6-1-2013

Kuhn vs. Popper on criticism and dogmatism in science, part II : how to strike the balance

Darrell Patrick ROWBOTTOM

Lingnan University, Hong Kong

Recommended Citation

Rowbottom, D. P. (2013). Kuhn vs. Popper on criticism and dogmatism in science, part II: How to strike the balance. Studies in History and Philosophy of Science Part A, 44(2), 161-168. doi: 10.1016/j.shpsa.2012.11.011

Kuhn vs. Popper on Criticism and Dogmatism in Science, Part II: How to Strike the Balance

Darrell P. Rowbottom

DarrellRowbottom@ln.edu.hk

This paper is a supplement to, and provides a proof of principle of, *Kuhn vs. Popper on Criticism and Dogmatism in Science: A Resolution at the Group Level*. It illustrates how calculations may be performed in order to determine how the balance between different functions in science—such as imaginative, critical, and dogmatic—should be struck, with respect to confirmation (or corroboration) functions and rules of scientific method.

1. Introduction

In a recent paper in this journal, I argued that the debate between Popper and Kuhn on whether scientists should be critical or dogmatic should be resolved by thinking at a higher level, namely the level of the group.[1] I argued that it is typically preferable for the two kinds of activity to co-exist, along with imaginative/creative processes, and for the balance between them to alter in response to the context. I noted that the

[1] I have now discovered that a similar move to group level considerations was suggested by Musgrave (1976), but as a critique of Lakatos's view of science. I highly recommend reading this paper, which I am grateful to Professor Musgrave for drawing to my attention.

crucial question then becomes 'How should the balance between functions be struck?' But I did not show how to tackle it.[2]

My task in the present paper is to remedy this situation, by providing a proof of concept of calculating how to strike the balance between functions. I will start with a simple model, and show how more complexity may be built in on a step-by-step basis. I will also outline a general procedure by which to generate competing models.

Before I do this, I will revisit a few of my previous claims in order to make it clear precisely what I am endeavouring to show. First, consider the following analogy:

> Imagine you, the chess player, are managing science. The pieces are the scientists under your command, and their capacities vary in accordance with their type (e.g. pawn or rook). The position on the board—nature is playing the opposing side—reflects the status quo. And now imagine you are told that, against the rules of normal chess, you are allowed to introduce a new [piece] (which you can place on any unoccupied square)… (This is akin to the introduction of a new scientist; pieces working in combination on your side can be thought of as research groups, and so on.) Some moves will be better than others, given your aim of winning the game, and in some circumstances it will be clear that one available move is best. (Rowbottom 2011a, p. 124)

[2] I also added that 'It may prove to be… beyond our power to answer satisfactorily except in highly idealized contexts.' (Rowbottom 2011a, p. 124) However, computer simulations allow us to introduce more complexity than we can handle via traditional philosophical methods.

I was at pains to emphasize that there is nothing relativistic or debilitating, from the point of view of mainstream philosophy of science, about this picture. On the contrary, my intent was to suggest that we can employ existing logical resources—by which I mean resources from the logical tradition in the philosophy of science, such as confirmation functions[3]—to inform our account of *what the group should do* rather than *what each (and every) individual should do*. I added that:

> I have not denied that there is a fact of the matter about what an individual scientist might best do (or best be directed to do) in a particular context of inquiry. Rather, I have suggested that determining what this is requires reference not only to the state of science *understood as a body of propositions (or as knowledge)* but also to *what other scientists are doing and the capacities of the individual scientist.* (Rowbottom 2011a, 123–124)

In summary, I therefore claimed not only that 'How should the balance between functions be stuck?' can be answered, but also that it can be answered in a way that is compatible with the logical tradition in the philosophy of science (and in particular, the view that theories have measurable degrees of confirmation or corroboration). So this, in full, is what I shall endeavour to show in what follows.

Before I continue, however, I should make it explicit that both epistemic and pragmatic ends underpin the 'how should' question. For example, some distributions of labour might maximise the probability of truth likeness of the theoretical products

[3] More specifically, I wrote: 'I take there to be measures… even if they are rough measures… of how theories (and/or research programmes, modelling procedures, etc.) are faring. And these, given the resources at our disposal, determine how we should respond.' (Rowbottom 2011a, p. 124)

of science (which may be seen as an epistemic good), while nevertheless rendering science terribly inefficient in some respects, e.g. slow and costly in theory production (*qua* endorsement). So answering the 'how should' question requires us to be clear about what we want from science (given the external constraints imposed in any given context, e.g. funding and available womanpower). Naturally, this will be delimited by what we think that science can ideally achieve (in said context); and hence, opinions will be divided along realist and anti-realist lines, among others.[4] So for present purposes, I wish to leave this question relatively open. All I shall assume is that everyone will agree that there are more or less efficient ways of organising science, and that the question has bite in this respect (at the bare minimum). To return to the earlier chess analogy, it is better to force checkmate in three moves rather than five moves, if it is possible to do so.

I also should emphasise that I am mainly concerned, for the present, with the question of how the balance should be struck *in principle*, rather than what should be done *in practice* to strike it (or to get as close as possible to striking it). Answering the *in principle* question is important, in my view, for answering the *in practice* question properly. Think of it this way. When creating a model for some specific purpose, it is useful to have some kind of idea of what you would include in a complete model (i.e., one where complexity were no obstacle). You can then think about the situations for which your model is intended; call these the *contexts of application*. You can leave out factors that would be in the complete model but won't, you anticipate, make much difference in the target contexts of application. You can also try to find ways to account for relevant factors that you aren't able to include in a computationally

[4] Personally, due to my instrumentalist tendencies, I am strongly inclined to thinking of the matter in purely pragmatic terms. But there is no need to do so.

tractable way. And so forth.[5] This fits with the way that Muldoon and Weisberg (2011, p. 162) describe the modelling process: 'we must rely on idealizations to reduce the complexity… [but] not… so extreme that we lose the ability to explain'.

I might also add that intervening in science, e.g. as a research manager, will often have direct consequences that I won't take into account; for example, it might cause resentment and therefore have a negative effect on output. When I talk of distribution of labour *in principle*, I am working with the idealisation that no consequences of this kind will occur. (If liked, to return to the chess analogy, I assume that what the pieces do won't be affected by who moves them and how. And in saying this, I am not presupposing that they are moved, as a matter of fact.) But this assumption must be relaxed if one is to tackle the *in practice* question seriously.

Finally, on a somewhat related note, I should say something about the scope of my subsequent discussion. I will write as if I am discussing only the whole of science. But similar considerations may instead be used to determine the proper way to strike the balance within smaller units of inquiry, such as sub-disciplines or research groups. The only difference is that extra considerations—or higher-order models—may also be required. If one is faced with the *in practice* question of how best to organise a specific research group, for example, one may consider the relative prospects of competing with other groups rather than attempting to carve out a new niche. Such issues have already been considered by Kitcher (1990, 1993) and Strevens (2003).

[5] Take building a model of planetary motion in our solar system as a case in point. A complete model would include the positions of all heavenly bodies that exert gravitational forces, *inter alia*. It is clear that leaving out some of these, e.g. cosmic dust and anything in other galaxies, is fine for arriving at reasonably accurate predictions of the orbital paths of the planets. Leaving out dwarf planets—such as Eris, Chiron, and Pluto—is more questionable.

2. A Simple Model

Let's assume that we've agreed on a confirmation or corroboration function, because the following account is intended to be compatible with all the contenders for this role. (In what follows, I will use 'confirmation' loosely, so as to cover anti-inductivist accounts of corroboration.) Such functions are typically defined in terms of conditional probabilities, which involve three distinct types of argument: hypotheses (*h*), evidence (*e*), and background statements (*b*).[6] Most obviously, this goes for Bayes's theorem; see, for example, Salmon (1990). But here are some other examples (along with references to their advocates):

$$\mathrm{C}(h, e, b) = \frac{\mathrm{P}(e|hb) - \mathrm{P}(e|b)}{\mathrm{P}(e|hb) - \mathrm{P}(eh|b) + \mathrm{P}(e|b)}$$

Popper (1983)[7]

$$r(h, e, b) = \log \frac{\mathrm{P}(h|e\&b)}{\mathrm{P}(h|b)}$$

Milne (1996)

$$l(h, e, b) = \log \frac{\mathrm{P}(e|h\&b)}{\mathrm{P}(e|{\sim}h\&b)}$$

Huber (2008)

[6] How to understand *b* is a rather fascinating question. Should it be indexed to the present, or to the past point at which the hypothesis *h* was generated? And should it be thought of as background *knowledge*, or as something else? On these topics, see Musgrave (1974), Williamson (2010, p. 4–6), and Rowbottom (2011b, p. 90–92; In Press B).

[7] For some of the problems with this function, in the broader context of Popper's philosophy, see Rowbottom (In Press A).

I should like to point out, however, that none of the following *depends* on the existence of precise valued confirmation ratings of theories, or even absolute, rather than relative, measures of confirmation. One may instead imagine that we only have a way to rank theories or research programmes with respect to fruitfulness to date. Then it will be possible to consider the following in terms of numerical intervals. (If we can say that one theory is between two and three times as confirmed as another, or even just more confirmed than any other, for example, then this may provide *constraints* on how to organise science.)

Now let's consider a toy methodological rule:

> Distribute labour such that the work performed on each alternative (theory or research programme) is proportional to its relative degree of confirmation.[8]

I am not endorsing this rule. I just wish to show how one can balance on this rule. In fact, I believe that such rules are context sensitive (or conditional) and numerous. But I don't want to overcomplicate the discussion until later. Suffice it to say that there may be some points, in science, at which it is reasonable to follow this rule. This becomes rather more plausible if we consider it as a *ceteris paribus* affair, in the same

[8] Note that this leaves room for including 'catch alls', i.e. degrees of confirmation for 'no available theory is true' or equivalently 'some *unavailable* theory is true', although I do not include these in the following example for reasons of simplicity. It seems plausible that distributing labour to these alternatives would amount to performing imaginative/theory-construction work, as discussed below in section 3.1. This recognition makes the rule appear more plausible than it otherwise might. (Thanks to a referee for raising the idea of 'catch alls' here.)

way that we do many of our scientific laws.[9] To draw an analogy, we say that 'All iron bars expand when heated' although there are clearly circumstances in which it is possible to heat an iron bar without any expansion occurring. (It is true that scientific laws are descriptive, whereas methodological rules are normative, but this is no obstacle. For instance, there may be times at which a Christian must breach one of the Ten Commandments in order to obey another.) In any event, I shall later return to the subject of methodological rules, and show that nothing rides on this particular rule *ever* being appropriate.[10] To foreshadow that subsequent discussion, what ultimately matters is that the appropriate rules are sufficiently similar in character to also be applicable; and all this takes is that they rule out a relevant class of ways to organise science.

With our toy rule in place, let's consider a simple situation in which there are just two competing (mutually exclusive) theories, T_1 and T_2. Let $C(T_1)$ represent the absolute

---

[9] On *ceteris paribus* laws (and provisos), see Canfield and Lehrer (1961), Hempel (1988), Cartwright (1983), and Earman and Roberts (1999).

[10] One might argue against the toy rule by pointing to historical situations in which theories with little or no confirmation, over a considerable period of time, eventually emerged as superior to their rivals. Take Wegener's plate tectonics theory, mentioned by one of the referees of this paper, as a case in point. Doesn't the success of this, over that of earlier alternatives, show that following the toy rule is unwise? I think not, because the rule may be understood as a proposed means to improve efficiency in the long run, in a particular class of situations. And the fact that following the rule reduces efficiency in some situations does not mean that it does not improve efficiency overall, in the class of situations in which it is to be applied. (That is, for example, as against not employing the rule.) Similarly, the rule of 'bet on a heads result' may help you to win in the long run, on bets of whether a biased coin lands on heads or tails. But in some cases, following the rule will lead to losses. Nevertheless, there may be no better rule available. (One might also think there is something that makes the plate tectonics case 'special', such that the rule should not be applied; it is unclear to me, however, what that might be.) My worry about the toy rule is therefore slightly different. I wonder if we have empirical evidence, from the history of science, that following it would improve efficiency *overall* in some clear kind of situations.

confirmation value of T_1, and $C(T_2)$ represent the absolute confirmation value of T_2. Let's first define a relative confirmation/corroboration value, in order to help us to apply our rule:

$$R(T_1, T_2) = \frac{C(T_1)}{C(T_2)}$$

In general, if T_1 is *n* times as confirmed as T_2 then $R(T_1, T_2)$ will be *n*. So the fraction of total work we will want performed on T_1, according to our methodological rule, will be:

$$\frac{R(T_1, T_2)}{1 + R(T_1, T_2)}$$

We will want the remainder of the total work to be performed on T_2, as it is the only available alternative. The next step is to consider how to bring the appropriate balance into being. To do this, we need to think about the scientists we have available and their properties.

Let's say that we have *m* scientists. To a first approximation, we might add that each scientist has a criticism skill, CR, and a puzzle solving skill, PZ. Each may be assigned a value between unity and zero: 1 is the highest skill level possible, and 0 is the lowest skill level possible. But this is plausibly a bit too simplistic, because ability to be critical and solve puzzles is context dependent. So we really need criticism skills and puzzle solving skills in different domains. In this case, each scientist will have a $CR(T_1)$, $PZ(T_1)$, $CR(T_2)$, and $PZ(T_2)$ property. (This may still be too simplistic. One can sub-divide skills too. In line with Rowbottom (2011a), for example, puzzle solving involves doing more than one thing. We will return to this issue subsequently.)

We also need to think about how hard the scientists will work, in spite of their ability, i.e. how industrious they will be. Again, give this a value between unity and zero which is domain specific: $I(T_1)$ and $I(T_2)$. For the moment, assume that this value is fixed for any individual independent of their research environment, and that there's no variation between how industrious a scientist is when performing one function rather than another, e.g. criticising rather than puzzle solving. (These are further false oversimplifications. Yet again, we shall return to them. Such is the nature of model building.[11])

We can now define work output potentials for each scientist. Let's make the assumption that any given scientist can only be actively using either CR or PZ skills at any point in time, and can only be operating these skills on one theory at any point in time. This is reasonably realistic. But we must allow that the proper distribution of labour can change very quickly, indeed from moment to moment. Consider the earlier chess analogy. Time changing is like the position on the board changing. A piece that was previously performing a crucial role may now be contributing little or nothing. It may even be surplus to requirements.

[11] This is true in physics no less than here. Idealizations and approximations are rife in modelling, as discussed, for instance, in Hesse (1966), Gentner & Gentner (1983), Laymon (1990), Frigg & Reiss (2009), and Rowbottom (2009, 2011d). Often it is a sensible procedure to start simply and then modify slowly, as required, in order to render the model more realistic.

We may now define a scientist's work output potentials, which range from 0 to 1, as follows:

$I(T_1)_i CR(T_1)_i$

$I(T_1)_i PZ(T_1)_i$

$I(T_2)_i CR(T_2)_i$

$I(T_2)_i PZ(T_2)_i$.

The subscripts represent to whom each property belongs. So $I(T_1)_i$ represents how industrious scientist *i* will be when working on theory T_1, $PZ(T_2)_i$ represents scientist *i*'s puzzle solving skill with T_2, and so forth.

The work potentials, so defined, have several intuitively appealing features. Let's consider $I(T_1)_i CR(T_1)_i$ for illustrative purposes, and restrict the discussion to work on criticising T_1. The possible values of this potential range between zero and unity; unity corresponds to the maximum output potential possible for a scientist (which one could index to the best scientist available, or define in terms of the ideal scientist), whereas zero corresponds to no output potential. If the scientist's industry property—$I(T_1)_i$—is zero, then the work potential will be zero irrespective of her skill; no effort has no results. But if the industry property is positive, by contrast, then the work potential will be in direct proportion to the scientist's skill. Another interesting feature is as follows. Let one scientist be half as skilled but twice as industrious as another. On such a model, each scientist will have the same work output potential. (If this feature is not desired, it is easy to alter the function appropriately. For example, a threshold of skill, below which the function's value is zero, may be introduced. This fits with the idea that a very low skilled worker won't produce anything of value even if he works exceedingly hard.)

From the work output potentials of a scientist, we know where she would be best placed from the point of view of maximising her personal output. One thing we might already want to say on the prior assumptions—at least, if we add an additional plausible rule that *we want to maximise the total work performed*—is that we do not want to assign a scientist a task for which she has a zero work output potential. If we relax some of the prior assumptions, this may change. For example, the presence of an unproductive scientist in a research group might have a strong positive effect on the work output potentials of the other scientists therein. Thus there may be a net gain in including him in the group, even if he contributes little directly.

We can now imagine outlining finitely many possible distributions of labour, the number of which will depend on the number of scientists available. In each case, we can sum the personal work output potentials for each available task, to see how much work we can expect to be done in criticising T_1 ($WCRT_1$), in puzzle solving using T_1 ($WPZT_1$), in criticising T_2 ($WCRT_2$), and in puzzle solving using T_2 ($WPZT_2$). In mathematical notation, we have:

$$\mathrm{WCRT}_1 = \sum_{i=a}^{b} I(T_1)_i CR(T_1)_i$$

$$\mathrm{WPZT}_1 = \sum_{i=c}^{d} I(T_1)_i PZ(T_1)_i$$

$$\mathrm{WCRT}_2 = \sum_{i=e}^{f} I(T_2)_i CR(T_2)_i$$

$$\text{WPZT}_2 = \sum_{i=g}^{h} I(T_2)_i PZ(T_2)_i$$

where the group of scientists criticising T_1 is $\{a, \ldots, b\}$, the group of scientists puzzle solving with T_1 is $\{c, \ldots, d\}$, and so forth.

Now in general, a sub-set of the possible divisions of labour will come closest to obeying the rule:

$$\frac{WCRT_1 + WPZT_1}{WCRT_2 + WPZT_2} = \frac{R(T_1,T_2)}{1 + R(T_1,T_2)}$$

Alternatively, in terms of the more fundamental quantities, this rule may be expressed as:

$$\frac{\sum_{i=a}^{b} I(T_1)_i CR(T_1)_i + \sum_{i=c}^{d} I(T_1)_i PZ(T_1)_i}{\sum_{i=e}^{f} I(T_2)_i CR(T_2)_i + \sum_{i=g}^{h} I(T_2)_i PZ(T_2)_i} = \frac{C(T_1)}{C(T_1) + C(T_2)}$$

Indeed, there may be a unique balance between dogmatism and criticism that obeys, or comes closest to obeying, this particular rule. If there is not, there will be a range of allowable balances and a free choice between them.

In the next section, I want to consider how the model can be improved, and rendered more realistic, in line with some of the previous asides concerning the assumptions used above. In the final section, I will discuss how the situation with respect to rules is/should be much more complex.

However, the mechanism by which to determine the appropriate balance, even when further complexity is added, will be the same:

(1) Work out the ideal distribution of labour (with reference to data such as confirmation values, and other aspects of the state of inquiry).

(2) List all the possible ways that science could be set up.

(3) Use a function to select those closest to—or acceptably close to—the ideal.

3. Refinements to the Model: A Smörgåsbord

The model outlined above can be refined in numerous ways, many of which involve relaxing assumptions made in the previous section. In what follows, in a number of sub-sections, I will cover a variety of potential alterations that I consider to be of special interest in making the model more realistic. I will not be concerned, until later, with the *rules* of inquiry (such as the toy rules used previously); rather, I will be concerned with what the rules may concern. For example, I introduce the imaginative (or theory-construction) function because there is plausibly a balance that must be struck between this and the other two functions—criticism and puzzle solving—discussed above.

My aim is to focus on reasonably broad factors that will often be significant not only in science in general, but also in smaller research groups. The values of interaction coefficients (discussed in 3.3), for instance, may be determined by a wide variety of psychological considerations. But I do not delve deeply into such possible considerations; instead, I provide examples designed to illustrate how the output of some group may not be a simple sum of the output that each individual would give when working in isolation.

3.1. Additional Functions: Imaginative/Theory-Constructing

The previous treatment assumes that scientists only perform two different kinds of function. If science were truly so limited, however, then it is plausible that little progress, if any, would be possible. There would be no means (within science itself) by which to generate new theories. Scientists would be stuck with using theories generated by others, e.g. folk theories.

This is why I have previously proposed, following Popper, that something akin to an imaginative function is crucial for science (Rowbottom 2011a). But it is important to understand this broadly. If preferred, this may be thought of as merely a *theory-constructing* function; and then we may leave open the question of the extent to which imagination is involved in this creative activity, which may prove somewhat divisive. I will continue to use the phrase ‘imaginative function’, however, for reasons of consistency.

Such a function may be represented by IM. Hence, in line with the prior treatment of CR and PZ, we may say that $IM(T_j)_i$ represents the imaginative skill of scientist *i* with respect to theory *j* (and multiplying this value by $I(T_j)_i$ will give a work potential). Naturally this does not, however, represent the ability of scientist *i* to create *j*; rather, it represents her skill when it comes to generating possible *replacements* for *j*. This skill need not, and plausibly should not, be construed merely as reflecting the ability to generate a particular *quantity* of potential replacements. The quality of the theories a scientist is able to generate, measured in terms of supra-empirical theoretical virtues such as simplicity and scope, may also be significant. (Accuracy and fruitfulness cannot generally be ascertained beforehand; the other functions explore these.[12])

Note that defining imaginative skill in such a way is consistent with adopting the view that variations in theory-construction ability are only domain specific. That's to say, we may imagine that $IM(T_j)=IM(T_k)$, for any given scientist, if theories *j* and *k* are in the same area of science, e.g. fluid mechanics or quantum chemistry. I leave open the question of whether this is so, and of how large the appropriate areas should be understood to be if it is.[13]

Some will think that there are further central functions in science—beyond CR, PZ, and IM—whereas I do not. I will not dwell on this, however, since I trust it is easy to see how further functions might be introduced.

---

[12] I use 'fruitfulness' in the sense of Kuhn (1977, p. 322): 'a theory should be fruitful of new research findings: it should, that is, disclose new phenomena or previously unnoted relationships among those already known'.

[13] My own view is that variations are often more subtle than this, and may depend on differences in internal features of the theories, e.g. the mathematics involved.

3.2. Sub-Functions: Offensive, Defensive, Evaluative, Classificatory/Predictive, Aligning, and Articulating

To say that two scientists are both puzzle solving (or contributing to the puzzle solving effort) is not to say that they are both doing precisely the same kinds of thing, because puzzle solving can involve a range of different activities. And the same is true, *mutatis mutandis*, for performing a critical function (or contributing to the critical effort). So if we think of the functions discussed up to this point as *primary*, we may also say that there are *sub-functions* that serve to constitute them. To use a simple example: to devote all one's effort to attacking a specific popular scientific theory may make no less a critical contribution to science than devoting all one's effort to evaluating the relative strengths and weaknesses of the competing theories in a specific area of science (such as the biomechanics of biped locomotion).

So what kind of sub-functions might there be? I have already mentioned two: offensive and evaluative. And if we add to this a defensive function, then I believe we have all the central aspects of critical activity. Let's denote the skill in performing each sub-function, for some scientist *i* with respect to theory (or group of theories) *j*, as follows: $OF(T_j)_i$; $DF(T_j)_i$; and $EV(T_j)_i$. As with the imaginative function, discussed above, it may be helpful to think in terms of areas of science in order to compare the values for these. For example, it is natural to think that $EV(T_1)_i = EV(T_2)_i$ for any *good* evaluator if T_1 and T_2 are competing scientific theories.

I take each of these activities to be reasonably self-explanatory, and discussed them in the parent paper (Rowbottom 2011a), so will not dwell too much on explaining them.

I will draw attention only to two things. First, those performing each kind of activity will often interact (even if only indirectly, i.e. by engaging with one another's work) in significant ways. Those defending will seek to address criticisms made by attackers, just as attackers will seek to challenge defensive work. Evaluators will try to take a balanced view on the state of the debate (whether or not they are also participants in it), with careful attention to the arguments advanced on either side. Good evaluators will succeed in doing so.

Second, those performing these sub-functions will be relying on the work of puzzle solvers. The success of puzzle solvers may be taken to reflect not merely on their own competence (*contra* what Kuhn (1962) suggests is proper in normal science), but also, to some extent, on the quality of the theories that they employ. (The estimated expertise of the relevant puzzle solvers can be taken into account in making a judgement.) If T_1 and T_2 are competing (mutually exclusive) theories, and the former can be used to solve a puzzle that the latter cannot (in the current state of play), then this may tell in the former's favour (e.g. if equal work has been done on trying to solve the puzzle with each). Indeed, some puzzle solving activity may be understood as fulfilling a testing role, even if it is not intended for that purpose, as we will see below.[14]

We are left with the sub-functions of puzzle solving. Following Kuhn (1962, chapter 3), I take these to be three in number: classificatory/predictive, aligning, and articulating. We can denote the skill in performing each of these sub-functions, for

[14] Nevertheless, there are grounds for thinking that actively seeking to refute theories is important. One reason is that this may serve to prevent the use of *selective* evidence (e.g. which focuses merely on confirming instances). See Rowbottom (2011b, pp. 94–95).

some scientist *i* with respect to theory (or group of theories) *j*, as follows: $CP(T_j)_i$; $AL(T_j)_i$; and $AR(T_j)_i$.

I have discussed each elsewhere—see Rowbottom (2011c)—so will here give only an overview. Classification involves the collection of data that is considered important on the basis of the dominant theories of the day; this could be anything from the spectrum of a star to the Young's modulus of a newly synthesized material. Prediction is a key role for this data (in principle if not always in practice). The spectrum of a star may be used (in conjunction with other data) to predict how far away it is from Earth. The Young's modulus of a newly synthesized material serves to predict what can be built with it. And so forth. (The case could be made that a scientist may be good at prediction but bad at classification, in some domain, or *vice versa*. This can easily be dealt with by splitting this sub-function into two component parts.)

Alignment involves 'fitting theories to facts', or more precisely extending the applicability of existing theories. A good way to understand this—although not explicitly Kuhn's—is in terms of model development. One begins with a relatively simple model, such as that of the simple pendulum, and then seeks to refine it by reducing the degree of idealization. The mass of the rod bearing the bob and the friction at the bearing may be taken into account, for instance. Naturally, exactly when one is generating a new model rather than a new theory may be somewhat murky, but we need not concern ourselves with this directly.

Articulation, the final activity, is somewhat more nebulous in character. One clear aspect is the measurement of central constants in theories, such as the gravitational constant, the permittivity of free space, and Planck's constant. However, articulation may also involve experimental attempts to explore ambiguous aspects of the dominant paradigm (*qua* disciplinary matrix), and more particularly the theories therein. In Rowbottom (2011c), I suggest that the EPR experiments may be cast in this light; what it means for variables to be non-commuting is investigated (if we try to frame matters in a way that doesn't presuppose realism). If this is incorrect, however, then we may think of articulation as considerably narrower than Kuhn suggests.

These activities are just as intertwined as their critical counterparts. Consider, for instance, how developing a better theoretical model of actual pendulums (alignment) may aid in geophysical measurements of *g* (classification), and how these better measurements might in turn help in determining G more accurately (articulation). Alternatively, rather more in the abstract, consider how classification and prediction can reveal new practical possibilities that enable the construction of new instruments. These instruments may in turn lead to developments in all areas of puzzle solving.

Progress in one activity may therefore depend on the products from another activity. And this brings me to a point that I failed to emphasize as much as I should have in the parent paper: namely, that *the correct way to maximize productivity with respect to one area/function, over an extended period, may not be to maximize the amount of work done in that area*. Needless to say, this goes for sub-functions as well as (so called) primary functions. Critical activity may sometimes be superior than it would otherwise be, for instance, when it draws upon recent (negative or positive) results in puzzle solving. In fact, diachronically speaking, I believe that most pairs of functions or sub-functions are symbiotic. Ideally, this should be factored in when generating balancing rules. (An alternative approach is to make total output from an area dependent on outputs from elsewhere, and not just work done directly in the area. There are a variety of mathematical ways that this could be done.)

There is one final question we should tackle before moving on. How do skill values for sub-functions relate to skill values for the functions of which they are a part? There are several modelling possibilities. One option is to understand the skill values for functions as intervals, with minima and maxima equal to the lowest and highest values of the relevant sub-functions. (So if the highest value of a critical sub-function for some scientist is unity, and the lowest value of a critical sub-function is one half, the interval will be between unity and one half.) A precise value to be used in calculations may be determined by direct reference to the active sub-function in some (actual or hypothetical) scenario. That is, when possibilities concerning division of labour are enumerated in such a way as to pay attention to the sub-functions being performed.

Another option is to fix function skill values to the highest value of the relevant sub-functions. This may be helpful when performing calculations concerned simply with balancing at the level of the functions (rather than their components). And one can imagine other alternatives, such as looking to the lowest values of the sub-functions. There is no 'right and wrong' here; it is simply about what one wishes to calculate. Since to introduce sub-functions is to admit that they are more fundamental than the functions of which they are a part, on an individual level, there is a sense in which talk of higher-level individual functions may be an artifice. Alternatively, the critical and puzzle solving functions may be *emergent*. And here is a final idea on this topic:

we may think that skill values for functions are properties of groups of scientists, rather than individuals, whereas individual skills concern sub-functions. This fits well with the introduction of interaction factors, as discussed below.

3.3. Interaction Coefficients

To determine how well a group will perform some task is not simply a matter of considering how well each member of the group would perform the task if working alone. We know this all-too-well from team sports, such as football. A team composed of elite players may be defeated by another team with significantly less able players, but superior teamwork. And the same is true in science; the skill of a group is not, in general, an additive function of the skills of its members.

In the first instance, one might therefore consider assigning a teamwork value, or some such, to each scientist. This would be misguided, however, because the relation of working well together is not transitive. To be more specific, let Mr. A work well with Mrs. B, and Mrs. B work well with Ms. C. It does not follow that Mr. A works well with Ms. C! The reasons may be rather mundane. Perhaps Mr. A finds Ms. C very attractive, and becomes flustered in her presence. The situation might be even worse, for science but not for love, if Mr. A's feelings are reciprocated!

It therefore appears that we should introduce an interaction factor for each possible group, which may be used when calculating its total work output. In general, any set of scientists, $\{1, \ldots, n\}$, will have an interaction coefficient which we can denote by

η_1^n; and we may allow such a coefficient to take any positive value, in principle.[15] We may then multiply our original expressions for total work outputs by these coefficients. For example, our expression for $WCRT_1$ for some group {*a*, … , *b*) will become:

$$WCRT_1 = \eta_a^b \sum_{i=a}^{b} I(T_1)_i CR(T_1)_i$$

Naturally, it is possible to understand the value of such a group interaction coefficient as dependent on a number of personal interaction coefficients, i.e. to think of the output of each and every scientist as being affected differently, in some specific way, by their group membership. And one might specify these if a more complex and sensitive model is desired.

Matters can be further complicated by the recognition that such interaction coefficients may be activity specific, e.g. that a group of scientists may work well together when performing one sub-function, but poorly together when performing another. The reasons may be many and varied. Some scientists may get bored when doing some things, and then behave in ways that distract others. Some scientists may be more prone to work avoidance, e.g. surfing the net or reading e-mails, when working alone rather than in close collaboration with others; and performing some sub-functions may involve close collaboration, whereas performing others may not.

3.4. Temporal and Experiential Variance

[15] Negative values may also be introduced, for similar reasons as those suggested in the final sub-section.

It is also possible to allow the properties of scientists to vary across time and in response to experience, i.e. for values such as $OF(T_j)_i$ to be functions of time and experience. Let's consider the career of a hypothetical scientist for illustrative purposes. In her youth, she may be daring and adventurous; she may be highly imaginative and a sharp critic (especially in attack). But as she ages, she may mellow. She may become rather more set in her ways, i.e. less willing to entertain alternatives to the theories that she has spent so much time working with. As a result, her imaginative and offensive skills (in her specific area of science) may diminish. But in their place, more refined defensive and puzzle solving skills—again, with her preferred theories—may emerge. Thus she may be just as talented towards the twilight of her research career as she was when she was beginning, but in rather different respects.

The lesson is that it is valuable, when considering what best to do from a diachronic perspective, to think not only about what effect a scientist will have if she performs some particular task, but also about what effect performing that task will have on the scientist.

3.5. Other Refinements

Many other refinements are possible. But I should like to mention just two more possibilities. The first, which may sometimes be an alternative to using interaction factors, is to allow for negative values for skills (and to make these skills context sensitive). The underlying idea is that some scientists may be rather worse than totally unproductive! A scientist working in an area in which they had a negative skill value would be (actively) detrimental to that area.

The second notion is that the extent to which an individual is industrious may be task relative. This could be handled relatively easily, simply by making I values function (or even sub-function) specific: I_{CR}; I_{PZ}; I_{IM}; etc. A critical work output potential for scientist *i* relative to theory *j*, for example, might then be: $I_{CR}(T_j)_i CR(T_j)_i$. One significant reason underlying differences in industry will be motivation; and motivation may be affected by rewards systems. This provides an obvious place to link the present modelling approach to the work of, say, Strevens (2003).

4. Methodological Rules for the Distribution of Labour

Next, I should like to say a little about the ways that methodological rules might be employed in calculating the appropriate distribution of labour. First and foremost, I should like to emphasise that to understand science as properly guided by methodological rules is not to commit oneself to the view that there is a single list of rules that operate concurrently and continuously. Indeed, the problem with thinking of rules in such a way—as one might, for example, understand the Ten Commandments of the Old Testament—is that situations may occur where obeying one requires violating another (and vice versa), even if they can typically be obeyed simultaneously. For example, to honour my mother I may have to lie to her when she asks if she is fat. To respond to her question truthfully, on the other hand, would result in dishonouring her!

A partial solution to this kind of problem is to have an explicit priority order for the rules. In fact, this is implicit in my earlier mention of the idea that *we want to maximise the total work performed as well as to ensure that work done on each theory is in proportion to its relative degree of confirmation*. One way of handling this would be to say that when the possible distributions of labour with the correct proportions have been enumerated, in any given calculation, the preferred possibility should maximise the total work performed.

More subtle approaches are also possible. Perhaps sometimes maximising the total work performed is more important than having the correct proportions (and vice versa). So one may express priorities using conditional clauses, such as: if only approximately having the correct proportions rather than having precisely the right proportions can lead to more than *n* times as much work output, then select a distribution of labour that approximately has the correct proportions and maximises work output. (Note this is quite different from declaring a general rule: select the distribution of labour that approximately or precisely has the correct proportions and maximises work output.) One may go on to specify what counts as 'approximately having the correct proportions' by stating a specific fraction by which the actual proportions may deviate from the ideal proportions.

Somewhat more radically, it is possible to think of the set of applicable methodological rules as being entirely context dependent. Priority considerations may then be sunk into the rules; e.g., to follow on from the previous example, the rule in some contexts may be 'Select the distribution of labour which has approximately the correct proportions and maximises work output', whereas the rule in others may be

'Select the distribution of labour which has precisely the correct proportions and otherwise maximises work output'. Alternatively, this could be thought of as a situation where a single meta-rule applies. This would look something like this: In context class C, apply rule-set R, with priority ordering PR; in context class C_1, apply rule-set R_1, with priority ordering PR_1; and so on.

This may all seem terribly messy. But the reasons to make methodological considerations messy are compelling. Critiques of the very idea of any 'scientific method', such as Feyerabend's (1975), appear so effective only because they attack a naïve approach to the notion, namely one where there are context-invariant rules (like the aforementioned Ten Commandments). If we instead allow that the rules can change radically across contexts of inquiry, then the power of the historical examples marshaled by Feyerabend is diminished. To be specific, we may understand his work as helping to delineate the proper contexts of applicability of the rules he discusses. We may avoid the retreat to relativism. In a nutshell, we may agree that 'anything goes' (or accept that 'almost anything goes') across all contexts of inquiry, but deny that 'anything goes' (or 'almost anything goes') in any specific context of inquiry.[16]

5. Conclusion: Bounded Rationality and Striking the Balance

[16] By 'context', I mean something more fine-grained than a paradigm *qua* disciplinary matrix; for even within the boundaries of 'normal science', accepting that something like this exists, there will be times at which it is proper to devote more puzzle solving effort in one direction rather than another. See Rowbottom (2011c). Thus by a context I mean not only available theories, instruments, and personnel, but also other resources (e.g. financial) and research findings (e.g. classifications and reports on tests of models).

So what is the result of this paper? We have an overview of the kind of factors to build into an elementary (social epistemological) model of a scientific community, further factors that might be introduced to make it less idealized, and how the rules of method to be employed might be understood in a context dependent fashion. But this leaves the big question of whether such a model is implementable, say at the level of a research group (rather than the whole of science), and hence of any practical value whatsoever. In closing, I will tackle various aspects of this issue.

First and foremost, the complexity required for a reasonably realistic model—both with respect to relevant factors in the situation, and the methodological rules to be employed—is considerable. Hence, I think that there will be few situations, at best, in which solving balancing problems (even assuming the accuracy of the data) will be a pen and paper exercise. Rather, I believe that computer simulations might be a suitable tool. Indeed my interest in the modelling approach outlined here is partly as a result of the increasing use of such simulations in social epistemology.[17]

Second, however, there is a more general question. Is finding an *optimal* solution to balancing problems possible even when computational problems are disregarded? Or is our rationality bounded in such a way, following Simon (1955), that this expectation is entirely unrealistic? One worry here, in particular, is that any computer model we build will only ever reflect the limited choices that we are able to consider, and that simulations will only be run on the data that we are able to gather (or automate the gathering of); so that in a special sense, *qua* problem solvers rather than

[17] See, for example, Hegselmann and Krause (2006), Zollman (2007), Douven (2009), Weisberg and Muldoon (2009), and Olsson (2011).

calculators, computers are only ever as good as their programmers.[18] My response, in due humility, is simply to admit the validity of this concern. Perhaps, indeed, optimal solutions are generally—or, rather more optimistically, typically—beyond our grasp. But this is compatible with the modelling strategy advocated here being sufficient for achieving less troublesome tasks, such as providing solutions that are sufficiently *proximal* to optimal to be better than those we could otherwise (reliably) arrive at. And as an empiricist, I think we should find out what these models can do, providing this is not prohibitively expensive, before writing them off.

Of course, some might think that the whole approach is wrong-headed; that doing social epistemology by considering fundamental factors, and then adding more in order to make higher-order corrections, is undesirable even if possible. (Ideally we shall want to consider the finances and equipment available, *inter alia*, for example.) Ditto for considering multiple rules of inquiry, and then meta-rules concerning when those should be applied. Perhaps it would be better to use heuristics, which can be accurate enough to be fit for purpose (on a case by case basis) but much more economical? This line of objection is inspired by Wimsatt (2007). My response is twofold. On the one hand, there may very well be some situations in which accuracy is of considerably greater import than economy, and in which the extra expense of employing a computer simulation is acceptable. (That is to say, the cost-benefit analysis *may* favour heuristics in most cases. But it does not follow that it does in all. This is an empirical question to which I do not wish to guess the answer; my sample

[18] Simon (1955, p. 99) aims at an account of 'rational behavior that is compatible with the access to information and the computational capacities that are actually possessed by organisms, including man, in the kinds of environments in which such organisms exist'. Here we are concerned not with 'computational capacities', which we have seen can be expanded by the use of computers, but with 'access to information'.

size is too small and selective.) On the other, experimentation *in silico* may be of service in the development of new, and better, heuristics for organizing research. It may disclose previously unnoticed correlations between variables, the existence of which we may corroborate via socio-historical studies, for instance. So in summary, I do not want to engage in a methodological dispute at this level. I simply think that we should investigate the avenue, provided that the opportunity cost is not too great. It may be a dead end, of course. Such is the nature of research.

Third, however, a more specific criticism of the proposed modelling strategy is that many of the factors mentioned above—e.g. skills, such as problem solving, and dispositions, such as industry—are not operationalised. Hence, one might think it is unclear about how they could be used in any practical model, computational or otherwise. I admit that there is some justice to this worry. But I think legitimate disagreements over such measurement issues are natural, and I do not want to become embroiled in them in the present context (because doing so would lead to lengthy digressions). For example, one might measure a scientist's industry by looking at how many papers she produces, or how many experiments she manages to perform, in a complexity-weighted fashion. Different forms of measurement will have different advantages, and clearly some will be appropriate in some cases whereas others will not. (Measuring industry by looking at experiments performed is hopeless when one is interested in a theoretician, for example. This is just common sense.) So I prefer just to leave matters open at this level; the important point is that the approach is flexible enough to cope with different forms of measurement. Only if one takes a strong operationalist stance like (the early) Bridgman (1927, p. 5), where 'we mean by any concept nothing more than a set of operations; the concept is synonymous with the corresponding set of operations' might one see this as a deep problem. In effect, this would mean that terms like 'industry', and related ones like 'hardworking' or 'lazy', may fail to be meaningful; and, more precisely, that I am presently unable to show that they are meaningful by specifying the sets of operations with which they are synonymous (according to linguistic convention).[19]

In summary, in this piece I provide the basis for a research programme. I can provide no guarantee that it will prove fruitful, for I am no oracle. But I also see no reason to doubt that we can do somewhat better than we presently do, when it comes to considerations of scientific method, by giving this programme a go. To be frank, we could hardly do much worse. It is hardly as if we have some well-developed and highly successful approach—as against the overgeneralizations of Kuhn and Popper—that it is intended to replace!

Acknowledgements

Many thanks to Paisley Livingston for several sharp comments on the initial version of this piece. I am also grateful to two unusually diligent referees for insightful criticism of the version initially submitted.

References

[19] Nonetheless, a reader with such a strong operationalist stance may feel able to show what I do not believe I can, or should, show. So she need not reject the approach outlined here. As Chang (2009, section 2.2) notes, moreover, even Bridgman appeared to admit at points that 'All that was required was that the theoretical system touched the operational ground somewhere, eventually.' See also Gillies (1972).

Bridgman, P. W. 1927. *The Logic of Modern Physics*. New York: Macmillan

Canfield, J. and K. Lehrer. 1961. 'A Note on Prediction and Deduction', *Philosophy of Science* 28, 204–208.

Cartwright, N. 1983. *How the Laws of Physics Lie.* Oxford: Oxford University Press.

Chang, H. 2009. 'Operationalism', in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy,* URL: http://plato.stanford.edu/archives/fall2009/entries/operationalism/

Douven, I. 2009. 'Introduction: Computer Simulations in Social Epistemology', *Episteme: A Journal of Social Epistemology* 6, 107–109.

Earman, J. and J. Roberts. 1999. 'Ceteris Paribus, There is no Problem of Provisos', *Synthese* 118, 439–478.

Feyerabend, P. K. 1975. *Against Method: Outline of an Anarchistic Theory of Knowledge*. London: Verso.

Frigg, R. and J. Reiss. 2009. 'The Philosophy of Simulation: Hot New Issues or Same Old Stew?', *Synthese* 169, 593–613.

Gentner, D. and D. R. Gentner. 1983. 'Flowing Waters or Teeming Crowds: Mental Models of Electricity', in D. Gentner and A. Stevens (eds), *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum.

Gillies, D. A. 1972. 'Operationalism', *Synthese* 25, 1–24.

Hegselmann, R. and U. Krause. 2006. 'Truth and Cognitive Division of Labour: First Steps Towards a Computer-Aided Social Epistemology', *Journal of Artificial Societies and Social Simulation* 9.

Hempel, C. 1988. 'Provisoes: A Problem Concerning the Inferential Function of Scientific Theories', *Erkenntnis* 28, 147–164.

Hesse, M. 1966. *Models and Analogies in Science*. Notre Dame: University of Indiana Press.

Huber, F. 2008. 'Milne's Argument for the Log-Ratio Measure', *Philosophy of Science* 75, 413–420.

Kitcher, P. 1990. 'The Division of Cognitive Labour', *Journal of Philosophy* 87, 5–22.

Kitcher, P. 1993. *The Advancement of Science*. New York: Oxford University Press.

Kuhn, T. S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kuhn, T. S. 1977. *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press.

Laymon, R. 1990. 'Computer Simulations, Idealizations and Approximations'. In *PSA: Proceedings of the biennial meeting of the philosophy of science association* (Vol. 2, pp. 519–534).

Milne, P. 1996. 'Log[P(h/eb)/P(h/b)] Is the One True Measure of Confirmation', *Philosophy of Science* 63, 21–26.

Muldoon, R. and M. Weisberg. 2011. 'Robustness and Idealization in Models of Cognitive Labor', *Synthese*, 183, 161-174

Musgrave, A. 1974. 'Logical Versus Historical Theories of Confirmation', *British Journal for the Philosophy of Science* 25, 1–23.

Musgrave, A. 1976. 'Method or Madness', in R. S. Cohen (ed.), *Essays in Memory of Imre Lakatos*, 479–482. Dordrecht: D. Reidel.

Olsson, E. J. 2011. 'A Simulation Approach to Veritistic Social Epistemology', *Episteme: Journal of Social Epistemology* 8, 127–143.

Popper, K. R. 1983. *Realism and the Aim of Science*. London: Routledge.

Rowbottom, D. P. 2009. ‘Models in Physics and Biology: What’s the Difference?’, *Foundations of Science* 14, 281–294.

Rowbottom, D. P. 2011a. ‘Kuhn vs. Popper on Criticism and Dogmatism in Science: A Resolution at the Group Level’, *Studies in History and Philosophy of Science* 42, 117–124.

Rowbottom, D. P. 2011b. *Popper’s Critical Rationalism: A Philosophical Investigation*. London: Routledge.

Rowbottom, D. P. 2011c. ‘Stances and Paradigms: A Reflection’, *Synthese* 178(1), 111–119.

Rowbottom D. P. 2011d. ‘Approximations, Idealizations and “Experiments” at the Physics–Biology Interface’, *Studies in History and Philosophy of Biological and Biomedical Sciences* 42, 145–154.

Rowbottom, D. P. In Press A. ‘Popper’s Measure of Corroboration and P($h|b$)’, *British Journal for the Philosophy of Science*.

Rowbottom, D. P. In Press B. ‘Information Versus Knowledge in Confirmation Theory’, *Logique et Analyse.*

Salmon, W. C. 1990. 'Rationality and Objectivity in Science or Tom Kuhn Meets Tom Bayes', in C. W. Savage (ed.), *Scientific Theories*, 175–204. Minneapolis: University of Minnesota Press.

Simon, H. A. 1955. 'A Behavioral Model of Rational Choice', *The Quarterly Journal of Economics* 69, 99–118.

Strevens, M. 2003. 'The Role of the Priority Rule in Science', *Journal of Philosophy* 100, 55–79.

Weisberg, M. and Muldoon, R. 2009. 'Epistemic Landscapes and the Division of Cognitive Labour', *Philosophy of Science* 76, 225–252.

Williamson, J. 2010. *In Defence of Objective Bayesianism*. Oxford: Oxford University Press.

Wimsatt, W. C. 2007. *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge, MA: Harvard University Press.

Zollman, K. J. 2007. 'The Communication Structure of Epistemic Communities', *Philosophy of Science* 74, 574-587.