https://plato.stanford.edu/entries/scientific-reproducibility/

The terms “reproducibility crisis” and “replication crisis” gained currency in conversation and in print over the last decade (e.g., Pashler & Wagenmakers 2012),

n 2016, a poll conducted by the journal Nature reported that more than half (52%) of scientists surveyed believed science was facing a “replication crisis” (Baker 2016)

for example, Vazire (2018) refers instead to a “credibility revolution” highlighting the improved methods and open science practices it has motivated.

The crisis often refers collectively to at least the following things:
the virtual absence of replication studies in the published literature in many scientific fields (e.g., Makel, Plucker, & Hegarty 2012),
widespread failure to reproduce results of published studies in large systematic replication projects (e.g., OSC 2015; Begley & Ellis 2012),
evidence of publication bias (Fanelli 2010a),
a high prevalence of “questionable research practices”, which inflate the rate of false positives in the literature (Simmons, Nelson, & Simonsohn 2011; John, Loewenstein, & Prelec 2012; Agnoli et al. 2017; Fraser et al. 2018), and
the documented lack of transparency and completeness in the reporting of methods, data and analysis in scientific publication (Bakker & Wicherts 2011; Nuijten et al. 2016).

rectify conditions that led to the crisis: promoting data sharing and public pre-registration of studies, advocating stricter editorial policies aroud statistical reporting including replication studies and statistically non-significant results

First look at term „reproducibility“ and related terms like „repeatability“ and „replication“
Second describe meta-science research that established and characterised the reproducibility crisis
Third attempts to address epistemological questions about the limitations of replication, and its value for scientific inquiry and accumulation of knowledge.
Fourth describes some initiatives the open science reform movement has proposed to improve reproducibility

1. Replicating, Repeating, and Reproducing Scientific Results
According to Cartwight 1991, the term „replication“, „reproduction“ and „repetition“ denote distinct cincepts, others (e.g, Atmanspacher & Maasen 2016) use it interchangeably. Different disciplines can have different understanding of these terms. In computational disciplines, reproducibility oftern refers to the ability to reproduce computations alonem it relates exclusively to sharing and sufficiently annotating data and code (e.g Peng 2011,2015). Replication descirbes the redoing of the whole experiments (Barba 2017). In contrast in psychology and other social and life sciences, reproducibility ma yrefer to either the redoing of computations or the redoing of experiemtns.

Describing a study as „replicable“, people could have in mind two different things:
First: replicable in principle, it can be carried out again, methods, procedures and analysis are described in sufficiently detailed and transparent way.
Second: carried out again, successfully produce the same or sufficiently similiar results as the original.
Arguably, most typologies of replication make more or less fine-grained distinctions between direct replication (which closely follow the original study to verify results) and conceptual replications (which deliberately alter important features of the study to generalize findings or to test the underlying hypothesis in a new way)
Computational reproducibility is most often direct (reproducing particular analysis outcomes from the same data set using the same code and software), but it can also be conceptual (analysing the same raw data set with alternative approaches, different models or statistical frameworks). For an example of a conceptual computational reproducibility study, see Silberzahn and Uhlmann 2015

4. Open Science Reforms: Values, Tone, and Scientific Norms

meta-science has unearthed a range of problems which give rise to the reproducibility crisis
the open science movement has propsoed or promoted various solutions.
Reforms grouped in four categories: (a) methods and training, (b) reporting and dissemination, (c) peer review processes, (d) evaluating new incentive structures

4.1 Methods and Training

Combating bias. The development of methods for combating bias, for example, masked or blind analysis techniques to combat confirmation bias (e.g., MacCoun & Perlmutter 2017).

Support. Providing methodological support for researchers, including published guidelines and statistical consultancy (for example, as offered by the Center for Open Science) and large online courses such as that developed by Daniel Lakens

Collaboration. Promoting collaboration and team/crowd sourced science to combat low power and other methodological limitations of single studies. The Reproducibility Projects themselves are an example of this, but there are other initiatives too such StudySwap in psychology and the Collective Replication and Education Project (CREP, see Other Internet Resources for both of these , see also Munafò et al. for a more detailed description) which aims to increase the prevalence of replications through undergraduate education.

4.2 Reporting and Dissemination

The TOP Guidelines. The Transparency and Openness Promotion (TOP) guidelines (Nosek et al. 2015) have, as at the end of May, 2018, almost 5,000 journals and organizations as signatories. Developed within psychology, TOP guidelines have formed the basis of other disciplinary specific guidelines, such as the Tools for Transparency in Ecology and Evolution (TTEE). As the name suggests, these guidelines promote more complete and transparent reporting of methodological and statistical practices. This in turn enables authors, reviewers and editors to consider detailed aspects of their sample size planning and design decisions, and to clearly distinguish between confirmatory (planned) analysis and exploratory (post hoc) analysis.

Pre-registration In its simplest form, pre-registration involves making a public, date-stamped statement of predictions and/or hypotheses, either before data is collected, viewed or analysed. The purpose is to distinguish prediction from postdiction (Nosek et al. 2018), or what is elsewhere referred to as confirmatory research from exploratory research (Wagenmakers et al. 2012) and a distinction perhaps more commonly known as hypothesis testing versus hypothesis generating research. Pre-registration of predictive research helps control for HARKing (Kerr 1998) and hindsight bias, and within the frequentist Null Hypothesis Significance Testing, helps contain the false positive error rate to the set alpha level. There are several platforms that host pre-registrations, such as the Open Science Framework (osf.io) and As Predicted (aspredicted.org). The Open Science Framework also hosts a “pre-registration challenge” offering monetary rewards for publishing pre-registered work.

Specific Journal Initiatives. Some high impact journals, having been singled out in the science media as having particularly problematic publishing practices (e.g., Schekman 2013), have taken exceptional steps to improve the completeness, transparency and reproducibility of the research they publish. For example, since 2013, Nature and Nature research journals have engaged in a range of editorial activities aimed at improving reproducibility of research published in their journals (see the editorial announcement, Nature 496, 398, 25 April 2013, doi:10.1038/496398a). In 2017, they introduced checklists and reporting summaries (published alongside articles) in an effort to improve transparency and reproducibility. In 2018, they produced discipline specific versions for Nature Human Behaviour and Nature Ecology & Evolution. Within psychology, the journal Psychological Science (flagship journal of the Association of Psychological Science) was the first to adopt open science practices, such the COS Open Science badges described below. Following a meeting of ecology and evolution journal editors in 2015, a number of journals in these fields have run editorials on this topic, often committing to TTEE guidelines (discussed above). Conservation Biology has in addition adopted a checklist for associate editors (Parker et al. 2016).

4.5 Values, Tone, and Scientific Norms in Open Science Reform

long philosophical debate about what role values do and shoould play in science (Churchman 1948; Rudner 1953; Douglas 2016)
In particular, Nosek et al. (2017) argue that there is a tension between truth and publishability.
the accuracy of scientific results are compromised by the value which journals place on novel and positive results and, consequently, by scientists who value career success to seek to exclusively publish such results in these journals. Many others in addition to Nosek et al. (Hackett 2005; Martin 1992; Sovacool 2008) have taken also take issue with the value which journals and funding bodies have placed on novelty.
Some might interpret the tension as a manifestation of how epistemic values (such as truth and replicability) can be compromised by (arguably) non-epistemic values, such the value

of novel, interesting or surprising results.
Longino, for example, claims that, other things being equal, novelty counts in favour of accepting a theory, and convincingly argues that, in some contexts, it can serve as a "protection against unconscious perpetuation of the sexism and androcentrism" in traditional science (1997: 22). However, she does not discuss novelty specifically in the context of the reproducibility crisis.
Giner-Sorolla (2012), however, does discuss novelty in the context of the crisis, and he offers another perspective on its value. He claims that one reason novelty has been used to define what is publishable or fundable is that it is relatively easy for researchers to establish and for reviewers and editors to detect. Yet, Giner-Sorolla argues, novelty for its own sake perhaps should not be valued, and should in fact be recognized as merely an operationalisation of a deeper concept, such as "ability to advance the field" (567). Giner-Sorolla goes on to point out how such shallow operationalisations of important concepts often lead to problems, for example, using statistical significance to measure the importance of results, or measuring the quality of research by how well outcomes fit with experimenters' prior expectations.

Values are closely connected to discussions about norms in the open science movement. Vazier (2018) and others invoke norms of science – communality, universalism, disinterestedness and organised skepticism- in setting the goals for open science, norms originally articulated by Robert Merton (1942). Each such norm arguably reflects a value which Merton advocated, and each norm may be opposed by a counternorm which denotes behaviour that is in conflict with the norm. For example norm of communality (Mertin called it „communism") reflects the value of collaboration and common ownership of scientific goods. Advocates of open science see such norms, and their values, as an aim for open science. Norm of communality is reflected in sharing and making data open, in open access publishing. In contrast the counternorm of secrecy is associated with a closed, for profit publishing system (Anderson et al. 2010).
Vazier (2018) and others have argued that, at the moment, scientific practice is dominated by counternorms and that a move to Mertonian nors is a goal of the open science reform movement. In particular, self-interestedness, as opposed to the norm of disinterestedness, motivates p-hacking and other Questionable Research Practices. Similarly, a desire to protect one's professional reputation motivates resistance to having one's work replicated by others (Vazire 2018).
This in turn reinforces a counternorm of organized dogmatism rather than organized skepticism which, according to Merton, involves the "temporary suspension of judgment and the detached scrutiny of beliefs" (Merton, 1973).
Anderson et al.'s (2010) focus groups and surveys of scientists suggest that scientists do want to adhere to Merton's norms but that the current incentive structure of science makes this difficult. Changing the structure of penalty and reward systems within science to promote communality, universalism, disinterestedness and organized skepticism instead of their counternorms is an ongoing challenge for the open science reform movement. As Pashler and Wagenmakers (2012) have said:
replicability problems will not be so easily overcome, as they reflect deep-seated human biases and well-entrenched incentives that shape the behavior of individuals and institutions. (2012: 529)

Calls for openness were interpreted as reflecting mistrust, and attempts to replicate others' work as personal attacks (e.g., Schnail 2014 in Other Internet Resources). Nosek, Spies, & Motyl (2012) argue that calls for openness should not be interepreted as mistrust: Opening our research process will make us feel accountable to do our best to get it right; and, if we do not get it right, to increase the opportunities for others to detect the problems and correct them. Openness is not needed because we are untrustworthy; it is needed because we are human. (2012: 626)

5. Conclusion

Reproducibily is associated with a turbulent period in contemporary science.
Called for a re-evaluation of the valuesm incentives, practices and structures which underpin scientifc inquiry.
Meta-science painted a bleak picture of reproducibility in some fields, it also inspired a parallel movement to strengthen the foundations of science.
More progress is to be made, especially in understanding the solutions to the reproducibility crisis