

Modeling the invention of a new inference rule: The case of ‘Randomized Clinical Trial’ as an argument scheme for medical science

Jodi Schneider^{a,*} and Sally Jackson^b

^a *School of Information Sciences, University of Illinois at Urbana–Champaign, Champaign, IL, USA*

^b *Department of Communication, University of Illinois at Urbana–Champaign, Urbana, IL, USA*

Abstract. A background assumption of this paper is that the repertoire of inference schemes available to humanity is not fixed, but subject to change as new schemes are invented or refined and as old ones are obsolesced or abandoned. This is particularly visible in areas like health and environmental sciences, where enormous societal investment has been made in finding ways to reach more dependable conclusions. Computational modeling of argumentation, at least for the discourse in expert fields, will require the possibility of modeling change in a stock of schemes that may be applied to generate conclusions from data. We examine Randomized Clinical Trial, an inference scheme established within medical science in the mid-20th Century, and show that its successful defense by means of practical reasoning allowed for its black-boxing as an inference scheme that generates (and warrants belief in) conclusions about the effects of medical treatments. Modeling the *use* of a scheme is well-understood; here we focus on modeling how the scheme comes to be *established* so that it is available for use.

Keywords: Warrant, practical reasoning scheme, warranting devices, inference rules, Argument Interchange Format, Forms Ontology, Randomized Clinical Trials, expert reasoning

1. Introduction

A fact about argumentation that deserves more notice is the extensibility of humanity’s store of tools for coming to conclusions and defending those conclusions to one another. People invent new methods for inference and they innovate in how they use these methods to come to shared conclusions. That is, they do not just devise novel arguments to meet specific needs but propose and advocate new *forms* of argument, intentionally designed for reusability. Invention and innovation have been occurring intermittently for millennia, but in the past century, these processes have become far more continuous, far more intentional, and far more visible than before. New methods of inference may appear at any time as arguers themselves invent and seek to justify new ways of generating dependable conclusions from new (or even old) kinds of data.

At present, many such inventions are emerging in scientific fields, and this burst of inventiveness is dramatically visible in certain domains – such as human health and environmental change – where enormous societal investment has been made in finding ways to reach more dependable conclusions about

*Corresponding author: Jodi Schneider, School of Information Sciences, University of Illinois at Urbana–Champaign, 501 E. Daniel Street, MC-493, Champaign, IL 61820-6211, USA. Tel.: +1 (217) 300-4328; Fax: +1 (217) 244-3302; E-mail: jschneider@pobox.com.

matters of great importance. But newly invented inference methods, as they emerge, require justification or defense before they can themselves be used to justify or defend particular conclusions. For example, use of computational simulations of natural processes becomes more convincing as particular models accumulate track records of success in describing or predicting the events they simulate. When much is at stake, new ways of coming to and justifying conclusions may themselves be highly controversial, deeply entangled with political interests or conflicting goals. This paper explores how to incorporate the contentious process of invention into the general project of modeling naturally occurring argument.

We focus here on a class of inventions that are generally understood by their originators and users as “methods.” They function both as a way to arrive at conclusions and as a way to justify those conclusions to skeptics and critics. With some modifications to be discussed below, these inventions serve the functions Toulmin [17] attributed to warrants, that of allowing for inference from data to conclusion. Because they are designed for repeated application to any number of inference problems, they may be seen as argumentation schemes in Walton’s [19] sense. From any of these perspectives, they are *inference rules* that are invented for the purpose of drawing and defending certain kinds of conclusions.

In the sections below, we first discuss inventions of new inference rules in health science, then describe the practical reasoning structures that are involved in establishing such rules within that field. Next we argue that no matter how complex the internal structure of such a rule, once the rule is “black-boxed” within a field, it is applied without explicit justification to individual instances of inference. Finally, since black-boxed inference rules can always be re-opened for critical questioning, and since any such rule may be superseded by something better, approaches are needed for representing relationships of conflict, preference, and inference among inference rules themselves (and not just for their application in particular cases).

2. Inventions in health reasoning

Working on health-related controversies, we have observed a steady stream of inventions and innovations in reasoning since about the middle of the twentieth century, accelerating rapidly as the idea of inventing becomes ever more fully naturalized. Jackson and Schneider [10] introduce the notion of a “warranting device” as a field-dependent inference rule, invented to serve a specific argumentative purpose. The warranting device is conceived as a conclusion-drawing rule that can be applied in many circumstances, plus whatever assurances may be available to back the dependability of the device. That is, the device consists of the warrant and its backing. Many such devices are being proposed and developed to handle various forms of aggregation problems: problems that involve constructing coherent conclusions from large bodies of potentially relevant information.

Jackson and Schneider [11] identified Randomized Clinical Trial and Cochrane Review as two distinct but closely intertwined examples of warranting devices appearing recently within reasoning about health, and Jackson and Schneider [10] gave a more detailed analysis of Cochrane Review. Experimental investigation of medical treatments is itself still a relatively new way of coming to conclusions about the effectiveness of these treatments, still having been in need of explicit defense through the 1950s. Subsequent to widespread adoption of Randomized Clinical Trial in medical science, many topics in medical research have generated dozens or even hundreds of individual experiments. Interpreting the body of evidence is not a matter of choosing among the various conclusions drawn by individual experimenters, but a matter of making sense of the whole evidence base. Until very recently, this sense-making was generally accomplished through whatever ad hoc procedures an individual literature reviewer considered

reasonable. The Cochrane Review appeared in the 1990s as a newly-invented method for aggregating evidence from many individual experiments on medical treatments to arrive at a conclusion that can guide medical practice.

In performing a Cochrane Review, the reviewer must follow well-defined procedures for locating potentially relevant evidence, evaluating the relevance of each source located, systematically examining the substance and the quality of each piece of evidence, and aggregating the results either qualitatively or quantitatively into a carefully-qualified statement of what is (and is not) shown by the body of research as a whole. The Cochrane Review and its closest competitors have obsolesced certain prior practices, and both scientists and nonscientists regard Cochrane Reviews as the most rigorous method currently known for making inferences about the effects of medical treatments. However, the reasons for trusting Cochrane Review typically have no presence in an individual instance of its use.

This is not unusual: Warrants rarely, if ever, have a specific textual anchor in an argument. Freeman [7] made the same point and went further to say that they should not be included at all in diagrams of argument (as products). At best, an argument may refer to the rule applied in making an inference from data to conclusion, with rule understood here as an instruction for how to draw conclusions from data of that kind. The warrant in an argument from expert opinion, often represented as a kind of assertion, is actually better understood as a directive (“Trust experts unless there is some reason to doubt them in a particular case”). And the warrant in an argument based on Cochrane Review is (analogously) a set of instructions, very detailed, for what to do, given the availability of a certain kind of information. Why the inference rule is trusted is not part of the argument in any one instance of its use.

A published Cochrane Review will typically describe itself as a Cochrane Review but it will not enumerate all of the reasons that Cochrane Reviews are considered to be a strong basis for conclusions. An individual Cochrane Review does not contain a complete description of how Cochrane procedures justify the inference from a pile of studies to a generalized conclusion about a medical treatment, any more than an argument from expert opinion contains whatever it is that leads people in a particular place and time to assume that if an expert says something, it is likely true.

A very important quality of these invented inference rules is that the rules themselves may change over time. Since its invention (just a quarter-century ago), the Cochrane device has been strengthened through refinement of its procedures and through accumulation of material and institutional resources that make it technically superior to prior and competing reviewing methods. Individual reviews are thus being conducted under conditions of “fluidity” [16]: While all such reviews produce conclusions warranted by the Cochrane device, the device itself is still changing, generally becoming less vulnerable to criticism, but occasionally making trade-offs between scientific rigor and other goals (such as timeliness of response to clinical need). The Cochrane community implicitly recognizes that this fluidity means that conclusions drawn under one version of the device may be weaker than conclusions drawn under later versions of the device; in fact, a Cochrane Review may be “withdrawn” as it ages, not only because the evidence base may have changed but also because the procedures that were used may fail to satisfy a rising standard of proof associated with improvements in the device. An argument warranted by a device may include interpretation of how the device was applied given the particular circumstances of application. A similar observation has been made by Tans [16] for the case of legal (judicial) reasoning; part of the fluidity of legal warrants is the work that goes into establishing whether some particular rule can be applied to a case. Device-based reasoning in science is at least as complex as judicial reasoning in this respect.

The work required to justify a newly invented warranting device appears to be one case of what Toulmin [17] had in mind when he distinguished warrant-establishing arguments from warrant-using

arguments. Unfortunately, Toulmin offered no serious analysis of how a novel warrant might come to be established, providing only a casual and suggestive description of how a warrant might be established: “Warrant-establishing arguments will be... such arguments as one might find in a scientific paper, in which the acceptability of a novel warrant is made clear by applying it successively in a number of cases in which both ‘data’ and ‘conclusion’ have been independently verified” [17]. Passing references elsewhere in Toulmin’s work suggest that he had in mind newly established empirical generalizations, which could properly be represented as conditionals, resulting in a defeasible *modus ponens* structure as shown in Bex and Reed [2]. Empirical generalizations phrased as conditionals are better understood not as anything like an inference rule but as part of the data. Warranting devices are not empirical generalizations but packages of procedures used to generate conclusions, and to try to model them as premises would mean enumerating a very large number of premises in modeling every conclusion drawn through application of the device. Bex and Reed discuss the general disadvantages of treating inference rules as conditional premises [2], and those disadvantages are massively amplified as the rule itself becomes more complex.

Toulmin provided no realistic examples of warrant-establishing argument to clarify what he meant, but our own search for cases of warranting-establishing argument have not, so far, borne much resemblance to Toulmin’s speculations. Where Toulmin hinted that warrants might be established by showing their ability to connect premises known independently to be true with conclusions known independently to be true, what we have observed is that warranting devices are often invented to generate conclusions for which there is no other known method of drawing a conclusion, let alone an independent method for determining whether the conclusion is true. Our very preliminary examinations of new methods for data aggregation suggest that one motivation for invention of a warranting device is to draw conclusions from observations that, previously, were considered too confusing to allow conclusions to be drawn at all, as for example when variations within a set of observations simply appear to be random. Warranting devices may also be invented in an effort to improve on whatever might previously have been the best available way of drawing a conclusion, where improvement might involve something like reducing the proportion of incorrect conclusions. As we plan to argue in the next section, a warrant-establishing argument will have a conclusion consisting of a new warrant (or possibly an endorsement of a new warrant), and data consisting partly of what would commonly be considered backing for the new warrant. To date, computational modeling of argument has mostly tackled the analysis of what Toulmin called warrant-using argument; this is most evident in the design of the Argument Interchange Format [1,5] around theory-specific inventories of argumentation schemes that are used to characterize and evaluate individual arguments. Evaluation of schemes themselves, including in the process of establishment (and possibly dis-establishment), must also be accomplished through argumentation, and must therefore be in the scope of computational modeling of argument.

3. Warrant-establishing arguments: Defense of a new inference rule

In the real world of argumentation, defense of a new inference rule must itself be accomplished through argumentation, so it depends on use of whatever prior reasoning resources are available to arguers. In different theoretical terms, this means that it must be possible for new inference schemes to be devised as extensions or replacements or improvements of old schemes, and this process must be possible by using trusted schemes to justify newly proposed ones. This suggests that any proposed modeling system with a finite catalog of inference rules should likewise be capable of extending itself

through use of the existing rules to build new rules. We show here that a catalog containing a Practical Reasoning scheme as described by Walton [18] can generate new schemes of the kinds we have been studying, and there may be other paths to extension of a catalog of schemes. Very likely, one aspect of field-dependence in argumentation may be the paths used for establishment of new warrants or new schemes.

Practical reasoning is reasoning about what to do in a situation. The search for new inference rules is often a search for what to do with certain kinds of information that present unresolvable dilemmas for established inferential practices. One way to establish a new inference rule is to construct a practical argument that advocates application of the rule as a way to meet the field's standing goals in some particular class of circumstances (such as the availability or unavailability of a certain kind of information). The conclusion of such an argument is an instruction about how to proceed in the relevant circumstances. A warrant-establishing argument must have an inference rule in (or as) its conclusion: an instruction or a set of instructions for how to make an inference, possibly including instructions on gathering or producing data.

A number of different analyses of practical reasoning are already available, but none, so far as we know, of what practical reasoning looks like when applied to the justification of an inference rule. Fixed elements in nearly all accounts of practical reasoning are a goal, a set of circumstances in which the goal has not been fulfilled, a means of attaining the goal (urging of which will be the conclusion), and some reasons for believing that the goal will in fact be achieved by the means proposed. Practical reasoning in particular domains can be further specified in terms of the content of the various structural elements (goal, circumstances, means-end premise, etc.). Fairclough and Fairclough [6] provide a useful variation on Walton's account, specifying the substantive components of practical reasoning in defense of policy proposals or governance actions. Practical reasoning in politics is characteristically comparative, involving simultaneous consideration of several competing actions, and Fairclough and Fairclough have created a "dialectical profile" for practical argument [6] that facilitates analysis of this kind of situation. This feature of their work is highly relevant to the use of practical reasoning to establish inference rules in science, since like alternative policies or alternative governance decisions, inference rules must often be defended in relation to other possible ways of drawing inferences in the same circumstances. Specifically, new inference rules typically appear in response to some recognition of weakness or vulnerability in a prior way of arriving at beliefs within the field. At least initially, a newly proposed inference rule may be seen as competing with a prior rule.

Establishing a new inference rule in science depends on showing that the rule satisfies some disciplinary goal (such as forming correct beliefs in the discipline's domain), given whatever is known about the potential obstacles to forming correct beliefs (here included in what Fairclough and Fairclough term "circumstances"). All practical reasoning includes (implicitly or explicitly) a means-end premise that is a reason for believing that taking an action will in fact achieve a goal, and in the case of arguments in support of an inference rule, the means-end premise will concern the ability of a rule to generate conclusions that are consistent with the field's goals and values. When multiple rules are available, for instance when a new rule is formulated as an alternative to some prior inference rule, the means-end premise is satisfied by showing that the new rule will produce results that, over many instances of use, will be more often correct than results produced by the old rule. Figure 1 shows our adaptation of Fairclough and Fairclough's practical reasoning scheme to the purpose of establishing a new inference rule by giving the rule itself as the conclusion of the argument.

To illustrate, consider Randomized Clinical Trials as an inference rule (or in our terminology, as a device for warranting beliefs about medical treatments). Randomized Clinical Trials (or RCTs, as they are

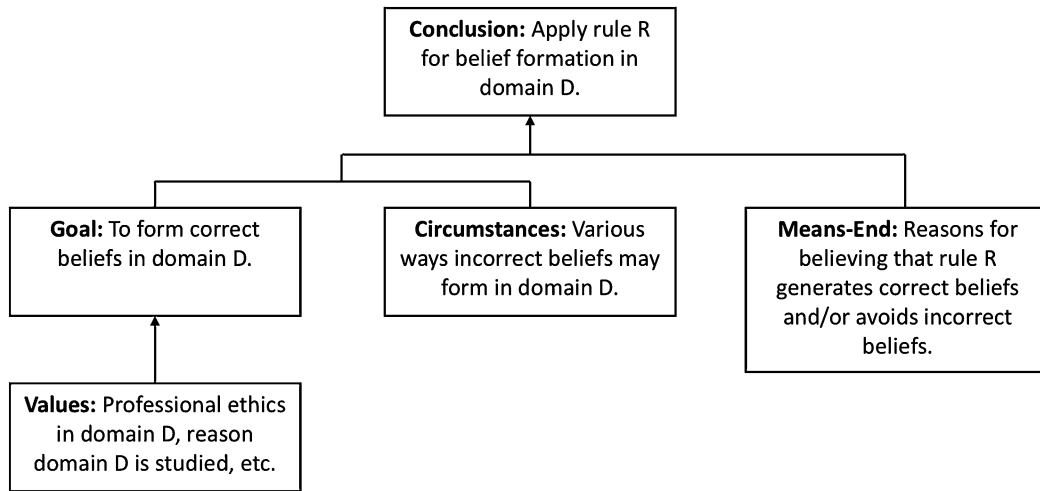


Fig. 1. The general structure of practical reasoning adapted from [6] to apply to establishment of inference rules.

commonly called) involve administering experimental treatments to patients who have been randomly assigned to one treatment condition or another, while attempting to control the influence of anything other than the experimental treatments on outcomes for the comparison groups. Many variations of RCT can be devised to adapt to specific scientific circumstances, but in the simplest possible case, two randomly assigned groups of patients are compared on outcomes that a treatment is expected to affect. For example, to determine whether a new drug is an effective remedy for some disease, an RCT would randomly divide a pool of patients into two groups with the disease, giving the new drug to one group and a different drug or a placebo to the other group. Patients in both groups would be observed to determine whether (or how quickly) they got better (or worse), and the aggregate results for the groups would be compared using statistical techniques to draw a conclusion about whether the new drug improved (or worsened) patient outcomes. In more complicated designs (factorial designs, designs with multiple outcomes, designs with pretest measurements as well as outcome measurements, etc.), random assignment has the same purpose and offers the same basis for fair comparison.

RCTs are experiments done to provide the data required by an inference rule, different from one another in many particulars but relying on the same principles to justify movement from the data generated in the experiment to an empirical conclusion. We will use RCT to refer to the rule in what follows. The path to establishment of RCT as a widely-used warranting device must be given complete treatment elsewhere [14], but can be briefly summarized here. RCT has some counterintuitive elements that required extensive defense by many advocates before it could become the taken-for-granted “gold standard” for medical evidence [3]. For example, Austin Bradford Hill, a statistician, explicitly argued before medical audiences [4] for the superiority of controlled experimentation over clinician experience as a basis for belief, and he rebutted ethical objections to experimenting on humans by insisting that *discovering* which treatments really work was among the medical field’s ethical responsibilities. His claim might be abstracted many ways (e.g., as “RCT is superior to clinician experience in assessing the effectiveness of medical treatments” or as “Medical science should increase its use of RCT”), but since his clear intent, and that of others arguing similarly, was to shift the basis for belief, we represent the conclusion in the form of an inference rule in directive form: “Apply RCT for belief formation about medical treatments”. Bradford Hill’s argument, as written, maps quite precisely onto the practical argument scheme shown in

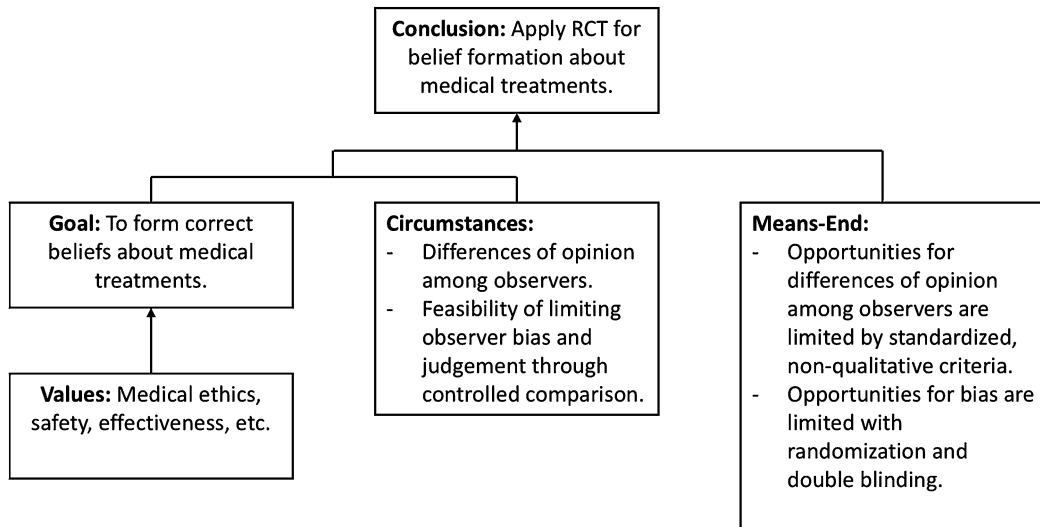


Fig. 2. Bradford Hill's argument for RCT represented as practical reasoning.

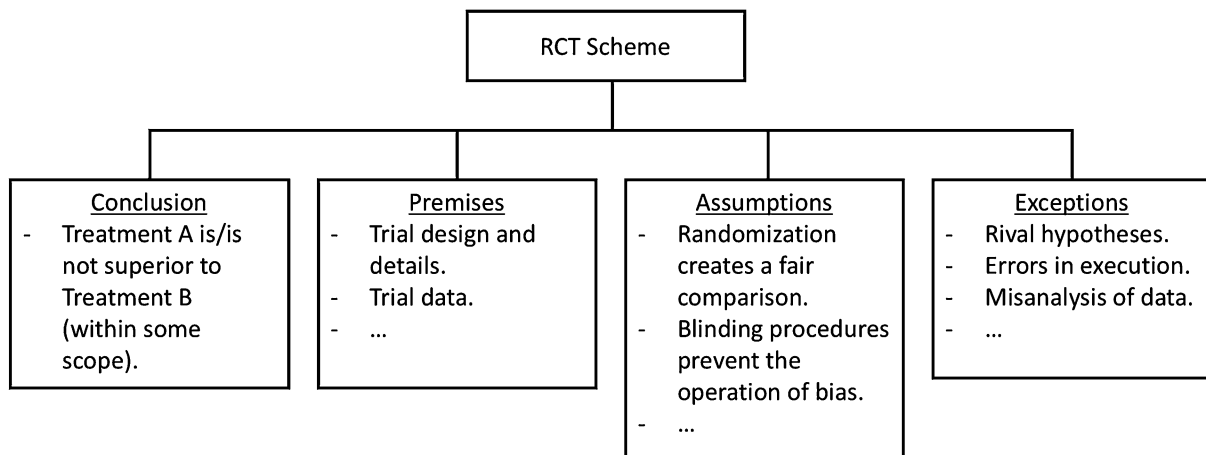


Fig. 3. A scheme definition for the RCT Scheme, to enable adding it as a new AIF Form.

Figure 1. In Figure 2, his warrant-establishing argument for using RCT to derive beliefs about medical treatments is diagrammed in the style of Fairclough and Fairclough [6].

The establishment of a new inference rule would be modelled in any AIF-compliant approach as the addition of a new scheme definition. For RCT, as noted earlier, a family of schemes would likely be required, involving variations in the design of experiments. Figure 3 shows a very generic version of a scheme definition for RCT, based on a simple treatment-versus-treatment comparison.

The difference between the warrant-establishing argument and subsequent arguments using the newly established warrant is illustrated in Figures 4 and 5. None of this defense of RCT (or other such inference rules) will normally appear in everyday use. Rather, nowadays RCT would be used as an inference rule, as shown in Figure 5.

Within any computational system with a set catalog of forms, it should be possible to derive new forms by applying some combination of existing forms. If the system acknowledges practical reasoning,

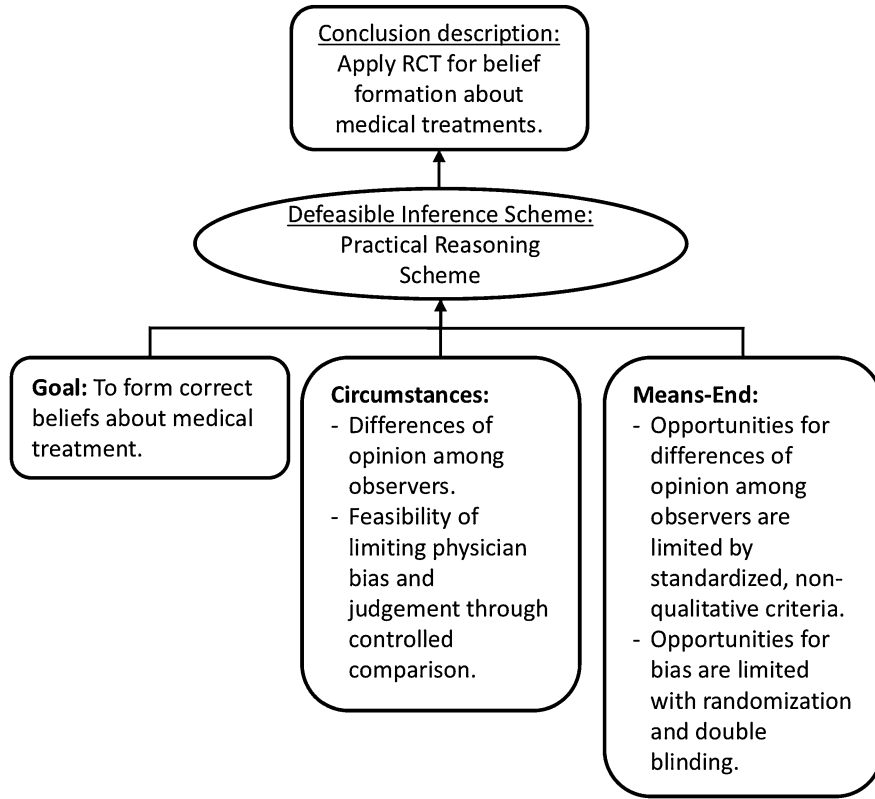


Fig. 4. Establishing RCT as an inference rule using the practical reasoning scheme.

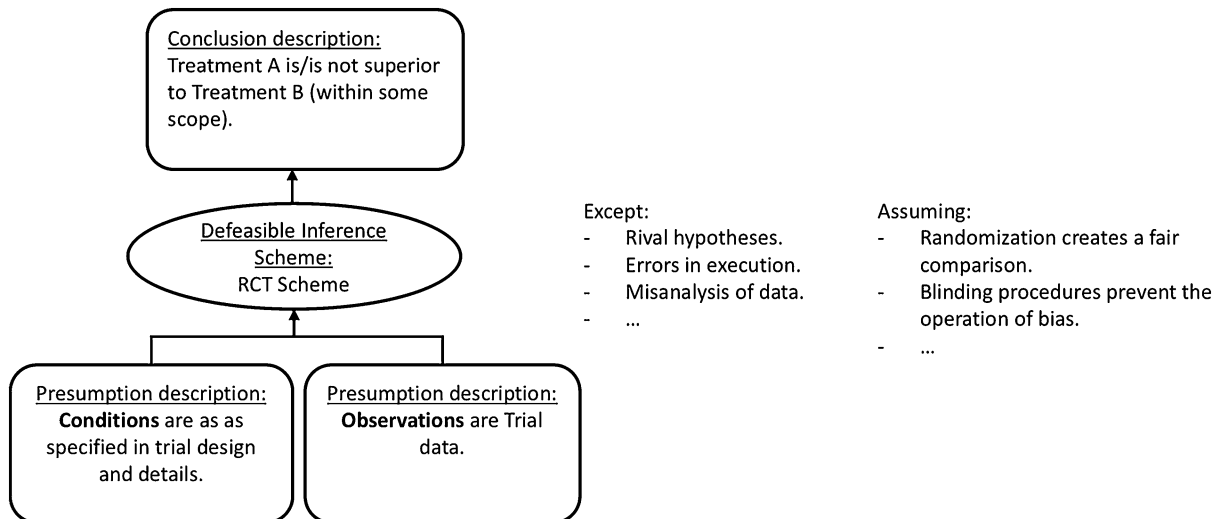


Fig. 5. Using RCT as an inference rule.

it can generate indefinitely many inference rules by specifying the circumstances, values, and goal, under which the rule would provide an effective means to achieve the desired end. Then by packaging the name of the rule together with a description of its premises, assumptions, exceptions, and conclusions, the rule can be specified as a new form. A finite catalog of forms can thus grow through recombination. For reasoning in medical science, RCT is such a form.

Although some warranting devices have no competitors, RCT does. That is, RCT is not the only path to a belief about a treatment. For many centuries, doctors formed beliefs about the effectiveness of particular treatments by observing individual patients to whom the treatments were given. A clinician's intimate familiarity with a patient's history and direct observational access to improvements (or worsenings) following a treatment may seem to be good bases for inferences about the effectiveness of the treatment – but they may also be sources of bias, including confirmation bias. Part of the defense of RCT was exposure of the weaknesses and vulnerabilities in these inferences, for example noting that observation of what happened following a treatment does not take into account what would have happened in the complete absence of treatment. In effect, the use of RCT to form beliefs about treatment effectiveness had to be compared with individual clinician experience as a basis for forming beliefs about treatment effectiveness, and establishing RCT as the preferred method for forming beliefs about medical treatment required practical argumentation organized around showing that RCT produces *better* grounding for a conclusion than that produced by clinician interpretation of patients' progress following a given treatment. The role of practical argumentation here is in attempting to get clinicians to give up one rule for forming beliefs about treatment effectiveness for a different rule (one that by its nature would be counterintuitive for the clinician).

Every use of the newly established warrant inherits vulnerabilities that may later be discovered in the inference rule itself. RCT's superiority over case-based reasoning by clinicians has long been assumed; RCT's strengths and limitations as compared with big data methods are matters of contemporary discussion. This means that an invented inference rule should have explicitly encoded, modifiable preference relationships with other inference rules in use within the same domain.

In the next section, we consider modeling the use of an invented inference rule that has become black-boxed – the kind of object we call a warranting device.

4. Warrant-using argument with black-boxed inventions

A successful defense of a new inference rule will lead to argumentative “black-boxing” of the rule [9]. Black-boxing in argumentation means modularization of a complex structure so that the entire structure may be used without requiring that it be rebuilt for each occasion of use. If for any reason arguers should begin to suspect that a black box is an unreliable source of conclusions, it can be re-opened, re-inspected, and even re-designed. But that generally means returning to argumentation over the rule itself, rather than critiquing the rule only in the context of one instance of its use.

Having been black-boxed within the field, the procedures of the RCT itself are not *defended* in each occasion of use, though certain details of its application may need to be explained. An investigator must show that felicitous conditions exist (e.g., that there are treatments that can be ethically compared) but need not argue for the RCT itself. For this situation, a warranting device can function like any other rule available for application. This conforms well with how empirical argumentation actually occurs within a literature consisting of research reports. The ability of the rule to deliver a believable conclusion, having been established previously within the same domain of discourse, is assumed rather than argued. At this

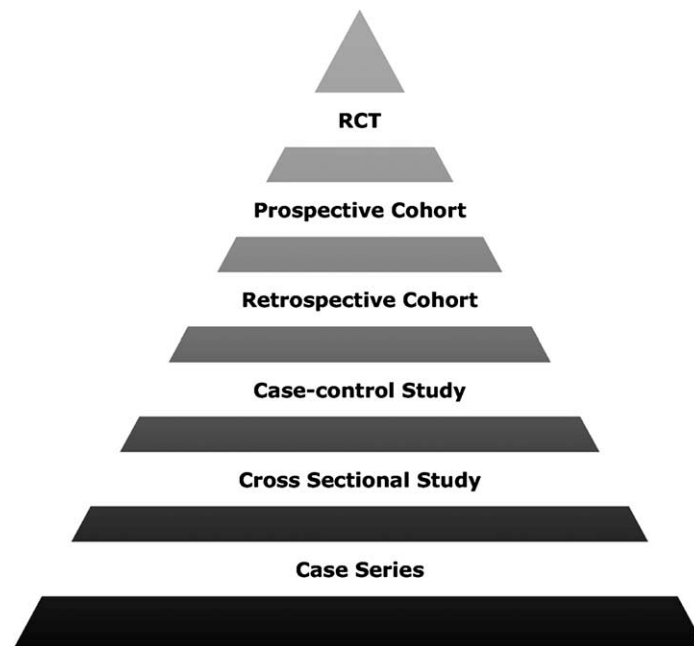


Fig. 6. An example of an evidence hierarchy in medicine, from [8].

stage, once the device is established, it can be applied matter-of-factly without defense: it is no longer necessary to locate and analyze the arguments that underwrite the inference rule.

An important qualification, though, is that within expert fields, there are very often alternative methods for arriving a conclusion, and they need not produce the same result. Consider three ways of building an argument for concluding that a treatment is effective: clinicians' records on their own patients, RCT, and Cochrane Review (of multiple applications of RCT or other rules). In health sciences, these and many other methods are available, and the confidence to be placed in each is discussed quite explicitly, often being represented as an evidence hierarchy that shows preference relationships among different ways of answering questions in medical science. These preference relationships are themselves contentious matters [8,15], so the establishment of preference schemes for sets of inference rules is also of interest.

Figures 6 and 7 show a common way of discussing preferences among different methods for drawing conclusions about medical treatments: as a pyramid, where the top is reserved for what is considered the strongest available form of evidence. RCT has been considered the top of the pyramid for establishing causality, as shown in Figure 6, but there are important qualifications on this having to do with limitations of any one evidence type [8]. For example, although an RCT can undercut causal arguments supported by other evidence, it is not quite right to say that the RCT-based conclusions will always win over conclusions drawn from (say) a cross-sectional study. An RCT may show that (in a trial setting) a treatment produces better outcomes than any of its alternatives; however, this conclusion can still be rebutted by showing that outside the trial setting, large numbers of patients receiving the treatment fare no better than those who do not receive the treatment.

As should be clear from Figure 7 (and from [15]), the positioning at the top of the pyramid may be contested. Increasingly, both researchers and scientists regard systematic reviews as preferable to even the most rigorously conducted individual studies, and for good reason. Results from a Cochrane Review or other systematic review of many RCTs cannot be refuted by any one of the individual studies, while

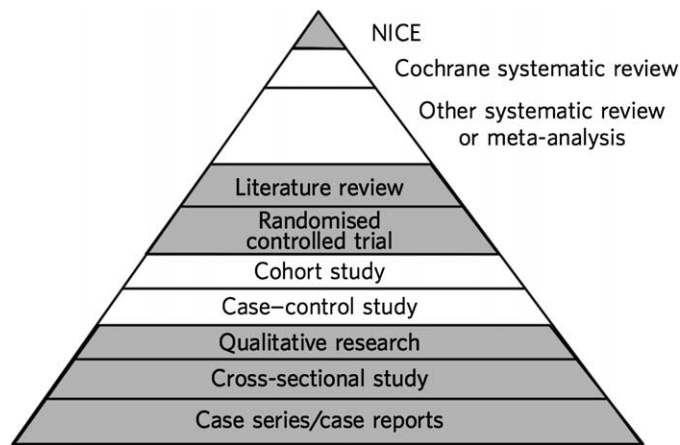


Fig. 7. An example of an evidence hierarchy in medicine, from [15].

every individual study is open to possible rebuttal on the grounds that it is inconsistent with the whole body of evidence reviewed. In this sense, the review conclusion is preferred to the conclusion of any individual study – leading to “demotion” of RCT from the position it occupies in Figure 6. But both individual circumstances and future innovations may result in challenges to what currently appears at the top of the pyramid. Figure 7 gives preference to “NICE guidelines” (clinical practice guidelines from the UK National Institute for Health and Care Excellence), based on the argument [15] that some purposes are better served by qualifying conclusions drawn from systematic review with other sources of guidance.

Representing preference relationships among individual instantiations of schemes is already a familiar task within computational modeling of argumentation, and it seems obvious that there must be preference schemes that compare inference schemes. As Figs 6 and 7 show, there is discourse within the field that can be modeled as preference schemes that apply to alternative inference schemes, and just as there is discourse aimed at establishing new inference schemes, there is discourse aimed at establishing preference among schemes.

Like any inference rule, a warranting device functions differently for performers and evaluators of arguments: For the performer, it functions as a set of instructions for how to arrive at a conclusion. For the evaluator of the conclusion, it functions as a warrant for believing the conclusion drawn by the performer or as a scheme for evaluating that conclusion. Within an expert field, the device itself is trusted to produce sound conclusions, and although the black-boxing of the device is reversible, this is rarely done in the context of challenging a single conclusion returned by the device. Similar considerations apply outside the field, with the additional complication that non-experts must decide whether the field’s trust in a device is also a reason for non-members to trust the device. In our study of Cochrane Review as a warranting device [10], we describe how Cochrane Review has earned trust within medical and public health professions, while still being open to some level of distrust by the non-expert public (who may not share the goals, values, and means-end assumptions that are taken-for-granted within the expert field).

A complicated arrangement like RCT or Cochrane Review, once established as a reliable basis for inference, can be invoked like any other scheme. These devices *feel* different from other common inference rules because they involve complex and explicit instructions for how to arrive at a conclusion. In any application of the device, representing the complexity of the device itself can be off-loaded to the arguments that establish it as reliable.

5. Conclusions and perspectives

Toulmin's distinction between warrant-using and warrant-establishing argument gains clarity from the insights of computational argument. A warrant-establishing argument is not a conclusion from induction (as Toulmin's examples seemed to suggest), but an instruction for how to draw a conclusion (including how to *perform* inductions). A new warrant can be established by practical argumentation that has a proposed inference rule as its action claim, and no doubt by other means in other fields.

In this conclusion, we want to draw out several implications for both empirical work in argumentation and for computational modeling as an approach to empirical work.

First, computational modeling of argument can represent more than one level of granularity in argumentation. Figures 4 and 5 illustrate what we mean: whereas Figure 4 represents an explicit trans-situational defense of the rule, Figure 5 represents the matter-of-fact application of the black-boxed rule in some particular situation. The latter involves less granularity in representation of what commitments the arguers take on in any application of the rule. But that can be expanded again since every argument applying the rule depends on the continued viability of the rule's own defense. Even though these commitments do not have any presence in the argument that invokes an established rule, they can be dredged up under critical questioning; all of the rule's "inner workings" can become part of the disagreement space any time an interlocutor chooses to open up the black box. Whatever reasons there are for the re-opening of a black box may or may not become a basis for critiquing any particular conclusion the black box has generated. In other words, re-examination of a device like RCT around a newly discovered issue may mean propagating that new issue through all prior and future uses of the device.

Second, it appears to be quite feasible for computational models to handle new types of arguments characterized by newly invented warrants, overcoming the limitations of a finite catalog of forms. This is important because new inference rules – specific practices, each proposed and validated for a particular use – are continually invented, and we expect them to continue to be invented regularly and without limit, so long as expert fields strive for greater knowledge, and so long as anyone strives for more rational deliberative processes. Modeling argumentation within expert fields will require extensibility in any list of schemes to allow for whatever might be invented to support expert reasoning. Besides allowing for new inference rules to be modeled and represented, there must be some approach for modeling how this actually happens. In scientific fields, where new inference methods are generally sought as solutions to inference problems, practical argument is one obvious path to justification of a new inference scheme.

Finally, the notion that new inference rules may be invented and established suggests at once that old ones may also be disestablished – either abandoned or replaced. The objects we have called warranting devices are obvious candidates for disestablishment, either because experience in their use exposes an unnoticed problem or because some new invention obsolesces what was previously the best-known way to reach a conclusion in a given domain. But even familiar presumptive reasoning schemes, like Argument from Expert Opinion, can be the target of a warrant-disestablishing argument; such an argument would focus on new reasons to doubt the reliability of the formerly trusted scheme, as in Mizrahi's recent reexamination of trust in experts [12,13]. For computational modeling, interesting challenges surround the process by which arguers maintain a repertoire of forms over time through additions, transformations, and even removals. That is, something new to model is the change in the (natural) system that is being modeled.

Acknowledgements

Thank you to Kiel Gilleade for assistance in preparing figures.

References

- [1] Argument Interchange Format 2 Specification, available in various formats from <http://www.argumentinterchange.org/aif2spec>, e.g. OWL: <http://www.arg.dundee.ac.uk/wp-content/uploads/AIF.owl>, Jan. 2013.
- [2] F. Bex and C. Reed, Schemes of inference, conflict, and preference in a computational model of argument, *Studies in Logic, Grammar and Rhetoric* **23**(36) (2011), 39–58.
- [3] L.E. Bothwell, J.A. Greene, S.H. Podolsky and D.S. Jones, Assessing the gold standard – Lessons from the history of RCTs, *New England Journal of Medicine* **374** (2016), 2175–2181. doi:[10.1056/NEJMms1604593](https://doi.org/10.1056/NEJMms1604593).
- [4] A. Bradford Hill, The clinical trial, *New England Journal of Medicine* **247**(4) (1952), 113–119. doi:[10.1056/NEJM195207242470401](https://doi.org/10.1056/NEJM195207242470401).
- [5] C. Chesñevar, S. Modgil, I. Rahwan, C. Reed, G. Simari, M. South, G. Vreeswijk and S. Willmott, Towards an Argument Interchange Format. *The Knowledge Engineering Review* **21**(4) (2006), 293–316. doi:[10.1017/S0269888906001044](https://doi.org/10.1017/S0269888906001044).
- [6] I. Fairclough and N. Fairclough, *Political Discourse Analysis: A Method for Advanced Students*, Routledge, New York, 2012.
- [7] J.B. Freeman, *Argument Structure: Representation and Theory*, Springer, Dordrecht, 2011. doi:[10.1007/978-94-007-0357-5](https://doi.org/10.1007/978-94-007-0357-5).
- [8] P.M. Ho, P.N. Peterson and F.A. Masoudi, Evaluating the evidence: Is there a rigid hierarchy?, *Circulation* **118**(16) (2008), 1675–1684. doi:[10.1161/CIRCULATIONAHA.107.721357](https://doi.org/10.1161/CIRCULATIONAHA.107.721357).
- [9] S. Jackson, Black box arguments. *Argumentation* **22**(3) (2008), 437–446. doi:[10.1007/s10503-008-9094-y](https://doi.org/10.1007/s10503-008-9094-y).
- [10] S. Jackson and J. Schneider, Cochrane Review as a “warranting device” for reasoning about health. *Argumentation* **32**(2) (2018), 241–272. doi:[10.1007/s10503-017-9440-z](https://doi.org/10.1007/s10503-017-9440-z).
- [11] S. Jackson and J. Schneider, Argumentation devices in reasoning about health, in: *16th Workshop on Computational Models of Natural Argument (CMNA) Collocated with the International Joint Conference on Artificial Intelligence (IJCAI)*, CEUR Workshop Proceedings, Vol. 1876, 2016, <http://ceur-ws.org/Vol-1876/paper10.pdf>.
- [12] M. Mizrahi, Why arguments from expert opinion are weak arguments, *Informal Logic*. **33** (2013), 57–79. doi:[10.22329/il.v33i1.3656](https://doi.org/10.22329/il.v33i1.3656).
- [13] M. Mizrahi, Argument from expert opinion and persistent bias, *Argumentation* **32**(2) (2018), 175–195. doi:[10.1007/s10503-017-9434-x](https://doi.org/10.1007/s10503-017-9434-x).
- [14] J. Schneider and S. Jackson, Innovations in reasoning about health: The case of the Randomized Controlled Trial, Paper presented at *ISSA 2018. International Society for the Study of Argumentation*, Amsterdam, Netherlands, Jul. 3–6, 2018.
- [15] R. Steele and P.A. Tiffin, ‘Personalised evidence’ for personalised healthcare: Integration of a clinical librarian into mental health services – A feasibility study, *The Psychiatric Bulletin* **38**(1) (2014), 29–35. doi:[10.1192/pb.bp.112.042382](https://doi.org/10.1192/pb.bp.112.042382).
- [16] O. Tans, The fluidity of warrants: Using the Toulmin model to analyse practical discourse, in: *Arguing on the Toulmin Model*, 2006, pp. 219–230. doi:[10.1007/978-1-4020-4938-5_14](https://doi.org/10.1007/978-1-4020-4938-5_14).
- [17] S.E. Toulmin, *The Uses of Argument*, Cambridge University Press, London, 1958.
- [18] D.N. Walton, *Practical Reasoning*, Rowan & Littlefield, Savage, MD, 1990.
- [19] D.N. Walton, *Argumentation Schemes for Presumptive Reasoning*, Lawrence Erlbaum Associates, Mahwah, NJ, 1996.