# <Literature Review>

**Donghyun Kang**

## The Emergence and Dominance of Team Science

A crucial trend in contemporary science is the increase in collaboration. This is somewhat contrasting to a popular image of solitary scientists in the 17th-18th centuries, struggling with experimental equipment alone. Collaboration became an essential part of contemporary science partly due to the amounts of capital - both in terms of experimental instruments and various level of researchers including postdocs and graduate students - were required for doing scientific research (Hagstrom 1964). Especially the wartime science during the second world war and the subsequent period of the cold war are known to have contributed to the emergence of "big science" (de Solla Price 1963, Wolfe 2012) supported by the funding organizations such as the National Science Foundation (NSF) and the Defense Advanced Research Projects Agency (DARPA) in the U.S. Since the "violent transition" (de Solla Price 1963) from "solo investigators" (Leahey 2016) to team scientists in the mid-20th century, the trend has been continuing; for example, recent studies have documented that the average number of authors per paper rose to 3.5 in 2000 from 1.9 in 1955 (Wuchty et. al. 2007); the proportion of coauthored papers in JSTOR increased to over 60% in 2011 from 6% in 1900 (West et al. 2013); and the trend holds across almost of all academic disciplines (Leahey, 2016).

Beyond the historical context facilitating collaboration in scientific research, it is pointed out that the development of the communication technology (e.g., the Internet) reduced the cost of collaboration (Binz-Scharf et al. 2015); the intensified competition among scientists is also considered as one of driving forces towards more collaborations. However, the oft-cited fundamental benefit from collaborative research spans to the potential synergies based on the "division of cognitive labors" (i.e., specialization) of scientists with different types of expertise (Leahey 2016). In addition to this, it has been argued that collaborative research activities from a larger team are in a better position for discovery (Esparza & Yamada 2007) and they produce more reliable research in a sense that errors and bad ideas could be filtered out in advance (Clark & Llorens 2012, Bikard et al. 2015). In this context, various policy recommendations and methods to promote collaboration in science have been proposed (Stokols et al., 2008, Croyle 2008, Bennett et al. 2012).

Despite this enthusiasm, recent studies demonstrate some inherent issues of doing team science as well. For example, embedding the co-authorship network of physicists of the $20^{th}$-century into a 2-dimensional Poincare disk at the institution level, Wu et al. (2018) find that the dense connection among physicists created by the growing collaboration has resulted in less diversity (or contraction) in research agendas. In the same vein, drawing upon scientific publications, patents, and software codes spanning from 1954 to 2014, Wu et al. (2018) argue that large teams tend to conduct research conservatively in a sense of choosing their research topics, whereas small teams tend to produce disruptive work more frequently than larger teams.

The issue I would like to address goes one step further from the mentioned above potential disadvantages of team science: the issue of non-reproducibility of scientific research. This is certainly contrasting to the assertion that team science promotes more reliable research outcomes; however, the empirical evidence tells us that while the team science dominates the contemporary science, it also faces with the issue of non-reproducibility, which will be discussed in the next section.

**Replication Crisis of Science**

A puzzling hot issue in the contemporary scientific enterprise is the high level of concern for reproducibility/replicability[1] of published studies. Of course, it is not unusual to observe some degrees of errors or debates in the field of science. The modern organization of science is believed to handle with the self-correcting institution: the peer review process and the open discussion after the publication of the results. The former aims to control the quality of scientific outlets before publication and the latter often leads to open debates regarding certain topics and sometimes results in retraction of papers, if the errors are found to be serious.[2] These practices render modern science authority over other types of knowledge.

However, the issue of non-reproducible science is far beyond dealing with the rates of (natural) error. One recent survey conducted by a leading scientific academic journal Nature in 2016 suggests that 90% of respondents comprising of biologists, chemists, physicists, etc., expressed their concern about the reproducibility crisis. Despite the survey samples do not strictly represent the whole working scientists, the result is quite surprising: for example, approximately 70% of respondents who identified themselves conducted research in medicine had failed to reproduce someone else's experiment at least once; approximately 90% of respondents from chemistry answered that they had an experience of failing in reproduce the other researchers' experiments (Baker, 2016).

[1] In this paper, the two terms are used interchangeably. More nuanced differentiation is of course possible. For example, one might want to assert that replicability is achieved when we can obtain the identical results as any scientific studies instructed whereas reproducibility can be achieved only when more diverse sets of (experimental) conditions produce the consistent result. However, I decided not to differentiate the term given that articles released through Nature and Science do not differentiate the two terms in a strict sense.

[2] Recently, a database of retracted research papers was released in Oct 2018. See http://retractiondatabase.org/.

Social sciences are not an exception. Even though the concern had been around academia, the issue of non-reproducible research was strongly raised from psychology (especially in the field of social psychology) starting from the early 2010s. In 2012, Pashler et al., the editors in *Perspectives on Psychological Science* at that time called serious attention to the non-replicability in psychology. Moreover, the "Reproducibility Project: Psychology" which was launched by the Center for Open Science in 2011. After their attempts to replicate 100 psychological studies published in 2008 through three leading journals[3], the project group concluded that only 35 studies out of the 97 studies with significant results were replicated (Open Science Collaboration, 2015).

One tightly related issue with the replicability of research outcomes is the application of statistics in research practice. The standard frequentist's statistical procedure allows two types of errors - the type 1 error (false-positive) and the type 2 error (false-negative; i.e., statistical power). Each discipline or specific research field has own standard in terms of evaluating statistical significance; while experimental physicists impose a stricter standard [4] in terms of reporting research outcomes to prevent false-positive findings, medical and social science research generally accepts the 5% significance level (i.e., the type 1 error) and this practice is often blamed one of the underlying reasons leading to the non-replicability crisis (Ioannids 2005,  2016).

Several statistical techniques called meta-analysis have been of course developed to overcome the limitation of a single study with small sample size and generate more precise ranges of effect sizes. The multilevel regression (Hox et al., 2017) and the False Discovery Rate (Benjamini & Yosef 1995)

[3] *Psychological Science, the Journal of Personality and Social Psychology, the Journal of Experimental Psychology: Learning, Memory, and Cognition.*

[4] The threshold for "evidence of a particle," corresponds to p = 0.003, and the standard for "discovery" is p = 0.0000003. (citation needed)

approach are often employed in medical research. Meta-analysis technique, however, is still not free from the so-called "file drawer problem" (Rosenthal, 1979) that leads to the non-report of the insignificant result. The "publish or perish" culture in modern science often pushes researchers to neglect the insignificant research outcomes; thus, the null findings tend not to get neither published nor recognized in many cases, which could distort the result from a meta-analysis by lowering the proportion of published research articles with null findings.

To address the "crisis" of science, many suggestions have been made from various sides: the large research funding organizations such as NIH and NSF and leading academic journals in each discipline have begun to require researchers to publicize the data and document a better pre-research design (and pre-registering) to increase the transparency (Collins & Tabak, 2014); a reformation to shift the incentive structure for publication more favorable towards null results has been suggested (Begley & Ellis, 2012); the criticism against the traditional 5% threshold has been growing[5] and there is even an academic journal[6] that stopped using the p-value as scientific evidence; a new approach such as "blind analysis" was suggested to counter possible bias and errors (MacCoun & Perlmutter 2015). Admittedly, all the suggestions are a worthwhile consideration. However, they do not explicitly consider that the modern science is collective intellectual enterprise (Crane, 1972), which relies on the social mechanism of the "wisdom of crowds" to reach judgment (Surowiecki, 2005).

[5] *The American Statistician* released a special issue regarding the misuse of the p-value on Mar 2019.

[6] *Basic and Applied Social Psychology* stopped publishing paper relying on the p-value from Jan 2015.

**The Wisdom of Crowd and Its Implication on Science**

The "wisdom of crowds" refers to a social phenomenon in which the aggregated judgments or decisions from independent individuals often leads to better results than those made by a single individual (Surowiecki, 2005). The wisdom of crowds was first documented by Francis Galton (1907), who was surprised by the fact that the collective guessing about weights of an ox was almost close to the actual weight when "the middlemost estimate" (i.e., the median) was considered. In his report, Galton concluded that this showed the "more creditable to the trust-worthiness of a democratic judgement than might have been expected" (Galton 1907: 451).

One of the assumptions in the original ox weight-guessing that Galton observed is the independence of the judgment which prevents the group thinking or social influence that can result in bias. In addition to this, the diversity of opinions, thoughts, or guessing matters as well; to wit, the wisdom of crowds is better than an individual guess on average when their guesses are aggregated in a macro scale. In other words, the wisdom of crowds works properly when decentralized individuals with diverse viewpoints make their judgments independently and aggregate them into a collective one.[7] While the debates regarding the relationship between social influence and the quality of collective judgment are proceeding recently (see Lorenz et al. (2010) and Becker et al. (2016) for the negative impact of social influence on collective judgment; Almaatouq et al. (2017) for the conditions that mitigate the negativity), the research regarding the wisdom of crowds basically argues that a group of people can make a good decision when appropriate social conditions are met.

At first glance, the wisdom of crowds does not seem related to the field of science considering that scientists are supposed to have expertise in their fields: they are not mere "crowds". But the

[7] More formally, group error = $\sum \text{individual error} - \text{diversity}$, according to Page (2008).

empirical evidence suggests that the collective knowledge production system of science is also conceived similarly to the wisdom of crowds. Collaboration such as coauthoring a paper is not the only case; beyond this, modern science constitutes “Invisible colleges” (Crane, 1972) as an autonomous self-regulating system of knowledge production. In other words, the modern process of the collective scientific endeavor is based on the “collective wisdom of scientists” (Surowiecki, 2005: 169).

Evidence, however, also suggests that scientists are not free from social influence and bias. For example, studies have shown low inter-rater reliability during the peer review processes in scientific publications (Mahoney, 1977; Emerson et al., 2010, Lee et al. 2013). Moreover, a recent study on the grant reviews process at the National Institute of Health showed that the trade-off between the expertise and bias of the grant reviewers (Li, 2017). These empirical studies display different pictures from what Merton (1942) hypothesized about modern science with the normative ethos of universalism, communism, disinterestedness, and organized skepticism.

In sum, the reliability of science can be also affected by the social structure of the field of science. Thus, we need to take into accounts social aspects of science to explain how contemporary science is organized and to evaluate the robustness of science.

**References**

A. Almaatouq, A. Noriega & A. Pentland. “Wisdom of Dynamic Networks: Selective Social Learning and the Adaptive Wisdom of Crowds.” Collective Intelligence Conference, NYU, New York, USA. June 2017

Becker, Joshua, Devon Brackbill, and Damon Centola. "Network dynamics of social influence in the wisdom of crowds." *Proceedings of the national academy of sciences 114*, no. 26 (2017): E5070-E5076.

Binz-Scharf, Maria C., Yuval Kalish, and Leslie Paik. "Making science: New generations of collaborative knowledge production." *American Behavioral Scientist* 59, no. 5 (2015): 531-547.

Begley, C. Glenn, and Lee M. Ellis. "Raise standards for preclinical cancer research." *Nature* 483, no. 7391 (2012): 531-533.

Benjamini, Yoav, and Yosef Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal statistical society: series B (Methodological)* 57, no. 1 (1995): 289-300.

Bennett, L. Michelle, and Howard Gadlin. "Collaboration and team science: from theory to practice." Journal of Investigative Medicine 60, no. 5 (2012): 768-775.

Clark, Benjamin Y., and Jared J. Llorens. "Investments in scientific research: Examining the funding threshold effects on scientific collaboration and variation by academic discipline." Policy Studies Journal 40, no. 4 (2012): 698-729.

Collins, Francis S., and Lawrence A. Tabak. "Policy: NIH plans to enhance reproducibility." *Nature News* 505, no. 7485 (2014): 612.

Crane, Diana. "Invisible colleges; diffusion of knowledge in scientific communities." (1972).

Croyle, Robert T. "The National Cancer Institute's transdisciplinary centers initiatives and the need for building a science of team science." *American journal of preventive medicine* 35, no. 2 (2008): S90-S93.

Danchev, Valentin, Andrey Rzhetsky, and James A. Evans. "Centralized" big science" communities more likely generate non-replicable results." *arXiv preprint arXiv:1801.05042*(2018).

Fortunato, Santo, Carl T. Bergstrom, Katy Börner, James A. Evans, Dirk Helbing, Staša Milojević, Alexander M. Petersen et al. "Science of science." *Science* 359, no. 6379 (2018): eaao0185.

Galton, Francis. "Vox populi (the wisdom of crowds)." *Nature*75, no. 7 (1907): 450-451.

Hagstrom, Warren O. "Traditional and modern forms of scientific teamwork." *Administrative Science Quarterly* (1964): 241-263.

Hicks, Diana, and J. Sylvan Katz. "Science policy for a highly collaborative science system." Science and public policy 23, no. 1 (1996): 39-44.

Hox, Joop J., Mirjam Moerbeek, and Rens Van de Schoot. *Multilevel analysis: Techniques and applications*. Routledge, 2017.

Ioannidis, John PA. "Why most published research findings are false." *PLoS medicine* 2, no. 8 (2005): e124.

Ioannidis, John PA. "Why most clinical research is not useful." *PLoS medicine* 13, no. 6 (2016): e1002049.

Leahey, Erin. "From sole investigator to team scientist: Trends in the practice and study of research collaboration." Annual review of sociology 42 (2016): 81-100.

Lee, Carole J., Cassidy R. Sugimoto, Guo Zhang, and Blaise Cronin. "Bias in peer review." *Journal of the American Society for Information Science and Technology* 64, no. 1 (2013): 2-17.

Latour, Bruno. *Science in action: How to follow scientists and engineers through society*. Harvard university press, 1987.

Li, Danielle. "Expertise versus Bias in Evaluation: Evidence from the NIH." *American Economic Journal: Applied Economics* 9, no. 2 (2017): 60-92.

Lorenz, Jan, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing. "How social influence can undermine the wisdom of crowd effect." *Proceedings of the national academy of sciences 108*, no. 22 (2011): 9020-9025.

MacCoun, Robert, and Saul Perlmutter. "Blind analysis: hide results to seek the truth." *Nature News* 526, no. 7572 (2015): 187.

Mahoney, Michael J. "Publication prejudices: An experimental study of confirmatory bias in the peer review system." *Cognitive therapy and research* 1, no. 2 (1977): 161-175.

Merton, Robert K. "Singletons and multiples in scientific discovery: A chapter in the sociology of science." *Proceedings of the American Philosophical Society* 105, no. 5 (1961): 470-486.

Merton, Robert K. "The Normative Structure of Science", in Merton, Robert K. (ed.), *The Sociology of Science: Theoretical and Empirical Investigations*, Chicago: University of Chicago Press (1973) [1942]

Nissen, Silas Boye, Tali Magidson, Kevin Gross, and Carl T. Bergstrom. "Publication bias and the canonization of false facts." *Elife* 5 (2016): e21451.

O'Brien, Timothy L. "Change in academic coauthorship, 1953–2003." Science, Technology, & Human Values 37, no. 3 (2012): 210-234.

Open Science Collaboration. "Estimating the reproducibility of psychological science." *Science* 349, no. 6251 (2015).

Page, Scott E. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies-New Edition*. Princeton University Press, 2008.

Pashler, Harold, and Eric–Jan Wagenmakers. "Editors' introduction to the special section on replicability in psychological science: A crisis of confidence?." *Perspectives on Psychological Science* 7, no. 6 (2012): 528-530.

Price, Derek J. de Solla (Derek John de Solla). *Little Science, Big Science-- and Beyond*. New York: Columbia University Press, 1986 [1963].

Prinz, Florian, Thomas Schlange, and Khusru Asadullah. "Believe it or not: how much can we rely on published data on potential drug targets?." *Nature reviews Drug discovery* 10, no. 9 (2011): 712.

Rosenthal, Robert. "The file drawer problem and tolerance for null results." *Psychological bulletin* 86, no. 3 (1979): 638.

Stokols, Daniel, Kara L. Hall, Brandie K. Taylor, and Richard P. Moser. "The science of team science: overview of the field and introduction to the supplement." American journal of preventive medicine 35, no. 2 (2008): S77-S89.

Shenhav, Liat, Ruth Heller, and Yoav Benjamini. "Quantifying replicability in systematic reviews: the r-value." *arXiv preprint arXiv:1502.00088* (2015).

Surowiecki, James. *The wisdom of crowds*. Anchor, 2005.

Wolfe, Audra J. *Competing with the Soviets: Science, technology, and the state in Cold War America*. JHU Press, 2012.

Wu, Lingfei, Dashun Wang, and James A. Evans. "Large teams develop and small teams disrupt science and technology." *Nature* 566, no. 7744 (2019): 378.

Wu, Lingfei, Linzhuo Li, and James Evans. "Social connection induces cultural contraction: Evidence from hyperbolic embeddings of social and semantic networks." *arXiv preprint arXiv:1807.10216* (2018).

Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi. "The increasing dominance of teams in production of knowledge." *Science 316, no. 5827* (2007): 1036-1039.