RASK 50
AUTUMN 2019

Improbable Fairness: Reviewing under the lenses of Impact Factor

DAVIDE SECCHI AND STEPHEN J. COWLEY

Centre for Computational & Organisational Cognition, Department of Lanaguage and Communication, University of Southern Denmark

ABSTRACT
This study presents an agent-based simulation model of a peer-review process to study how bibliometric measurements may affect the assessment of a paper. The Impact Factor (IF) is selected, for it is diffused and a well-known source of judgement bias. By using affiliation to community of researchers that are more or less tied to an understanding of *quality qua IF*, the model simulates how much peer evaluations connect to scientific value of a paper, thus affecting the review process. Simulation findings show that the effect of IF extends to those reviewers affiliated with an IF-agnostic group as well.

KEYWORDS
peer-review, impact factor, distributed/systemic e-cognition,organisational cognition, agent-based simulation model

1. Introduction

This article is concerned with a relatively straightforward research question: how do attitudes toward the formerly Thomson Reuter's Impact Factor (IF) affect the reliability of the review process? Consistent with the literature (e.g., Callaway 2016; Frey & Rost 2010; Moed & Van Leeuwen 1996; Seglen 1997), the angle with which we tackle the issue is the assumption that IF has no relation with the quality of either a journal nor the articles published in it. Given this assumption, we study cognitive and psychological biases (e.g., Gigerenzer & Selten 2001; Kahneman 2003) related to the perception that quality is represented by IF, held by some reviewers. In addition to that we also attempt to understand the influence that reviewers receive from professional academic communities with stronger or weaker conviction on IF as a quality measure.

Consistently with a recent article on the perception of scientific value and IF (Secchi & Cowley 2018), we develop an agent-based simulation model to reason on some of the critical aspects of peer review. In the following pages, we introduce a theoretical framework that is rooted in cognitive biases and offer a distributed cognition perspective. We then introduce our simulation model and analyse a selection of results before offering some concluding remarks.

secchi@sdu.dk

2. Theoretical framework

The influence of IF on the reviewing process is analysed here through a quick overview of some of the issues that afflict it. The index is then compared to some of the biases that affect judgement and decision making. Some of these mechanisms are socially construed and are supported by socially-binding cognitive processes. By the use of a *distributed e-cognition* approach – a perspective that considers cognition not solely limited to a functioning brain but to a system of interacting resources internal and external to the brain (Cowley & Vallée-Tourangeau 2017) – , we outline how organisations may play a role in reinforcing the use of IF as a bias.

2.1 Issues concerning the Impact Factor

According to its creator, "[a] journal's impact factor is based on 2 elements: the numerator, which is the number of citations in the current year to any items published in a journal in the previous 2 years, and the denominator, which is the number of substantive articles (source items) published in the same 2 years" (Garfield 1999: 979). Every year, the *Journal Citation Report* provides this calculation for a selection of journals from a number of different fields. When the index is considered to represent journal reputability or it is used to judge the quality of specific articles (Garfield 2003), several issues arise (Callaway 2016; Vanclay 2011). We mention three: reliability, disciplinary specificity, and distribution.

Reliability. The IF is sometimes used to assess whether an article has some inherent quality. This is because high IF journals are supposed to be more restrictive in their assessments, hence articles appearing in their volumes and issues are believed to be tied to higher scientific standards (Garfield 2003: 365). However, this does not mean that any specific article will be cited, as also Garfield noted. But, if this is the case, then the high IF of that journal is unreliable as far as single articles are concerned.[1] In a noticeable analysis, Seglen (1997) indicates how individual articles' citations do not correlate consistently with the IF of the journal they are published in. The implication is that one should not use IF to assess scientific research outcomes (Osterloh & Frey 2020). Along these lines but from a different perspective, it is worrisome that some journal editors may develop the tendency to skim papers having possible citation counts and IF in mind rather than scientific quality, hence affecting reliability quite significantly.

Disciplinary specificity. The two-year horizon of the original IF index may be well suited to some disciplines as opposed to others. For example, those tied to high-tech and, traditionally, medicine and biology, might have citation scores that reflect quickly on a two-year time span. It may take a much longer time for the number of citations to increase for articles in other disciplines—such as economics and business. For this reason, a five-year IF has been introduced by Thomson at a certain point in the history of the index. However, it is unclear how this five-year index compares to the two- and when one or the other should be used. The two-year reference index remains the most commonly used standard. This uncertainty has been highlighted at

1 A clarification remark is necessary here. The idea that a journal with higher scientific standard is more likely to publish, on average, higher standard articles may sound logically consistent. However, even if we could agree on what higher scientific standards are in each discipline—and we do not—there is no known metrics to identify them. In the text, we use expressions such as "supposed to" and "believed to" to highlight this issue.

various times, by denouncing that the two-year time span is arbitrary (Walter et al. 2003) and that adjustments by discipline may lead to changes in the journal rankings (Curry 2012).

Distribution. IF users also tend to associate all articles published in a journal to the "reputability" of that journal. This is usually a conspicuous mistake because there are only a few articles that attract a high-to-very-high number of citations while the vast majority attracts much less citations. For example, as Callaway (2016) shows in his study, only a very limited number of articles published in *Nature* attract a number of citations that are up to or more than the journal's IF. In other words, there are only a very limited number of articles that appear to be the "drivers" of IF (Colquhoun 2003). Other issues have been raised such as the limited number of journals in the ISI Thomson database (Bloch & Walter 2001), the influence of self-citations (Seglen 1997), and the assumption that all citations carry a positive assessment of the cited material (Walter et al. 2003). However, the three concerns above are probably the most relevant to the topics of this article, because they are more tightly related to judgemental biases.

2.2 The bounded rationality of reviewing

From what written above, an attempt to attach the perception of research quality (or *perceived scientific value*; PSV in our model) to IF lies on shaky grounds. And yet, it seems that governmental agencies, research institutes, and universities are very much likely to employ IF as a shortcut towards more fine-grained judgement of research. The ABS list in the UK, the ABDC ranking in Australia, or the Danish BFI list are all examples of some use of the IF, or of a combination of IF and its surrogates. As indicated in a recent study (Osterloh & Frey 2020), these decisions affect thousands of academics and their careers, with effects that are beyond just a poor assessment of research. In order to understand *why* (and *how*) the use of IF as a proxy for decision making is particularly problematic, we turn our attention to the limitations of human decision making.

The literature on biases and limits of decision making originated from the work of Simon (1955, 1997) when he posited that the rational decision maker is limited from a set of cognitive internal bounds and from access to external information. Cognitive limitations are due to the ability to process a set amount of information, while external limitations relate to how much information is available on any decision topic. To overcome these limits, decision makers exploit heuristics (e.g., Gigerenzer & Selten 2001), are subject to biases (e.g., Kahneman & Tversky 1979), and use the surrounding social environment to adjust their judgement (e.g., Simon 1993; Secchi 2011).

If we take a step back to the cognitive mechanisms affecting these three strategies, we may refer to the so-called *logical fallacies* (Woods 2004). One that applies to the case of IF is the *composition and division*, where an argument is made by attributing the system's characteristics to each one of its parts. When a journal has a high/low IF then all of its articles are necessarily of high/low quality. This assumes that IF is a measure of quality (characteristic of the system...) and that this quality is inherent to each and every published article (...reflected in its parts). A biased reviewer might evaluate a submitted paper based on the IF of the journal it was submitted to. This reviewer's perception of scientific value will be reflected in the way the assessment is performed, probably by stricter requirements as IF goes up.

The second fallacy applied to our case is the *ad verecundiam* that is the attribution of value to something because of the authority it represents (Woods 2004; Secchi 2011). This "appeal to authority" is used every time one claims that an argument is good because of the outlet in which it appears as opposed to the logical soundness of it. A sentence such as "it appeared in *Nature*, hence it must be excellent!" is a logically flawed argument because its only strength is a very weak connection to the content of the article. The perception of scientific value is, for a biased journal reviewer, reflected on to the journal by force of its IF. Of course, given institutional constraints it is rational in a social sense for scholars to follow the conventions that surround IF for career purposes and/or for one's own survival in contemporary academia. A relatively recent stream of literature building on Bayesian models (e.g., Hahn et al. 2013; Harris et al. 2016) indicates that this strategy is successful when the credibility of the source can be established (e.g., Bovens & Hartmann 2003). And yet, this does not rule out the fact of IF's unreliability. We can call this the *Oracle of Delphi effect*. The majority believed the Oracle (IF for us here) was a reliable and credible source of information, and it made sense for most to perceive it as such, because it would have been awkward to go against the social conventions of that time. But, the fact is that the Oracle (IF) was not credible neither reliable.

In a more pragmatic approach, one may suggest that IF works as an *anchor bias* (Kahneman & Tversky 1979). This happens when individuals refer to a number, a concept, or a process as a—conscious or, most of the times, unconscious—benchmark for making a judgement. This ends up limiting the depth and meaning attributable to alternative choice options and, eventually, the final decision. A *status quo bias* (Silver 1990)may also be seen to play a role in the process of using the IF as a proxy to evaluate papers. Depending on the community of scholars in which one feels committed to, a peer-reviewer may wish to keep the entry barrier high for new comers in the discipline, in order to preserve the established "quality". This reflects on the lack of innovation or methodological variability present in specific areas of science.

2.3 Applying organisational cognition

From what written above, it is plausible to claim that different dispositions towards IF derive from personal attitudes while some others can be tracked down to the professional community in which one belongs. Theoretical developments in the area of *systemic/distributed e-cognition* (Hutchins 1995; Cowley & Vallée-Tourangeau 2017) help us define these aspects more clearly.

2.3.1 INDIVIDUAL COGNITIVE MECHANISMS

The distributed cognition approach was developed in continuity with the tradition of bounded rationality (Simon 1997; Gavetti et al. 2007) in that it still postulates that there are limits of a person's abilities. However, it also breaks with some of its basic assumptions in that it does not demarcate so clearly between internal (brain) and external cognitive resources (Bardone 2011; Secchi 2011). The core of this approach is that cognitive processes are not bound to one's brain but expand depending on the interaction between internal (brain) and external (environmental) resources (Hutchins 1995; Clark & Chalmers 1998; Magnani 2007; Cowley & Valleé-Tourangeau 2017). This leads to abandon the "internal vs external" dichotomy typical of Simon-like views, and to embrace a more dynamic and open approach where the decision maker may move his/her own rationality in co-evolution with the resources at hand (Secchi 2011).

By adapting this approach to the work of a reviewer, the following are examples of elements that may be analysed in terms of a distribution of cognitive processes: a socially construed belief that IF represents quality (implemented in the model as agents belonging to intelligence unit 2, or IU2); the production of a report that is then shared with an editor (in the mode, each agent produces a report per paper assessed); and the eventual discussions the reviewer may have with colleagues and/or the editor about intriguing points of the paper under review (this is the second—social—evaluation stage in the model).

2.3.2 THE SOCIAL ORIENTATION

In another article (Secchi & Cowley 2018) we argued that most of the interactions happening during a review process may be framed in a *distributed organisational* cognitive view. The fact that, in a peer review process, some of these organisations are informal while some others are formal sets the case for an even more intriguing and interesting study. In fact, each one of the actors of the reviewing process has several belongings, roughly represented by one's own university affiliation, the professional academic group, the editorial board, and the publisher, to name a few. All these groups are relevant because they define how cognitive processes that enable one to perform the review are distributed and shaped (Secchi & Adamsen 2017). The use of *socially* relevant cognitive resources becomes crucial (Hollan et al. 2000) to understand these processes. And a key feature of the individual that is tied to these social processes has been called "docility" in recent and past literature (Bardone & Secchi 2017; Knudsen 2003; Secchi 2011, 2016; Secchi & Bardone 2009, 2013; Simon 1993).

The attitude with which a decision maker leans on others for information, recommendations, advice, and suggestions that are used to make a decision, goes under the name "docility" (Simon 1993). In this respect, the word is true to its Latin root and it literally means the "willingness to be taught" (Secchi 2011). By looking at this aspect, some have indicated that this passive mechanism could be added to a more active disposition that makes highly docile individuals share their points of view with others (Secchi & Bardone 2009). Highly docile individuals would be more likely to be affected by information coming from social channels or, in a distributed cognitive language, by social cognitive resources (Secchi 2011; Secchi & Bardone 2009).

A peer reviewer that shows docility is one who updates his/her own evaluation by discussing the paper with colleagues. Moreover, such docile individual is more likely to align with the academic professional community of reference. There are clear limits to being docile in that conformity may be valued more than divergence (Secchi & Gullekson 2016). However, docile individuals may also be "inquisitive" and reach outside of their own community to assess different standards and sources of information (Bardone & Secchi 2017).

3. The Model

The simulation is performed by using an agent-based computational model (ABM). This technique is particularly well suited to model complex adaptive systems such as societies (e.g., Edmonds & Meyer 2017), organisations (e.g., Fioretti 2013), and smaller communities such as teams (e.g., Secchi 2015), for example. ABM have begun to be used in the social sciences during the mid-Nineties (Gilbert & Troitzsch 2005), and have sparked an increasingly wide interest ever since, to the point that some have started to indicate simulation as a third leg of science, next to empirical and theoretical enquiries (Axelrod 1997). This class of models is centered on specific entities—called agents—that are defined computationally as they interact with other entities and with the environment. Every single aspect of an agent-based simulation can be specified to its finest details or kept more abstract (Edmonds & Moss 2005). In addition to that, random components can be always accounted for in ABM, making researchers overcome the typical limits of equation-based modeling (Madsen et al. 2019), finding solutions (if any) through the means of computational power rather than analytics (for a short introduction, see Gilbert 2008).

In the following pages, we first provide a succinct description of the parameters and some of the characteristics of the simulation and then present interaction dynamics by introducing the processes implemented in this simulation.[2]

All parameters of the simulation are shown in Table 1. For this simulation, there are two types of agents: *papers p* and *reviewers*, leaving editors, publishers, and others affect the process only through indirect mechanisms. Reviewers are initially categorised in IF-lovers, IU2, and IF-agnostics, IU1, the first are those whose *IF attitude* IF_a[3] is higher than a threshold level defined by the modeller IF_t so that $IU2 : IF_a > IF_t$ and $IU1 : IF_a \leq IF_t$. After assigning attitudes at random to reviewers, the threshold IF_t is only used to separate those who are more favorable towards high-IF journals and those who are not. As it will be explained below, articles are assigned to reviewers at random but behavior depends on *IF attitudes*.

Every reviewer is also characterised by a level of *docility* $\sim N(m_d, 0.2)$[4], where the mean m_d can be allocated through the interface of the model and is currently set to 0.60. As already mentioned above, docility is a cognitive mechanism that is socially distributed and it provides an overall idea of how much the decision maker uses information coming from peers to make decisions (Simon 1993; Secchi 2011; Bardone & Secchi 2017). It is, in this simulation, a measure of the malleability with which an agent evolves, modifies, and updates one's own views. Higher values of the parameter correspond to an agent that is more susceptible to listen to other points of view while lower values describe a relatively a-social agent.

The third characteristic of the reviewer is the *perception of scientific value*, or PSV, distributed $\sim N(0,0.15)$ for IU1 and $\sim N(0+\alpha_p, 0.15)$ for IU2. The value α_p [0,1] is just the different mean that an IU1 or an IU2 reviewer have, on average, on PSV. This is the idealistic way with which a reviewer approaches science in general and it reflects on his or her actions when behaving as a scholar (e.g., writing, reading, reviewing). IU2 reviewers care for IF and believe that journals with high IF have higher standards for science, hence their articles have a higher PSV.

2 Additional information and the model is available on the OpenABM platform, downloadable for free here: https://www.comses.net/codebases/c913254a-0fdc-4298-b304-13890c6049ab/releases/1.1.0/ .

3 IF_a is allocated randomly on a normal distribution $\sim N(0.5, 0.25)$; precautions are taken so that $0 \leq IF_a \leq 1$

4 The symbol for mean docility is represented in Table 1 by d bar.

TABLE 1: PARAMETER NOTATIONS AND VALUES

Parameter	Notation	Values	Description
steps	s	100	The maximum number of opportunities that agents have to interact with each other when dealing with reviews.
runs	N	25	Number of times a simulation is performed with each given configuration of parameters.
Impact Factor	IF	$\sim [0, 5]$	This is the IF associated with the journal where the publication is submitted, hence it is attributed to the submission as well. It is randomly assigned and varies between 0 and 5.
intrinsic value	p_v	$\sim [0, 1]$	The value of a scientific contribution assigned to each submission on a random basis.
IF attitudes	IF_a	$\sim \mathcal{N}(0, 0.5)$	This is the attitude each agent has towards the impact factor (IF) of a journal publication.
IF threshold	IF_t	$\sim [0, 1]$	This is the threshold between IU1 and IU2 in IF_a levels.
docility	d	$\sim \mathcal{N}(\bar{d}, \sigma_d)$	This is the docility level associated randomly to each agent in the simulation — higher values indicate higher probability to adapt to the respective IU as well as less autonomy from the respective IU.
perc. sc. value	PSV	$\sim \mathcal{N}(1 + \alpha_p, 0.15)$	This is the disposition toward scientific value—i.e. how much a contribution is thought it is worth—used by a reviewer that in the IU1 category ($\alpha_p = 0$) and IU2 ($\alpha_p \in [0, 1]$).
PSV diff.	α_p	$0.25, 0.50$	The average difference in the perceived scientific value that members of IU2 have in relation to those from IU1—this value affects directly the mean of the random-normal distribution for IU2; st.dev. is unaffected.
group	G	[true, false]	The tendency to be socio-cognitively closer to the other members of the IU the agent is affiliated with.
review	r	5	This is the value used to explore the environment that surrounds each agent.

Papers have an *intrinsic value* p_v [0, 1] that is the value that the overall community of scholars would assign to a paper if they were to evaluate it. In other words, it is some sort of anticipated scientific value of each contribution to science because it reflects how the community of scholars (and society) would react to it once/if published. It is immanent to the review but it is almost never clearly visible by reviewers. This does not mean that reviews are meaningless, it only means that it is very difficult for reviewers to estimate the possible impact of a paper on the scientific enterprise. To use a strange but probably effective analogy, a paper stays to a reviewer like a business start-up stays to a venture capitalist. At time zero, it is very difficult to assess if a start-up will become the next success story. One may make informed guesses, but there is no certainty. This general ambiguity is reflected by the perception of p_v in our simulation. Each paper is also associated with an Impact Factor (*IF*) that derives from the journal in which it has been submitted. The two characteristics are independent[5], one is attributed using a random

5 This decision is consistent with the critiques on IF expressed briefly at the beginning of the article. In order to test this assumption further, one could build a model where *IF* and p_v are correlated to each other to understand to what extent some of the effects in the simulation depend on the independence assumption or not. We believe it would be too much to test it here, but are already considering this option in our work elsewhere. We thank one of the reviewers for this suggestion.

floating number to avoid any particular distribution while $IF : \sim N(2.5, 1)$. IF is field specific, in some disciplines it is very high while in others it has particularly low values; we decided to take an arbitrary distribution of values since the simulation is relative to any value set for IF.

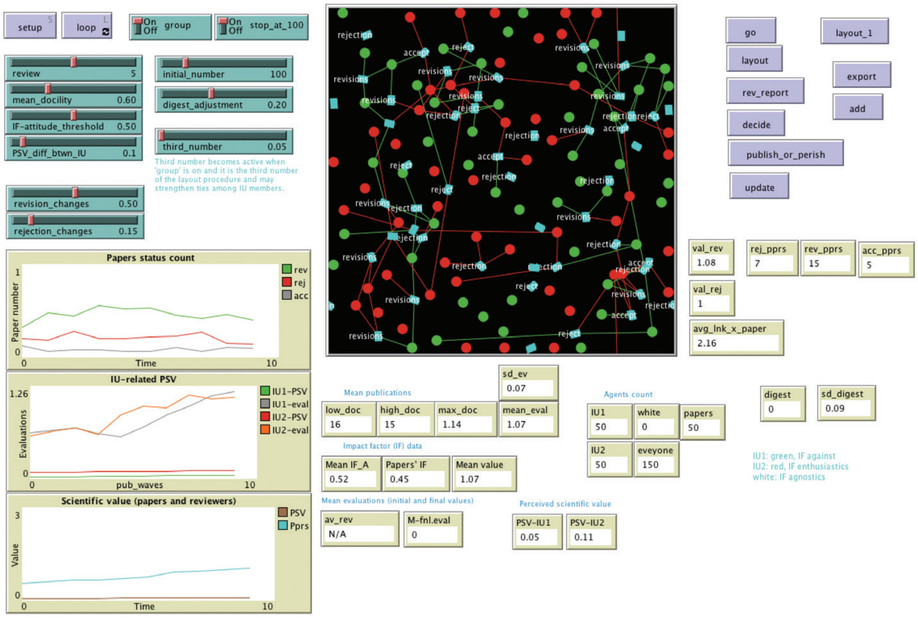


Figure 1: The PRIF Model ABM interface

At the start, all agents appear on a 2-dimensional environment at random. Each agent has a range (*review* $r[1, 10]$) with which it is assigned papers to review. The larger the range, the more likely it is that one has more papers to review. The peer-reviewing under the lenses of Impact Factor (PRIF) simulation is implemented using the software NetLogo (Wilensky 1999). The simulation interface is replicated in Figure 1.

4. Procedures

In the following pages, we describe the more dynamic aspects of this model, beginning with those mechanisms that apply at the setup, then moving to explain how reviewer assessments are produced and how an editorial decision is reached. The section ends with the presentation of more socially-oriented mechanisms that lead to update a reviewer's opinions.

4.1 Start mechanisms and uncertainty

At every round (step *s*) of interaction, every reviewer looks around its review opportunities (range *r*) and connects to papers that are not already under revision. This assures that only papers that have either been previously "rejected" or not yet submitted anywhere can be reviewed. Papers evaluated as "revise and resubmit" stick with the same reviewer. All rejected papers are resubmitted to journals with a potentially different IF, higher or lower depending on a randomly generated number $\sim N(0, 1)$. When a paper is "accepted" it then disappears from the system. At the same time, a random number (up to three) of new papers p_n are generated whenever one is accepted. A paper could have only up to three reviewers but papers with only one reviewer are not evaluated; this mimics "desk-rejection".

In order to represent some of the limitations typical of uncertainty in rational decision making (Simon 1997), all agents move in the space following a function of attraction-repulsion to each other and their links (Fruchterman & Reingold 1991).

4.2 Production of reviewer reports

A first evaluation—a reviewer's report—is generated following different rules for IF- lovers and IF-agnostics. The basic mechanism is that every reviewer uses some function of PSV to assess the value of scientific work. For IU2 reviewers, when the assessment of a paper is made for a journal with high IF and intrinsic value $p_v > PSV_{IU2}$,[6] then the evaluation is $\in = p_v + |PSV_{IU2}| + e$. The last term is an error that can be positive or negative and is such that e $[0, \mathrm{PSV}_{\mathrm{IU2}}]$. When the journal has a high IF and $p_v \leq PSV_{IU2}$, then the evaluation is $\in = p_v\,(\,|PSV_{IU2}| + e)$. In the opposite case when $p_v \leq PSV_{IU2}$, the evaluation becomes $\in = p_v + PSV_{IU2} + e$. These evaluative mechanisms here represent a reviewer that adopts different standards depending on the journal. Under conditions of high IF, the IU2 reviewer tends to give better evaluations, on average, to papers that are perceived to have some intrinsic value p_v while harsher evaluations when the paper has lower p_v. When the agent reviews for low-IF journals then the evaluation is a free-float, in the sense that it could be higher or lower than the intrinsic value p_v. No clear criterion for being higher or lower is set in this case. IU1 always follow this last approach to paper evaluations.

4.3 Decision making

The decision on whether a paper is accepted, rejected, or needs revisions is made through a simple mechanism. When the mean of reviewers' reports are such that $m_{\in} > p_v$ and the distance (standard deviation) in judgement is $\sigma_{\in} \leq k_a$, then a paper is "accepted". The constant k_a is

6 The two values are obviously harmonised in order to allow comparisons.

smaller for papers submitted to journals with higher IF as opposed to those submitted to lower IF journals—0.01 as opposed to 0.02. A decision to reject a paper is made every time the mean evaluations of reviewers is $m_\epsilon < p_v$ and the reports diverge so that $\sigma_\epsilon > k_r$, where $k_r = 0.075$ for high IF journals and $k_r > 0.05$ for the others. A rejected paper severs the links with the reviewers and is relocated at random in the environment so that it is newly submitted to another journal, with new IF and new links. All papers that are not accepted nor rejected are to "revise and resubmit".

4.4 "Publish or perish"

When revisions are requested for a paper, authors are assumed to adjust its contents. As academics know, this is not, unfortunately and painfully, always going to please reviewers. This uncertainty is reflected in a change in p_v that adjusts itself by adding a random value from a normal distribution $\sim N(m_{rc}, \sigma_{rvc})$ with mean $m_{rc} = 0.5$ when $p_v \leq 0$ and $m_{rc} = 0$ if $p_v > 0$. The standard deviation σ_{rvc} [0, 1] is a parameter for variability of revision changes that can be determined by the modeller. Something similar happens to rejected papers, with a random normal distribution $\sim N(m_{rc}, \sigma_{\mathrm{rjc}})$, where σ_{rjc} [0, 1] is the parameter controlling average changes after rejection. Rounds of review are also counted in the system with the purpose of limiting them to 3. It is assumed that, when a paper still receives a revision "verdict" after three rounds of review, it becomes very unlikely that it will improve satisfactorily in the following round(s). This is a somewhat arbitrary limit but it mimics an editor's intervention on a rather difficult situation that appears to be stuck.

4.5 Update mechanisms

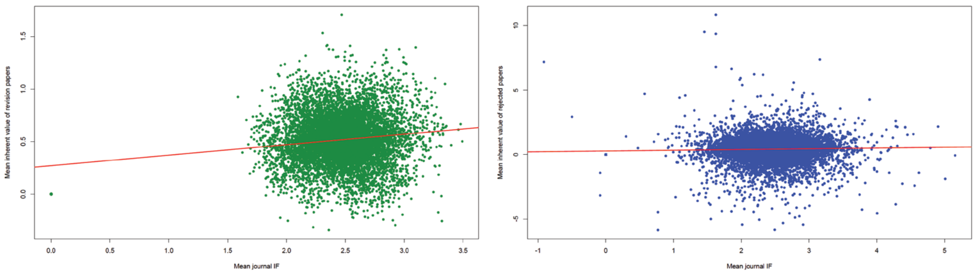
All reviewers have a system to update their evaluation preferences. This is based on the cognitive mechanism grounded in Simon's (Simon 1993) "docility" (Bardone 2011; Secchi 2011; Secchi & Bardone 2009), at the core of organisational cognition (Secchi & Cowley 2018). For highly docile individuals i, with $d_i > m_d + \sigma_d$, the difference between one's own evaluation and the mean of the evaluations of the other agents is such that $\epsilon_i > m_\epsilon + \sigma_\epsilon$, then their own PSV_i becomes $PSV_i = PSV_i + \epsilon_{adj} \times \epsilon_i$, where ϵ_{adj} [0, 1] is another parameter that updates and conforms the highly docile's assessments to the way the system is heading. If $\epsilon_i < m_\epsilon - \sigma_\epsilon$ then the adjustment is done downwards: $PSV_i = PSV_i - \epsilon_{adj} \times \epsilon_i$.

A final adjustment materialises with the "group" option. This makes agents with similar attitudes to IF—allegedly part of the same research community—to adjust their judgment by aligning to community standards rather than to the system. When this condition is set to 'on' then the starting point is similar to what just described in the paragraph above for highly docile individuals, with the only difference that, in this case, an agent sets its benchmark to other agents with the same IF dispositions. In more formal terms, if $d_{IU1} > m_{d_{IU1}} + \sigma_{d_{IU1}}$, and $\epsilon_{IU1} > m_{\epsilon_{IU1}} + \sigma_{\epsilon_{IU1}}$, then PSV_{IU1} becomes $PSV_{IU1} = PSV_{IU1} + \epsilon_{adj} \times \epsilon_{IU1}$. The same as above happens in the case $\epsilon_{IU1} < m_{\epsilon_{IU1}} - \sigma_{\epsilon_{IU1}}$ and the agents in the other group IU 2 behave similarly relative to their own community.

5. Findings, discussion and concluding remarks

We performed a few pilot runs to test code and conditions and identify a factorial design of $2^4 \times 3^3$ for the simulation settings. These are: α_p = [0, 0.05, 0.1], IF_t = [0.25, 0.50], α_{rvc} = [0.15, 0.25, 0.50], σ_{rjc} = [0.05, 0.15], ϵ_{adj} = [0.05, 0.15, 0.2], *group* = [true, false], *m* = [0, 0.05][7]. We then used statistical power to determine the number of runs (Secchi & Seri 2017; Seri & Secchi 2017), and settled on 2 runs per configuration of parameters using a formula for regression analysis (as explained in Seri & Secchi 2018). In the following pages, we present a summary of results, while more detailed findings are available online[8].

Figure 2 shows journal IF for submitted papers per intrinsic value of papers p_v, under all conditions of the simulation (*group* = false). Figure 2a shows all the papers with revisions while Figure 2b shows those rejected. The number of "revision" papers seem to increase very slightly as IF becomes higher at a rate that is not reflected in rejected papers. This is somehow in line with expectations, since the cloud of points showing rejected papers (Figure 2b) is more dispersed. While we have rejected papers for journal IF of 3.5 and higher, we do not have the same for papers with outcome "revisions." Simulation data report that the model works as someone may expect, i.e. making it very hard to receive a "revise and resubmit" outcome in journals with high IF. These results are partially confirmed by looking at the regression estimates reported in Table 2, where the number of rejections is affected by an increasing journal's IF (Model 2: $\beta = 0.09$, $p < 0.05$) while this same pattern is not visible for "revision" papers (Model 1: $\beta = 0.10$, $p = 0.1051$).



(a) Papers with outcome "revisions" (b) Papers with outcome "rejection"

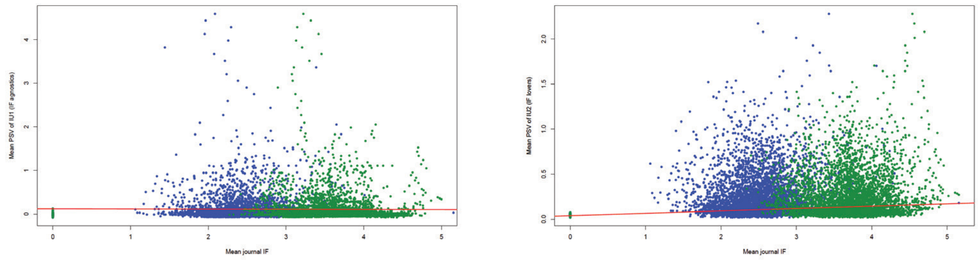
Figure 2: Journal IF and p_v for "rejected" and "revise and resubmit" papers

Figure 3 shows papers revised or rejected by PSV from academics affiliated with IF agnostics (IU1) or lovers (IU2). As Figure 3b indicates, PSV is slightly higher in reviewers from IU2. Also, it

7 The last notation has not been introduced above and it is a movement constant, to make agents move more or less away from each other.

8 Supplementary documentations, findings, technical detailed description of the procedures, the code, data, and the full model in on the OpenABM platform, downloadable for free here: https://www.comses.net/codebases/c913254a-0fdc-4298-b304-13890c6049ab/releases/1.1.0/.

does grow when IU2 academics review for higher IF journals. In both cases, papers submitted to lower IF journals tend to be revised although the distinction between high/low IF submissions seem to be clearer for IU2. A paper submitted to a journal with IF ranging between 2.5 and 3.5 may receive an outcome of rejection or revision by IU1 academics (Figure 3a). Instead, a paper submitted to a journal with $IF \leq 3.0$ is very much likely to receive a "revisions" outcome if the reviewer is from IU2 (Figure 3b). Again, these results reflect the assumptions of the model. However, they appear more robust than expected in that they hold independent of the different configuration of parameters.

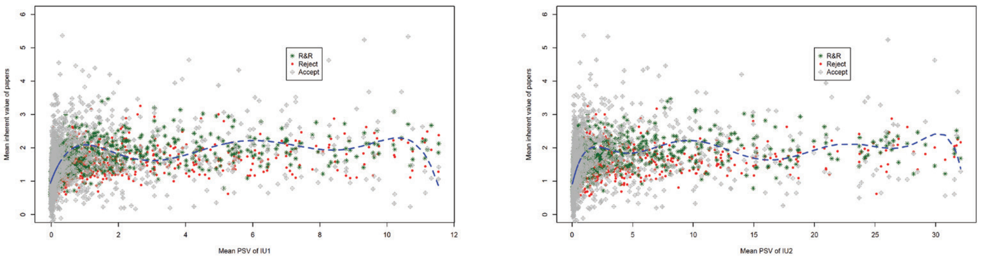


(a) IU1: IF agnostics

(b) IU2: IF lovers

Figure 3: Journal IF and PSV (revisions, rejections)

The two groups of reviewers, IU1 (IF-agnostics) and IU2 (IF-lovers), seem to have a similar overall behavior when it comes to evaluate papers. Figure 4 shows the perceived scientific value (PSV) as a potential explanation for the inherent value of reviewed papers. However, when one looks at the differences between Figure 4a and 4b, it is clear that IU2 have much higher PSV (up to over 30) while UI1 seem to have more realistic expectations on articles (up to 12). It is worth noting that reviewers from the different groups may be reviewing the same paper, with a likelihood of producing rather dissimilar assessments. On the left part of the two graphs, it is also interesting to notice that there are a number of papers that are either rejected, revised, or accepted when PSV is low (and inherent value is also on the lower end of the scale). However, IU2 are more conservative: what happens for values between 0 and 2 for IU1, happens for values between 0 and 5 for IU2.



(a) IU1: IF agnostics (b) IU2: IF lovers

Figure 4: Mean inherent value of rejected, R&R, and accepted papers as predicted by the mean perceived scientific value (PSV)

In the following two figures, we express the number of IU2 (IF-lovers) reviewers as a ratio of IU1. This means that, when the number in the *x*-axis is 1, the number of active reviewers from the two groups is the same, when it is < 1 then active IU1 are more than IU2 and vice versa, when > 1 IU2 are more than IU1. The purpose of Figure 5 is to understand whether article assessments change as more IU2 enter in the reviewing panel. In general, it seems that numbers show a slight prevalence of IU2 over IU1 and this, overall, affects the number of rejections and revisions. These results require a search for further explanations of the differences observed in the previous figures. In Figure 6 we still plot IU2 reviewers as a ratio of IU1 to explain whether papers are assessed differently depending on the inherent value p_v of the articles. In so doing, and since we are discussing IF-lovers (IU2), we have decided to present results by factoring out the IF of the journal in which the article has been submitted. This determines an interesting scale. Lower negative values indicate cases in which IF is higher than p_v, the more negative the value the larger the gap. Higher positive values define cases in which the article has been submitted to a journal whose IF is lower than the article's p_v. While an assessment of revise and resubmit is not affected by the number of active IU1 and IU2, both rejections and acceptance decrease as IF increases, together with the number of active IU2. This confirms that IU2 reviews become harsher as IF increases and, at the same time, they become more lenient towards some articles. In short, Figure 6 is particularly revealing because it highlights that both IU1 and IU2 have biased judgements that depend on a journal's IF. In short, independent of one's beliefs, IF seems to cloud a reviewer's judgement.

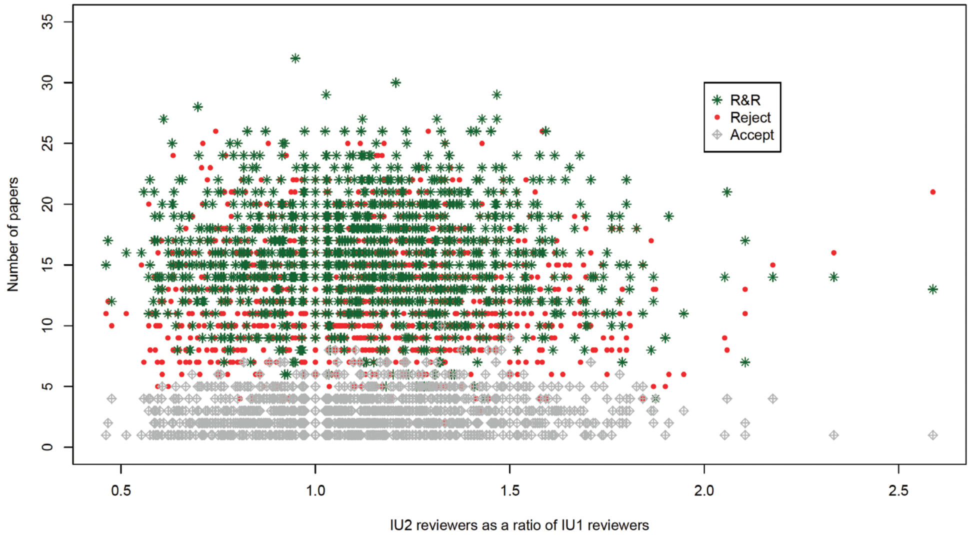


Figure 5: Number of papers rejected, R&R and accepted compared to the number of IU2 reviewers calculated as a ratio of IU1 reviewers

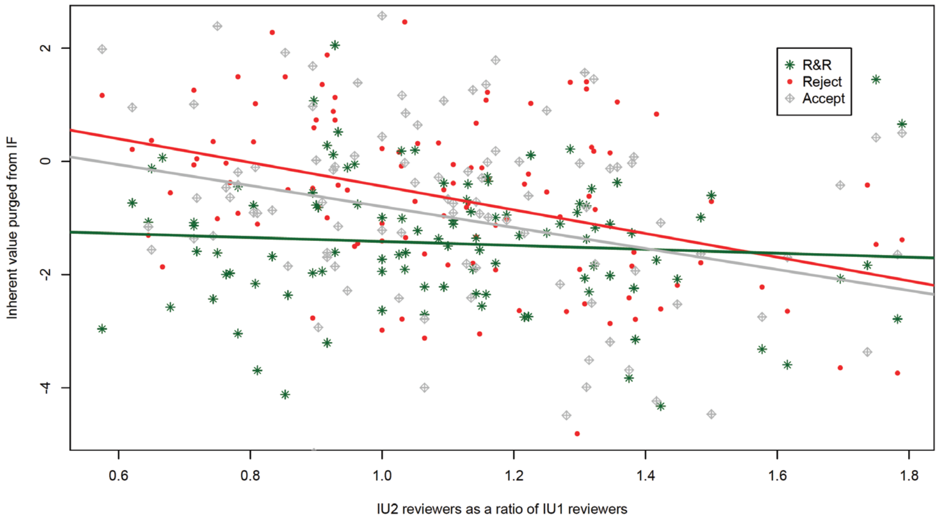


Figure 6: Inherent value (value - IF) compared to the number of IU2 reviewers calculated as a ratio of IU1 reviewer (in the last 20% of time when PSV increases)

In order to reason on a more detailed set of results, we have produced two fixed panel regressions[9] where we can study the effect of selected variables and parameters on the number of papers with outcome "revisions" and "rejection". Table 2 shows that there is no difference between the number of IU1 or IU2 reviewers involved in peer review and the way rejection or revisions are affected. Mean PSV of both IU1 and IU2 reviewers have a stronger effect on rejections, as it is foreseeable. These results are counterbalanced and overcome by PSV differential, with a very strong effect on rejections (Model 2: $\beta = 5.57$, $p < 0.001$). The higher the PSV differential, the more likely that a paper is rejected (or revised; see Model 1 also). Changes in a paper decrease the likelihood that is revised or rejected. Effects are, again, higher for rejected papers (Model 2: $\beta = 3.32$, $p < 0.001$). The mean intrinsic value p_v of papers is not significant for rejected papers while it has an effect for revised papers (Model 1: $\beta = 0.15$, $p < 0.001$). Finally, mean IF of papers to which the article is submitted does not affect revised papers while it does affect rejections (Model 2: $\beta = 0.09$, $p < 0.05$).

5.1 Implications and conclusions

Overall, it seems that there is an impact of IF on publications. On the one hand, IF-lovers (IU2) develop higher PSV and show some problems in the interpretation of inherent value (p_v). This is particularly concerning in terms of the fairness and equitability of the review process because it makes reviews lean more towards rejections and revisions independent of p_v but in relation to a journal's IF.

On the other hand, we could detect an organizational/community impact on the distortion that both groups show in the evaluation of papers. This is an effect that allows individuals from both groups, IU1 and IU2, to work as proper intelligence units. In turn, this indicates that most of the elements that affect reviewers' judgement are reinforced by the professional academic community they belong to. It is the practice and shared interpretation of *what makes science* that make reviewers identify with a certain stream. And we have shown that this is reflected in their understanding of PSV.

9 We performed Hausman tests to compare them with respective random effects panel regressions; in both cases the fixed effects resulted a better fit for the data. We left out the group effects, and the *m* parameter from this one, since they would require more extensive comments.

TABLE 2: FIXED EFFECT PANEL REGRESSION MODELS

	Model 1 *DV: Num. pap. rev.*		**Model 2** *DV: Num. pap. rej.*	
	β	(st.err.)	β	(st.err.)
Active IU1 reviewers	0.85^{***}	(0.00)	0.85^{***}	(0.00)
Active IU2 reviewers	0.83^{***}	(0.00)	0.83^{***}	(0.00)
Mean PSV$_{IU1}$	0.25^{***}	(0.02)	0.46^{***}	(0.02)
Mean PSV$_{IU2}$	0.25^{***}	(0.03)	0.44^{***}	(0.03)
PSV difference (α_p)	4.65^{***}	(0.40)	5.87^{***}	(0.49)
IF threshold (IF$_t$)	2.75^{***}	(0.34)	2.73^{***}	(0.42)
Revision paper changes	-0.77^{***}	(0.11)		
Mean rev. paper p_v	0.15^{***}	(0.04)		
Mean IF of revision papers	0.10	(0.06)		
Rejection paper changes			-3.32^{***}	(0.77)
Mean rej. paper p_v			0.02	(0.04)
Mean IF of rejected papers			0.09^{*}	(0.05)
R^2	0.73		0.73	
F-test	27531^{***}		18109.8^{***}	
Degrees of freedom	$9, 89891$		$9, 59891$	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

In other words, the way reviewers perceive scientific value is a function of their feeling of belongingness to a particular community. Their cognition is shaped by it.

In summary, the article presents an agent-based model (that we called PRIF Model) of how IF may be considered among reviewers' biases during peer-review. Even though they are preliminary, our results show that IF affects reviewers behaviour in a rather unsuspected way in that it (a) has similar impact to those who "love" it as well as to those that are indifferent to it, and (b) it seems not to affect papers to "revise and resubmit" in a direct way.

More experiments are needed, both from within this model and through the means of different research, to fully disclose the effects of how cognitive limitations influence the review process.

References

Axelrod, Robert. 1997. Advancing the art of simulation in the social sciences. In R. Conte, R. Hegselmann, & P. Terna (Eds.), *Simulating Social Phenomena.* New York: Springer-Verlag. 21–40.

Bardone, Emanuele. 2011. *Seeking chances. From biased rationality to distributed cognition. Cognitive Systems Monographs vol. 13.* New York: Springer.

Bardone, Emauele & Davide Secchi. 2017. Inquisitiveness: Distributing rational thinking. *Team Performance Management* 23(1/2). 66–81.

Bloch, Sidney & Garry Walter. 2001. The Impact Factor: time for change. *Australia New Zealand Journal of Psychiatry* 35. 563–568.

Bovens, Luc, & Stephan Hartmann. 2003. Solving the riddle of coherence. *Mind* 112(448). 601–633.

Callaway, Ewen. 2016. Beat it, Impact Factor! Publishing elite turns against controversial metric. *Nature* 535(7611). 210–211.

Clark, Andy & David Chalmers. 1998. The extended mind. *Analysis* 58. 7–19.

Colquhoun, David. 2003. Challenging the tyranny of impact factors. *Nature* 423. 479.

Cowley, Stephen J. & Frédéric Vallée-Tourangeau (Eds.). 2017. *Cognition beyond the brain. Computation, interactivity and human artifice.* 2nd ed. London: Springer

Curry, Stephen. 2012. Sick of impact factors. blog post: Occam's typewriter. http://occamstypewriter.org/scurry/2012/08/13/sick-of-impact-factors/.

Edmonds, Bruce & Ruth Meyer (Eds.). 2017. *Simulating Social Complexity. A Handbook.* 2nd ed. Heidelberg: Springer.

Edmonds, Bruce & Scott Moss. 2005. From KISS to KIDS—an 'anti-simplistic' modelling approach. In P. Davidson (Ed.), *Multi Agent Based Simulation. Lecture Notes in Artificial Intelligence vol. 3415.* New York: Springer. 130–144.

Fioretti, Guido. 2013. Agent-based simulation models in organization science. *Organizational Research Methods* 16(2). 227–242.

Frey, Bruno S. & Katja Rost. 2010. Do rankings reflect research quality? *Journal of Applied Economics* 13(1). 1–38.

Fruchterman, Thomas M. J. & Edward M. Reingold. 1991. Graph drawing by force-directed placement. *Software-Practice and experience* 21(1). 1129–1164.

Garfield, Eugene. 1999. Journal impact factor: a brief review. *Canadian Medical Association Journal* 161(8). 979–980.

Garfield, Eugene. 2003. The meaning of the Impact Factor. *International Journal of Clinical and Health Psychology* 3(2). 363–369.

Gavetti, Giovanni, Daniel Levinthal & William Ocasio. 2007. Neo-carnegie: The carnegie school's past, present, and reconstructing for the future. *Organization Science* 18(3). 523–536.

Gigerenzer, Gerd & Reinhard Selten. 2001. *Bounded Rationality. The Adaptive Toolbox.* Cambridge, MA: MIT Press.

Gilbert, Nigel. 2008. *Agent-Based Models. Quantitative Applications in the Social Sciences vol. 153.* Thousand Oaks, CA: Sage.

Gilbert, Nigel & Klaus G. Troitzsch. 2005. *Simulation for the social s cientist*. Maidenhead UK: Open University Press.
Hahn, Ulrike, Mike Oaksford, & Adam JL Harris. 2013. Testimony and argument: A bayesian perspective. In *Bayesian argumentation* (15–38). Springer.
Harris, Adam JL, Ulrike Hahn, Jens K. Madsen, and Anne S. Hsu. 2016. The appeal to expert opinion: quantitative support for a bayesian network approach. *Cognitive Science* 40(6). 1496–1533.
Hollan, James, Edwin Hutchins & David Kirsh. 2000. Distributed cognition: toward a new foundation for human-computer interaction research. In *Proceedings of the ACM Transactions on Computer-Human Interaction (TOCHI) vol. 7.* New York: ACM Digital Library. 174–196.
Hutchins, Edwin. 1995. *Cognition in the wild*. Cambridge, MA: MIT Press.
Kahneman, Daniel. 2003. A perspective of judgement and choice. mapping bounded rationality. *American Psychologist* 58(9). 697–721.
Kahneman, Daniel & Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47(2). 263–292.
Knudsen, Thorbjørn. 2003. Simon's selection theory: Why docility evolves to breed successful altruism. *Journal of Economic Psychology* 24. 229–244.
Madsen, Jens K., Richard Bailey, Ernesto Carrella & Philipp Koralus. 2019. Analytic versus computational cognitive models: Agent-based modeling as a tool in cognitive sciences. *Current Directions in Psychological Science* 28(3). 299–305.
Magnani, Lorenzo. 2007. *Morality in a technological world. Knowledge as a duty*. New York: Cambridge University Press.
Osterloh, Margit & Bruno S. Frey. 2020. How to avoid borrowed plumes in academia. *Research Policy* 49(1). 103831.
Secchi, Davide. 2011. *Extendable rationality. Understanding decision making in organizations*. New York: Springer.
Secchi, Davide. 2015. A case for agent-based models in organizational behavior and team research. *Team Performance Management* 21(1/2). 37–50.
Secchi, Davide. 2016. Boundary conditions for the emergence of 'docility:' An agentbased model and simulation. In D. Secchi & M. Neumann (Eds.), *Agent-Based Simulation of Organizational Behavior. New Frontiers of Social Science Research*. New York: Springer. 175–200.
Secchi, Davide & Billy Adamsen. 2017. Organizational cognition: A critical perspective on the theory in use. In S. J. Cowley & F. Vallee-Tourangeau (Eds.), *Cognition Beyond the Brain: Computation, Interactivity and Human Artifice*. Heidelberg: Springer, 2nd edition. 305–331.
Secchi, Davide & Emanuele Bardone. 2009. Super-docility in organizations. An evolutionary model. *International Journal of Organization Theory and Behavior* 12(3). 339– 379.
Secchi, Davide & Emanuele Bardone. 2013. Socially distributed cognition and intraorganizational bandwagons: Theoretical framework, model, and simulation. *International Journal of Organization Theory and Behavior* 16(4). 521–572.
Secchi, Davide & Stephen J. Cowley. 2018. Modeling organizational cognition: the case of impact factor. *Journal of Artificial Societies and Social Simulation* 21(1). 13.

Secchi, Davide & Nicole L. Gullekson. 2016. Individual and organizational conditions for the emergence and evolution of bandwagons. *Computational and Mathematical Organization Theory* 22(1). 88–133.

Secchi, Davide & Raffaello Seri. 2017. Controlling for 'false negatives' in agent-based models: A review of power analysis in organizational research. *Computational and Mathematical Organization Theory* 23(1). 94–121.

Seglen, Per O. 1997. Why the impact factor of journals should not be used for evaluating research. *British Medical Journal* 314(7079). 498.

Seri, Raffaello & Davide Secchi. 2017. How many times should one run a computational simulation? In B. Edmonds & R. Meyer (Eds.), *Simulating Social Complexity. A Handbook*. 2nd ed. Heidelberg: Springer. 229–251.

Seri, Raffaello & Davide Secchi. 2018. A power primer for agent-based simulation models. Determining the number of runs in linear and polynomial regression. In *European Academy of Management Annual Conference.* Reykjavik, Island.

Silver, William S.& Terence R. Mitchell. 1990. The status quo tendency in decision making. *Organizational Dynamics* 18(4). 34–46.

Simon, Herbert A. 1955. A behavioral theory of rational choice. *Quarterly Journal of Economics* 69(1). 99–118.

Simon, Herbert A. (1993). Altruism and economics. *American Economic Review* 83(2). 156–161.

Simon, Herbert A. 1997. *Administrative behavior*. 4th ed. New York: The Free Press.

Vanclay, Jerome K. 2011. Impact factor: outdated artefact or stepping-stone to journal certification? *Scientometrics* 92(2). 211–238.

Walter, Garry, Sidney Bloch, Glenn Hunt & Karen Fisher. 2003. Counting on citations: a flawed way to measure quality. *Medical Journal of Australia* 178(6). 280–281.

Wilensky, Uri. 1999. Netlogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Woods, John. 2004. *The death of argument: Fallacies in agent-based reasoning.* Dordrecht: Kluwer. Axelrod, Robert. 1997. Advancing the art of simulation in the social sciences. In R. Conte, R. Hegselmann, & P. Terna (Eds.), Simulating Social Phenomena. New York: Springer-Verlag. 21–40.