What Can Be Learned about Argument Schemes from Other Fields' Inventions?

SALLY JACKSON AND JODI SCHNEIDER

Department of Communication
University of Illinois at Urbana-Champaign
USA
sallyj@illinois.edu

School of Information Sciences
University of Illinois at Urbana-Champaign
USA
jodi@illinois.edu

ABSTRACT: Argumentation in specialized fields cannot be adequately described in terms of vernacular schemes. As Toulmin (1958) observed, specialized fields invent new inference methods and innovate in their use. We argue that studying these inventions and innovations is important in itself, but it can also challenge current theoretical assumptions about vernacular schemes as well.

KEYWORDS: argumentation schemes, critical questions, warrant-establishing argument, randomized clinical trials

1. INTRODUCTION

For several years, we have been closely studying innovations in reasoning about health. (Jackson & Schneider, 2018; Schneider & Jackson, 2018; Schneider & Jackson, 2020). Combining the strengths of our different backgrounds, we have been trying to understand a cluster of phenomena that have to do with new ways of knowing things that become resources for managing disagreement, while also being contestable practices that define new disagreement management challenges. The innovations we study compete with prior reasoning practices-–for example, with the patterns of argument we all recognize as schemes. For that reason, we are inclined to think of them as scheme-like, having a similar function but much more complex structure than vernacular schemes. This paper draws out four major findings about innovative forms of reasoning and speculates on what that might mean for all other schemes.

2. NEW ARGUMENT SCHEMES CAN BE INVENTED

First, we have learned that new schemes are being invented all the time in fields whose core business is knowledge production. Without using the language of schemes, Toulmin said almost the same thing long ago (1958), but he did not really show it. We take Toulmin's remarks not as finished theory but as a suggestion for how argumentation theory might develop a set of concepts applicable to argument as it occurs in realistic contexts. If we want a realistic theory of argumentation, we must pay attention to how the practice develops and changes over time.

Our first collaborative project (Jackson & Schneider, 2018) was a study of a hugely influential innovation of the 1990s, known as a systematic review, one version of which is the now very well-known Cochrane Review. Early on we recognized this as the sort of thing Toulmin had in mind when he talked about specialized warrants that fields invent to augment their pre-existing stock of reasoning tools. A Cochrane Review takes a pile of published research as its data and moves from the pile of studies to a conclusion about "what research shows."

We coined the term "warranting device" to try to convey two things. First, inventions like Cochrane Reviews play the same role in arguments as is played by a propositionalized warrant like "An expert's opinion that A is true is a reason to believe A." Second, this particular invention, Cochrane Review, is a bundle of procedures, resources, and institutional commitments that are all part of the warranting. The bundle is device-like in various ways, including in being open to improvements that preserve its core role in reasoning.

Working on Cochrane Reviews led us immediately to a prior invention, an even more significant one: the Randomized Clinical Trial (Schneider & Jackson, 2018). RCTs were invented in the twentieth century to solve the problem of how to evaluate causal relationships between medical treatments and health outcomes. By the end of the century tens of thousands of RCTs were being published in the medical literature every year.

RCT-based arguments are typically expressed in research reports that describe exactly how the experiment was conducted. To arrive at a scheme definition, we must abstract from those research reports. For example, we need to see that these RCT-based arguments all have the same kind of conclusion, an assertion about effects of a medical treatment. Further, we need to see that the scheme requires at least two distinct kinds of premises: the observations made, obviously, but also, a specification of how these observations were obtained. Both the observations and the specification are needed for evaluation of the conclusion. And not all RCTs produce strong arguments. The exceptions shown in Figure 1 serve the same role as critical questions.

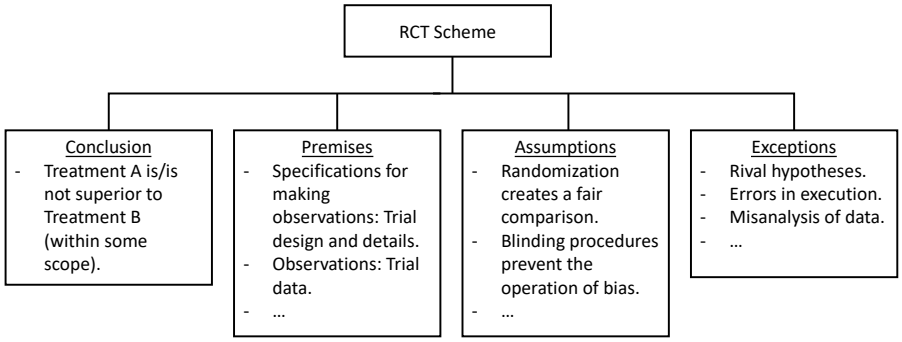

Figure 1: A scheme definition for the RCT Scheme from Schneider & Jackson (2018).

Specification of how observations were made is an unusual premise type, compared with more familiar schemes. It seems quite natural that the actual observations should serve as a premise, but describing the procedures that produced the data has no analogue in the schemes included in the user's compendium compiled by Walton, Reed, and Macagno (2008). Instructions on how to perform (and report) RCTs are common: numerous articles and books describe how to conduct an RCT in every medical specialty (e.g., Zabor et al.,

2020, for pulmonary medicine). Cochrane Reviews, likewise, are produced according to a carefully curated handbook (Higgins et al., 2020), and the report must affirm that the handbook was followed. Affirming that observations were obtained properly is what distinguishes arguments as "from RCT" or "from Cochrane Review."

Such explicit operationalization is not a feature of the vernacular scheme definitions, like the Expert Opinion scheme given by Walton, Reed, and Macagno (2008, p310), shown in Figure 2. The critical questions point to vulnerabilities one would want to avoid if possible, but exactly how are they to be avoided? Is there anything that can be said about when and how to rely on expertise in reasoning and in arguing? To describe how to instantiate Expert Opinion well requires more than critical questions. It would require instructions on how to instantiate the scheme type or guidance on what one should do to produce a strong instantiation rather than a weak one.

Argument from Expert Opinion	**Critical Questions**
Major Premise: Source E is an expert in subject domain S containing proposition A. Minor Premise: E asserts that proposition A is true. Conclusion: A is true.	• How credible is E as an expert source? • Is E an expert in the field that A is in? • What did E assert that implies A? • Is E personally reliable as a source? • Is A consistent with what other experts assert? • Is E's assertion based on evidence?

Figure 2: Expert Opinion Scheme and Critical Questions from Walton, Reed, and Macagno (2008, p310).

This contrast leads us to ask: Should we really consider RCTs as schemes? What makes warranting devices like RCT and Cochrane Review seem different from ordinary schemes is that they include explicit instructions for instantiating the scheme, while other schemes are abstractions from instantiated schemes—without any allusion to the instantiation process. Vernacular schemes do get instantiated, but seemingly without effort or plan.

What might be involved in investigating the methods of instantiation for vernacular schemes? A beginning point would be to consider the many different ways a person ends up saying something that can be reconstructed as an argument from expert opinion. Just a few examples will suffice to show that something might be gained by trying to specify how best to make use of expertise:

a) If any expert says anything, believe it (and repeat it if an opportunity arises).
b) If someone challenges something you believe, go find an expert who contradicts your challenger.
c) In making a decision about what to do or what to believe, locate any relevant expert fields and find out what experts seem to agree on.

Notice that only one of these is a method any argumentation theorist would be likely to recommend. Although the argumentation theory literature is rich with suggestions on how to evaluate arguments based on expertise, it is poor in advice on how to strengthen them. We might be able to do much better, as we'll suggest momentarily.

So let's return to what we learn from studying invented schemes and ask if any new directions open up for studying familiar schemes. New schemes are being invented in many

specialized fields involved in knowledge production, and their unexpected properties can lead us to unexpected possibilities for further investigation of schemes considered to be well-understood. Specifically, all schemes must be instantiated to become arguments and for some, it will be worthwhile to propose methods of instantiation.

3. INVENTED SCHEMES CAN BE REVISED IN LIGHT OF EXPERIENCE

The story of RCTs is a twentieth century story, but people did not start from scratch inventing the RCT. Some form of medical experimentation has been going on for a very long time. Many building blocks for RCTs are ancient. Without any of the tools of modern science, we can imagine that many physicians lost to history tried things on their patients and drew inferences from the results. Anyone can conduct their own experiment of this kind. What happens, you might ask, if you take melatonin as a preventive for jet lag? You may learn something, but you will not generate particularly strong evidence that melatonin does or does not work.

Applying Mill's (1843) method of difference improves the situation a little bit. Observing an untreated patient provides at least some idea of how much a person might improve without treatment, so the difference in outcomes, at least conjecturally, could be the effect of the treatment. But in contemporary terms, we would say that even this improved design fails to provide strong evidence because it confounds treatment effects with uncontrolled individual differences between the two patients.

Observing many patients given the treatment or denied the treatment is a further improvement, because it eliminates the confounding. Yet this is still highly vulnerable to doubt, especially if there is any possibility that the two groups of patients started out different.

Two twentieth-century inventions, closely connected to one another, attempt to protect conclusions drawn from experiments from any suspicion that something other than the contrasting treatments resulted in different outcomes for the two groups. Both inventions originated in agricultural research (Fisher, 1935) and diffused into research on human subjects. The first invention was random allocation—the idea being to create a fair comparison, not to guarantee equivalence between the two treatment groups. The second invention was statistical comparison, a technique for assessing how confident we can be that a difference in outcome was not due to chance variation among randomly allocated patients.

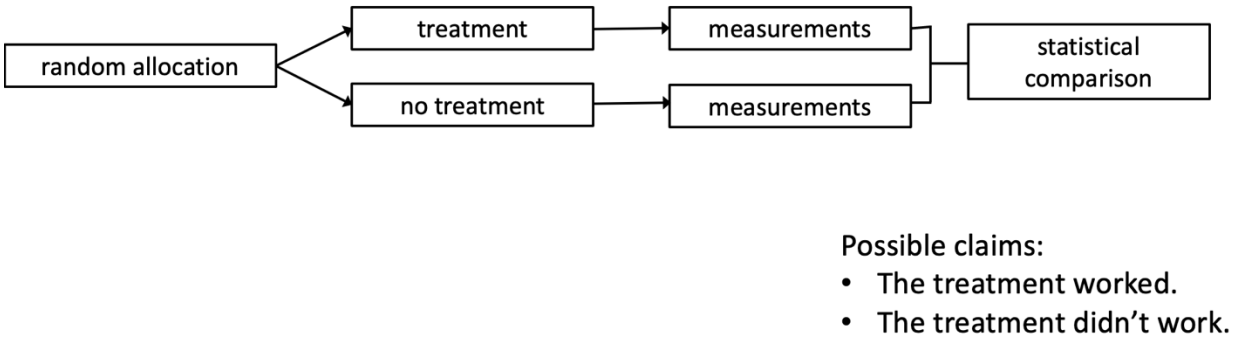

Figure 3 A basic design for experimenting on medical treatments: .

The design shown in Figure 3 can be found with slight variations in many books on experimental design. As compared with earlier ways of reasoning about effects of interventions, this design is much stronger, having built in answers to many questions that could have been asked about the nonrandomized version. This much can be found directly in Fisher's classic work—but his concern was experimenting on plants and soil, not on humans. To tailor randomized experiments to human subjects, researchers simply had to start *conducting* experiments to discover some of the things that can go wrong when human subjects are involved. One thing that can go wrong is that a patient's expectations about a treatment can make them more hopeful or more pessimistic, and these expectations can affect their response to treatment. This might be solved by attempting to prevent the patient from knowing which treatment they are receiving—for example, by giving a medically worthless pill to one group and a visually similar pill with the experimental drug to another group. This is a way to create "blinding." And precautions can also be taken to assure that those who administer the treatment or make observations are also blind to which group the subject is in. Designs comparing treatment to treatment are alternatives to designs comparing treatment to no treatment.

As inventions layer on top of one another, the iterative revision in design often comes about as an effort to build in answers to possible objections to conclusions drawn from the measurements collected in the experiment. Those objections are both critical questions that draw blood against prior designs and components of the improved RCT scheme. When schemes are designed, a recurrent critical question is often handled by tweaking the procedures.

Figure 4 shows the modern RCT structure, the expected standard for medical evidence. An experiment must have these features built in from the start to pass peer review or to be used for regulatory decisions from agencies like the US Food and Drug Administration (Schneider & Jackson, 2018). Critical questions about the equivalence of the conditions cannot be adequately addressed if randomization has not been part of the procedure.

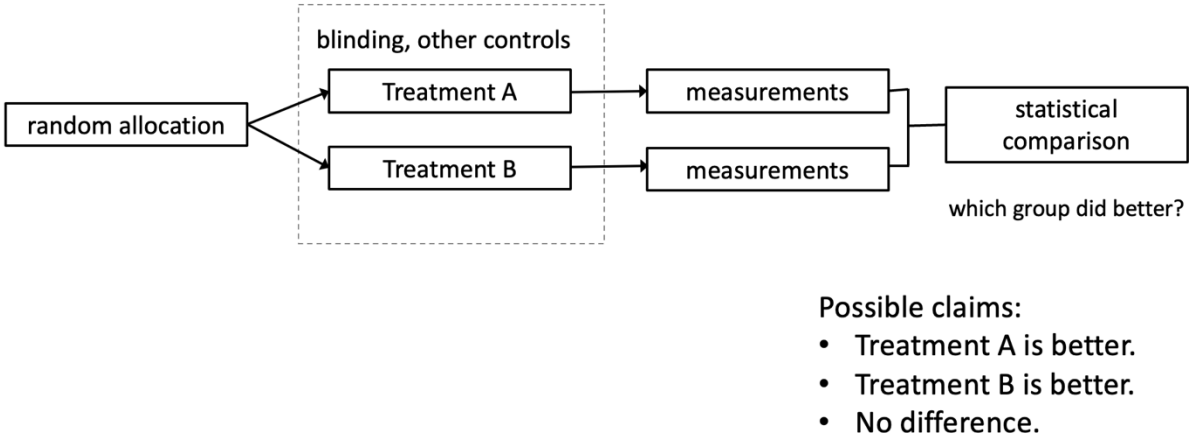

Figure 4: Layering blinding and other controls over the previous inference structure strengthens any claim about the effect of the treatment.

Figure 5 shows an argument constructed on the logic of RCT. Observations drawn from a properly performed RCT create some level of plausibility for a conclusion about

the effect of the treatment. Its plausibility depends on exactly what was done and exactly what was observed.

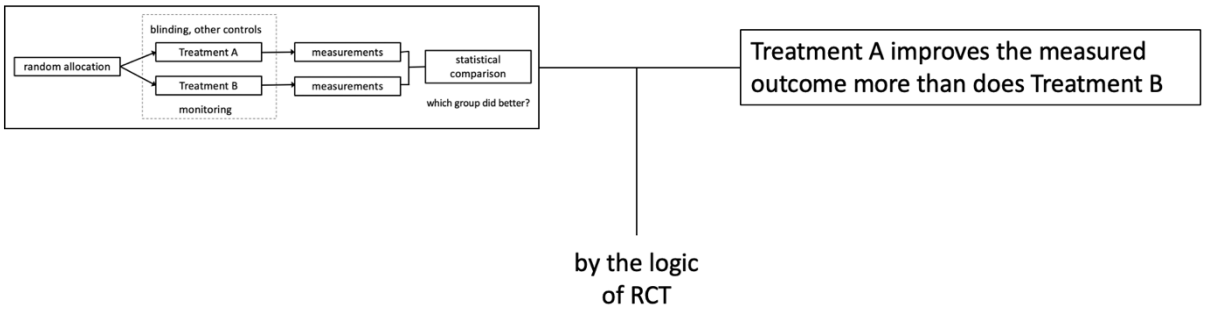

Figure 5: *The RCT scheme with a sample instantiation.*

Pretty obviously, an experiment without randomization is less convincing than one with randomization, and an experiment with a single observation per condition is less convincing than one with many replications. People have known for a very long time that it might be worthwhile to make observations under contrasting conditions. But starting in the early twentieth century people started inventing techniques to develop this basic strategy for experiments on human subjects.

Here again, what we observe for recently invented schemes seems to set them apart from vernacular schemes like Argument from Expert Opinion. The revision history of an invented scheme like RCT or Cochrane Review is a matter of record, openly debated and easily located. Is the same true for vernacular schemes? We can't as easily see shifts that may have taken place unnoticed over millennia. The best contemporary instantiations of a scheme like Argument from Expert Opinion may closely resemble arguments made hundreds or thousands of years ago. But the resemblance could be deceiving. Expertise as an argumentative resource has certainly not remained static, nor have ideas about how best to manage this resource. If a vernacular scheme absorbs invented resources, is it still the same scheme?

Consider, for comparison, recent inventions designed on purpose to make an appeal to expert opinion as strong as it can be, as invulnerable to critical questions as possible. Science Court was a speculative 1970s-era proposal for a way of improving policy debate by extracting issues requiring technical expertise (Aakhus, 1999), and using a quasi-judicial process to return just one adjudicated expert opinion to the policy debaters. Aakhus explored the design of Science Court, treating the separation of factual questions from broader argument structures as a design hypothesis. Consensus reports work on a similar design hypothesis: They are increasingly used to translate technical expertise into propositions usable in practical deliberation. Consensus reports assume that no single expert can speak for expert fields anymore, and we should not be asking whether any particular expert has credibility but rather what consensus experts form when tasked with hashing out their disagreements.

Accepting the output from black-box devices of these kinds is still reasoning from expert opinion, but we suggest that this can be seen as a "revision" of this reasoning principle that tries to deal with the fact that experts in the same field may support opposite conclusions based on their opinions.

4. CRITICAL QUESTIONS, AND THEIR OWN LIMITATIONS, ARE DISCOVERED THROUGH EXPERIENCE

We have already seen that invented schemes undergo revision, sometimes very rapidly, as the first efforts at instantiating the scheme reveal vulnerabilities. The third thing we have learned from our prior work is that methods for evaluating instantiations of these schemes also emerge over time. They are not invented but discovered through experience in use of the scheme.

When RCTs were introduced into medical research—around mid-twentieth century—other fields already had quite a lot of experience in conducting experiments on human subjects. For example, experimental persuasion research was already well-established, and widely recognized "threats to validity" became a routine part of peer review (Campbell & Stanley, 1963). Some of these threats were attached to faulty design, such as failing to assign subjects to treatment at random, but others were out of the experimenter's control, like differential loss of subjects after randomization.

There does not seem to be a finite list of critical questions for RCTs—just an aggregation of everything learned in experience, so far, using this scheme. Figure 6 lists a few very typical questions that social scientists and medical scientists ask when evaluating an experiment.

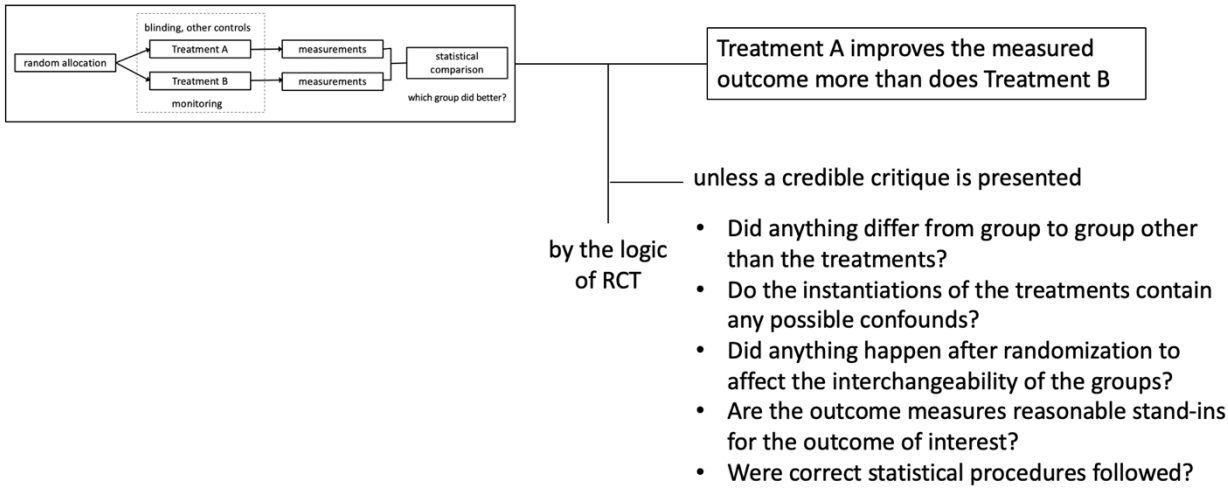

Figure 6: *Experience-based critical questions for evaluating claims supported by RCTs.*

Yet efforts to grade the evidence provided by an experiment, as separate from whatever the experimenter claimed, do seem truly innovative. The rise of systematic review methods like Cochrane's required methods for aggregating the results of many experiments. This created a need for grading the individual studies in order to decide whether all should be weighted equally. Figure 7 shows a common element of Cochrane Reviews—a table of assessments of the individual studies reviewed against a set of biasing factors such as failure of randomization.

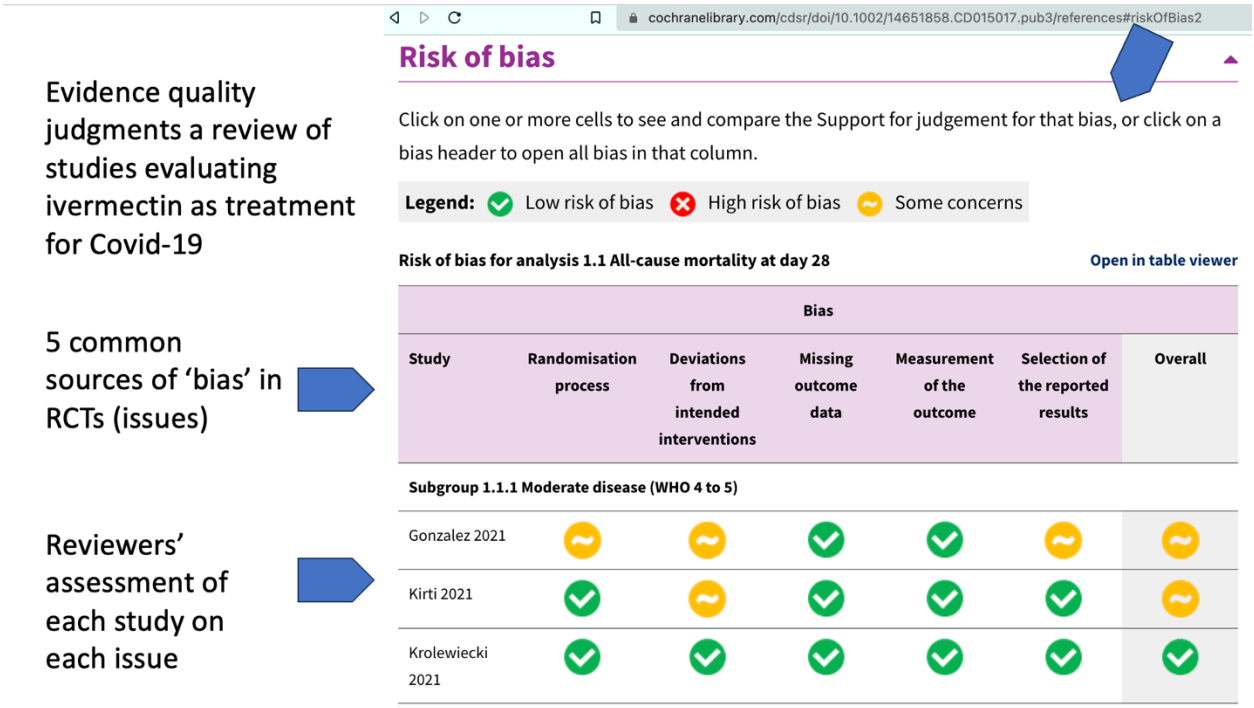

Figure 7: *Risk of bias as assessed in a Cochrane Review (Popp, 2022).*

Much more sophisticated grading methods have emerged in preparing research evidence for incorporation into policy deliberation. These are based on the realization that even a very well-conducted RCT can fail to produce evidence for anything actionable. To separate judgments of study quality from judgments of evidence quality, people began creating rubrics for extracting and evaluating bits of evidence from research reports (Guyatt et al., 1995).

Critical questions are often structured to support binary decisions between accepting a conclusion as presumptive and rejecting it as fallacious or weak. A critic may be tempted to run down a list of critical questions and find one whose answer justifies rejecting an argument. With invented schemes, we see two quite innovative ideas added to the idea of critical questions. First, in evaluating an argument we can do much better than simply accepting a conclusion as presumptive or rejecting it out of hand. We can grade its strength or weakness and pool it with other conclusions also varying in strength. Second, even if we reject a *conclusion*, inspection of the data may support some alternative conclusion that follows quite convincingly.

We have seen that critical questions for invented schemes are discovered gradually in the use of invented schemes, and that we can innovate in assessment methods in ways that look nothing like critical questions. Innovation in assessment methods for vernacular schemes is no less possible than it is for newly invented ones.

5. SCHEMES AND OTHER REASONING RESOURCES ACQUIRE TRACK RECORDS

For vernacular schemes, we have seen this point made in Mizrahi's series of papers on the poor track record of Arguments from Expert Opinion (2013, 2016). But our notion of track

record is a little more expansive, including not only the scheme's tendency to produce bad arguments but also the practical issues surrounding the use of the scheme.

We have noticed that warranting-establishing arguments like those that persuaded medical professionals to adopt RCTs (Schneider & Jackson, 2018) and those that established Cochrane reviews as preferable to prior review methods (Jackson & Schneider, 2018) often rest on future projections of track records. Essential to establishing a new warrant is reason to believe that it will produce results not possible otherwise, better results than current methods, the same quality of results more efficiently, or some other general advantage. For example, Bradford Hill (1952) attempted to persuade clinicians to give up autonomous experiments of their own in favor of cooperative clinical trials in which clinician judgment was replaced by random assignment of patients to treatments. Most people regard this as a huge advance, believing that RCTs have served us better than what preceded them. The comparison he made is shown diagrammatically in Figure 8.

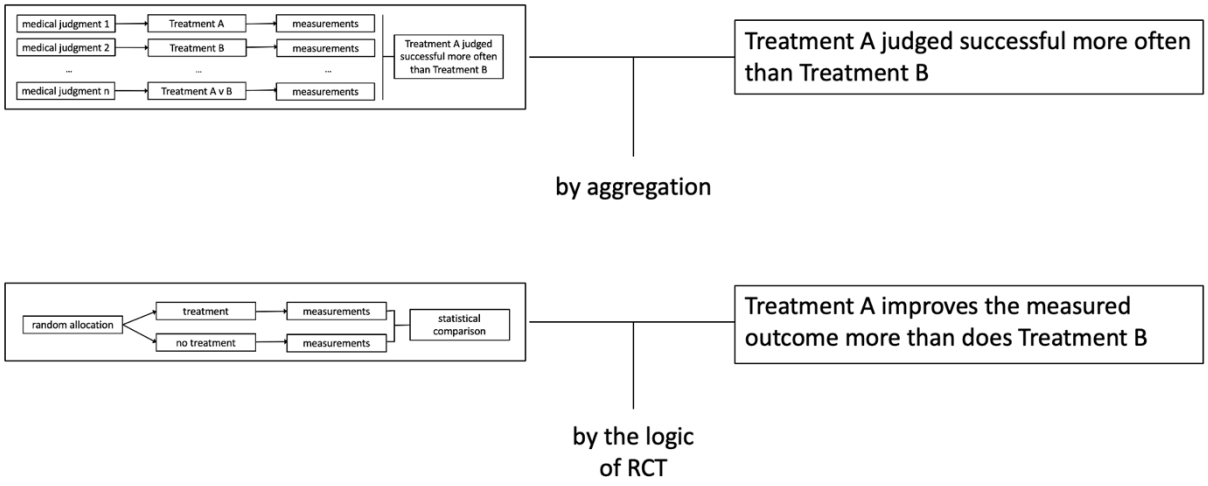

Figure 8: Aggregation of individual doctors' clinical experience versus reliance on RCTs.

But despite the reverence accorded to RCTs by mainstream public health authorities, the RCT enterprise has also met with a certain amount of disillusionment. There is a growing realization with policy-making and clinical care that evidence from RCTs always falls far short of what is needed to make a confident decision (Schneider & Jackson, 2020). For example, a critique of RCTs that first appeared in 1967 was centered on the fact that RCTs structured to best support causal claims are unsuited for their actual purposes in public health (Schwartz and Lellouch, 1967). As shown in Figure 9, RCTs can supply a means-end premise for a practical argument about how to treat a patient, but all of the familiar questions one would naturally pose about this conclusion are unanswered. For example, if multiple treatments exist or if new ones appear, the options can "outrun the evidence." More critically, the fact that a treatment is efficacious on average does not mean that it will work for every patient. It took decades for Schwartz and Lellouch's arguments to attract any real notice (Schneider & Jackson, 2020); the RCT scheme had to acquire a disappointing track record before their arguments could be appreciated.

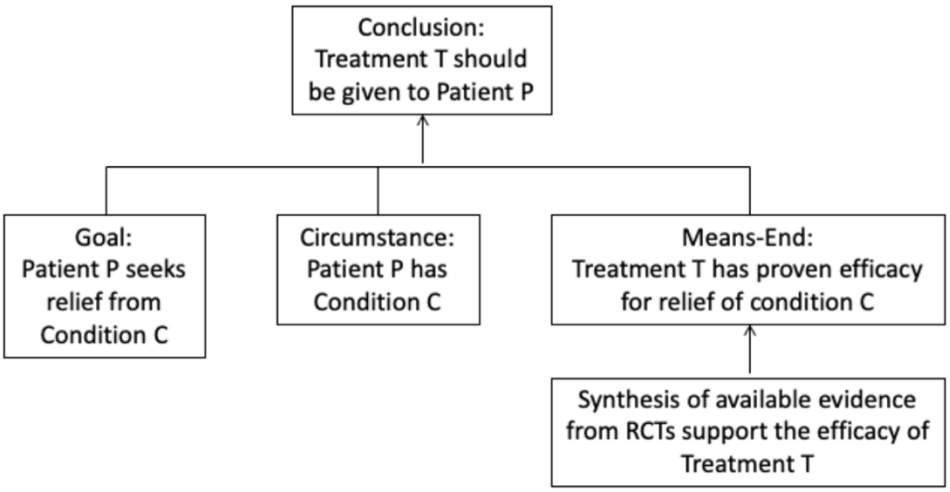

Figure 9: Practical reasoning diagram with Means-End premise drawn from explanatory RCTs (Schneider & Jackson, 2020).

RCTs will certainly not be obsolesced all at once, but at least three developments on the frontiers of medical science provide concepts to which they can be compared (Schneider & Jackson, 2020). These are innovations like single-subject experiments (Nikles et al., 2011), pragmatic trials using large numbers of patients in realistic care settings (Tunis et al., 2003), and causal modeling of treatment effects using huge datasets generated in clinical practice (Pearl & Mackenzie, 2018).

For vernacular schemes, it is not quite apparent who will care about any studies we may do on their track records in ordinary discourse. For example, it is hard to imagine ordinary people in ordinary communication contexts resolving to give up Argument from Expert Opinion despite Mizrahi's efforts to demonstrate the poor track record of this form of arguing. But invented schemes with poor track records quickly get replaced, and empirical assessment of their track records can be part of a design methodology (Jackson & Aakhus, 2014).

6. CONCLUSION

Studying the emergence of new forms of reasoning has practical importance, since these new forms are responsible for more and more of what circulates as knowledge. And as we have tried to suggest here, whatever we learn about innovations in reasoning and argument can reflect back on well-established argumentative practices, drawing our attention to things missed in the past.

One of the things we've seen quite clearly in our own work is that specialized schemes often start out producing arguments with vulnerabilities that even laypersons can easily spot, but the process of revision soon makes the arguments produced too technically complex for anyone but experts to meaningfully challenge them. One compelling reason to study them from an argument theoretic perspective is to contribute to critical assessment *from outside* the specialist field.

REFERENCES

Aakhus, M. (1999). Science court: A case study in designing discourse to manage policy controversy. *Knowledge, Technology & Policy*, *12*(2), 20–37. https://doi.org/10.1007/s12130-999-1020-6

Bradford Hill, A. (1952). The clinical trial. *New England Journal of Medicine*, *247*(4), 113–119. https://doi.org/10.1056/NEJM195207242470401

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Houghton Mifflin.

Fisher, R. A. (1935). *The Design of Experiments*. Oliver and Boyd.

Guyatt, G. H., Sackett, D. L., Sinclair, J. C., Hayward, R., Cook, D. J., Cook, R. J., Bass, E., Gerstein, H., Haynes, B., Holbrook, A., Jaeschke, R., Laupacls, A., Moyer, V., & Wilson, M. (1995). Users' guides to the medical literature: IX. A method for grading health care recommendations. *JAMA*, *274*(22), 1800–1804. https://doi.org/10.1001/jama.1995.03530220066035

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2020). *Cochrane Handbook for Systematic Reviews of Interventions* (Second edition). Wiley-Blackwell.

Jackson, S., & Aakhus, M. (2014). Becoming more reflective about the role of design in communication. *Journal of Applied Communication Research*, *42*(2), 125–134. https://doi.org/10.1080/00909882.2014.882009

Jackson, S., & Schneider, J. (2018). Cochrane Review as a "Warranting Device" for reasoning about health. *Argumentation*, *32*(2), 241–272. https://doi.org/10.1007/s10503-017-9440-z

Mizrahi, M. (2013). Why arguments from expert opinion are weak arguments. *Informal Logic*, *33*(1), 57. https://doi.org/10.22329/il.v33i1.3656

Mizrahi, M. (2016). Why arguments from expert opinion are still weak: A reply to Seidel. *Informal Logic*, *36*(2), 238. https://doi.org/10.22329/il.v36i2.4670

Mill, J. S. (1843). *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence, and Methods of Scientific Investigation*. J.W. Parker.

Nikles, J., Mitchell, G. K., Schluter, P., Good, P., Hardy, J., Rowett, D., Shelby-James, T., Vohra, S., & Currow, D. (2011). Aggregating single patient (n-of-1) trials in populations where recruitment and retention was difficult: The case of palliative care. Journal of Clinical Epidemiology, 64(5), 471–480. https://doi.org/10.1016/j.jclinepi.2010.05.009

Pearl, J., & Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books.

Popp, M., Reis, S., Schießer, S., Hausinger, R. I., Stegemann, M., Metzendorf, M.-I., Kranke, P., Meybohm, P., Skoetz, N., & Weibel, S. (2022). Ivermectin for preventing and treating COVID-19. *Cochrane Database of Systematic Reviews*, *2022*(6). https://doi.org/10.1002/14651858.CD015017.pub3

Schneider, J., & Jackson, S. (2018). Modeling the invention of a new inference rule: The case of 'Randomized Clinical Trial' as an argument scheme for medical science. *Argument & Computation*, *9*(2), 77–89. https://doi.org/10.3233/AAC-180036

Schneider, J., & Jackson, S. (2020). Beyond Randomized Clinical Trials: Emerging innovations in reasoning about health. In C. D. Novaes, H. Jansen, & J. A. V. Laar (Eds.), *Reason to Dissent: Proceedings of the 3rd European Conference on Argumentation, Volume III* (pp. 311–324). College Publications.

Schwartz, D., & Lellouch, J. (1967). Explanatory and pragmatic attitudes in therapeutical trials. *Journal of Chronic Diseases*, *20*(8), 637–648.

Toulmin, S. E. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.

Tunis, S. R., Stryer, D. B., & Clancy, C. M. (2003). Practical clinical trials: Increasing the value of clinical research for decision making in clinical and health policy. *JAMA*, *290*(12), 1624–1632. https://doi.org/10.1001/jama.290.12.1624

Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation Schemes*. Cambridge University Press.

Zabor, E. C., Kaizer, A. M., & Hobbs, B. P. (2020). Randomized Controlled Trials. *Chest*, *158*(1S), S79–S87. https://doi.org/10.1016/j.chest.2020.03.013