Accepted version—please cite the published version, which will appear in *Synthese*.

Improving the Quality of Case-Based Research in the Philosophy of Contemporary Sciences

Karen Yan (first author, corresponding author)
Institute of Philosophy of Mind and Cognition
National Yang-Ming University
Taiwan

Meng-Li Tsai
Department of Biomechatronic Engineering
National Ilan University
Taiwan

Tsung-Ren Huang
Department of Psychology
National Taiwan University
Taiwan

Abstract

This paper aims to address some methodological issues related to case-based research in the philosophy of contemporary sciences. We focus on the selection processes by which philosophers pick or generate a particular set of papers to conduct their case-based research. We illustrate how to use various quantitative and qualitative methods to improve the epistemic features of the selection processes, and help generate some potential case-based hypotheses for further philosophical investigation.

1. Introduction

This paper aims to address some methodological issues related to case-based research in the philosophy of contemporary sciences. Methodological issues regarding case-based research have been well-discussed among social scientists, but historians and philosophers of science have paid less attention to them (Morgan 2019). Recently, some have addressed methodological issues encountered by historians and philosophers of science doing case-based research (Chang 2012; Morgan 2012, 2014 and 2019; Ankeny 2007, 2011 and 2014; Pietsch 2016; Currie 2015). Among them, the most discussed methodological issue concerns how knowledge generated from a single case can be generalized or transferred to new sites. However, in this body of literature, less attention is paid to how philosophers of contemporary sciences pick or generate a particular set of papers to begin their case-based research. Presumably, there is some selection process through which a set of papers is selected as the starting point for their case-based investigation. But there is very little philosophical discussion about the epistemic features of this selection process.

To see the problem within the context of the philosophy of contemporary sciences, let's compare the following two situations. If you want to get started on a scientific case regarding the Chemical Revolution, it's relatively clear how to get

started. One way to go about it is to read some of the recent secondary literature, and then read some of the primary literature cited in that secondary literature. But if you want to get started on a contemporary scientific case, you may not be able to follow the same procedure, i.e., if there is no secondary literature. Thus, depending on one's target field/period and research questions, sometimes it is obvious which set of work is the starting point for case-based investigation, and sometimes it is not. Temporal differences between the two sorts of situations are also important. The Chemical Revolution happened over 200 years ago, and the set of primary sources that you would have to examine is relatively fixed at this point. Although scholars may disagree about various issues, there is some consensus regarding how to characterize this event, its historical significance, and its aftermath. However, ongoing research in the contemporary sciences is (obviously) ongoing. It is therefore unclear what the historical significance of this work will be. So without the benefit of hindsight, how do philosophers of contemporary sciences choose the primary source material in this kind of case? Moreover, philosophers of contemporary sciences face another difficult challenge regarding the large number of scholarly papers in a given contemporary scientific field. It is quite often the case that if one searches some keyword in a database, it returns more than 10,000 relevant papers. In such a case, it is not possible to read through all the papers and judge which papers are more relevant or provide more useful information than the others in the database. Even if one were to read all the papers, the result of the analysis might not be robust across different readers due to interpretative differences. Philosophers of contemporary sciences certainly need some strategies and tools to handle large bodies of literature without the benefit of hindsight. There are obviously many strategies and tools one can use to select some papers from the database and ignore others, and it is philosophically important to ask how philosophers justify their selection of papers. Why is it that a particular set of papers, and not some other set, is selected as the starting point of a particular case-based investigation? Is the result of a given selection process robust across different individuals? What relationship does this set of papers bear to the targeted body of literature and research question? These questions are examples of the types of questions we aim to raise regarding the epistemic features of the selection processes.

In this paper, we aim to address this neglected issue concerning how philosophers of contemporary sciences select a set of papers to conduct their case-based research. Our strategy is to illustrate how quantitative and qualitative methods can help philosophers improve the quality of the selection processes and the ways in which they develop philosophical accounts of their targeted cases. We adopt Morgan's (2012) characterization of case study (or case-based) research. According to her, case study research is a type of epistemic genre, i.e., a way of doing science, that offers a generic way of studying a whole event, episode, or situation within its own context. Moreover, case-based research can be conducted by employing various methods, such as survey methods, ethnographic methods, statistical methods, and historical methods. Morgan's characterization helps us to clarify our methodological goal in this paper. We aim to illustrate three existing quantitative methods: citation analysis, classification analysis, and text analysis; and a practice-based approach to qualitative methods, i.e., using scientific practices as units of analysis for various types of qualitative methods, which we illustrate in terms of a method for performing literature review. The goal is to show how they can improve the ways in which philosophers conduct their case-based research. Specifically, we aim to show how the

methods and the approach can improve the epistemic features of the selection processes.[1]

Having positioned our paper in terms of the literature on case-based research within the philosophy of contemporary sciences, here are four more specific ways to state our goals in this paper. Our paper will show how to

1. robustly identify non-transparent patterns regarding a large body of literature by using existing quantitative methods,
2. justify the selection of a set of papers based on the identified patterns,
3. further analyze the relationship between the selected paper(s) and the identified patterns and generate research questions by using a practice-based literature review, and
4. use the quantitative and qualitative results to generate potential hypotheses.

The quantitative methods here involve using various computational tools to perform a large-scale literature analysis. This kind of analysis helps us identify non-transparent patterns in a given body of literature in the sense that the patterns are not easily discernible by a traditional close-reading method. Moreover, the patterns are identified in a robust way in the sense that, given the same dataset and the same computational procedure, the result of the quantitative analysis is the same regardless who performs it. The practice-based approach to qualitative methods here involves using scientific practices as units of analysis when one is employing some type of qualitative method. This approach is inspired by the recent practice-turn in the philosophy of science.[2] Proponents of this approach take scientific practices as their units of analysis when addressing philosophical questions. Scientific practices are activities performed by scientists in order to achieve certain goals. The practice-based approach thus focuses philosophers' attention on the relationships that obtain among three things: scientists, their activities, and their goals (Chang 2011). The practice-based approach aims to correct the default tendency among some philosophers of science, which is to focus almost exclusively on theories as the relevant units of analysis and pay less attention to, or even ignore, issues concerning scientists' goals and activities. The inadequacy of this tendency has been shown to impede philosophers' understanding of the relevant philosophical issues and debates.[3]

In the following, we illustrate how philosophers can use existing quantitative methods and the practice-based approach to analyze the heart-rate variability (HRV) field and select a set of papers for their case-based research. To that end, the paper will proceed as follows. In Section 2, we show how to use three quantitative methods and present three sets of data: one from citation analysis, one from classification analysis, and one from text analysis using word frequency analysis. Based on these

[1] It is important to note that we do not aim to propose original scientific methods for doing scientometric or bibliometric analysis.

[2] See Soler, Zwart, Lynch, and Israel-Jost (2014) for a recent edited collection regarding this approach.

[3] Examples of such issues and debates include: how reductionistic research strategies actually work in biology (Wimsatt 2006), how scientists perform modeling and use models (Giere 1988), how biologists use the molecular concept of gene (Waters 2014), how the metaphysical issues of individuation and individuality are connected to various sciences (Bueno, Chen, and Fagan 2018), and how Kuhn himself moved away from the novel contributions he made in his seminal book *The Structure of Scientific Revolutions* (Shan 2020).

three sets of data, we identify a key review article and a significant relationship between the publication of this review article and the subsequent quantitative and qualitative development of HRV research. In Section 3, we show how to use the practice-based approach to analyze the review article and argue that its main achievement is standardization. In Section 4, we demonstrate how to generate some potential hypotheses based on the quantitative and qualitative results. In Section 5, we conclude that the quantitative methods and the practice-based approach to qualitative methods can improve the quality of the selection processes for philosophers doing case-based research.

2. Illustrating Three Quantitative Methods

In this section and the next section, we will illustrate how to apply some existing quantitative methods and the practice-based approach to a specific biomedical field and how the applications improve the quality of the selection processes. This strategy helps us to illustrate how we achieve the four goals listed in Section 1. Before we go into the details, we will first introduce our target research topic, biomedical field, period, and body of literature, all of which center on the study of a physiological phenomenon called heart-rate variability (HRV) and its clinical and non-clinical applications. HRV refers to "the oscillation in the interval between consecutive heart beats as well as the oscillations between consecutive instantaneous heart rates" (Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology 1996, p. 354). It is worth noting that HRV is a physiological phenomenon that is about variations of the intervals that are constituted by heart beats or heart rates, not about heart rates per se. The clinical relevance of HRV was first reported in Hon and Lee's (1963) investigation of fetal distress. Since then, many researchers have recognized HRV's potential in clinical and non-clinical applications. For example, scientists found reliable experimental evidence to support the correlation between a propensity for lethal heart arrhythmias and signs of autonomic nervous activity, e.g., increased sympathetic activity or reduced vagal activity (Lown and Verrier 1976; Corr, Yamada, and Witkowski 1986; Schwartz and Priori 1990; Levy and Schwartz 1994). Measuring HRV is considered to be a promising way of observing how the sympathetic and parasympathetic nervous systems interact with each other (Rajendra Acharya et al. 2006, p. 1031). Scientists have been working on developing HRV as quantitative markers of autonomic activity in order to predict cardiovascular and other types of diseases (Kleiger, Miller, Bigger, and Moss 1987; Malik, Farrell, Cripps, and Camm 1989; Bigger et al. 1992).

Imagine now that you are unfamiliar with the HRV field and aim to develop some case-based research from this field to address some philosophical issues. What are the difficulties you will encounter when you are attempting to justify your selection of a particular set of papers as your starting point? Where and how do you begin your analysis? You could ask some HRV scientists about important papers in the field, but you might get different answers from different scientists. How do you evaluate these different opinions? You could pick all the review articles in the field, but you might not be able to evaluate their influences on the field by just reading the review articles. How do you show the connections between the review articles and all the other papers in the field? You could just pick some highly cited papers in the field, but you need to show what the highly cited papers mean to the field. These are examples of the difficulties philosophers might face when they are attempting to

justify their selection of some HRV papers as the starting point of their case-based investigation. In this section, we aim to show that using existing quantitative methods can generate robust quantitative reasons to justify one's selection of some papers and thus address some of the above difficulties.

The quantitative methods in this section involve using a database and computational tools for doing literature analysis. Databases and computational tools are very useful in this regard. We can search for all of the articles relevant to HRV and perform different quantitative analyses on them. Of course, there are many available databases and tools one can use. We do not think there is only one legitimate way to conduct quantitative analyses on the relevant body of literature for philosophical purposes**.** What really matters, philosophically, is how one uses existing databases and tools to robustly identify non-transparent patterns from quantitative analyses, and use the identified patterns to justify one's selection of a particular set of papers. Hence, we focus on the epistemic features of this selection process, and in this section, we aim to make a methodological contribution by supporting the following two claims. First, using these quantitative methods increases the robustness of the selection process. This robustness can be characterized in terms of what Leonelli (2018) calls computational replicability, which is the property of "being able to obtain the same outcomes when running a given dataset through the same algorithms" (p. 135). Second, these quantitative methods can also generate reasons for justifying one's selection of a set of papers based on the identified robust patterns in a body of literature.

2.1 First Quantitative Method: Citation Analysis

2.1.1 Method

We queried Web of Science on August 7, 2017 with 'heart rate variability' (without quotation marks),[4] collected papers published between 1970 and 2016 (articles assigned to an issue of a journal, not including articles published online first without being assigned to an issue of a journal), and exported the results as a Microsoft Excel file. We chose Web of Science because it has citation information; and because, in contrast to Scopus, it covers articles published before 1995 (Falagas, Pitsouni, Malietzis, and Pappas 2008). We then manually selected articles on HRV research that satisfy the following threshold of an article's average number of citations per year:

$$\frac{\text{Total citation number}}{\text{2017 — year of publication}} \geq 11$$

The threshold 11 is determined by the following two considerations. First, we want the resulting sample size to be an adequate size, i.e., not too small to miss influential papers, but not too big to include papers that are only cited a few times. Second, after

[4] Please see Section 3, footnote 8 for our explanation of why we queried 'heart rate variability' without quotation marks.

trying different thresholds, we found that the threshold 11 provides the most informative visualization (see Figure 1 and Section 2.1.2 for more details).

2.1.2 Results

Using this method, we arrived at a list of 40 highly cited articles. Among them are 25 original research articles, numbered as OA-1, OA-2, OA-3, and so on; and 15 review articles, numbered as RA-1, RA-2, RA-3, and so on. Figure 1 contains two graphs that show the number of citations of all 40 articles. The top graph in Figure 1 shows the quantitative differences in citation number between the 25 original research articles and RA-11. The bottom graph in Figure 1 shows the quantitative differences in citation number between RA-11 and the other 14 review articles. In both graphs, there are three rough patterns: (1) a cluster of papers with low threshold, (2) a cluster of papers with medium threshold, and (3) the fact that RA-11 stands out regardless of whether it is plotted against all the original articles (the top graph) or all the review articles (the bottom graph).

RA-11 is the review article titled "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," authored by the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology.[5] It is cited 6460 times and is the most cited article on our list. As Figure 1 shows, RA-11 is cited much more than other highly cited original research and review articles.

[5] We will use 'Task Force 1996' in subsequent citations to this article.

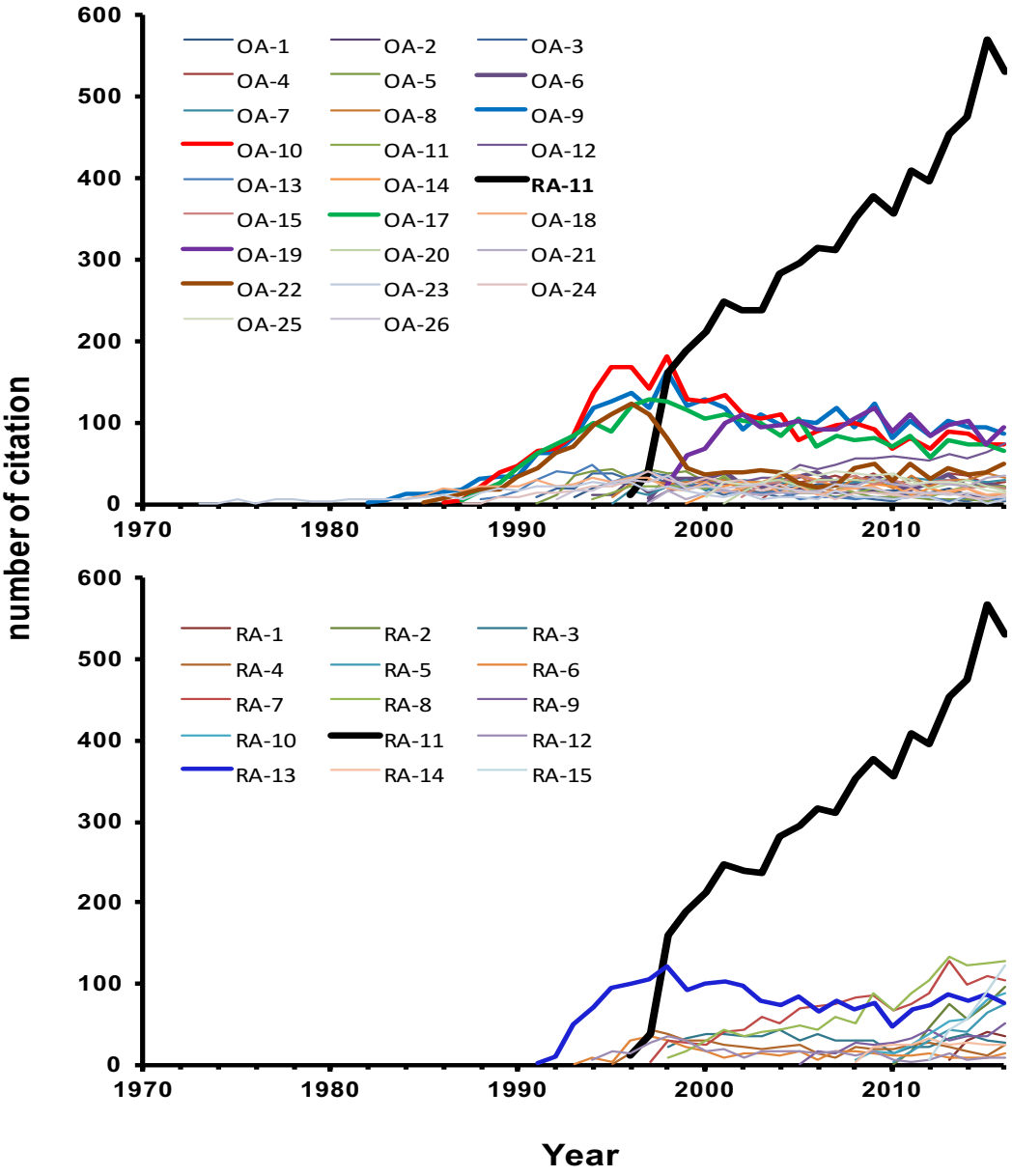

Fig. 1 Number of citations of the 25 original articles and RA-11 (top) and of the 15 review articles (bottom). Fig. 1 shows that RA-11 is cited significantly more compared to both the original research articles and the rest of the review articles

It is important to acknowledge the limitations of this method and of the information provided by citation number. But at the very least, this quantitative information gives us a starting point for further philosophical analysis because it is based on screening all the relevant papers in a given database. The purpose of this citation analysis is to provide non-arbitrary quantitative reasons to guide philosophers through the process of selecting some papers as the starting point of their case-based investigation. In the following subsections, we report the results of two additional analyses to reveal more quantitative and qualitative information about the development of the HRV field in relation to RA-11, specifically before and after 1996, the year RA-11 was published.

2.2 Second Quantitative Method: Classification Analysis

2.2.1 Method

We queried PubMed on May 12, 2017 with 'heart rate variability' (without quotation marks),[6] collected papers published between 1970 and 2016 (articles assigned to an issue of a journal, not including articles published online first without being assigned to an issue of a journal), exported the results as a Microsoft Excel file, and manually deleted papers not related to HRV research and duplicate records. This process resulted in 19,795 articles. We chose PubMed because its keyword searching function offers optimal update frequency (Falagas, Pitsouni, Malietzis, and Pappas 2008). We then manually classified these 19,795 articles into one of the following three categories: (A) foundational research on relevant mechanisms or technological advances, (B) correlation to cardiovascular disease, and (C) correlation to non-cardiovascular disease. The classification was done by an HRV expert based on the titles of the articles. When the title of an article was insufficient for him to make a judgment, he read the abstract or main text to make the judgment.

2.2.2 Results

Figure 2 shows the percentage of each type of article in every thousand articles arranged chronologically. The results show that, before RA-11 was published in 1996, the percentage of category A research decreased and then increased, the percentage of category B research was increasing, and the percentage of category C research was decreasing. After RA-11 was published in 1996, these trends changed. The percentage of category C articles increased rapidly between 1996 and 2001, and has remained above 60% since 2001. The percentages of category A and B articles, by contrast, steadily decreased between 1996 and 2001, and since 2001 have remained less than 10% for category A, and roughly between 15% and 25% for category B. Thus, Figure 2 shows that the HRV field transformed in both quantitative and qualitative ways after the publication of RA-11.

[6] Please see Section 3, footnote 8 for our explanation of why we queried 'heart rate variability' without quotation marks.

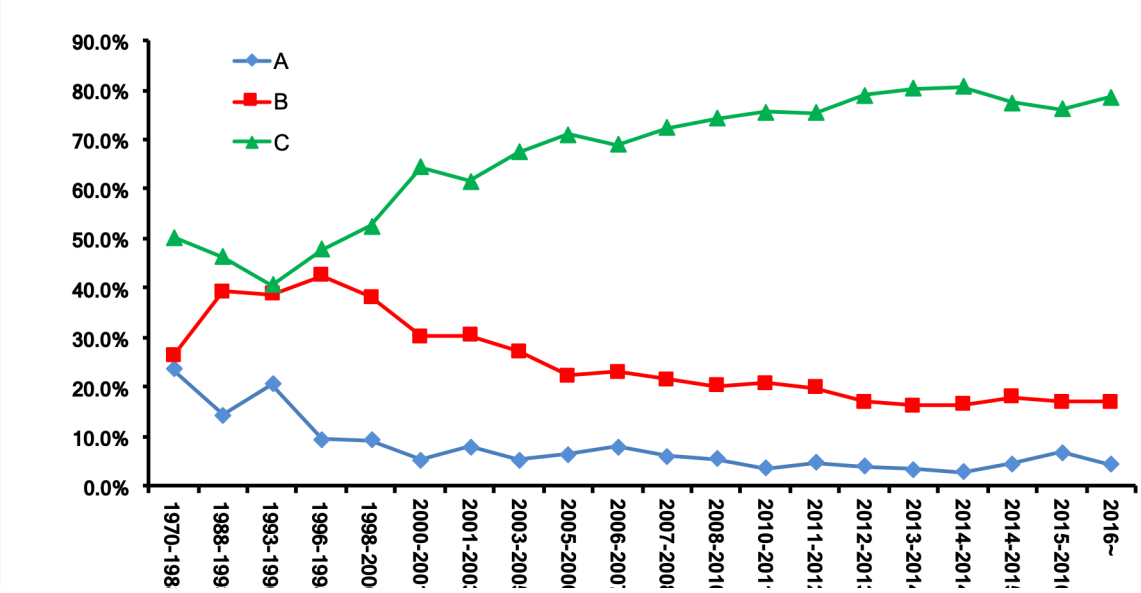

Fig. 2 Percentages of articles in each of the three categories per group of one thousand articles, arranged chronologically

2.3 Third Quantitative Method: Text Analysis

2.3.1 Method

Text analysis can create sets of structured data out of heaps of unstructured text. One can then employ text analytics, which involves a set of computational techniques, to analyze the data in order to derive valuable information and discover patterns. Specifically, we calculated the frequency of each word in all the titles of the 19,795 articles used in our classification analysis. The most frequent words discovered herein were then highlighted by word clouds to provide insightful visualization (Felix, Franconeri, and Bertini 2017). Some might worry about the limitations of analyzing papers' titles only (Borgerson 2009). This is why we use the text analysis of article titles to provide an independent way of confirming the results of the classification analysis in Section 2.2, not to directly infer something from the analysis of titles alone. Also, we focused on a text analysis of titles because titles often indicate the topic of a paper, and because our purpose was to identify how the research topics evolved between 1970 and 2016. We did not aim to perform any finer-grained analyses regarding the contents of the papers. We plotted the five most frequent words in every thousand articles arranged chronologically. The text size in Figure 3 reflects the absolute frequency of occurrence of each word.

2.3.2 Results

Figure 3 shows that, before RA-11 was published, the most frequent words included fetal, ventricular, spectral, myocardial, and infarction. Between 1996 and 2001, sleep starts showing up. After 2001, exercise, sleep, and stress became the three most frequent words. Exercise, sleep, and stress are three important topics in the clinical applications of HRV. This shows that, after RA-11 was published, the HRV field shifted its main research focus toward clinical applications. This shift is consistent

with what our second set of data shows: Between 1996 and 2001, the percentage of category C research rapidly increased, and after 2001, it consistently remained above 60%.

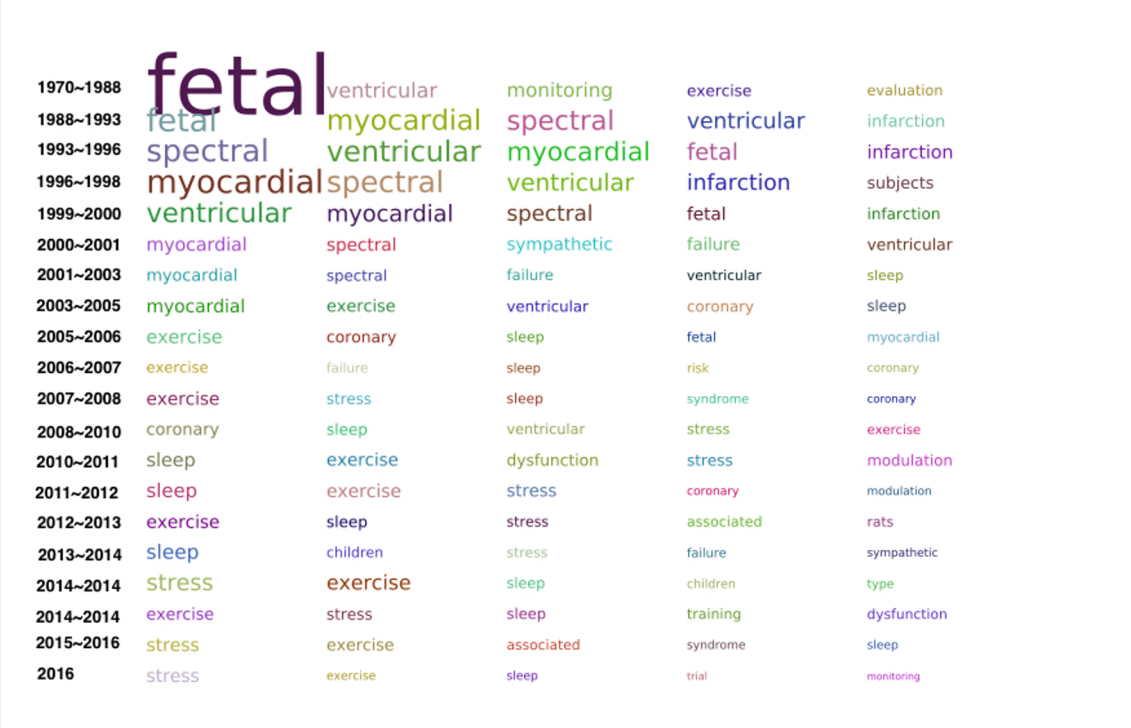

Fig. 3 The five most frequent words in the article titles of every thousand articles, arranged chronologically

2.4 Discussion of the Results

To sum up, our first set of data shows that RA-11 is the most cited article in the field. Our second set of data shows the quantitative and qualitative development of HRV research before and after RA-11 was published in 1996. Our third set of data shows a pattern of development that is consistent with our second set of data. Based on these three sets of quantitative data, we can justify using RA-11 as the starting point of our case-based research and investigate the following research question (which is also generated from the three datasets): Why did the HRV field exhibit the identified quantitative and qualitative changes after RA-11 was published in 1996?

In the subsections above, we have shown how quantitative methods can help philosophers to identify an intriguing correlation between RA-11 and a specific pattern of change in the HRV field, especially the flourishing and steady output of category C research. By doing so, we aim to provide support for the following two points.

The first point concerns how philosophers can robustly identify non-transparent patterns in the literature. Note that the identified correlation and pattern of change would not be transparent to those who use a traditional close-reading method. Moreover, the citation analysis and the text analysis methods provide a strong degree of robustness for this particular selection process because of the degree of

computational replicability they achieve. That is, by using the same dataset and the same computational procedure, different scholars are very likely to produce the same quantitative results. Of course, our classification analysis may not yield the highest possible degree of robustness for the selection process because it was done manually rather than following a computational procedure. However, we do not aim to argue that our approach provides the strongest degree of robustness for the selection process. There are obviously other ways of performing quantitative analyses, and they might achieve a degree of robustness in this particular case that is higher than what we have shown. The point we aim to make here is that using quantitative methods can provide some degree of robustness for the selection process because of the degree of computational replicability they can achieve. Philosophers of contemporary sciences can improve the quality of their selection process by establishing some degree of robustness for their selection process.

The second point concerns how philosophers can generate verifiable and non-private reasons to justify their selection processes. We have shown that philosophers can cite the identified non-transparent patterns as reasons to support their choice of a set of papers, as we have done with RA-11. These reasons are verifiable in the sense that different people can run the same dataset with the same computational procedure to verify whether they get the same result. These reasons are also non-private in the sense that they are not based on scientists' common sense or intuition. That said, it is often the case that an experienced scientist does have some good sense or intuition about how other scientists in her communities perform their research. For example, it might strike her to be obvious, or even trivial, that the review work done by a task force has a strong influence on the field. She might think this is just how science works. However, in order to justify their selection processes, philosophers of science are expected to provide reasons that enable them to engage in the social process of argumentation with other philosophers. These reasons function as epistemic objects that other philosophers can examine, verify, challenge, or revise when they examine whether or not a given selection process is really a good one for addressing a specific research question. An experienced scientist's common sense or intuition certainly can provide good insights for philosophers. But common sense and intuition cannot perform the epistemic role played by verifiable and non-private reasons in the social process of argumentation. Thus, a selection process grounded on verifiable and non-private reasons is much better than one grounded on a scientist's common sense or intuition.

By using these quantitative methods, we are able to identify an intriguing correlation between RA-11 and the specific pattern of change in the HRV field. Without the above quantitative analyses, we, as philosophers, cannot robustly identify non-transparent patterns in the literature and justify our selection of RA-11 with evidence-based, quantitative reasons that can be subject to further empirical examination.

It is perhaps worth noting at this point that it is not the aim of the quantitative methods we employ to yield a fine-grained analysis of the practices or claims in any specific article. To be sure, this detailed, content-based type of analysis plays an important role in our approach. But it does so only after we have used the quantitative methods to identify important patterns in the literature, generate our research

question, and justify why a particular set of papers can be used as a foundation for our case study.

3. Illustrating a Practice-Based Approach to Qualitative Methods

However, at this point, we do not have an in-depth understanding of why this correlation holds, what RA-11 means to the field, or how to use our knowledge of this correlation for further investigation. This is where the practice-based approach to qualitative methods comes in. It is important to note that, because of the quantitative analyses we have reported in Section 2, we are able to focus our attention on RA-11 and its relationship to the rest of the papers in the field. Without the guidance from the quantitative analyses, we would not know where to apply the practice-based approach to subsequent qualitative analysis.

As we have mentioned earlier, the practice-based approach is inspired by the recent practice-turn in the philosophy of science. The key feature of the practice-based approach is to use scientific practices as the units of analysis. Scientific practices are activities that are performed by scientists in order to achieve certain goals (Chang 2011). In order to understand scientific practices, it is therefore important to understand the relationships that hold among scientists, their activities, and their goals. In this section, we aim to apply the practice-based approach to our qualitative analysis of the content of RA-11. Specifically, we have two goals. First, we will use scientific practices as our units of analysis to analyze what the task force accomplished. We will identify the goals that the members of the task force set for themselves. Identifying these goals will help us to identify the activities they performed in order to achieve the goals. Second, we will analyze how the accomplishments of the task force relate to the subsequent patterns of change in the field. It is important to note that we are able to specify the above two goals to further our case-based investigation because we have used the above quantitative analyses to justify our selection of RA-11 and identify the correlation between RA-11 and the subsequent patterns of change in the field.

The task force states five goals they aim to achieve: "standardize nomenclature and develop definitions of terms; specify standard methods of measurement; define physiological and pathophysiological correlates; describe currently appropriate clinical applications, and identify areas for future research" (Task Force 1996, p. 354).[7] In 1996, the task force published their results as the review article labeled as RA-11 in Figure 1. In what follows, we will first review the relevant details about how the task force achieved each goal, and then show that many of the scientific activities the task force performed involved standardization.

The first goal of RA-11 is to standardize nomenclature. Before RA-11 was published, many terms were used to refer to the phenomenon of interest, i.e., variations of the intervals between consecutive heart beats or consecutive instantaneous heart rates. For example, terms like 'cycle length variability,' 'heart

[7] Citations in the remainder of this section will be to Task Force (1996), and will only include the page number.

period variability,' 'RR variability,' and 'RR interval tachogram' were used (p. 354).[8] When compared with 'heart rate variability,' these terms better convey the fact that the target of analysis is variations of the intervals, not heart rate per se. Nonetheless, the term 'heart rate variability' gained popularity over other terms. RA-11 thus adopts the term 'heart-rate variability' and standardizes its definition as "the oscillation in the interval between consecutive heart beats as well as the oscillations between consecutive instantaneous heart rates" (p. 354). In addition to standardizing the term used to describe the phenomenon under investigation, they also standardize other related technical terms in the context of standardizing methods of measurement and clinical applications, which we will turn to now.

The second goal of RA-11 is to standardize methods of measurement. Out of the 22 pages of the main text, the part of the article devoted to standardizing measurement methods takes up 10 pages, almost 50% of the article. At the very least, this shows that the task force judged that it was important to devote more space in the article to standardizing measurement methods than to other issues.[9] In RA-11, the task force discusses various methods for measuring HRV, including time domain methods, frequency domain methods, rhythm pattern analysis, and non-linear methods. They standardize time domain methods (p. 358, Table 1) and frequency domain methods (p. 360, Table 2), and specify how to correlate these two kinds of measures (p. 362, Table 3). They also emphasize the value of rhythm pattern analysis, specify the appropriate situations for using it, and recommend using it when it is appropriate. But they don't recommend using non-linear methods on the grounds that more systematic studies are still needed in order to assess these methods. In addition to standardizing methods for measuring HRV, the task force also specifies recording requirements when collecting data for HRV analysis, especially when using commercial equipment to do so. They also standardize the procedure of recording and processing signals from electrocardiography (p. 365, Figure 7). The standardizing work of the task force established practical standards for HRV researchers. By adopting these standards, researchers could conduct their research on the same common ground, more or less. This in turn made fruitful comparisons and competitions possible and sustainable.

The third goal of RA-11 is to specify physiological correlates of HRV components and pathophysiological correlates of changes of HRV. The task force reviews the literature on the relevant mechanistic details (pp. 365-66). And they recommend using a frequency domain method to interpret the high frequency component of HRV (HF) as correlating with vagal activity in the autonomic nervous system. They note the disagreement regarding how to interpret the low frequency component of HRV (LF). Some correlate LF with sympathetic modulations, while others correlate it with both sympathetic and vagal activity. Different interpretations of LF thus lead to different interpretations of the correlate of the LF/HF ratio. The task force notes the disagreement and recommends doing further work to clarify the

[8] Because of these terminological variations, we intentionally queried the databases with the term 'heart rate variability' without quotation marks, so that we could include papers that use terms like 'cycle length variability,' 'heart period variability,' and 'RR variability.' However, because we did not use quotation marks, our results included papers that contain the terms 'heart,' 'rate,' or 'variability,' but are not related to HRV research. We then manually deleted these irrelevant papers.

[9] Incidentally, the fact that the task force didn't concern itself with proposing a central theory to guide future HRV research shows that our practice-based approach is appropriate, and that it would be misguided to adopt some sort of theory-based approach for analyzing this article.

physiological correlate of LF. They also specify five pathophysiological correlates of five types of changes of HRV (pp. 366-67). One of them concerns mental stress, which is one of the most frequent words after 2001 in Figure 3. They also review five interventions that can modify HRV. But they note that it is problematic to assume that modifying HRV can protect patients from cardiac mortality and sudden cardiac death, and that more research is still needed on this issue (pp. 367-68).

The fourth goal of RA-11 is to describe the clinical applications of HRV that the task force considered appropriate at the time. The task force states that, though there had been many attempts to use HRV to investigate cardiological and non-cardiological diseases and other clinical conditions, only two clinical applications were widely accepted by the HRV community. One is to predict the risk of mortality after acute myocardial infarction. The other is to assess early warning signs of diabetic neuropathy. The task force then standardizes the relevant measures of HRV suitable for the two clinical applications and specifies how to use them (pp. 368-371). They also summarize a set of studies that apply HRV to cardiological conditions other than myocardial infarction and note their potential without doing any standardizing work regarding these applications (pp. 372-373).

The fifth goal of RA-11 is to identify areas for future research. The task force points out what work needs to be done in order to improve HRV measurement and our understanding of the physiological correlates of HRV. They also point to various possibilities of future clinical utility, specifically about sleep and exercise (p. 347), which are two of the most frequent words after 2001 in Figure 3.

Based on the above, it is clear that the task force carried out a great amount of standardizing work. They standardized the nomenclature regarding HRV, methods for measuring HRV, procedures for recording and processing HRV signals, and procedures and methods for measuring HRV for the two clinical applications discussed above. The other achievements of RA-11 are either implications of this standardizing work or complementary to it. For example, the future research possibilities are suggested based on what was left unaddressed or insufficiently addressed after the standardizing work had been completed.

To recap: The HRV field has undergone significant quantitative and qualitative changes since the publication of RA-11. Moreover, RA-11 is the most cited article in the field, and the task force that published RA-11 undertook various types of standardizing work. Given all of that, how do we understand the relationship between the standardization achievements of this heavily cited article and the subsequent changes in the field?

4. Using Case Study Methods to Generate Hypotheses

In this section, we will illustrate how the above question can guide philosophers to formulate potential hypotheses to inform further case-based investigation. We will show what a potential hypothesis would look like and how it could be developed in terms of the case-based understanding generated through the quantitative and qualitative results from the previous two sections. It is important to note that we do not aim to directly argue for the validity of the hypothesis in this paper since the evidence we have provided in this paper is obviously not sufficient to substantiate the

validity claim. What we aim to show is the potential of our quantitative methods and the practice-based approach for doing case-based research in the philosophy of science, specifically, how they can help philosophers generate potential hypotheses for further case-based investigation.

In the remainder of the section, we will elaborate one potential hypothesis regarding the relationship between the standardization achievements of RA-11 and the subsequent changes in the field. We develop this hypothesis in terms of the concept of exemplar (Kuhn 1970; Rouse 2002; Shan 2020). It is worth emphasizing that the concept of exemplar we will use has been expanded to go beyond Kuhn's original articulation of that concept, and we do not aim to defend the whole package of Kuhn's work here. Our goal here is to use the concept of exemplar to show how it helps generate a hypothesis. Having shown that, we will discuss some limitations of this potential hypothesis, specifically, that it cannot provide the whole story regarding the targeted relationship. We will briefly indicate how other concepts, e.g., the concept of repertoire (Ankeny and Leonelli 2016), might be useful to probe other aspects of the relationship.

4.1 Exemplars

Kuhn originally introduced exemplars as "accepted examples of actual scientific practice—examples which include law, theory, application, and instrumentation together—[that] provide models from which spring particular coherent traditions of scientific research" (1970, p. 10). Kuhn himself stated explicitly that the concept of exemplar is "the central element of what I now take to be the most novel and least understood aspect of this book," namely, *The Structure of Scientific Revolutions* (Kuhn 1970, p. 187). Kuhn's articulation of the notion of an exemplar is strongly practice-oriented. An exemplar is not merely an accepted theory within a research community, but a multi-dimensional entity that includes non-theoretical aspects of practices as well.

Both Rouse (2002) and Shan (2020) articulate and expand upon Kuhn's original concept of exemplar and propose to use it to analyze the history of scientific practice. Rouse (2002) articulates and expands upon these non-theoretical aspects by emphasizing that accepting exemplars amounts to mastering "exemplary ways of conceptualizing and intervening in particular situations" (2002, p. 107). He lists some typical skills acquired through accepting an exemplar: (1) applying concepts to specific situations; (2) deploying mathematical tools, applying them correctly to the situation at hand, knowing their limitations, and knowing ways to circumvent those limitations; (3) using instrumental and experimental techniques and procedures; and (4) recognizing significant opportunities to extend these skills to new situations that can extend the research potential of an exemplar (Rouse 2002, pp. 107-108).

Similarly, Shan (2020) also articulates and expands upon the non-theoretical aspects of exemplars. Kuhn had also introduced exemplars as "concrete problem-solutions" that "show [scientists] by example how their job is to be done" (1970, p. 187). But, as Shan points out (2020, pp. 389-90), though Kuhn's original characterization of exemplar is richer than merely taking exemplars as problem-solutions, he didn't further articulate exemplars sufficiently. Shan thinks the practice of defining research problems is just as important as the practice of puzzle-solving,

and often the two are intertwined activities (2020, p. 391). Because of this, he sees the need to expand Kuhn's concept of exemplar to include, not just a set of problem solutions, but also the practice of defining research problems. Thus, for Shan, an exemplar is a set of problems and their solutions, which contains vocabulary, practical guides, hypotheses, and patterns of reasoning (2020, p. 392).

We think that Rouse's and Shan's expanded articulations of the notion of an exemplar are two different ways of characterizing the same thing. Rouse focuses more on characterizing various practices that constitute an exemplar, whereas Shan focuses more on elaborating how the practice of defining research problems is intertwined with the practice of puzzle-solving within the context of other practices that constitute an exemplar. In the following subsection, we make use of Rouse's and Shan's work on the concept of exemplar and propose our hypothesis regarding the relationship between the standardization achievements of RA-11 and the subsequent changes in the HRV field.

4.2 An Exemplar-Based Hypothesis

Recall that the task force (1) defines how to apply terms and concepts in specific situations, (2) specifies the standards for methods of measuring HRV, (3) outlines procedures for recording and processing HRV signals, (4) specifies procedures and methods of measuring HRV for the two accepted clinical applications (problem-solutions), and (5) also identifies ways of extending these skills to new research possibilities.

These achievements can all be subsumed under the concept of exemplar as both Rouse and Shan have articulated it. Rouse takes an exemplar to be constituted by exemplary ways of conceptualizing, recording, measuring, analyzing, and intervening. (1)-(5) can all be interpreted as the constituents of an exemplar in Rouse's sense. Shan further articulates an important constituent of an exemplar: a set of well-defined research problems. Since the task force standardized many practices, especially ways of applying terms and concepts in specific situations, these standardized practices provide the common currency for formulating well-defined research problems. Moreover, the task force indeed points out potential clinical applications as future research problems for the field.

If we apply the concept of exemplar to interpret the significance of the task force's standardization achievements, we can generate the following exemplar-based hypothesis:

> (Exemplar-Based Hypothesis) An exemplar was established by the task force that published RA-11, and RA-11 bears the identified relationship to the changes of the HRV field because it plays the role of an exemplar for the field.

This hypothesis provides a potential understanding of the relationship between the standardization achievements of the task force and the subsequent changes in the field. By introducing an exemplar for the field, the task force establishes the exemplary ways of conceptualizing, recording, measuring, analyzing, and intervening.

This can partly account for why RA-11 bears a strong correlation to the identified patterns of change that the HRV field exhibits.

To complicate the matter a bit, we want to introduce Morgan's (2019) notion of an exemplary account of a case phenomenon. According to her, an exemplary account of a phenomenon is an account developed by scholars that "illuminates and changes the way that the community thinks about that phenomenon" (p. 2). Moreover, an exemplary account can be developed based on a mundane case. The relevance of Morgan's notion here is that we take our exemplar-based hypothesis to be a potential instance of an exemplary account. Since the core of our HRV case study involves a review article written by a task force, one might think that this case is very mundane. It is expected that review articles in a scientific field influence the field in certain ways. This observation is fair. But we are not claiming that our HRV case itself is special in the sense that it reveals a novel, previously unstudied way in which review articles influence their fields. Instead, we aim to use the exemplar-based hypothesis to show how one could build a potential exemplary account of a mundane case, such as a review article in a biomedical field.

4.3 A Repertoire-Based Hypothesis

If philosophers use the above hypothesis to guide their further research, they can explore more about the epistemic dimension of the relationship between the standardization achievements and the pattern of change in the HRV field. But it is also reasonable to suspect that the developmental pattern of the HRV field could also have been influenced by non-epistemic factors. For example, researchers in biomedical sciences are regularly pressured to demonstrate the practical relevance of their research. It could be the case that RA-11 bears the identified relationship with the HRV field because the task force endorses the two instances of clinical applications of HRV research (Task Force 1996, pp. 368-371) and the prospect of exploring more clinical applications (Task Force 1996, pp. 372-373). In this possible scenario, the need to demonstrate and pursue clinical relevance of HRV research is also one of the factors that shapes the subsequent pattern in the HRV field. One could argue that the standardization achievements as exemplars are epistemic tools the task force uses to guide the HRV field to develop more clinical applications based on the two established clinical applications before 1996.

In order to capture the above possible scenario, we will introduce Ankeny and Leonelli's (2016) concept of research repertoire. Ankeny and Leonelli seek to frame scientific change in terms of administrative, material, technological, and institutional innovations, not merely in terms of theoretical innovations. They also seek to track the development of research practices in a given discipline without using the concepts of paradigm shift and scientific revolution. They thus propose the concept of a research repertoire. In a way, they take their concept of a research repertoire to be similar to Kuhn's concept of an exemplar (p. 20), but they include social, institutional, and economic features that one does not find in Kuhn's concept of exemplar.

This concept places more emphasis on the non-epistemic conditions in which a certain research style is established, evolves, becomes entrenched, or is transferred. The concept is a performative concept in the sense that it aims to capture the skills that are required to produce the relevant products and the products themselves. With

respect to the skill aspect, Ankeny and Leonelli emphasize the importance of looking into performative, social, financial, and organizational conditions in which a certain style of performing research is established, evolves, and reproduces (p. 21). They use the style of doing research with model organisms in biological sciences as an example of what a research repertoire is. They argue that the use of model organisms arises mainly because the proponents of this research style are able to convince their peers and large-scale government funders that this style of research provides critical opportunities for sharing knowledge, materials, and technologies across different biological research groups (Ankeny and Leonelli 2011). They aim to use this case to show that some non-epistemic conditions are critical for establishing the style of research using model organisms.

If we apply the concept of a research repertoire to interpret the development of the HRV field, we can generate the following repertoire-based hypothesis:

> (Repertoire-Based Hypothesis) RA-11 bears the identified relationship to the changes of the HRV field because the task force was able to convince their peers and government funders that using the standardized HRV practices provides critical opportunities for exploring possible clinical applications of HRV research.

It is important to note that the exemplar-based hypothesis and the repertoire-based hypothesis direct philosophers to investigate different dimensions of the relationship between the standardization achievements of RA-11 and the HRV field. The exemplar-based hypothesis might capture various epistemic roles the task force plays and could potentially account for why RA-11 bears the identified relationship to the pattern of change in the field. On the other hand, the repertoire-based hypothesis might capture various non-epistemic roles the task force plays and could also potentially account for the identified relationship. Whether these two hypotheses are competing or complementary is an empirical issue that requires more case-based investigations to clarify. We acknowledge that the task force's influence on the HRV field is a complex episode and hence requires multiple perspectives to investigate it. Perhaps approaching this complexity from different angles for different purposes is the best strategy for philosophers of biomedical sciences in this case.

5. A Methodological Conclusion: The Illustrated Methods and Approach Can Improve the Quality of the Selection Process

This paper addresses a neglected methodological issue in the literature on case-based research. The issue concerns the epistemic features of the selection process by which philosophers of contemporary sciences begin their case-based investigations. We address this issue by illustrating how quantitative methods and the practice-based approach to qualitative methods can help improve the quality of the selection process. Our illustrated quantitative methods can improve the degree of robustness of the selection process, and identify non-transparent patterns in a large body of literature, which in turn function as verifiable and non-private reasons to justify philosophers' selection of a case and enable them to engage in the social process of argumentation. Our illustrated practice-based approach can help generate more specific research questions through analyzing the selected case in detail, and potential hypotheses that can guide further philosophical investigation regarding the selected case. We thus

recommend that philosophers of contemporary sciences employ them to epistemically improve their selection processes.

Acknowledgements

We'd like to thank audiences at the 2017 meeting of the European Philosophy of Science Association, the Institute of Philosophy of Mind and Cognition at National Yang-Ming University, the 7th and 8th Asia-Pacific Philosophy of Science Conferences, the Scientonomy 2019 Conference, the Workshop on Interdisciplinary Philosophy at National Yang-Ming University, and the 2019 Annual Meeting of the Korean Society for the Philosophy of Science. And we'd like to thank three anonymous reviewers for comments that drastically improved the content of this paper on every page. We'd also like to thank Jonathon Hricko and Linton Wang for helpful questions and comments. This work is supported by the Ministry of Science and Technology in Taiwan (MOST 103-2633-H-197-001; MOST 107-2410-H-010-008-MY2).

References

Ankeny, R. A. (2007). Model organisms as case-based reasoning. In A. Creager, M. Norton Wise, & E. Lunbeck (Eds.), *Science without laws: Model systems, cases, exemplary narratives* (pp. 46–58). Durham and London: Duke University Press.

Ankeny, R. A. (2011). Using cases to establish novel diagnoses: Creating generic facts by making particular facts travel together. In P. Howlett & M. S. Morgan (Eds.), *How well do facts travel?* (pp. 252–272). Cambridge: Cambridge University Press.

Ankeny, R. A. (2014). The overlooked role of cases in causal attribution in medicine. *Philosophy of Science*, *81*(5), 999–1011. https://doi.org/10.1086/677693

Ankeny, R. A., & Leonelli, S. (2011). What's so special about model organisms? *Studies in History and Philosophy of Science Part A*, *42*(2), 313–323. https://doi.org/10.1016/j.shpsa.2010.11.039

Ankeny, R. A., & Leonelli, S. (2016). Repertoires: A post-Kuhnian perspective on scientific change and collaborative research. *Studies in History and Philosophy of Science Part A*, *60*, 18–28. https://doi.org/10.1016/j.shpsa.2016.08.003

Bigger, J. T., Fleiss, J. L., Steinman, R. C., Rolnitzky, L. M., Kleiger, R. E., & Rottman, J. N. (1992). Frequency domain measures of heart period variability and mortality after myocardial infarction. *Circulation*, *85*(1), 164–171. https://doi.org/10.1161/01.CIR.85.1.164

Borgerson, K. (2009). Why reading the title isn't good enough: An evaluation of the 4S approach to evidence-based medicine. *IJFAB: International Journal of Feminist Approaches to Bioethics* 2(2), 152-175. https://www.muse.jhu.edu/article/315496.

Bueno, O., Chen, R., & Fagan, M. B. (2018). *Individuation, process, and scientific practices*. New York: Oxford University Press. https://doi.org/10.1093/oso/9780190636814.001.0001

Chang, H. (2011). The philosophical grammar of scientific practice. *International Studies in the Philosophy of Science*, *25*(3), 205–221. https://doi.org/10.1080/02698595.2011.605244

Chang, H. (2012). Beyond case-studies: History as philosophy. In S. Mauskopf & T. Schmaltz (Eds.), *Integrating history and philosophy of science: Problems and prospects* (pp. 109–124). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-1745-9_8

Corr, P. B., Yamada, K. A., & Witkowski, F. X. (1986). Mechanisms controlling cardiac autonomic function and their relation to arrhythmogenesis. In H. A. Fozzard, E. Haber, R. B. Jennings, A. N. Katz, & H. E. Morgan (Eds.), *The Heart and Cardiovascular System* (pp. 1343–1403). New York: Raven Press.

Currie, A. (2015). Philosophy of science and the curse of the case study. In C. Daly (Ed.), *The Palgrave handbook of philosophical methods* (pp. 553–572). London: Palgrave Macmillan. https://doi.org/10.1057/9781137344557_22

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, 22(2), 338–342. https://doi.org/10.1096/fj.07-9492LSF

Felix, C., Franconeri, S., & Bertini, E. (2018). Taking word clouds apart: An empirical investigation of the design space for keyword summaries. *IEEE*

Transactions on Visualization and Computer Graphics, 24(1), 657–666. https://doi.org/10.1109/TVCG.2017.2746018

Giere, R. N. (1988). *Explaining science: A cognitive approach*. Chicago: University Of Chicago Press.

Hon, E. H., & Lee, S. T. (1963). Electronic evaluations of the fetal heart rate. VIII. Patterns preceding fetal death, further observations. *American Journal of Obstetrics and Gynecology*, *87*, 814–826.

Kleiger, R. E., Miller, J. P., Bigger, J. T., & Moss, A. J. (1987). Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *The American Journal of Cardiology*, *59*(4), 256–262. https://doi.org/10.1016/0002-9149(87)90795-8

Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Leonelli, S. (2018). Rethinking reproducibility as a criterion for research quality. In L. Fiorito, S. Scheall, & C. E. Suprinyak (Eds.), *Research in the history of economic thought and methodology: Including a symposium on Mary Morgan: Curiosity, Imagination, and Surprise* (pp. 129–146). Bingley: Emerald Publishing.

Levy, M. N., & Schwartz, P. J. (1994). *Vagal control of the heart: Experimental basis and clinical implications*. Armonk, NY: Futura Publishing Company.

Lown, B., & Verrier, R. L. (1976). Neural activity and ventricular fibrillation. *The New England Journal of Medicine*, *294*(21), 1165–1170. https://doi.org/10.1056/NEJM197605202942107

Malik, M., Farrell, T., Cripps, T., & Camm, A. J. (1989). Heart rate variability in relation to prognosis after myocardial infarction: Selection of optimal processing techniques. *European Heart Journal*, *10*(12), 1060–1074. https://doi.org/10.1093/oxfordjournals.eurheartj.a059428

Morgan, M. S. (2012). Case studies: One observation or many? Justification or discovery? *Philosophy of Science*, *79*(5), 667–677. https://doi.org/10.1086/667848

Morgan, M. S. (2014). Resituating knowledge: Generic strategies and case studies. *Philosophy of Science*, *81*(5), 1012–1024. https://doi.org/10.1086/677888

Morgan, M. S. (2019). Exemplification and the use-values of cases and case studies. *Studies in History and Philosophy of Science Part A*, *78*, 5–13. https://doi.org/10.1016/j.shpsa.2018.12.008

Pietsch, W. (2016). Two modes of reasoning with case studies. In T. Sauer & R. Scholl (Eds.), *The philosophy of historical case studies* (pp. 49–67). Cham: Switzerland. https://doi.org/10.1007/978-3-319-30229-4_4

Rajendra Acharya, U., Paul Joseph, K., Kannathal, N., Lim, C. M., & Suri, J. S. (2006). Heart rate variability: A review. *Medical & Biological Engineering & Computing*, *44*(12), 1031–1051. https://doi.org/10.1007/s11517-006-0119-0

Rouse, J. (2002). Kuhn's philosophy of scientific practice. In T. Nickles (Ed.), *Thomas Kuhn* (pp. 101–121). Cambridge: Cambridge University Press.

Schwartz, P. J., & Priori, S. G. (1990). Sympathetic nervous system and cardiac arrhythmias. In D. P. Zipes & J. Jalife (Eds.), *Cardiac electrophysiology: From cell to bedside* (pp. 330–343). Philadelphia: W.B. Saunders.

Shan, Y. (2020). Kuhn's "wrong turning" and legacy today. *Synthese*, *197*(1), 381–406. https://doi.org/10.1007/s11229-018-1740-9

Soler, L., Zwart, S., Lynch, M., & Israel-Jost, V. (2014). *Science after the practice turn in the philosophy, history, and social studies of science*. New York and London: Routledge.

Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, *17*(3), 354–381. https://doi.org/10.1093/oxfordjournals.eurheartj.a014868

Waters, C. K. (2014). Shifting attention from theory to practice in philosophy of biology. In M. C. Galavotti, D. Dieks, W. J. Gonzalez, S. Hartmann, T. Uebel, & M. Weber (Eds.), *New directions in the philosophy of science* (pp. 121–139). Cham: Springer. https://doi.org/10.1007/978-3-319-04382-1_9

Wimsatt, W. C. (2006). Reductionism and its heuristics: Making methodological reductionism honest. *Synthese*, *151*(3), 445–475. https://doi.org/10.1007/s11229-006-9017-0