV is a way to implement this result. To make the argument clearer:

- i. EDD': A democratic voting system is better than other ones if it produces more correct outcomes.
- ii. CJT: The higher the competence involved, the likelier the correctness of the outcome.
- iii. By definition, PV assigns more influence to competence.

---

Therefore, *ceteris paribus*, PV democracy is better than non-PV democracy.

Note that the only assumptions needed are CJT and EDD'. For this reason, it appears that, if the legitimacy of democracy is argued only in terms of the quality of the outcomes it produces, then weighted-voting systems like PV should not only be accepted as legitimate forms of democracy, but also advocated for as more reliable truth-trackers.

- 3. The Objection from External Authorities

It appears that, from an EDD perspective, the legitimacy of democracy hinges on its relation with some external, procedure-independent truth. A democratic decision is a legitimate source of political authority insofar as it is more likely to track that truth than a decision issued by another procedure. This means that, if democracy is the most reliable truth-tracker over a matter of fact *d*, then it is also a source of epistemic authority<sup>5</sup> on *d*.

Yet, it may often be the case that there exists an alternative epistemic authority on that matter. This would seem to be particularly plausible if *d* concerned scientific issues like climate change, or technicalities like drilling concessions. In such cases, there could be a different and domain-specific form of decision-making, which would possibly be a better truth-tracker than democracy.

If that is so, then on the basis of EDD democracy would be dispensable. If what legitimises democracy is its epistemic value, then whenever there is a third-party authority with better epistemic reliability democracy is redundant, as that authority is a better truth-tracker. Under EDD, such independently legitimised epistemic authority should also count as a source of political authority on *d*.

Therefore, it seems that EDD can justify democracy only in those cases where there is no better epistemic authority. As soon as there is one, democracy cannot be legitimised by EDD. Which means that, in these cases, either we dispense of democracy, or we legitimise it on non-epistemic grounds. In the next section, I will explore a non-epistemic argument for democracy, and assess its import for the legitimacy of Mill's

---

<sup>5</sup> For a definition of political and epistemic authority, see Peter (2016).

PV.

- 4. Equal Opportunity

If we want to legitimise democracy even in presence of an external epistemic authority, then the legitimising argument must be non-epistemic. One such argument could run on the following lines.

Most real-life policy decisions involve value judgements. It seems reasonable to assume that at least some, if not all, of these judgements are such that they do not have a truth value. If a judgement does not have a truth value, then a procedure to decide on it cannot be justified epistemically, since it is hard to determine which procedure is a better truth-tracker with respect to a judgement that does not have a truth value. Moreover, the population may disagree not only on the judgements, but also on the identification of any relevant expert for the matter (Christiano 2008). Therefore, it seems that divergences in value judgements may be irreducible – either because there is widespread disagreement that cannot be settled by any recognisable moral authority, or even, as Berlin (1958) claims, because values are intrinsically plural. Yet, individual divergences still need to be settled collectively.

In absence of some externally legitimated authority, citizens themselves should be considered equal authorities over values, and therefore over the evaluation of political decisions. For this reason, individuals should be given an equal opportunity of participation, as this equality is the only way for them to avoid being governed with laws justified on values that they do not recognise, and that therefore are illegitimate. Democracy is justified in virtue of it being the only form of government that recognises equal agency to citizens.

It is clear that a non-epistemic argument constructed on these lines

cannot support a weighted-voting system as the one proposed by Mill. PV is grounded precisely on the fact that better outcomes come from an unequal distribution of opportunities to influence political agency. Therefore, if we justify democracy to the extent that it guarantees equal opportunity of influence, then weighted-voting systems like PV are illegitimate

One could accept that equal participation is required for value judgements, and that external authorities may be epistemically more reliable than voting to deal with empirical judgements. But then one could argue that, since most policy decisions involve a mixture of the two, a procedure like PV could provide a good compromise between epistemic accuracy and equal treatment. However, an argument on these lines is not purely epistemic anymore. It is not concerned uniquely with the quality of the outcomes, as Mill's argument is, but it accepts the relevance of other, purely procedural parameters.

## • 5. Conclusions

Mill's proposal of assigning more votes to more competent citizens is justified within a general epistemic defence of democracy. If quality of the outcomes is the standard by which forms of government are assessed, then CJT shows not only the legitimacy of democracy, but also that of PV as a democratic system. However, EDD can legitimise democracy only in the absence of an external epistemic authority. In any other case, democracy needs to be legitimised with another argument. This could be based on value-laden decisions: in the lack of legitimate moral authority, citizens should enjoy equal opportunity to participate in political agency. But a justification for democracy grounded on the value of equal participation, by definition, does not justify systems based on unequal influence like Mill's PV.

As a side speculation, this does not mean that democracy is always the best decision system. Purely technical and scientific issues, like the limits of drilling concessions or climate change, should be resolved by experts, without wasting public resources to democratically deliberate over matters on which the threshold of minimum competence required by CJT is unlikely to be reached. A more promising way to conciliate epistemic and democratic values could be to delineate the respective spheres of competence and identify the right contexts of applications of different decision-making procedures, rather than looking for the perfect blend in a single voting system.

#### REFERENCES

- Berlin, I. (1958). Two Concepts of Liberty. In Hardy H. (ed.), (2002). *Liberty*. Oxford: Oxford University Press.
- Brilhante, Á. A., & Rocha, F. J. S. (2013). Democracy and Plural Voting in John Stuart Mill's Political Thought. *Ethica@*, 12(1), 53-65.
- Cerovac, I. (2016). Plural Voting and JS Mill's Account of Democratic Legitimacy. *Croatian Journal of Philosophy*, 16(1 (46)), 91-106.
- Christiano, T. (2008). *The constitution of equality: Democratic authority and its limits*. Oxford: Oxford University Press.
- Condorcet, N.C. (1785) *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris: Imprimerie Royale.
- Cozzi, A. (2016): [Italy, the \(imperfect\) oil referendum](#). In *Mediterranean Affairs*, 9/05/2016.
- Estlund, D. M. (2009). *Democratic authority: A philosophical framework*.

Princeton: Princeton University Press.

List, C., & Goodin, R. E. (2001). Epistemic democracy: generalizing the Condorcet jury theorem. *Journal of Political Philosophy*, 9(3), 277-306.

Mill, J. S. (1859). On Liberty. In Gray, J. (ed.), (1998). *On Liberty and Other Essays*. Oxford: Oxford University Press.

Mill, J. S. (1861). Considerations on Representative Government. In Robson, J. M. (ed.), *The Collected Works of John Stuart Mill, Volume XIX – Essays on Politics and Society Part II*. Toronto: University of Toronto Press.

Peter, F. (2016). The epistemic circumstances of democracy. In Fricker, M., & Brady, M. (ed.), *The Epistemic Life of Groups*, 133-149.

Peter, F. (2017). Political Legitimacy. In Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition).

**In sustain of M. Devitt’s “Resurrecting Biological Essentialism”**

By Francesca Bellazzi

**Abstract**

In this essay, I analyse Devitt’s paper’s “Resurrecting Biological Essentialism” (2008) to show that essentialism is a proper way to address the problem of the definition of species. In particular, I consider Devitt’s thesis: *Linnean taxa* have essences that are, at least partially, underlying intrinsic essential properties. These essential properties are largely, but not entirely genetic and determine the belonging to a species. This “resurrection” of biological essentialism is proposed in relation to the insufficiency of the definition of species in current philosophy of biology. The major objection to essentialism challenges the notion of essence in view of variation and change and evolutionary theory. I propose two lines of response to this objection: the first concerning the concept of evolution, the second the concept of essential properties. In conclusion, I state that Devitt’s essentialism stands up to scrutiny and that the essentialism account is an interesting position for further research about the differences in species.

**Introduction**

The debate about what makes an organism member of a species has a long history in philosophy and represents a central theme in the philosophy of biology. In particular, contemporary authors have abandoned the classic essentialist conception of species because it is deemed as incompatible with the Darwinian theory of evolution. However, Devitt (2008) suggests the adoption of an essentialist conception in the definition of species.

In this essay, I will consider Devitt’s article and thesis. Then, I will expose the objection of variation and change as the significant critique to essentialism (Sober 1980, Griffith 2002). I will try to give a preliminary response to this objection in two ways. In conclusion, I will present that the shortcomings of essentialism do not undermine it, but represent a stimulus for future studies.

### Devitt's argument

*Essentialism about species is today a dead issue.*

*(Sober 1980, p.353)*

Essentialism has already been dead for twenty-eight years when Devitt resurrected it. In his paper, *Resurrecting biological essentialism* (2008) he aims at reintroducing essences in relation to the definition of biological species in order to give a structural explanation of biological generalisations. In particular, his concern is to identify the properties in virtue of which an organism is a member of a species. He argues that essential properties have this role. More precisely, *Linnean taxa*<sup>6</sup> have essences that are, at least partially, underlying intrinsic essential properties; these essential properties are mostly, but not entirely genetic, and determine the belonging to a species rather than another. His view, and in general the essentialist position, has been criticised by many authors (Sober 1980, Richards 2010, Wilson *et al.* 2007). The most substantial objection to essentialism is based on evolutionary theory and change and variations in species. Indeed, evolution challenges the idea of an essence that remains stable within time and define the species.

Let us now consider more in detail Devitt's argument. Firstly, he separates the *category question* from the *taxon question* about the definition of species (Devitt 2008, 2010; Ereshefsky 2010, 2017). These can be summarised as follows:

- The category question concerns the definition of the species category, i.e. what is for a group or kind to be a species. It gives the *ratio cognoscendi* or *classificandi* of the species, that is, respectively, why we *know* that *Xenopus laevis* is a species and why we *classify* groups of frogs as species.
- The taxon question addresses the properties in virtue of which an organism belongs to a particular species, i.e. “what is for an organism to be a member of a particular kind” (Devitt 2008, p.334). It gives the *ratio essendi* of being a member of a species, that is why a *Xenopus laevis*

---

<sup>6</sup> *Linnaean taxa* are biological kinds that reflect the biological categorisations of Carl N. Linnaeus' hierarchy (1735-58): domains, kingdoms, phyla, classes, orders, families, genera, species, and subspecies.

*is* a member of the species *Xenopus laevis* rather than, e.g. *Xenopus tropicalis*.

Devitt believes that the current positions in philosophy of biology about species are answers to the category question rather than to the taxon question. These positions can be summarised as the following:

- Phenetic concepts theory: species are based on overall similarities between phenotypic traits. This position is now generally considered naïve.
- Biological species concepts theory (BSC): species are groups of interbreeding natural populations that are reproductively isolated from other such groups.
- Ecological niche concepts theory (ENC): species occupy a certain ecological niche, and their members exploit the same set of environmental resources and habitats.
- Phylogenetic-Cladistic concepts theory (P-CC): species are identified regarding evolutionary history.

However, none of them seems to explain why, for example, a horse is a horse rather than a cat. Indeed, these positions are able to give a definition of the *category* species, i.e. they are able to define the category species, but they do not provide any further structural explanation on its members. As far BSC as is concerned, it is possible to identify the *category species* because it is a group of interbreeding natural populations. However, this does not give us any explanation of why a member of a species interbreeds just with a member of the same species. In consequence, Devitt claims that essentialism is the only position that can answer properly to the *taxon* question. In fact, essentialism individuates the structures thanks to which an organism is a member of a particular species. He presents his intrinsic essentialism in the following way. At first, he individuates intrinsic properties as the ones that are possessed by an entity independently from outside forces. Then, he defines essential properties as those intrinsic properties that characterise the entity as a member of the species. In *Linnean Taxa*, intrinsic essential properties are at least partially

genetic. In conclusion, he states that an individual is a member of a particular species, partially, in virtue of such essential genetic makeup that it possesses. It is important to notice that his essentialism is not committed to a fully intrinsic essence position. Devitt's essentialism wants to underline that there are common shared genetic properties between members of the same species and those, with others such as a specific relation with the environment, can explain, i.e., why a horse is a horse rather than a cat. Essential properties are explanatory structures not used to obtain the definition of the individual itself, but to the definition of the individual as a member of a species.

Moreover, this kind of essentialism is interesting because it is a *taxon-answer* compatible with different *categories-answers*. Indeed, it is compatible with BSC: species are groups of interbreeding natural populations that are reproductively isolated from other such groups because members have the same essential properties. The same holds for with ENC: members share the same ecological niche because members share the same essential properties. As P-CC is concerned, essentialism can be in agreement also with this position. In fact, essentialism is not committed to a fully intrinsic essence, and the interaction with the environment is acceptable.

### **Variation and change**

Devitt's argument is interesting and well explained. Indeed, I agree that current positions in philosophy of biology are insufficient and not satisfactory in the explanation of why an organism is a member of a species rather than another. However, there are some points in the argumentation that need further development and can be subjected to critiques. The primary critique concerns change and variations in species and the evolution of species gradually in time. Indeed, it seems difficult to identify essences that will not be affected by changes. In particular, due to the brevity of this essay, I will focus on Sober's (1980) and Griffith's (2002) arguments against biological essentialism. Sober (1980) claimed that the genetic make-up of a species is something that continually changes and evolves. The variety of the genetic composition is central to the concept of evolution and adaptation itself. He assumes that

essences that are central in the definition of species are stable and fixed. Thus, this implies that intrinsic genetic essences are in contradiction with the contemporary theory of evolution. Secondly, Sober (1980) and Griffith (2002) affirmed that essentialism does not entail the importance of genetic variation in the Darwinian selection and evolution. In particular, Sober (1980) claimed that “Aristotelian” essentialism interprets variation as taking the organism away from its original nature and from the ideal, according to the theory of “Natural State Model”. In this way, differences caused by variations are unpleasant effects of the modifications of the organisms (Sober 1980, p. 362). This, in his opinion, is strictly in opposition to Darwinian Theory. Indeed, according to the evolutionary theory differences are not effects, but causes of evolution and better adaptation to the environment. An essentialist conception would be necessary in contradiction with this because essences are presented as stable and ideal properties that must be preserved. Devitt himself replies to these objections stating that essentialism is not committed to a wholly intrinsic essence position that avoids any form of change. This “weak” version of essentialism has two implications: only a part of the genetic material is essential and the essence of the species is influenced by external factors without. In consequence, it is possible to combine small variations in the genetic set and still preserve the belonging to a particular species.

### **Two possible answers to the objection of variation and change**

*Whatever it was that Darwin was up against, it was not Aristotelian essentialism.*

*(Lennox 2001, p.162)*

In this section, I will try to propose two other possible answers to the “evolutionary objection”. The first one is based on the definition of evolution in this context. The second one is based on a better explanation and understanding of what an essential property is. Let us now consider the first answer to the objection. In the contemporary philosophy of biology, there are two broad conceptions concerning the concept of evolution: “evolution **within** the specie or *microevolution*” and “evolution **of** the species or *macroevolution*”

(Godfrey-Smith 2014). If evolution is considered only as “evolution **within** the species”, then essentialism can be problematic. Indeed, the species need to be the same even if the essential genetic properties are changing and this is contradictory. However, if we conceive evolution as an “evolution **of** the species”, then evolution is the substitution of a species by another and essentialism can be saved. In particular, it can be possible to identify two families of changes in the genetic make-up. The first one is a change of the essential part of the genetic set so that the species is substituted by another and this refers to *macroevolution*. It can be possible to identify a “critic threshold” in the modification of genetic material beyond which it is correct to identify the substitution of the species with another. The second one is a change in genetic set that is not essential: it merely diversifies the individual from another member of the species and this is the case of *microevolution*. Indeed, it would be a simple variation that differentiates an individual from others, without undermining its membership to a particular species. In this way, essentialism entails for both micro and macro evolution.

Let us now move to the second answer that is based on a deeper understanding of the concept of essential properties. At first, individuals of the world have some properties, and these can be divided into essential properties and non-essential properties. Essential properties are the ones that characterise the species of the individual, and they are a sort of “necessary condition” for the identity of it *as a member of a species*. Non-essential properties are extrinsic or functional properties that don’t define the identity of the individual *as a member of a species*. These distinctions were made by Aristoteles in the *Categories*, where he divided the properties that explain the *essence of an object* from the ones that are *merely* in the object. In the *Metaphysics* he separated what is *accidental* (συμβεβηγός) and what is *essential* (καθ’ αὐτό). Being as accidental is something that is “not forever nor mostly”, which means it can change in the individual under consideration without any affection to the identity of it. Being as essential, on the other hand, is what the individual is “for itself”. The essential properties represent the content of the definition and, in the classic

conception<sup>7</sup>, essential properties are what characterise the kind or the species of an individual. In this context, species can be identified as groups of individuals that share the same essence. The classical conception of essential properties is consistent with Devitt's thesis. Indeed, species can be identified as groups that share, at least partially, the same set of genetic makeup that represents the intrinsic genetic property of the species. Furthermore, a deep understanding of this conception of essential properties can undermine the objections reported in the previous section. At first, Sober's (1980) presentation of the Aristotelian conception of essential properties seems incorrect and a result of a combination of the Platonic conception of essences with the Aristotelian one as a result of Mayr's analysis of essentialism (1959). Indeed, in the Platonic conceptual framework, essences are the ideal of the objects, they are immutable and stable, and they have an ontological priority compared to the individual which instantiates them. This version is incompatible with evolution. On the opposite, in the "Aristotelian conception", the concrete individual with all of its characteristics has an ontological priority. The essential properties define it as a member of a species because these properties are instantiated in all the individuals that are the member of that species. In this second view, variation and change in the essential properties are possible because they are a consequence of variation and change within the concrete biological organisms. These changes can lead to a substitution of a species by another and so macroevolution. Thus, this kind of essentialism seems not to be in contradiction with evolutionary theory. Furthermore, let us consider more precisely Sober's (1980) critique about the incompatibility between the Aristotelian conception and the evolutionary theory one. At first, in the Aristotelian conception changes in the individual are not conceived as effects of external factors. On the contrary, variation is interpreted as a cause, specifically, the final one, and it is a movement from potentiality to act. The theory of potentiality and actuality is crucial in order to understand the development of biological entities, and the idea of movement and variation is of fundamental importance in the development of biological life (cfr. *Generation*

---

<sup>7</sup> With classical conception I am referring to Aristoteles and Thomas Aquinas, in particular *De ente et essentia*.

*of Animals, Parts of Animals*). Secondly, to mention the theory of “Natural State Model” in this context is incorrect. Indeed, the Natural State Model is presented in the *Physics* and in *On the Heavens* and concerns the movement of physical objects. This theory does not apply to Aristoteles’s biological theory, where the idea of change is central about potentiality and actuality. Aristotle himself separates the method of study physical objects from the method of study biological ones (*Parts of Animal I,1*), and he recognises that the two domains describe different kinds of necessity. Indeed, if physical objects respect the Natural State Model, biological objects change according to a form of hypothetical necessity. This hypothetical necessity refers to the fact that biological beings are not directed toward a specific form or species, but rather to the fact that they have to be alive and survive (Lennox, 2001a). Thus, in natural beings, modifications are not seen as a deviation from ideal concepts of species, but rather as something that permits the survival of the being. In conclusion, I think that a conception of essential intrinsic properties that are close to the Aristotelian conception of essences is compatible with evolutionary theory. Furthermore, the identification of at least partially genetic essential properties in the definition of species can help us in answer to the taxon question.

In conclusion, I find that Devitt’s argument stands up to scrutiny. I agree with him about the inadequacy of other accounts in the answer of what he calls the “*taxon question*” and the efficacy of the essentialist position. However, I find that his argument needs further explanation about what he intends with genetic makeup and further specifications about the commitment to a “not fully determined essence”. Indeed, the identification of a specific genetic set and of a specific “threshold” beyond which evolution leads to a new species can give a crucial advantage to the argumentation. Nonetheless, I do not think that this lack of further explanations undermines the argument, but it represents a reason to develop this position further.

## **Conclusion**

In this essay, I have exposed Devitt's paper "Resurrecting Biological Essentialism" (2008) to show that it is possible to sustain essentialism about the definition of biological species. The result of this analysis is that species have essences that are at least partially genetic. Then, I have exposed the major objection to essentialism: variation and change in evolutionary theory. I have tried to answer this objection in two ways. The first one concerns the concept of evolution, the second one the concept of essential properties. In conclusion, Devitt's thesis is sound and that the lack of relation to the identification of "specific genetic set" represents a stimulus for further researches.

### **Bibliography**

- Aristotle: *L'anima (On the Soul)*, translated by G. Moravia. Bompiani, 2001, Milan
- Aristotle: *Generations of Animals*, translated by W. D. Ross, J. A. Smith. *The Works of Aristotle*, The Clarendon Press, 1908, Oxford
- Aristotle: *Metafisica (Metaphysics)*, Books A(2), Z(4-7), translated by G. Reale. Bompiani, 2014, Milano
- Aristotle: *Motion of Animals*, translated by W. D. Ross, J. A. Smith. *The Works of Aristotle*, The Clarendon Press, 1908, Oxford
- Aristotle: *Parts of Animals*, translated by W. D. Ross, J. A. Smith. *The Works of Aristotle*, The Clarendon Press, 1908, Oxford Devitt M., 2008, *Resurrecting biological essentialism*, University of Chicago Press
- Devitt M. (2008): *Resurrecting biological essentialism. Philosophy of Science*, 75, 344-382
- Devitt M. (2010): *Species Have Partly Intrinsic Essences. Philosophy of Science*, 648-661. The University of Chicago Press on behalf of the Philosophy of Science Association
- Ereshefsky M. (2010a): *What's Wrong with the New Biological Essentialism. Philosophy of Science*, 77 (5), 674-685

- Ereshefsky M. (2017): Species. *Stanford Encyclopedia of Philosophy* <https://plato.stanford.edu/entries/species/>
- Godfrey-Smith P., (2014a): Evolution and Natural Selection. *Philosophy of biology*, Chapter 3, Princeton Foundations of Contemporary Philosophy, Princeton University Press, Princeton, (2016)
- Griffith P. (2002): What is Innateness. *The Monist* 85, 70-85
- Lennox J. G., (2001): Are Aristotelian Species Eternal?. *Aristotle's Philosophy of Biology*, Cambridge University Press, Cambridge
- Lewens T., (2012a): Species, essence and explanation. *Studies in History and Philosophy of Biological and Biomedical Sciences*. 43, (2012), 751-757
- Mayr E., (1959): Typological versus Population Thinking. *Conceptual Issues in Evolutionary Biology*, ed. Sober E., (1984), 14-17, A Bradford book, The MIT Press, Cambridge (MA); London, England
- Richards A., (2010): The species problem: A Philosophical Analysis. *Cambridge University Press*, Cambridge
- Sober E., (1980): "Evolution, Population Thinking and Essentialism", *Philosophy of science* 47: 350-383. *The unites of Evolution: Essays on The Nature of Species*, ed. M. Ereshefsky (ed.). Cambridge, MA: MIT Press, 247-278
- *The Cambridge Dictionary of Philosophy* (III edition) 1995, General Editor Robert Audi, Cambridge University Press, New York, 2015
- Wilson R.A, Barker M. J., Brigandt I., (2007): When Traditional Essentialism Fails: Biological Natural Kinds. *Philosophical Topics* 35 (1-2), 189-215

## **Can there be a trade-off between internal and external validity?**

By Martin Vaeth

*Abstract:*

Proponents of randomised controlled trials (RCTs) see in them the potential to revolutionise empirical methods in the social sciences. RCTs are experiments that randomly allocate participants into two groups, one that receives a treatment (the treatment group) and one that receives no treatment or a placebo (the control group). This design extracts the causal effect of the treatment. RCTs can be compared to other empirical methods along the two dimensions of internal validity, i.e. how good they are at finding the true causal relationship in the studied population, and external validity, i.e. how good the discovered causal relationship can be generalised to other populations. While there is broad consensus that RCTs are better in internal validity than orthodox empirical methods such as regressions, there is much controversy whether this advantage comes at a cost to external validity. Even more, there are competing views about the relationship between internal and external validity that revolve around the question whether there can be a trade-off between them. In this paper, I use a formal approach to define internal and external validity and show that a trade-off is conceptually possible and how it might arise in practice.

Recent decades have seen a rise in new experimental methods such as randomised controlled trials (RCTs) in economics and other social sciences. RCTs promise to lead to better causal inferences than orthodox methods such as regression. Duflo et al., for example, write (2004: 8) that “creating a culture in which rigorous randomised evaluations [RCTs] are promoted, encouraged, and financed has the potential to revolutionise social policy during the 21st century, just as randomised trials revolutionised medicine during the 20th.” There has been criticism, however, that the power of RCTs is being overestimated due to their lack in external validity (Cartwright 2007: 12). It is often held that there is a trade-off between internal and external validity in

empirical methods (ibid: 11). According to this view, RCTs may have higher internal validity than methods such as regressions but RCTs have lower external validity. A contrasting position is that internal validity is a prerequisite for external validity (Lucas 2003: 248; Guala 2003: 1198; Hogarth 2005: 262). In this case, there can be no trade-off because it is not possible that external validity increases while internal decreases. I think these contrasting views result from a lack of clarity about what internal and external validity are and what their relationship is.

To clarify the matter, I propose a formal definition of internal and external validity. Using these definitions, I show that a trade-off between internal and external validity is possible and how it might arise. Section 1 introduces my notion of causality and gives formal definitions of internal, external and overall validity of an empirical method. These definitions will be measures of how close a coefficient is to the real causal coefficient, namely a mean squared error. Section 2 applies the formal definitions to a hypothetical example comparing an RCT with a regression and shows how the regression might be better in external but worse in internal validity than the RCT. I essentially give a formalisation of the idea that a regression can be better in external validity because it is based on a more heterogeneous sample.

## **Section 1**

**For the purpose of this article, I understand causality as a probabilistic relation between variables. Variable A having a causal effect on variable B means that if all other variables stay the same, a change in A leads to a change in the probabilistic distribution of B (see Hitchcock 1997: section 2). For example, A could be the price of a distributed insecticide-treated bed net in sub-Saharan Africa and B the usage of that bed net or the effect on malaria prevention, to give an example from the literature (Cohen and Pascaline, 2010). The causal effect of A on B can depend on the value of other variables and therefore be different for different regions and it can also depend on the value of A. To simplify the following analysis, I suppose the causal effect of A on B, given the**

**values of the set of other variables,  $V$ , has a linear form:**

$$B = f(V) + \beta(V) \cdot A$$

This means that  $B$  is function of  $V$  plus a coefficient  $\beta$  that may depend on the variables  $V$  times the value of  $A$ . I use this simplification because it allows us to capture the causal effect in a single coefficient  $\beta$  and because such a form is commonly assumed in practice.

For the rest of this article, I suppose that the purpose of an empirical method is to find the true causal coefficient of one variable on another variable in a ‘target population’. An empirical method does so by studying a population which I will call ‘population of study’. According to the classical definitions by Cook and Campbell (1979: 37), internal validity “refers to the approximate validity with which we infer that a relationship between two variables is causal or that the absence of a relationship implies the absence of cause”, and external validity “refers to the approximate validity with which we can infer that the presumed causal relationship can be generalized to and across alternate measures of the cause and effect and across different types of persons, settings, and times.” Let us call  $\beta$  the causal coefficient that our empirical method yields,  $\beta_1$  the true causal coefficient of the population of study and  $\beta_0$  the true coefficient of the target population. Intuitively, internal validity should be a measure of how close  $\beta$  is to  $\beta_1$  and external validity should be a measure of how close  $\beta_1$  is to  $\beta_0$ . When measuring the proximity of these coefficients, we should keep in mind that  $\beta$  is a random variable (and later I will argue that  $\beta_1$  and  $\beta_0$  can be seen as random variables as well). A first measure of internal validity would be the bias  $\mathbf{E}(\beta - \beta_1)$ . However, even if the bias is very small,  $\beta$  can usually be very far from  $\beta_1$  if the variance of  $\beta$  is large. A better measure for internal validity that takes into account both the variance of our estimator and the bias is the mean squared error  $\mathbf{E}((\beta - \beta_1)^2)$  which is commonly used in probability theory.

**Internal Validity:**  $\mathbf{E}((\beta - \beta_1)^2)$

It can be shown that the mean squared error is exactly the sum of the squared bias and the variance of  $\beta$ . The mean squared error is always non-negative. If it is zero, we find exactly the right coefficient. The bigger it is, the

less precise is our estimator  $\beta$ , i.e. the weaker is our internal validity.

Like written above, external validity is generally understood as a measure of how good the causal effect can be generalised from the population of study to the target population. That is why I take external validity as how close the coefficient of the population of study,  $\beta_1$ , is to the coefficient of the target population,  $\beta_0$ . This does not depend on the empirical method itself (it is independent of internal validity) but on how similar the two populations are. So I define external validity analogously to above as:

**External Validity:**  $\mathbf{E}((\beta_1 - \beta_0)^2)$

Finally, I define overall validity as the as the expectation of the squared distance of our estimator  $\beta$  and the true coefficient of our target population  $\beta_0$ :

**Overall Validity:**  $\mathbf{E}((\beta - \beta_0)^2)$

These definitions acknowledge the fact that we can never find the true coefficient with perfect precision but that we only have estimates. While we are using probabilistic expectations, we do not establish whether the underlying probabilities are objective or epistemic. My formal definition of validity allows both interpretations.

Now we can turn to the connection between internal, external validity and overall validity. We can mathematically derive the following equation (called equation (1) from now on):

$$\begin{aligned} \mathbf{E}((\beta - \beta_0)^2) &= \mathbf{E}(((\beta - \beta_1) + (\beta_1 - \beta_0))^2) = \mathbf{E}((\beta - \beta_1)^2) + \mathbf{E}((\beta_1 - \beta_0)^2) + \\ &2 \cdot \mathbf{E}((\beta - \beta_1)(\beta_1 - \beta_0)) \end{aligned} \quad (1)$$

Here I used that  $\mathbf{E}(A+B) = \mathbf{E}(A) + \mathbf{E}(B)$  for random variables A and B. (1) shows that overall validity is the sum of internal validity, external validity and a third term. This third term can either be positive (weakening overall validity) or negative (strengthening overall validity). It strengthens overall validity if it happens that  $\beta$  overestimates  $\beta_1$  but (incidentally)  $\beta_1$  is smaller than  $\beta_0$ , thus  $\beta$  ends up being close to  $\beta_0$ . The third term captures this combination effect that even poor external and internal validity can theoretically result in strong overall validity. As we cannot depend on this combination effect to improve overall

validity, we can see from (1) that we need both a strong internal and a strong external validity for a strong overall validity. Furthermore, (1) shows that the role that internal and external validity play in determining overall validity is symmetric. Thus, a trade-off is conceptually possible.

## **Section 2**

To see how there could be a trade-off between internal and external validity in practice, let us look a hypothetical example. In this example, a regression will have a weaker internal but stronger external validity than an RCT. Let us suppose that we want to find the causal effect of the price of insecticide-treated bed-nets on their usage. These bed-nets are distributed by aid programmes to prevent malaria. According to one hypothesis, people value the bed-nets more if they paid something for them and therefore use them more. Our target population is region X in sub-Saharan Africa where we want to know the causal coefficient. We have an RCT in another region Y in sub-Saharan Africa and a statistical regression from data of regions  $Z_1, \dots, Z_n$  in sub-Saharan Africa. They yield the coefficients  $\beta_{\text{rct}}$  and  $\beta_{\text{reg}}$ , respectively. The true coefficient of region Y is  $\beta(Y)$ , the average coefficient of regions  $Z_1, \dots, Z_n$  is  $\beta(Z)$  and the true coefficient of the target population region X is  $\beta_0$ .

It is commonly held that RCTs tend to have a stronger internal validity than regression analyses. The basic reason is that in a RCT the experimenter can select the control group and treatment group randomly while in a regression they cannot be chosen. In the latter case, different forms of selection bias might arise. Thus, I suppose that the internal validity of the RCT is stronger than the one of the regression.

My main aim is to argue that in some cases we have reason to believe that the external validity of a regression is stronger than the one of an RCT. This can be the case because regressions typically base their estimate of the causal coefficient on a more heterogeneous population than RCTs. Let us suppose we could divide sub-Saharan Africa into regions of equal size such that each

region has one homogeneous causal structure (and thus one causal coefficient).<sup>8</sup> This is of course wrong, but that does not matter as this example is purely hypothetical. Furthermore, we assume that we have no information at all about whether  $\beta(Y)$ ,  $\beta(Z_i)$  and  $\beta_0$  are at the high or low spectrum of the distribution of causal coefficients in different regions. To say something about the external validity I make a crucial step: We can think about the coefficients  $\beta_0$ ,  $\beta(Y)$  and  $\beta(Z_i)$  (for every region  $Z_i$  in  $Z$ ) as drawn randomly and independently from the distribution of coefficients in regions of Africa. Thus, they all have the same distribution and especially, same expected value and same variance  $\sigma^2$ . We should interpret this not as an assumption about objective probabilities of the values of these coefficients but as one about epistemic probabilities. In the context of external validity, objective probabilities are hard to apply as our choice of the target population and the population of study might not be the result of a random experiment. We probably choose them according to some criteria like availability of data or suitability for a RCT. As long as we do not believe these criteria are correlated with the causal coefficients in the region, we can motivate the assumption above by a principle of indifference or a flat prior Bayesian belief. Then we have the following external validity of the RCT:

$$\begin{aligned} \mathbf{E}((\beta(Y) - \beta_0)^2) &= \mathbf{E}(\{[\beta(Y) - \mathbf{E}(\beta(Y))] + [\mathbf{E}(\beta_0) - \beta_0]\}^2) = \\ &= \mathbf{E}([\beta(Y) - \mathbf{E}(\beta(Y))]^2) + \mathbf{E}([\mathbf{E}(\beta_0) - \beta_0]^2) + 2 \cdot \mathbf{E}([\mathbf{E}(\beta(Y)) - \beta(Y)][\mathbf{E}(\beta_0) - \beta_0]) = \\ &= \text{Var}(\beta(Y)) + \text{Var}(\beta_0) + 2 \cdot \mathbf{E}((\mathbf{E}(\beta_0) - \beta_0)) \cdot \mathbf{E}(\mathbf{E}(\beta(Y)) - \beta(Y)) = \\ &= \text{Var}(\beta(Y)) + \text{Var}(\beta_0) = 2 \cdot \sigma^2 \end{aligned}$$

The first equality holds because  $\mathbf{E}(\beta(Y)) = \mathbf{E}(\beta_0)$  and the third because  $\beta_0$  and  $\beta(Y)$  are independent.

Analogously, we obtain the following external validity for the regression:

$$\mathbf{E}((\beta(Z) - \beta_0)^2) = \text{Var}(\beta(Z)) + \text{Var}(\beta_0) = \text{Var}(\beta(Z)) + \sigma$$

---

<sup>8</sup>This means that we assume there is no unmodeled causal heterogeneity. If such unmodelled causal heterogeneity was present, then additional problems with the internal validity of the regression would occur (see Aronow and Samii, 2016).

To determine  $\text{Var}(\beta(Z))$ , we first remember that because the regions are of equal size, the true causal coefficient of the set of regions  $Z$  is the average over the coefficients of the different regions in  $Z$ , so  $\beta(Z) = \frac{1}{n} \sum_{i=1}^n \beta(Z_i)$ . Because  $\beta(Z)$  is an average, it has a lower variance than each  $\beta(Z_i)$ :

$$\text{Var}(\beta(Z)) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n \beta(Z_i)\right) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n \beta(Z_i)\right) = \frac{1}{n} \sigma^2$$

Here I used that  $\text{Var}(cA) = c^2 \text{Var}(A)$  and  $\text{Var}(A+B) = \text{Var}(A) + \text{Var}(B)$  for any constant  $c$  and independent random variables  $A, B$ . This means that the external validity of the regression is smaller than the one of the RCT:

$$\mathbf{E}((\beta(Z) - \beta_0)^2) = \left(1 + \frac{1}{n}\right) \sigma^2 < 2 \cdot \sigma^2 = \mathbf{E}((\beta(Y) - \beta_0)^2)$$

This effect can be interpreted as follows: The RCT looks only at one region while the regression averages over many regions. It is less probable for the average over many regions than just for the coefficient of one single region to be far from the average coefficient in sub-Saharan Africa.

Finally we have to consider the effect of our empirical method on the third term in overall validity, which is  $2 \cdot \mathbf{E}((\beta - \beta_1)(\beta_1 - \beta_0))$ . Here  $\beta$  stands for the coefficient that the RCT or the regression yields, respectively, and  $\beta_1$  stands for  $\beta(Y)$  or  $\beta(Z)$ , respectively. It is reasonable to assume that the internal error  $\beta - \beta_1$  and the difference  $\beta_1 - \beta_0$  are independent in both cases, so  $\mathbf{E}((\beta - \beta_1)(\beta_1 - \beta_0)) = \mathbf{E}(\beta - \beta_1) \cdot \mathbf{E}(\beta_1 - \beta_0)$ . Because  $\mathbf{E}(\beta(Y)) = \mathbf{E}(\beta(Z)) = \mathbf{E}(\beta_0)$ , the second factor and hence the whole term is zero in both cases.

To sum up, I have proposed a formal definition of internal, external and overall validity. Not only do these definitions capture the intuition behind these terms in an adequate way, they prove to be fruitful concepts to analyse the validity of an empirical method. In the example in section 2, I made a number of assumptions to isolate the effect that a regression can have a stronger external validity than an RCT because it is based on a more heterogeneous sample. The strength of the formal analysis is to highlight all the assumptions that need to hold for this effect and to hint at other possible

effects that might affect validity.

#### References

- Aronow, P. M., & Samii, C. (2016). Does regression produce representative estimates of causal effects?. *American Journal of Political Science*, 60 (1), 250-267.
- Cartwright, N. (2007). Are RCTs the gold standard?. *BioSocieties*, 2 (1), 11-20.
- Cohen, J., & Dupas, P. (2010). Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment. *The Quarterly Journal of Economics*, 1-45.
- Cook, T. D., Campbell, D. T., & Day, A. (1979). *Quasi-experimentation: Design & analysis issues for field settings* (Vol. 351). Boston: Houghton Mifflin.
- Duflo, E., Glennerster, R., & Kremer, M. (2004). Randomized evaluations of interventions in social service delivery. *Development Outreach*, 6 (1), 26-29.
- Guala, F. (2003). Experimental localism and external validity. *Philosophy of science*, 70 (5), 1195-1205.
- Hitchcock, C. (1997). Probabilistic causation. *The Stanford Encyclopedia of Philosophy*. Winter 2016 Edition. <https://plato.stanford.edu/entries/causation-probabilistic/#ProRaiTheCau>
- Hogarth, R. M. (2005). The challenge of representative design in psychology and economics. *Journal of Economic Methodology*, 12 (2), 253-263.
- Lucas, J. W. (2003). Theory-testing, generalization, and the problem of external validity. *Sociological Theory*, 21 (3), 236-253.

## Border Control and Relational Egalitarianism

Lewis Williams

*In this article, I examine Wellman's proposal in Immigration and Freedom of Association that wealthy states have a right to close their borders to migrants from poorer states. I pay particular attention to Wellman's discussion of relational egalitarianism, and his argument that even if the claim of relational egalitarianism is weighty enough to trump the wealthy state's right to closed borders, the lack of relational inequality between members of wealthy and poorer states means that no sufficient reason is generated to oblige wealthy states to close their borders. I argue that relational inequality does exist between members of wealthy and poorer states, and that opening borders could be one particularly effective way of overcoming this inequality. Therefore, I conclude that if we take the weightiness of the relational egalitarian claim seriously, then we should understand it as undermining the right of wealthy states to border control*

### **Introduction**

In *Immigration and Freedom of Association*, Wellman argues that wealthy states have a right to close their borders to migrants from poorer states. This proposition can be understood in one of two ways. Firstly, wealthy states may have an absolute right to close their borders to migrants from poorer states. Secondly, wealthy states may have a right to close their borders to migrants from poorer states, given that it is not trumped by competing considerations. In line with Wellman's discussion of border controls<sup>9</sup>, I will focus on the second interpretation: whether the right of wealthy states to border control is trumped by competing considerations. One such competing claim is that of the relational egalitarian: to avoid a level of inequality that facilitates domination through oppressive

---

<sup>9</sup> I will use the term 'border control' to refer to the state's right to close their borders.

relationships. Wellman accepts the pull of relational egalitarianism, yet argues that it does not suffice to outweigh the right of states to border control. This is because, argues Wellman, there is no relational inequality between members of wealthy and poorer states (2008: 123). Further, even if there were such inequality, there are better ways to combat it than opening one's borders.

In this essay, I will argue that there is in fact relational inequality between members of wealthy and poorer states, and that the alternative methods of combatting relational inequality are not clearly superior to open borders. I will remain agnostic as to whether the claim of the relational egalitarian is weighty enough to trump the state's right to border control. However, I will argue that given that *if* we accept that the claim to relational egalitarianism is sufficiently weighty, then this does undermine the claim of wealthy states to border control; firstly because relational inequality does exist between citizens of wealthy and poorer states, and secondly because there are no clearly superior alternative methods of combatting relational inequality.

### **Wellman's Argument**

While Wellman concedes that he is "*personally inclined towards more open borders*" (2008:116-117), Wellman argues that insofar as we have a right to freedom of association, states should be equally entitled to freedom of association, and therefore to the freedom to set their own immigration policy. Note that Wellman refrains from arguing that states *should* close their borders, just that states have a *right* to close their borders. Wellman's argument can be presented thus:

- (1) In the absence of sufficiently weighty competing claims, states have a right to self-determination
- (2) Freedom of association (and disassociation) is one component of self-determination

- (3) Therefore, in the absence of sufficiently weighty competing claims, states have the freedom to exclude

To understand the nuances of Wellman's position, his argument requires further unpacking. Firstly, the 'in the absence of sufficiently weighty competing claims' aspect of (1) is significant. In the context of border control, Wellman does not argue that the state's right to self-determination is absolute. Rather, Wellman only intends his support of the right to border control to be "*prima facie*" (2008:113), and this right could in theory be overridden by competing claims (Fine, 2013:160). Nonetheless, Wellman later argues that no competing claim is weighty enough to override the state's right to border control.

Secondly, Wellman offers further argumentation to support the link between freedom of association and self-determination. Wellman notes that without the right to freedom of association, it is difficult to explain the wrongness of forcing a country to join the EU (2008: 112). To force a country to join an association such as the EU would be to violate their freedom to associate and disassociate with whomever they please. This freedom to association is guaranteed by self-determination, for without the freedom to join and remove itself from such associations, a country cannot be free to determine its own statehood.

Finally, Wellman goes into more detail about this *prima facie* reason to respect the state's right to freedom of association. In doing so, Wellman draws a comparison between the state and the individual's right to freedom of association. In the same way that individuals have the right to choose who they do and do not marry (given the consent of the prospective partner), states have the right to choose who they invite into their community. Just as individuals have the right to remain single, states have the right to reject people from their community (2008:110-111). Wellman accepts that there may be some disanalogy here, as freedom of association in a marital context is more important for individuals than

freedom of association in a political context. Nevertheless, argues Wellman, his argument does not depend upon an equivalence of importance between these two types of freedom of association. It is enough for Wellman to establish the prima facie importance of the freedom of association, and later to argue that no competing considerations are as weighty as this prima facie importance.

To summarise, Wellman offers a logically valid argument, as illustrated in bullet-point format above. Given that states have a right to self-determination in the absence of sufficiently weighty competing considerations, and that the freedom of association is one component of this right, then states must have the freedom to exclude. Further, as explained above, Wellman additional presumptive support for this argument through his analogy with marriage.

In order to make my argument, I will not attack either of Wellman's premises or his conclusion. I will grant that states have a right to self-determination in the absence of sufficiently weighty competing considerations, that this right includes the right to freedom of association, and therefore the right to exclude. Further, I will not commit to arguing that relational egalitarianism is a sufficiently weighty consideration. Instead, I intend on objecting to Wellman's claim that even if we do accept the weightiness of relational egalitarianism, that this "*does not generate sufficient moral reasons to obligate the wealthy state to open its borders*" (208:122). I will argue that, given Wellman's argument, *if* we do accept the weightiness of relational egalitarianism, then this does in fact generate sufficient moral reasons to obligate wealthy states to open their borders.

### **Relational Egalitarianism**

As previously explained, Wellman's endorsement of the right to border control is contingent upon there being no competing considerations with

sufficient weight to override the right to border control. One proposed competing consideration is the claim of the egalitarian. The egalitarian position forks off in two directions, towards luck egalitarianism and towards relational egalitarianism. The luck egalitarian points out the vast inequalities between states, and argues that it is unjust that one should face lower life prospects due to the morally arbitrary fact of being born in one state rather than another. The luck egalitarian may therefore reject the right to border control, as such a right reinforces this bad brute luck. Wellman recognises the intuitive appeal of this position; It is unfair that some people's lives start off worse than others through no fault of their own, and it would be better if this were not the case (2008:121). Nevertheless, it is relational egalitarianism, not luck egalitarianism, which captures what is most important about inequality for Wellman. Relational egalitarianism is less concerned with absolute inequalities between persons than with the relationships that these inequalities entail. What is particularly problematic about inequality for the relational egalitarian is the domination that arises from the oppressive relationships facilitated by inequality. Wellman justifies his endorsement of relational egalitarianism as "*the most compelling understanding of inequality*" (2008:120) by appealing to Anderson's split-world case. In one world, there exist great inequalities between the citizens of A and B. However, living on opposite sides of the world, they know nothing of the other's existence. This inequality is mirrored in society C, where not only are the citizens aware of their unequal status, but their relationships are affected by these inequalities. Wellman's greater appreciation of relational egalitarianism above luck egalitarianism is reflected by the fact that it is "*uncontroversial that the inequality among the Cs is much more worrisome than the inequality between the As and Bs*" (2008:122).

### **The Significance of Relational Egalitarianism**

Despite his advocacy of relational egalitarianism as an explanation of what is wrong with inequality, Wellman states that “*the lack of a robust relationship between the constituents of a wealthy state and the citizens of a poorer country*” (2008:123) implies that the objectionable level of inequality between the states does not give sufficient moral reason to obligate the wealthy state to open their borders. For Wellman, even if the claim of the relational egalitarian could be sufficiently weighty to override the right to border control, there is no relational inequality between states, so such a claim cannot get off the ground.

However, it is not immediately clear that such a robust relationship is lacking between constituents of wealthy states and citizens of poorer countries. The type of robust relationship with which Wellman concerns himself is one in which “*subordinates are dominated in oppressive relationships*” (2008:121); Wellman refers to ‘oppressive relationships’ on numerous occasions, while also quoting Anderson’s appeal to the need to protect citizens’ “*entanglement in oppressive relationships*” (2008:122). While a precise definition of ‘oppressive relationships’ is not found in Wellman’s paper, Anderson sheds some light on this area, drawing upon Young in defining oppression as “*marginalization, status hierarchy, domination, exploitation and cultural imperialism*” (1999: 312). If we accept Wellman’s account of robust relationships as relationships that involve domination through oppressive relationships, then certain relationships between citizens of poorer and richer states appear to be robust in this relevant sense. One pertinent example involves the exploitation of migrant Nepali 2022 World Cup workers. Lynch of Amnesty International describes how Nepali workers are indebted by the huge fees charged by Qatari recruitment agencies, leaving them with no choice but to stay in low-paid and dangerous jobs. Here, Nepali workers are exploited by their Qatari employers – once they arrive in Qatar their employers take advantage of their inability to return home by paying them a lower monthly salary than

was originally promised (Kelner, 2007). Such a relationship fits Anderson's criteria of an oppressive relationship. Furthermore, this oppressive relationship exists between members of different states – specifically, between Nepalese workers and the Qatari recruiters who exploit and dominate Nepalese workers. Further examples include sweatshop labour; McVeigh writes of American sportswear company Nike recruiting Cambodian women to work 60-hour weeks in 37C heat, resulting in mass faintings as a result of their working conditions (McVeigh, 2017). If Wellman is to take seriously the claim of the relational egalitarian and the worrisome nature of such inequalities, then Wellman should take seriously the threat that relational egalitarianism poses to the right of wealthy states to border control.

Wellman, however, constructs a second defence against the relational egalitarian. Wellman argues that even if relational inequality does exist between members of different states, and the claim to alleviate this inequality is weightier than the state's claim to their right to border control, it need not follow that the right of the state to border control is overridden. Even under such a strong claim to alleviate relational inequalities, a policy of open borders is not the only mechanism, or even the most efficient mechanism, available to the wealthy state. Wellman draws upon Miller's criticism that a policy of open migration may not offer substantial benefits to the worst-off who would be unable to afford to migrate before concluding that "*sending aid abroad is a better way to rescue those most imperilled by poverty*" (2008:128). Wellman's point is that even if wealthy states have a duty to move to alleviate relational inequality, this need not involve opening their borders. However, it is unclear how Wellman reaches the conclusion that aid is any more effective in rescuing those most imperilled by poverty than open migration. Firstly, the efficiency of aid in helping the worst-off is far from uncontroversial. Mwenda notes how \$600bn was provided to Africa in aid between 1960

and 2003 (Mwenda, 2007), yet still almost half of Sub-Saharan African children live in absolute poverty (Hodal, 2016). Furthermore, while overstating the efficiency of aid in rescuing those most imperilled by poverty, Wellman understates the efficiency of open migration. A lack of resources does not suffice to render migration impossible for the worst-off; this is evidenced by the continued efforts of Libyan refugees to migrate to wealthier countries. As Carens points out, millions of poor and oppressed people would have much to gain from the opportunity of migrating to wealthier states (1987:264).

To argue that relational inequalities can be easily overcome without having to resort to open borders oversimplifies the complexity of the problems posed by relational inequality. Of course, it would be an even greater oversimplification to argue that open borders alone could solve the problems posed by relational inequality. Nevertheless, it is clear that migrants from poorer states would have something to gain from some form of open borders, and it is less clear that we would be able to completely eradicate inter-state relational inequality without wealthy states opening their borders to some degree.

Therefore, I have argued that if Wellman is correct that a state's right to border control can be overridden by sufficiently weighty competing considerations, and that the claim of the relational egalitarian has sufficient weight, then the right of wealthy states to border control is undermined by the claim of the relational egalitarian. I have justified this claim by arguing that relational inequality does exist between states, and that there is no clearly superior mechanism that would be able to wholly eradicate inter-state relational inequalities.

### **Bibliography**

Anderson, E., 1999. What Is the Point of Equality?. *Ethics*, 109(2), pp.287-337.

Carens, J.H., 1987. Aliens and citizens: the case for open borders. *The review of politics*, 49(2), pp.251-273.

Fine, S., 2013. The Ethics of Immigration: Self-Determination and the Right to Exclude. *Philosophy Compass*, 8(3), pp.254-268.

Hodal, K., 2016. Nearly half all children in sub-Saharan Africa in extreme poverty, report warns. [online] The Guardian. Available at: <https://www.theguardian.com/global-development/2016/oct/05/nearly-half-all-children-sub-saharan-africa-extreme-poverty-unesf-world-bank-report-warns> [Accessed 30 Nov. 2017].

Kelner, M., 2017. 2022 World Cup migrant workers 'systematically' exploited, Amnesty says. The Guardian. [online] Available at: <https://www.theguardian.com/football/2017/dec/17/2022-world-cup-migrant-workers-systematically-exploited-qatar-amnesty-international> [Accessed 18 Dec. 2017].

McVeigh, K., 2017. Cambodian female workers in Nike, Asics and Puma factories suffer mass faintings. *The Guardian*. [online] Available at: <https://www.theguardian.com/business/2017/jun/25/female-cambodian-garment-workers-mass-fainting> [Accessed 14 Feb. 2018].

Mwenda, A., 2007. Aid for Africa? No thanks. [online] Ted.com. Available at: [https://www.ted.com/talks/andrew\\_mwenda\\_takes\\_a\\_new\\_look\\_at\\_africa](https://www.ted.com/talks/andrew_mwenda_takes_a_new_look_at_africa) [Accessed 30 Nov. 2017].

Wellman, C.H., 2008. Immigration and freedom of association. *Ethics*, 119(1), pp.109-141.

Young, I., 1990. *Justice and the Politics of Difference*. Princeton, N.J.: Princeton University Press

## How convincing a case does Taurek make that one has no duty to save the greater number?

By Ragini Dua

***Abstract:** In this paper, I consider John Taurek's arguments against the common-sense intuition that when faced with a choice between saving different groups of people, special considerations aside, one ought to save as many people as possible. He is partly successful insofar as he highlights that our moral intuitions change when we consider a choice between saving one person that we know and like, and saving five strangers. If saving the greater number is a genuine moral duty, he argues, it should not be overridden so easily. However, Taurek's justification that a person is permitted to save one rather than five on the basis that the one would be permitted to choose his own life over the lives of five others also leads to some counterintuitive results. This paper does not attempt to defend a duty to save the greater number, but simply considers how to get out of the deadlock that results from Taurek's attack on this claim.*

In this essay I critically assess John Taurek's defence against Claim X: "Special considerations apart, one ought to save the greater number" (Taurek, 1977, p. 294), detailed in his paper, "Should the Numbers Count?" I focus on what I view to be Taurek's three core arguments, as well as corresponding objections, particularly those posed by Derek Parfit. Taurek makes a compelling argument, and though I do not believe he is fully successful in refuting this claim, his arguments do lead one to question the intuitions that sit behind it.

Taurek uses the following hypothetical scenario to illustrate his argument. Imagine that I have in my possession the supply of a life-saving drug. I am with six people, all of whom are on the verge of death. Five people need one-fifth of the drug and a sixth person, David, needs the entire dosage. Taurek denies that one has a moral duty to save the five rather than the one.

The first argument questions the nature of these "special considerations". If David were someone who I know and like, Taurek believes I would be morally permitted to save him. This varies from considerations that make David's death a worse thing than that of the five such as if he was close to discovering a cure for cancer or the five were old and senile. Further, I have no duty towards David, as I would if I had a contractual obligation to provide him this quantity of the drug, for example. Rather, I am permitted, though not obligated, to save David simply because I prefer to do so. If Claim X can be so easily overridden by a mere preference, it should be rejected.

In response to this argument, one might say that moral principles can be overridden without ceasing to have value. Taurek notes that parents can be thought of as having special obligations towards their children, and this explains the intuition that I would be

morally permitted or perhaps even obligated to save David if he was my son. However, if I had previously entered into contracts with each of the five to give them one-fifth of the drug, I would be obligated to do so and allow my son to die. Perhaps Taurek's response to this would be that although moral principles can be overridden by more stringent moral principles, they should not be overridden by mere preferences. However, if David was a dear friend, for example, I suspect few people would deny that I am permitted to save him, even though I cannot be said to have a strong duty towards him. In the absence of a duty, this too could pose a threat to Claim X.

Parfit's explanation of a preference to save a dear friend is that though contracts may give rise to special obligations which we must carry out "even at heavy cost to ourselves", this "may not be true of saving the lives of strangers" (Parfit, 1978, p. 290), and this explains why we would be permitted to save a loved one. This seems intuitively correct; it is uncontroversial that one would not have a duty to save the lives of strangers if it involves a grave danger, for example.

Parfit seems inclined to refute Taurek's argument by rejecting the claim that we are permitted to save David simply because we know and like him. However Taurek's next argument addresses such a sceptic. Imagine that David owns the drug. It would be inappropriate to try and convince him to give the drug to the five on the grounds that it is a worse thing for five people to die rather than one person. From David's perspective, it is a worse thing if he dies. It is permissible for him to use the drug because "he values his own life more than he values theirs" (Taurek, 1977, p. 300), and in doing so he doesn't violate their rights. Therefore, it must be permissible for me to give David the drug because I value his life more than that of the five. Taurek generalises this principle as follows: "if it would be morally permissible for a person B to choose to spare himself a certain loss, H, instead of sparing person C, a loss H', ... then it must be permissible for someone else ... to choose to secure the outcome most favourable to B instead of the outcome most favourable to C". (Taurek, 1977, p. 301). He is unfazed by the implication that A would therefore be permitted to choose B's arm over C's life on the basis that B would be permitted in choosing her own arm over C's life. Taurek claims this is the choice he would make if B's welfare was "more important" (Taurek, 1977, p. 302) to him than C's welfare, and does not believe it is unacceptable "unless it is for some reason impermissible for one person to take the same interest in another's welfare as he himself takes in it" (Taurek, 1977, p. 302).

Parfit provides a forceful objection to this argument. In this context, it is perhaps best to interpret "taking the same interest in" as "giving the same priority to". The common view is that one is permitted to give priority (though not absolute priority) to one's own

welfare as well as that of certain others, such as one's children, but this priority is agent-relative. It is to my arm or my child that I give priority. By contrast, Taurek argues that I can give priority to person B over person C not because I have a special relationship with B, but because there is no reason why I cannot adopt B's point of view. Taurek says that this is what he would choose to do if he was more concerned with B's welfare, but following his argument, it would also be permissible to adopt B's point of view for a totally arbitrary reason. Parfit, on the other hand, is arguing that I may sometimes be permitted to prioritise my own welfare or my child's welfare, but the justification for this is that it is *my* welfare and *my* child.

The final argument of Taurek's that I shall consider is the "impersonal, evaluative judgement" (Taurek, 1977, p. 306), Claim Y: "it is a worse thing that these five people should die than that this one should" (Taurek, 1977, p. 303). He would need to believe this in order to be motivated to choose the five rather than the one, when all are strangers to him. He believes it would be a "moral shortcoming not to prefer what is admittedly in itself a better thing to what is in itself a worse thing" (Taurek, 1977, p. 306). We do not view David as being morally deficient for preferring the outcome where he lives and that these five people should die, and therefore, Claim Y cannot be the impersonal judgement he is looking for to explain this preference.

In my view, Taurek is overstating the relationship between Claim X and Claim Y. Consider the following parallel. Claim X': "special considerations aside, one has a duty not to steal". It is not obvious why Claim Y': "it is a worse thing to steal than not to steal" must be true for everyone. One would not think a poor man morally deficient for preferring an outcome whereby he steals and has more money, even if we would think him morally deficient for acting on it. Wanting more money is not a special consideration that would typically override a moral principle. We accept Claim X' but its truth needn't be dependent on the "worseness" of various outcomes, either intrinsically or to different people. The same is true of Claim X. Parfit's objection is along similar lines. He believes it may be possible to prefer the worse of two outcomes without being morally deficient where the better outcome poses "too great a sacrifice" (Parfit, 1978, p. 292). Taurek doesn't account for this view.

In attempting to refute Claim X, Taurek does not explicitly commit himself to an alternative view although he does say that he would be inclined to flip a coin to "best express" his "equal concern and respect" (Taurek, 1977, p.303) for each person by giving them a fifty-fifty chance of surviving. This view is compelling, and I will return to it shortly. However, although he would be inclined to do this, he is not saying that one would be obliged to do this, and nevertheless seems to endorse the view that one can choose to save either the one or the

five for an arbitrary reason. Further, when Taurek considers the trade-off between A's arm and B's life, he does not outline any specific conditions to the permissibility of taking A's point of view. Therefore, I would be fully permitted to either flip a coin, or make a choice for a completely arbitrary reason, to choose A's arm over B's life, even if both were strangers to me. I feel this position is sufficiently counterintuitive so as to be rejected, particularly when we consider that it could be extrapolated further. Person A, who can swim, but is irrationally afraid of open water might be excused in choosing his or her own psychological comfort over saving the life of a drowning person. Would it therefore follow that I am permitted to save A, who is caught in open water and is very scared, but could safely swim herself to shore, over B, who will certainly drown if I do not save him? Choosing A's arm or A's psychological comfort over B's life would certainly not demonstrate "equal respect and concern" for each life.

Perhaps this view would not appear to be so counterintuitive if one also held that there is no moral duty at all to save anyone. In Taurek's original example, this would therefore indicate that one is permitted to simply throw the drug away and allow everyone to die. If that were one's moral position, perhaps the thought that you could, if you wished, choose to save A's arm over B's life would not be so absurd. But Taurek explicitly excludes the possibility of saving no one in a footnote of the article, so I will set it aside for now, and assume that it is false that if a person can spare another person significant harm at little or no cost to himself, he may nevertheless ignore them.

Although Taurek is not stating that one ought to flip a coin, he nevertheless provides a compelling argument for why he would be inclined to do this. Let us consider Claim Z as a rival position to Claim X about how one ought to behave when faced with such a situation, whereby Claim Z is as follows: "Special considerations aside, one ought to give each person an equal chance to survive". By flipping a coin and giving each individual a fifty-fifty chance of survival, Taurek questions who amongst them could complain that he has done wrong.

In explaining this, Taurek states that he is most concerned with "the loss to" rather than "the loss of" these people (Taurek, 1977, p. 307). No one person would suffer a loss five-times greater than any other person. Taurek denies that people have a "certain objective value" (Taurek, 1977, p. 307), however I disagree. Taurek emphasises that his concern for the loss of A's arm for instance, is rooted in empathy because he himself would be concerned with the loss of his own arm. But assume that our stream of consciousness ends with our life. All six people are concerned with their impending loss of life, but once they die, it is not possible for them to be concerned with the loss of life itself. Even if one takes a different view of death we would not be able to predict this feeling with any certainty. Admittedly, this

position is contingent on a controversial view of death, but there does seem to be something strange about Taurek empathising with a dead person about the loss of their life. In my view, from an impersonal perspective, the one I would adopt towards six strangers, the loss of their life matters to be not because it matters to them only, but because each human life has intrinsic value.

Nevertheless, it does not follow that we are permitted to say it is five times worse if five people were to die than one person, and Taurek's explanation of his desire to flip a coin is actually quite compelling – it is true that flipping a coin gives everyone an equal chance of survival, and for each individual person, it intuitively does seem to be the fairest way to make such a decision. However, if we were to accept Claim Z, it would still be true that I am permitted to override this principle and save David because I know and like him. If saving David is a valid special consideration for Claim Z, it should permissibly be a valid special consideration for Claim X.

In summary, the permissibility of saving David simply because I know and like him threatens Claim X, but it would also threaten Claim Z or any other feasible approach we might have to tackle conflict cases of saving different groups. I have not, in detail, discussed positive reasons for accepting either of these views in this essay. However, Taurek's justification of why we may save one rather than five, because the one would himself be morally permitted to choose his own life over the life of five strangers, leads to very counterintuitive results. This creates a deadlock, and to move forward, our options would be to either deny that I am permitted to save David simply because I know and like him, deny that there is a moral duty to save anyone even if it is at little cost to ourselves or find a different way of justifying why it is permissible to save either the one or the five.

Perhaps there are simply no positive moral duties in such a scenario, but only negative moral duties – that it is not permissible to base our decision of who to save on race or sex, for example. Though Taurek does not appear to attach any conditions to adopting B's point of view when he chooses B's arm over C's life, I suspect he would agree that adopting B's point of view for a reason that is itself immoral, such as racism, would render the act immoral. It would be interesting to further investigate his views on this however. For instance, would I be permitted to save B's arm over C's life because B reminds me of a deceased relative? There is nothing immoral about this preference and yet it seems curious that our moral decision-making can legitimately be shaped by such a consideration, and not by considerations of race or sex. These considerations do not threaten Taurek's position but highlight certain issues that such a view would need to account for.

To conclude, though I do not believe that Taurek offers a successful alternative position to Claim X, he leaves a great many

questions in his wake, which challenge us to question our most basic and obvious intuitions.

Works Cited

Parfit, D. (1978). Innumerate Ethics. *Philosophy & Public Affairs*, 7 (4), 285-301.

Taurek, J. M. (1977). Should the Numbers Count? *Philosophy & Public Affairs*, 6 (4), 293-316.

## **Searle's Chinese Room Reconsidered**

By Niccolò Aimone Pisano

### **Abstract**

In this essay, I will argue that Searle's (1980) argument in its original version does not debunk the computational theory of mind, but it does if adequately modified in the light of Dreyfus's (1992) argument. I will first outline the core ideas behind the view that is meant to be challenged by Searle's mental experiment; then I will describe the thought experiment itself, also exposing, and objecting to, Boden's (1987) reply which highlights two difficulties of the argument, namely SIR's (Searle-In-the-Room) understanding of English and Searle's biological chauvinism. Finally, I will show how taking into consideration Dreyfus's (1992) claim that the computational theory of mind entails some unavoidable regress about the rules to be applied allows Searle's argument to be a more effective attack to the computational theory of mind.

### **Introduction**

In this essay, I will argue that Searle's (1980) argument does not debunk the computational theory of mind, but it does if adequately modified. I will first outline the core ideas behind the view that is meant to be challenged by Searle's mental experiment. Then I will describe the thought experiment itself, also exposing Boden's (1987) reply which highlights two difficulties of the argument; namely, the alleged *petitio principii* due to Searle-In-the-Room's understanding of English, and the unwarranted reliance on the in principle impossibility for a machine to simulate human cognitive behaviours. In doing so, I will show how the former objection is really weaker than the latter. Finally, I will show how taking into consideration some ideas from Dreyfus (1992), concerning

the practical impossibility to specify in advance the rules to be followed by a computational machine in order for it to pass the Turing test, allows Searle's argument to be a more effective attack to the computational theory of mind, in that in this way it hinges upon a proper argument without simply relying on the intuitions conveyed by his thought experiment.

### **The computational theory of mind**

First, it will be useful to spend a few words to explain what the computational theory of mind challenged by Searle's argument consists of. Roughly, a computational machine is a device which formally, i.e. purely syntactically, manipulates symbols following a set of rules and taking into account the state the machine is in. In other words, on the basis of the inputs received and depending on its internal conditions, it can give outputs according to the instructions previously implemented (the programme). For instance, if there is a rule such as “if you read *how are you?*, answer *fine* if you are in good conditions or *not so good* if you are not in good conditions”, every time the machine will receive the input *how are you?*, it will give *fine* as an output if its internal state is regular, *not so good* otherwise. Now, the computational theory of mind consists in understanding the human mind as a computational machine. This means that every human behaviour can be interpreted as the output of the application of some conditional rule, where the inputs are a combination of external stimuli (e.g. the question *how are you?*) and pre-existing internal states (e.g. the actual health of the person the question is asked to).

The most important claim of the computational theory of mind addressed by Searle is that it is possible for a complex enough machine not only to simulate human cognitive performances, but to be considered endowed with the same cognitive skills as the human mind in virtue of this. Such a view can be articulated in two points:

- (1) Human mind can be considered a computational machine.
- (2) If behaviour B cannot be distinguished in type from a previously observed behaviour A, and we do not possess other information about B except for that gathered from its observation, we have to consider both as outputs of the same type of underlying processes.

I have already explained what (1) means. On the other hand, (2) can be clarified by an example: “(a) x has opened her umbrella” & “(b) x opens her umbrella whenever she notices that it is raining” → “(c) x has noticed that it is raining”; hence, (2) entails that whenever (a) occurs it is necessary to conclude that (c), unless instead of (b) we have, say, “(d) x told me that she wants to prove the falsity of the superstition according to which it is bad luck to open an umbrella when you are inside a building”.

From these assumptions it follows that if one were capable to build a machine complex enough to simulate all human cognitive processes, so that it were behaviourally indistinguishable from a human in every situation, we would have to consider such machine as endowed with a mind similar to the human one, in that it would always produce the same sort of phenomenal outputs. This is what the Turing test for cognition is: if a human, after a conversation with a machine, is not able to tell whether his interlocutor is another human or a machine, the machine can be said to have passed the test and can be considered to be thinking like a human.

### **Searle’s thought experiment**

Now that I have explained what position Searle's thought experiment aims to challenge, let us turn to it. Searle imagines that there

is an English speaker (SIR, Searle-In-the-Room) inside a room, completely isolated except for the possibility to receive messages from the outside, to which he is required to write replies using an English rulebook. These messages, as well as the answers, consist of strings of symbols completely meaningless to SIR. It turns out that they are Chinese characters, and the rulebook allows SIR to answer meaningful Chinese questions as correctly as a native Chinese speaker would. Now, given (1) and (2), the external Chinese questioners are induced to suppose that whoever is inside the room is understanding Chinese. But, Searle argues, this is not the case: SIR does not know that he is manipulating Chinese symbols, nor that he is having a conversation with someone: he is just following rules such as “if you see *squiggle*, write *squoggle*”. Thus, the conclusion of the thought experiment is that, since SIR's processing symbols in a strictly syntactical way can be exhaustively equated to a machine's activity, although a machine can pass the Turing test (as SIR does), it does not have any proper understanding of what it is doing. Therefore, since they differ at least in that they can understand things, human minds cannot be instantiations of computational machines. In other words, the fact that in such a scenario the machine proves that the Turing test is not a valid test for cognition, in that it can be passed without there being any proper understanding of the dialogue, leads to the conclusion that (2) fails. But the failure of the second claim of the computational theory leaves (1) unwarranted; hence, the computational theory of mind has to be abandoned.

### **Boden's objections**

There are two difficulties with this thought experiment, as Boden (1987) has pointed out. First, it is not true that there is no understanding at all: SIR does genuinely understand the English rules written in the

rulebook, so that, although his consequent behaviour may not be displaying any understanding, the equation “SIR = computer” is not correct.

Second, Searle claims that there are intuitively compelling reasons to hold that the mind has peculiar features in virtue of its biological realizers; therefore, no non-biological machine can possibly be endowed with equivalent cognitive skill, since different materials do not support them (Searle 1980, p. 421). In other words, Searle claims that even if his thought experiment failed, there are material reasons for rejecting the computational theory of mind, which unacceptably equates human minds and computers. However, Boden argues, there is no clear evidence that this is the case: (so far) we do not know why the brain allows the existence of human mind, thus not being available any explanation, not entirely based on intuition, of why other brain-like structures could not allow it as well. Therefore, holding that artificial brain-like machines could not support human minds seems not to be enough well-supported a claim to be acceptable.

Boden claims that the more important of these two objections is the former, since the latter concerns an easily dismissible claim. I disagree: what is more important is that the intuition behind the second of Searle's claims, if adequately backed up, is what allows to debunk the computational theory of mind. Indeed, it is not hard to modify Searle's thought experiment in order to avoid Boden's first objection, the “English reply”: it will suffice to replace SIR with a system of levers and pulleys that purely mechanically elaborates the output. In this way, no human would be involved, and yet we would not grant to the mechanism inside the room a proper understanding of Chinese. In other words, the English reply is focused on SIR's humanity, while its sole purpose is to more easily convey the intuition on the basis of which the thought experiment is concocted. We could as well replace SIR with a mechanism

whose ignorance of Chinese language we have ascertained before its implementation in the thought experiment. SIR's humanity only makes clearer that, even if inside the room there were something *per se* capable of proper understanding, that specific input-output dynamic would not be evidence for SIR's understanding of Chinese.

### **Dreyfus and the infinite rulebook**

Before replying to Boden's other objection, I will introduce some ideas elaborated by Dreyfus (1992), whose main argument against the computational theory of mind concerns some unavoidable regress about the rules to be applied. In what follows, I will make use of these ideas to reinterpret Searle's thought experiment to make it more effective an attack to the computational theory of mind.

In the first part of this essay I have defined a computational machine as a symbol-manipulating device which operates following syntactic rules that take into account the inputs and the current state of the machine. Simple as it may seem, this pattern is extremely difficult to be realized when it comes to machines able to pass the Turing test. For, as Dreyfus stresses, it is easily applicable only to the most recent form of human language, the formalised one, but it is practically impossible to specify a set of rules exhaustively covering the complexity of the use of natural languages. Since they extensively rely on context-dependent contributions, even a simple question such as "where is the cat?" (not to mention sentences involving indexical terms, that is, terms such as "this" or "here" which entirely depend on the context) would require exceedingly complex computations to be answered, in that it would be necessary to specify things such as what sort of object a cat is, which particular cat we are talking about, what kind of landmarks can be mentioned in the answer, and so on. Therefore, in the light of Dreyfus's argument, there are only two alternatives: either it is necessary to specify

a thorough regulation of all the infinitely many linguistic combinations; or it is necessary to specify an infinite set of rules about when and how to apply a finite set of comparatively simple rules. For instance, on the basis of the first alternative, in the case of the question “where is the cat?”, we might need a set of extremely specific rules such as “if *where is the cat*, and the person-shaped object knows that you own a cat-shaped object, and you are in a building with such and such features, ..., then answer “it is in the living room” if the cat is in the living room, or “it is in the kitchen” if it is in the kitchen, ...”.

Both these tasks cannot obviously be materially accomplished: the first would require an infinite number of “atomic” rules, since the number of the sentences that can be uttered in any natural language is infinite; the other would require a potentially infinite generative process of rules' specification, for analogous reasons.

Now, Searle's argument does not take into consideration this fundamental issue at all. The English rulebook is taken to be comprehensively covering all the answers to the questions asked by the external Chinese speakers, and the focus of the thought experiment is rather on the intuitive unacceptability of labelling SIR's activity as a proper form of understanding Chinese. Nonetheless, an appeal to intuitive disagreement does not count as a satisfactory rebuttal of a theory. This is the reason why I do not think that Searle's “Room” argument debunks the computational theory of mind. However, if it is adequately modified in order to accommodate the aforementioned issue, it does.

### **Searle’s thought experiment modified**

In what follows, I will present my alternative version of Searle’s Chinese Room, which also counts as a reply to Boden’s second objection, as it will emerge. Suppose that instead of a question-answer dynamic,

Searle described a more general command-execution one. It could be ordered something like “the next time I say  $x$ , answer  $y$ ”. In this case, instead of applying a previously encoded rule such as, “whenever you read  $x$ , write  $z$ ”, it would be more appropriate to apply a rule such as:

if *the next time I say  $x$ , answer  $y$* , write in the rulebook “if you read  $x$ , write  $y$ ” and apply

“if you read  $x$ , do not apply 'if you read  $x$ , write  $z$ ', apply 'if you read  $x$ , answer  $y$ ' and 'if

you have applied 'whenever you read  $x$ , write  $y$ ' delete that rule from the rulebook”.

It is clear that it is impossible to write such a rulebook, since, as I have shown before following Dreyfus, it would be needed either an infinite number of “atomic” rules, or an infinite set of rules about when and how to apply some simple rules. Therefore, in order to effectively attack the computational theory of mind, it is better to concede that the Turing test is a valid test for cognition (in accordance to what (2) prescribes), as the upholders of that theory claim, and instead argue that, since as a matter of fact human minds exist while no cognitively comparable computational machine could be built, nor even in principle, human minds are not computational machines (against (1)) In other words, it seems to be a more effective strategy to attack (1) while accepting (2), instead of accepting (1) and challenging (2) as Searle does.

Moreover, suppose that a supporter of the computational theory of mind decides to face the second of the two aforementioned infinite regresses. He could argue that the rules-related regress can be stopped by assuming that there is some pre-normative (i.e. antecedent to and independent from the rules) bottom-level where the input-internal state couple physically determines the outputs (in this case, the formulation of

the rules). That is, the regress mentioned by Dreyfus can be avoided by assuming that there is no need for either an exhaustive enumeration of all the rules or of all the rules for generating the rules: at some point, SIR's production of outputs would be determined by a purely physical, non-linguistic, causal process. But this is where Searle's point which is underestimated by Boden kicks in. If at this basic level the rules, understood as action-reaction patterns, are determined by the physical properties of the substances constituting the machine instead of by linguistically formulated rules, the possibility that cognitive performances might be carried out by a non-human system just as effectively as by a real human is ruled out in virtue of the different physico-chemical properties of distinct substances. It seems, then, that a supporter of the computational theory of mind can ground the rules-generating process only at the price of conceding that no computational machine can ultimately be able to have the same rules-generating process of human minds, because that process would strictly depend on the physical properties of the substances the brain is made of. But if minds are to be conceived of as a particular kind of computational machines, they cannot display properties that no machine can possess. However, if my argument is correct, they do. Therefore, (1), the most important of the two central tenets of the computational theory of mind, has to be abandoned.

### **Conclusion**

It is finally possible to re-assemble all the pieces. Searle's original argument does not debunk the computational theory of mind, intended as the view whose core ideas are (1) and (2), since it merely makes appeal to intuition in order to challenge (2) while accepting (1). However, some modifications in the set-up of the thought experiment, together with a

shift in its focal points justified by Dreyfus's (1992) argument, allow to more strongly attack the computational theory of mind by accepting (2) and rejecting (1), while at the same time avoiding Boden's objections. In fact, on the one hand, the English reply can be easily avoided by replacing SIR with a mechanical system; on the other hand, since on the basis of Dreyfus's argument it is impossible to specify in advance a set of rules that would enable a machine to pass the Turing test, the fact that human minds can actually pass it entails that there must be some pre-normative process (i.e. some process antecedent to the rules) in act. If that is the case, then Boden's second point, according to which there is no guarantee that the physico-chemical properties of the brain are what ultimately leads to the occurrence of proper human cognitive behaviours, must be abandoned. But this means that no artificial computational machine can actually pass the Turing test, given that in order to do so it is necessary to have the same physico-chemical properties of human brains. Therefore, the computational theory of mind can be defeated by Searle's Chinese Room.

## REFERENCES

- Boden, M. A. (1987). *Escaping from the Chinese Room*. University of Sussex, School of Cognitive Sciences.
- Dreyfus, H. L. (1992). *What computers still can't do: a critique of artificial reason*. MIT press.
- Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417-424.

## **Is Fahrbach right in thinking theory change is coming to an end?**

By Silvina Maestro

---

Ludwig Fahrbach puts forward a defense of realism in the form of the argument of the exponential growth of science. This view takes into account the quantity of scientific work done by scientist since around the 18<sup>th</sup> century and proposes that we focus on the increment of the amount of such work in time. In this essay I shall assess Fahrbach's view that theory change is coming to an end. I begin with an explanation of the exponential growth of science approach he uses to support realism against the pessimistic meta-induction argument. I will then assess his refutation of the pessimistic meta-induction and finally I shall follow with a critical assessment of such views. I shall conclude that he fails to show compelling evidence for his argument, that higher degrees of success at best might be associated with a period of theory stability but is not enough to prove the end of theory change. Furthermore, I shall attempt to show that a more critical assessment of figures reveals that the growth of science has not been exponential.

---

**1. Introduction.** Fahrbach defense of realism is based on his argument of an exponential growth of science. He observes that the majority of all scientific work ever done, at least 80%, has happened since 1950. Within this work Fahrbach includes activities such as the making of observations, the conduction of experiments and the creation and testing of theories. His considerations are based upon two criteria, namely the number of journal articles published and the number of people working as scientists in any given period. Taking into account the quantity of scientific work done by scientist since around the 18<sup>th</sup> century he suggests we notice that the amount of scientific work has exponentially increased with time. Using figures coming from bibliometrics, he concludes that the exponential growth in the number of journal articles published has

resulted in a doubling of the rate of publishing to 15-20 for the last 300 years. He does, however, concede that this exponential curve does not align itself with all scientific disciplines but believes that it has certainly done so for some.

In regard to his second criterion, the amount of scientists, he cites Derek de Solla Price when he claimed in 1963 that said amount had doubled every 15 years in the last 300 years and that 90% of scientists that ever lived were alive at that time (de Solla Price 1963, p. 141). Fahrback wonders if that statement is still correct. After considering a slow down in the number of scientists in Europe and America in the 1970's, but also considering the results of an increase in rate in Asia and the number of research doctorates twice doubling in the US from the 1960's to the 2000's he concludes that de Solla Price's claim must also be correct for today. With this in mind, he claims that such a strong increment in both scientists and scientific activities means that the last 15-20 years has seen half of all the scientific work ever done, with three quarters of all such work being done in the last 30-40 years. He will use these doubling rates as the basis for his view of theory change coming to an end and his refutation of the pessimistic meta-induction argument.

**2. The pessimistic meta-induction refutation.** Before Putnam and Laudan, Henri Poincaré might be seen as laying the groundwork for the pessimistic induction argument when he talks about science and its ruins piled upon ruins (Poincaré 1905/1952, p. 160). In support of anti-realism, the pessimistic meta-induction argument appeals to the history of science and the fact that we find many of the successful theories of the past superseded by new ones. The inductive inference to approximate truth that realists attach to empirically successful theories is undermined by the anti-realists who invite us to accept our current successful theories

as candidates for the same destiny of those refuted in the past. Thus, the anti-realists will advise to either disbelieve or remain agnostic about current theories, lest they be proven false in the future.

Those supporting the pessimistic meta-induction, Fahrbach claims, are not considering the exponential growth of science. To further his point he introduces a graded notion of success of theories, according to which our current successful theories benefit from a much higher degree of success than those that having enjoyed success in the past were ultimately refuted. Using this graded notion of success, he attempts to show that by correctly modifying the realist view, it must not be in discordance with the history of science (Fahrbach 2011a, p. 1285).

The increase in scientific work, he claims, has led to an increase in the opportunities for our current theories to be refuted. For him this has not been the case and suggests that a period of theory stability has arrived. The exponential growth of computer and manpower resulted in a steady improvement in the amount and precision of data. At the same time the production of better methods and instruments has resulted in better observations. Thus, the degrees of success of the theories have greatly increased, compared to those refuted in the past when scientist did not have such amount and precision of resources available. Predictions from the past, Fahrbach claims, were comparatively fewer and less precise.

By increasing the degrees of success of our theories he brings them closer to approximate truth and by doing so he creates a form of modified realism (Fahrbach, 2011a, p. 1291) reconcilable with the history of science and saved - or so he thinks - from the perils of the pessimistic meta-induction

**3. Critical assessment.** It cannot be denied that since the first editions of the Philosophical Transactions of the Royal Society or the Journal des sçavans the number of scientific publications has increased. However,

could this observation be enough to substantiate an argument such as that of Fahrbach? That most scientific activity has taken place in the last 60 years and that in said period most of our best theories have remained unchallenged is not sufficient an argument to justify a claim such as the end of theory change. Fahrbach's resourceful proposition only manages to illuminate the way towards higher degrees of success, an insufficient path to arrive at the justification of the end of theory change - his desired destination. Such higher degrees of theory success at best show a slow down in the process of theory refutation or a period of theory stability. Considering the time frames we are given to compare, theories only seem to enjoy a longer-lasting shelf life. This phase of stability is what people like John M. Ziman call the steady-state dynamics of modern science; a phenomenon, he says, followed the exponential growth Fahrbach refers to. Ziman points out that if the exponential growth of science were not to recede at some stage

it wouldn't be long before every man, woman and child would be engaged in research and writing scientific papers (Ziman, 1994 p. 10).

In this period of stability, namely since the 1960's, some claim the age of cited material has risen, with scientist referencing material with a longer life cycle - an increasingly old body of literature. For Larivière et al, bibliometric data from macro-level patterns of usage of scientific literature over the last 100 years show that the exponential growth started to level off at the end of the 1970's; the advent of both world wars and the consequent decrease in publishing resulted in an increase in periods where scientists had to rely in older literature. Interestingly he points out that, for subfields such as nuclear and particle physics and astronomy, after the creation of e-print servers like arXiv the trend shows an increase in new material being cited. Fahrbach leaves physics outside of his considerations, as it appears to him to need special treatment. All things

considered, this increase in use of new material, Larivière claims, only averages to a period of Kuhnian normal science, even with some countries like China still enjoying an exponential growth rate, the majority of countries display low rate growth or stagnation (Larivière et al, 2008, p.291). The increase in quantity and precision in observations that Fahrbach talks about can thus be seen to be counteracted.

A more critical assessment of the figures might show that the growth of science in the 20<sup>th</sup> century was neither exponential nor constant. Mabe and Amin describe a split in three phases with the first (1900–1944) and the last (1978–1996) having identical slower rates, for them showing a tendency of the scientific activity - not unlike organisms - towards equilibrium (Mabe and Amin 2001, p. 157).

They perform a filtration of the data found in the Ulrich's International Periodicals Directory, Summer 2001 Edition, based on a number of criteria, namely that the journals must be a serial publication, classified as academic/scholarly, that the journal is still publishing at present and hence classified as active and most importantly that they are refereed as stipulated by the Ulrich's classification scheme – a scheme recently amended and allowing for more realistic calculations. Once this filtration has been done the number of journals amounts to 14,694 for 2001. This figure is markedly different from the previous estimations performed by de Solla Price, who considered that the number of periodical titles published by the end of the 20<sup>th</sup> century would exceed the million (Mabe, 2003, p. 192).

When the calculation is repeated for every year in the Ulrich's database since 1665 a growth curve resulting of the number of scholarly and scientific journals created and still active for the 338 years studied can be built. It shows, using a logarithmic scale, a calculation of the average

increase that states an almost consistent growth of 3.46% per annum. Growth rates then, are shown to have been strikingly consistent in time from the 1800 to present day, with 3.25% growth from 1860/1900 to 1940 and 1976 to today, including a middle period from 1945 to 1975 with a higher growth.

Also not considered by Fahrbach is the influence that world events such as wars and the fluctuating willingness from governments and institutions to apportion funds to research has in scientific undertakings. For Mabe, the period since 1977 is a period of conservative growth where the science and technology investments of the past failed to match what was anticipated and overly enthusiastic expectations of a more naïve era result in dissatisfaction, disenchantment and scepticism (Mabe, 2003, p. 194).

Fahrbach is right in pointing out that the acquisition of data has benefited greatly from a steady improvement in precision as well as diversity, and that automation has resulted in huge amounts of data being generated. Laborious number crunching activities can now be performed at great speeds by increasingly more powerful computers that double their capacity almost every two years (Fahrbach, 2011a, p. 1290). However, this is not enough to support his view. That better and more observations, instruments and data assist in the substantiation of better theories is not enough to rescue them from eventual refutations. Moreover, some claim that important data in the conduction of research is not deemed important or valuable enough to be kept; such data, however precise, is inaccessible to the larger scientific community (Wynholds et al, 2012 P.22). Fahrbach fails to establish a significant difference between those theories created in the initial 20% of the history of science to those in the 80% he holds as evidence to the end of theory

change. If this is then only a matter of degree Laudan's argument about approximate truth and success still follows (Laudan, 1981, p.33). Even if in possession of more empirical success, our current best theories are not completely immune to future obsolescence.

There are other reservations one might have when assessing Fahrbach's view. His approach of utilizing rough estimates (Fahrbach, 2011b, p.146) might have lead him into generalizations that could have the power to wound his argument fatally. By not exploring in more depth the figures he fails to notice the difference between "real" and "artificial" growth in science. Michels and Schmoch, for example, point out that even though there seems to be a growth of roughly 34% in article numbers in the period of 2000–2008, at least 17% is accounted for the inclusion of older literature that had not previously been included in databases (Michels and Schmoch, 2012, p.845). Using data from the number of articles on the Web of Science (WoS) in the period of 2000-2008, they assert that the growth in the amount of papers might be linked to simply the additional coverage by this database source of journals that already existed.

They conclude that the number of publications cannot be interpreted as an increase in scientific activities or the growth of science itself but as a separate growth in the structures of a particular field, for example the enlarging of databases. The increase arises from the addition of old journals already published some time ago but that have not yet been incorporated in the database. They also cite as a reason for the growth of number of publications the updating of databases by including amended features for citing author's names by including their full first name rather than just their initials.

There seems to be no way of differentiating between new, additional and recurring journal categories, and hence no differentiation between an actual growth of science or just journal's policies of enlarging the

coverage of their database. This, they say, can account for the increase in numbers and not the growth of science.

Another worry is that of the problem of accurate counting of unique work. In their research Larsen and von Ins have come across technical problems in the counting of papers where different publications use diverse methods; the use of certain methods can lead to the double counting of material. They also conclude that after taking into account co-authored papers, the productivity of unique contributors seems to have decreased from 1 to 0.8. (Larsen, 2010, p.593).

It can also be argued that the nature of scientific publishing has changed over time. Gross & Gross could be said to be the first to point out the overflowing of scientific literature to non-academic sources (Gross & Gross, 1927, p.388), a trend that has continued with the advent of commercial publishers. Mabe thinks that there seems to be a shift, with the publishing system shifting its drive from learned societies to a mixture market composed of commercial and society participants (Mabe, 2003, p.194). One might wonder if commercial or indeed professional considerations have influence over the quantity and quality of the material published. With publications under commercial pressure and with professional reputations to be built, it could seem like a good idea to focus on a quantitative rather than a qualitative approach, something that can stymie a bibliometric assessment of the real growth in science.

#### **4. Conclusion.**

Fahrbach fails to show compelling evidence for his argument. That most scientific activity has taken place in the last 60 years and that in said

period most of our best theories have remained unchallenged is not sufficient an argument to justify a claim such as the end of theory change. Higher degrees of success at best might be associated with a period of theory stability but are not enough to prove the end of theory change. A more probing assessment of figures suggests that the growth of science has not been exponential and that its fluctuations are dependent on socio-economical and political factors unscrutinized by Fahrbach. I hope I have also advanced the idea that the increase in the number of publications cannot be interpreted as an increase in scientific activities or the growth of science itself since there are fundamental problems with considering the number of publications simpliciter as a measure of growth.

REFERENCES:

de Solla Price, D. (1963). *Little science, big science*. New York: Columbia University Press.

Fahrbach L. (2011a) Theory Change and Degrees of Success. *Philosophy of Science*, Vol. 78, No. 5 (December 2011), pp. 1283-1292. The University of Chicago Press on behalf of the Philosophy of Science Association

Fahrbach, L. (2011b). How the growth of science ends theory change. *Synthese*, 180, 139–155.

Larivière V, Archambault É, Gingras Y. (2007). Long-Term Variations in the Aging of Scientific Literature: From Exponential Growth to Steady-State Science (1900–2004). *Journal of the Association for*

Information Science and Technology. Volume 59, Issue 2, 15 January 2008, Pages 288–296.

Larsen, PO, von Ins M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics* 84 3 575–603

Laudan L. (1981). A Confutation of Convergent Realism .Philosophy of Science, Vol. 48, No. 1 (Mar., 1981), pp. 19-49Published by: The University of Chicago Press on behalf of the Philosophy of Science Association

Mabe, M. A., & Amin, M. (2001). Growth dynamics of scholarly and scientific journals. *Scientometrics*, 51, 147–162.

Mabe, M. (2003). The growth and number of journals. *Serials*, 16(2), 191–197.

Michels C. , Schmoch U. (2012) The growth of science and database coverage. *Scientometrics* (2012) 93:831–846

Poincaré, H. (1905/1952). *Science and hypothesis*. New York: Dover Publications, Inc.

Wynholds, L; Wallis, J ; Borgman, C ; Sands, A; Traweek, S. (2012) Data, data use, and scientific inquiry: two case studies of data practices. 8300 defect for UNSW Proceedings of the 12th ACM/IEEE-CS joint conference on digital libraries, 10 June 2012, pp.19-22

Ziman, J.M. (1994). *Prometheus bound: Science in a dynamic steady state*. Cambridge, England: Cambridge University Press.

## Denied Entry: a Rawlsian Defence of the Moral Right of the State to Close its Borders

By Liam Fields

### Abstract

I justify the moral right of states to close their borders to economic migrants on Rawlsian grounds: the common interest in security that would be chosen in the original position allows the state to prioritise that over an individual migrant's right to liberty. I do not make a normative claim that they *should*, but rather that morally they *can* opt to exclude migrants.

### I. Introduction.

In this paper, I will argue that states have a right to exclude economic migrants, and that this right can be justified on Rawlsian grounds. A right to exclude would give the state the right to close its borders and deny entry to economic migrants if it so wishes.

Many of the arguments will also apply to migration by those in dire need, for example refugees, and as argued by Wellman (2008) it may well be the case that states have a right to exclude even those. While there will inevitably be some overlap, my aim is solely to show that states have the right to exclude economic migrants.

I will begin by setting out the right to exclude itself, as argued for by Wellman (2008). I will then look at the competing claims of migrants, examining the theories of Fine (2013) and Carens (1987) which argue in favour of open borders. Finally, I will show that although these theories are persuasive, the right to exclude can be defended against them when practical realities are brought to bear.

### II. The right to exclude entry.

For Wellman (2008) nation states have a right to freedom of association analogous to that held by individuals, and this further entails a right to exclude entry to immigrants. He argues that freedom of association is a key aspect of “...*dominion over our self-regarding affairs*” (p.110), which entitles individuals to choose who to

associate with, particularly in important areas like marriage and religion. An intrinsic part of this freedom is a right to refuse association - one does not have to marry a particular person, or become a buddhist, unless one chooses to do so.

He contends that the state's right to exclude entry is no different to the individual's right to remain single in any relevant way. Drawing this analogy is straightforward - as a person might choose a husband, a collective might choose whom, if anyone, it would like to join its community. It seems intuitively apparent that it would be wrong for Iran to be forced to join NATO, or for the UK to annex France, and the wrongness of these cannot, Wellman posits, be explained without reference to the right to autonomy these states possess.

Pevnick (2011, 30) disputes Wellman's analogy, arguing that membership by birth does not equate to free association, but this argument is not entirely convincing. We are incapable of agency at birth, and we readily accept that our parents make life decisions for us until we are able to do so ourselves. Membership of a society is arguably equivalent; our parents simply make the decision to enter into association on our behalf. In liberal societies at least, parents have a choice of where their children become citizens of, albeit one limited by the availability of alternatives. With this being so, I will continue by assuming the validity of Wellman's analogy, foregoing a defence of Pevnick's (2011) institutional ownership argument in favour of a right to exclude entry.

Wellman's analogy therefore establishes that states do have a right to exclude economic migrants. He argues that as it is rooted in an entitlement that it is a deontological right, but emphasises that it is not absolute - it can be defeated by competing considerations (p.117). Fine (2013) builds on this point, asking why the state's right to exclude outweighs the competing claims of the immigrants to be admitted.

She argues that the difficulties and costs of immigrating, including learning languages, abandoning social support networks, and building a life in an alien culture imply that people who do it must have good reasons (p.263). She highlights that the right of a state to exclude migrants is far more significant in a person's life than the right held by voluntary associations; no one state can necessarily

replace another, nor can one readily found their own state, so exclusion of immigrants has far greater consequences. These are valid concerns, and necessitate a closer look at the competing claims of immigrants, and the moral argument for open borders.

### III. Competing claims: open borders in theory.

Fine (2013, 256) notes that the exclusion of immigrants by nation states is a recent development, only becoming commonplace in the late 20th Century. She argues that the right of states to do so is contrary to a “*..fundamental liberal commitment to the moral equality of all people*”. Per Cole (2000), exclusion of immigrants cannot be justified on liberal grounds, as it is inherently morally arbitrary.

Perhaps the strongest liberal argument along these lines is made by Carens (1987), who equated citizenship in Western liberal democracies to feudal privilege, an inherited status that enhances your life chances for morally arbitrary reasons. His argument is based on John Rawls’ (1971) theory of justice as fairness, which posits that just principles for governing a society would be a chosen from behind a ‘veil of ignorance’ in the ‘original position’. A person in the original position has no knowledge of their own characteristics, including their gender, age, race, place of birth, or concept of the good, and would therefore choose principles impartially and rationally. While Rawls himself expressly limits his thought experiment to a closed system where the question of immigration could not be addressed, Carens (1987, 256) makes a convincing argument that it is a relevant way of looking at it. By examining immigration from the original position we avoid self interested and partisan considerations, treat all humans as free and equal moral persons, and remove morally arbitrary ‘*..natural and social contingencies*’.

Carens (p.255) follows Rawls’ approach of first setting out an ideal theory, which makes a number of idealising assumptions about the world in order to reach a set of fundamental principles by which we might reform the non-ideal world. Under ideal theory, we assume everyone will adopt the principles we set out, and that there are no countervailing

negative social conditions (Wenar, 2017). Carens thereby sets out a standard by which we would judge an ideal society.

When evaluating a principle from the original position, we adopt the perspective of the person it most disadvantages, as once the veil of ignorance is lifted, we may find it is we who are so disadvantaged. If we accept, as per Fine (2013) above, that an immigrant has compelling reasons to migrate, then they are most disadvantaged by exclusion and it is their perspective we must adopt. Carens (1987, 258) assumes that the benefits of decentralisation might persuade us to maintain nation states in an ideal world, so assumes we would keep them. In such a perfect world, he argues, even where states refrain on Rawlsian grounds from restricting religious freedom or sustaining inequalities between states, there are still many valid reasons for individuals to become economic migrants. Individual economic opportunities will vary between states, one might fall in love with a person living on the other side of the world, or one might want to live among people who share one's religious beliefs. Migration might, therefore, be essential to the migrant's plan of life. This provides us with a particularly strong competing claim against the right to exclude.

How might we adjudicate between these claims? Under Rawls' 'justice as fairness', liberty may only be restricted to ensure the preservation of public order and security, as these are ultimately necessary for all other liberties (Rawls, 1971:p.187). If we were to follow this, as Carens (1987) does, it would need to be shown that open borders would lead to a reasonable expectation of chaos and the breakdown of order in order to decide in favour of the right to exclude. If we accept all the assumptions of an ideal world, such an expectation seems far from reasonable.

The pursuit of an ideal then, to create a world in which morally arbitrary contingencies like birthplace and parentage have no justificatory strength, gives us a compelling reason for states to open their borders. Is the pursuit of an ideal sufficient justification to force states to open their borders in the real world, where we have to face historical obstacles in the absence of idealising assumptions?

#### **IV. Competing claims: open borders in practice**

In practice, there are vast social and economic inequalities between nations, and these conditions make it far more likely that migration controls are needed to impose or ensure order. A common argument concerns the size of demand (Carens, 1987); that the number of people in poor countries who would migrate to wealthy countries with open borders far exceeds the number that could practically be taken. For Carens, if the goal is to prevent wealthy countries being overrun, present restrictions on migration are too restrictive.

We would struggle to use the size of demand to justify the right to exclude on Rawlsian grounds however. Under the original position we have to prioritise the worse off, which means that we would have to show that the economic well-being of current citizens would be lowered below that of the migrants in their home country. Further than that, liberty is prioritised over economic condition, so the financial concerns of current citizens are largely irrelevant.

To justify the right to exclude on Rawlsian grounds then, you would need to show that there was a reasonable expectation that opening borders to migrants would cause chaos and the breakdown of order, leading to a net loss of liberty. This is more plausible in the absence of idealising assumptions. Carens (1987) warns that to think of migrants in this way is the territory of bigotry and ignorance, but even if we accept this, evidence suggests it *is* a reasonable expectation.

For example, Boyes (2017) connects an increase in Muslim migration to Sweden and Germany with increases in anti-semitism, including the firebombing of a synagogue in Gothenburg and the vandalism of a cemetery in Malmö. While the immigrants Boyes refers to are refugees, the principle is the same for our purposes. A recent study by the Zurich University of Applied Sciences, funded by Germany's Ministry of Family Affairs, showed that Lower Saxony experienced an increase in violent crime of 10.4% in 2015-2016, 90% of which was attributed to migrants (see, e.g., Deutsche Welle, 2018). The same study found that migrants from war torn places, like Syria and Afghanistan, were significantly less likely to commit violent crime; we might infer from this that economic migrants are *more* likely to commit violent crimes than refugees. Without

a right to exclude then, a state becomes vulnerable to chaos and disorder, as evidenced by violence and citizens being unable to express their religious beliefs in public.

Note that the justification sought here is dependent on a breakdown of social order, not merely on a challenge to the distinctive national cultures of, say, Sweden and Germany. Fine (2013, 261) quotes Miller (1995), who argues that distinctive national culture is necessary to “*provide citizens with a sense of their collective identity and belonging, and a background against which more individual choices about how to live can be made*”. Fine notes that such considerations don’t count under the original position however, as nobody behind the veil of ignorance would risk the possibility of foregoing an important right of freedom for a cultural ideal that might be irrelevant to them.

Boyes’ example, however, demonstrates chaos and the breakdown of order on a local scale, and is indicative of how it might occur on a national scale. With open borders, it is plausible that a large number of immigrants with values antithetical to the indigenous citizens would move to the state. If a small number of refugees are minded to incite violence, and prevent the state from protecting the religious freedoms of its citizens in localised places, we can infer that a larger (and unlimited) number of immigrants could be so minded across a larger area. This is sufficient, I think, to give us a reasonable expectation of chaos and the breakdown of social order; we need not defend exclusion merely on the grounds of a threat to the host culture. The right to exclude can therefore be justified on Rawlsian grounds.

## **V. Conclusion**

It seems clear that a person’s birthplace is morally arbitrary, and the argument that you should be free to migrate to the place where you have the best opportunities in life is a powerful one. In a perfect world, it would be indefensible to close state borders to migrants, and it is extremely difficult to advocate for such a position from behind Rawls’ veil of ignorance. Nonetheless, all liberties are ultimately dependent on security, as without it the state cannot uphold them, and the permissibility of prioritising security from the original

position means that states have a moral right to exclude economic migrants.

## VI. Bibliography

- Boyes, R. (December 20 2017) Muslim migrants behind rise in antisemitism. *The Times*. Available from: <https://www.thetimes.co.uk/article/muslim-migrants-behind-rise-in-antisemitism-rxdsjx2vt> [Accessed December 20 2017].
- Carens, J. (1987) Aliens and Citizens: The Case for Open Borders. *The Review of Politics* 49(2), p251-273.
- Cole, P. (2000) *Philosophies of Exclusion: Liberal Political Theory and Immigration*. Edinburgh, Edinburgh University Press.
- Deutsche Welle (January 03 2018) German study links increased crime rate to migrant arrivals. *Deutsche Welle*. Available from: <http://www.dw.com/en/german-study-links-increased-crime-rate-to-migrant-arrivals/a-42006484>.
- Fine, S. (2013) The Ethics of Immigration: Self-Determination and the Right to Exclude. *Philosophy Compass* 8(3), p254–268.
- Penvick, R. (2011) *Immigration and the Constraints of Justice: Between Open Borders and Absolute Sovereignty*. New York, Cambridge University Press.
- Rawls (1971) *A Theory of Justice*. Revised edition. Cambridge, Massachusetts, The Belknap Press of Harvard University Press.
- Wellman, C. (2008) Immigration and Freedom of Association. *Ethics* 119(1), p109-41.
- Wenar, Leif, "John Rawls", The Stanford Encyclopedia of Philosophy (Spring 2017 Edition), Edward N. Zalta (ed.) Available from: <https://plato.stanford.edu/archives/spr2017/entries/rawls/>

## **The Practicality of an External World**

By Julius Schwarzwälder

### **Abstract**

This essay aims to identify the central epistemological argument of G.E. Moore's 'Proof of an External World' (hereinafter 'PEW') as well as its application in the following scholarly work on it and examines it on the realisation of Moore's three conditions for a good proof: First (section 1), it checks validity and concludes that both arguments are valid. Second (section 2.1), it considers circularity of the arguments and establishes that a sceptic can rightfully object to Moore as him failing to address her objections by begging the question. Third (section 2.2), it examines soundness and contends that, in spite of 2.1, Moore can know his premises, as, by way of a pragmatist notion of truth, a dissolution of the issue is offering itself cogently.

### **The Form of Moore's Arguments, Epistemologically**

Divergent readings of PEW have led to debate on whether the nature of its core argument is epistemological or metaphysical. [MP] This essay will investigate solely its epistemological implications and for that purpose, I<sup>10</sup> will frame the structure of Moore's 'proof' (hereinafter 'the proof') as follows:

- iv. Here is a hand.
  
- v. There is another.

---

<sup>10</sup> This interpretation is common. Cf. a shorter version in [Wright]

vi. If there is a hand here and another there, then two hands exist.

) (4) Two hands exist<sup>11</sup>

(5) Each specific hand is a specific thing.

) (6) Two things exist.

(7) If there are things, i.e. if things exist, there is an external world.

) (8) There is an external world.

Conclusions (4) and (8) are applications of modus ponens,<sup>12</sup> while (6) is derivable in predicate logic with identity. Translated into a formalised argument, the proof, stated in what I take to be the most explicit and text-immanent formulation, is clearly valid.

Whilst not demonstrated as such in PEW, scholars have often [MP] drawn Moore's argument up as a response to different sceptical arguments. There are many and indeed many different sceptical arguments, but the ones relevant to this discussion (hereinafter S) are those that challenge the knowability of Moore's premises and are schematically opposable to PEW in the following valid way:

(A) If S is true, then Moore does/can not know the veracity of the

---

<sup>11</sup>There are thousands of different things that would, instead of (4), be sufficient for the proof. Thus, (5

<sup>12</sup> Modus Ponens is a technical term for a logically valid form of argument and denotes the following basic inference: (1) if P then Q. (2) P is true. (3) Therefore, Q is true

proof. (B<sub>1</sub>) S is true.

) (C<sub>1</sub>) Moore does/can not know the veracity of the proof.

Exemplary for S would be that Moore could, unbeknownst to him, merely be dreaming to have his hands in front of him and stand in a lecture hall. The force of the sceptical challenge hinges not on whether or not Moore really is dreaming, it is not an empirical matter. It problematises that Moore *can* not know whether he really holds his hands in front of him. A response (hereinafter ‘the response’) to this is what has come to be known as a ‘Moorean Shift’ - changing the opponent’s modus ponens to a modus tollens,<sup>13</sup> which are, once more, both formally valid methods of argumentation:

(A) If sceptical argument S is true, then Moore does/can not know the veracity of the proof. (B<sub>2</sub>) Moore does (and, antecedently, can) know that the proof is true.

) (C<sub>2</sub>) There is an external world.

## 2 **Reevaluating the Problem**

In PEW, Moore asserts that ‘[his argument] would not have been a proof unless three conditions were satisfied’ and checks their satisfaction, positively. [PEW] I covered the third one, validity, in section 1. Two remain to be re-checked: whether the arguments are non-circular and whether Moore knew his premises.

---

<sup>13</sup> Modus tollens, similarly to modus ponens, is a logically valid form of argument, but ‘the other way around’: (1) If P then Q. (2) Q is not true. (3) Therefore, P is not true.

## 2.1 Circularity and Question Begging

Moore, in an attempt to demonstrate that his conditions are satisfied in PEW, claims that

(4) is different from (1) and (2), and he does so rightfully, for they are of exemplary nature: named references of (4) - mere instantiations of a general claim. Thus, the conclusion could be true, whilst the premises were not. But he omits that, as the proof is one of an external world, the first conclusion entails (5)-(8) and misses the issue at hand.

For truth to be transmitted from the premises to the conclusion (this is now no longer a formal matter), the former must establish the latter independently of it. If they fail to do so, the argument assumes what it is trying to prove. The central proposition (the aim) of the proof is (8) and as such, it must be both affirmable and deniable. So we can infer that (8) can be denied; by a sceptic, for instance.

But to be able to believe that (1) and (2) are true, one *first* has to believe that (8) is true. Since if one does not presuppose that (8) is true, the perceptual warrant - which is the lynchpin for Moore<sup>14</sup> - one has for (1) and (2) is naught, as it could, after all, be an hallucination or dream Moore falls prey to. But (8) is the very thing in contention; so, as Moore has to affirm what the sceptic denied *already* in his premises, his opponent can accuse him of begging the question - a circular way of argumentation.

---

<sup>14</sup> Cf. his emphasis on indicating by gesture

The response, whilst in itself not circular, is also inflicted by question begging. This holds, as the proof is contained in (B<sub>2</sub>), which was just shown to be circular.

## 2.2 On the Knowability of Hands

Moore does not specify what it is to know things. The analytical canon - post-Gettier - holds that for something (such as Moore's premises) to be known, it has to be a true, justified belief, with adequate links between these conditions. [Bieri p. 77–78] I will consider the first two of these requirements. As truth of the premises would also entail soundness of Moore's arguments, I will consider it separately. Thereafter, I will tackle the issue of justification of his claims.

### 2.2.1 Three Takes on Soundness

Moore, guardian of common-sense, defends what can be called loosely 'the layman's perspective'. To any non-philosopher, it is of such abundant clarity<sup>15</sup> that there is an external world that any debate seems superfluous. Moore, although engaging in that debate, too seems to be unwilling to give this stance up and will defend his intuitions as being sufficient for knowledge - implying that his premises are, i.a., true - no matter what: *How absurd it would be to suggest that I did not know it, but only believed it, and that perhaps it was not the case!* [PEW]

But if the rich history of philosophical scepticism has brought certainty

---

<sup>15</sup> "I am sitting with a philosopher in the garden; he says again and again 'I know that that's a tree', pointing to a tree that is near us. Someone else arrives and hears this, and I tell him: "This fellow isn't insane. We are only doing philosophy.'" [Wittgenstein §467]

to anything, it is the fact that not all concede Moore's intuitions as objectively true.

I will outline one such sceptical position (held by an imaginative sceptic, to which I will for now remain impartial) that poses a significant challenge to Moore in a bit more detail. My sceptic is in some form agnostic<sup>16</sup> concerning the question on whether there is an external world. Such a non-committal alleviates her from the burden to prove existence or non-existence of the world<sup>17</sup>. But the sceptic is thus also someone who purports to know an S, as outlined in section 1. Furthermore, this sort of scepticism is local, not global, as it does not doubt everything altogether, but only withholds committing itself to the truth of a proposition or its negation in exchange for asserting that this withholding is necessary. Local is preferred to global scepticism, as, in order for an objection of a sceptical challenge to be relevant, it must consider its propositional content to be coherent and meaningful. 'Real' (if there can be such) global scepticism ends in infinite regress, an endless doubt without footing.<sup>9</sup> To this sceptic, whose argument seems plausible so far, Moore's arguments seem both faulty because, as presented in section 2.1, they are indeed begging the question.

But this agnostic sounds eerily familiar, if one recalls one of Moore's ending remarks: *'How am I to prove now that "Here's one hand, and here's another"?' I do not believe I can do it. In order to do it, I should need to prove [...]*

---

<sup>16</sup> Exemplary for an agnostic position is, as in 1, that we can not know about the existence of the external world because we, unbeknownst to us, might be dreaming.

<sup>17</sup> It plays no role for present purposes whether the non-commitment is due to the belief that we can not *right now* know that there is an external world, or that we can not know *it at all*.

*that I am not now dreaming.*' [PEW] But if Moore was so acutely aware (even more explicitly so in [Reply]) of this challenge, how are we to assess his arguments<sup>18</sup>. Following Wittgenstein, I believe that Moore was not confused, but somehow '*perfectly right*' [Wittgenstein §397], as even after taking the strongest sceptic, the agnostic, into account, his arguments, still, 'look' sound. I offer a reconciliation of this tense situation by way of a pragmatist notion of 'truth'. For brevity's sake, I can only outline how the relevant aspects of such a theory can be *applied* here.

Instead of asking for the conditions of a good proof, one can, one step prior, demand for inspection whether the conditions for a good problem are satisfied. Typically, a good problem is one for which there is some challenge, causing doubt of hitherto accepted beliefs, to defeat. As, unless we meet a particularly mischievous sceptic (remember, this includes the agnostic), the outcome of her argument will be overruled by practical behaviour - one will carry on living *as if* there really was an external world. The *practical* becomes the *conceptual* [Peirce]. But now this acting *as if*, which is linking up experiences coherently and to the best explanation, has given sufficient reason to end the debate altogether, which inclines us to hold it, along with Moore's premises, to be, instrumentally, 'true'<sup>19</sup>. Whether we could demand more from a set of beliefs than to be functioning when seeking to find an account of known truths is a question beyond the scope of this essay.

## 2.2.2 Foundations and Justifications

---

<sup>18</sup> Cf. [Wittgenstein §115].

<sup>19</sup> Relevant are: [Wittgenstein §§2, 4, 7, §§218–222].

When pressed by questions such as ‘You can not know whether there is an external world because of S. How can you claim to?’ or simply ‘How do you know that there is an external world?’, the scholars’ Moore’s response essentially consists of saying ‘But I do know it, because I know it.’, which, as previously noted, seems like an unsatisfying reply.

I identify the issue in different accounts of what is foundational. A sceptic demands a rational causal justification [Bieri p. 293–296] proving a necessary (non-)existence, while (my proto-pragmatist) Moore is content with conceptualised content of his action (or, dependent on the reading, perception) as foundational knowledge to justify further knowledge.

These different criteria for justification clearly hinge on what ‘truth’ is taken to mean, as shown in the previous question. Which to adopt decides whether we can move a step closer to Moore’s second condition (remember, the ‘adequate links’-condition has not been approached yet).

### 3 **Conclusion**

Moore’s arguments remain ineffective vis-à-vis an agnostic, as the proof is circular and the response, or what it was reduced to in 2.2.1, tautologous.

However, there is a strong case to be made that, in spite of this, one need not get convinced by objections to PEW, as Moore’s insistence on the truth of practically unchallenged intuitions is, in the eyes of a pragmatist, the sensible thing to do. (Dis-)agreement is deeply contingent on the epistemological approach one takes to the arguments.