Robustness and Idealization in Models of Cognitive Labor

Ryan Muldoon
Michael Weisberg

University of Pennsylvania

Introduction

Scientific research is almost always conducted by communities of scientists of varying size and complexity. Such communities are effective, in part, because they divide their cognitive labor: not every scientist works on the same project. Scientists manage to do this without a central authority allocating them to different projects. Thanks largely to the pioneering studies of Philip Kitcher[1] and Michael Strevens[2], understanding this self-organization has become an important area of research in the philosophy of science.

One fruitful way to study how scientists divide their cognitive labor is by constructing and analyzing mathematical models of the social structure of science. Such analyses can help us understand what an optimal distribution of cognitive labor would be and how scientists can organize themselves to promote such a distribution. In order to derive these benefits, models of cognitive labor must be simple enough for us to understand their dynamics, but faithful enough to reality that we can use them to analyze real scientific communities. To satisfy the first requirement, we must rely on idealizations to reduce the complexity found in real-world scientific communities. However, the second requirement demands that these idealizations not be so extreme that we lose the ability to explain the actual benefit of divided cognitive labor. These two requirements are not impossible to satisfy simultaneously, but

[1] See Chapter 8 of *The Advancement of Science* (New York: Oxford, 1993) and "The Division of Cognitive Labor" Journal of Philosophy, LXXXVII, 1 (January 1990): 5-22
[2] "The Role of the Priority Rule in Science" Journal of Philosophy, C, 2 (February 2003): 55-79

they do impose significant restrictions on model choice. In particular, they suggest that when idealizing assumptions are made for the sake of simplicity, it is incumbent upon the modeler to show that these assumptions are either approximately true or that they themselves do not drive the major results of the model.

Kitcher and Strevens attempt to balance realism and tractability by constructing representative agent models of cognitive labor that employ what we call the *marginal contribution/reward* (MCR) approach. This approach has been fruitfully employed in economics and seems like a reasonable place to begin studying cognitive labor. However, we will argue that at least two assumptions of the MCR approach are neither approximately true of the scientific community nor are models containing them robust against perturbations of these assumptions. On this basis, we conclude that the MCR approach as developed by Kitcher and Strevens is not fully sufficient for analyzing the division of cognitive labor.

The MCR Approach and its Assumptions

According to Kitcher and Strevens, the problem of optimally distributing cognitive labor is equivalent to a resource allocation problem. There is a certain good (cognitive labor) that can be allocated among different consumers (scientific projects). Each project has a marginal utility curve, called a *return function* by Kitcher and a *success function* by Strevens, which represents the ability of the project to productively utilize the cognitive resources of scientists and turn those resources into the possibility of a successful outcome. Such functions take the number of contributing scientists as an input, and output the probability of the project being successfully completed. When the success functions are considered simultaneously and the number of agents is known, the optimal distribution of scientists to projects can be calculated. In economics, the procedure for solving such a problem is called constrained maximization.

One of Kitcher's and Strevens's main arguments is that classic epistemic norms for individuals are likely to cause scientists to misallocate themselves across projects. Consider the following very simple example: There are two possible approaches to synthesizing some new chemotherapeutic molecule. The first approach has a high probability of success when a reasonable number of scientists are working on the project. The second approach has a relatively low probability of success, but this probability of success can be realized if a small number of scientists work on the project. If every individual scientist followed a classical epistemic norm such as "take the approach most likely to lead to the truth," no scientist would choose the second approach. Yet the scientific community would be better off if a small but significant number of scientists chose the second project because of the possibility that the first approach would not be successful no matter how many scientists worked on it. The optimum distribution of cognitive labor thus deviates from what classical epistemic agents would choose.

Kitcher and Strevens further argue that the optimal division of cognitive labor might be achieved when scientists act according to self-interest instead of epistemic norms. This is accomplished by means of community-level reward schemes that assign money or credit to scientists working on whichever program happens to be successful in the end. For example, in the scheme Strevens calls *Marge*, scientists are rewarded according to the marginal contribution that they make to their project's probability of success. If the community followed this reward scheme, individual scientists determine their expected reward by calculating their marginal contribution to the probability of the project's success and multiplying this by the total reward.

When scientists maximize their chance at reward, as opposed to their chance of working on the successful project, then they will more closely approximate the optimal community distribution of workers to projects. Strevens further argues that the *Priority* reward scheme, where the first scientist to successfully complete the project gets the whole reward, pushes the community of scientists to a nearly optimal distribution.

At the core of MCR models is a procedure by which scientists calculate their expected rewards. Individually, each scientist calculates his or her marginal contribution to the probability of a project's success and then uses this information to calculate his or her expected reward. MCR models can differ in their reward schemes, their success functions, the maximum probabilities of projects' success, and so forth, but they all embody this core assumption.

To instantiate the MCR approach in specific models, a number of further idealizing assumptions are made by Kitcher and Strevens. They first assume that scientists are utility-maximizers, responding rationally to the established incentive system. Second, they assume that the division of cognitive labor can be represented as a choice amongst a number of pre-defined projects. Third, every scientist knows the distribution of cognitive labor before she chooses what project to work on.[3] We call this the *distribution assumption*. Finally, each project has a success function, which takes as input units of cognitive labor (work from scientists) and outputs objective probabilities of success. The *success function assumption* is the assumption that these functions are known by all of the scientists in the model. In the remainder of this paper we will show that when either the distribution assumption or the success function assumption is relaxed, MCR models misallocate scientists to projects.[4]

Methodology of Evaluation

Evaluating the distribution and success function assumptions involves constructing more flexible MCR models than the ones offered by Kitcher and Strevens. In their own versions of MCR, Kitcher and Strevens use a representative-agent approach, where the calculations are done from the point of view of a single agent who is assumed to be in exactly the same situation as all the others. In

[3] Strevens (pp. 64-65); Kitcher 1990 (pp. 10-12, 14)

[4] Following Kitcher, Bill Brock and Steve Durlaf developed a similar economic model of the distribution of cognitive labor in "A formal model of theory choice in science" *Economic Theory* XIV, (1999) 113-130. While they abandon "success functions", they do retain the distribution assumption. As a result, our first critique, though not our second, applies to their work as well.

order to evaluate the status of the distribution and success function assumptions, we could not make this assumption because we wanted to study what happens when scientists have different information at their disposal. We therefore adopted an agent-based approach[5], where every scientist is represented explicitly.[6] While this approach allows us to relax key assumptions, it comes at the cost of requiring us to study our models by computer simulation rather than symbolic manipulation of closed-form mathematics.

Since our models are agent-based, we were required to make a number of assumptions about the nature of individuals and projects that could simply be left abstract in Kitcher's and Strevens' models. The first assumption we had to make was about the exact form of the success functions. Since this term canceled out in Strevens' analysis, he left the form of the function abstract in order to gain greater generality. However, our models require that each scientist-agent actually calculate their marginal contribution to their project's success using a specific function. Following Kitcher, we used the logistic growth equation[7] for the success function. Let *K* be the maximum probability of success, *N* the number of scientists working on the project, and *r* the *easiness* of the project. Here "easiness" determines both how much cognitive labor is required to realize the maximum probability of success and to the marginal probability that each each new agent contributes. Easier projects require fewer cognitive resources. The probability that a project will be successful is calculated with the following function:

$$P = \frac{K}{1 + e^{-rN}}$$

---

5 A standard discussion of the advantages of agent-based modeling can be found in Thomas Schelling's classic book *Micromotives and Macrobehavior* (New York: Norton, 1978). More recent discussion of these issues can be found in John H. Miller and Page, Scott, *Complex Adaptive Systems* (Princeton: Princeton University Press, 2007).

6 All of the simulations discussed in this paper were developed in NetLogo 3.1.4. Code for the models, including the parameter sets used in this paper, is available from the authors upon request.

7 This deviates slightly from Strevens' requirement that success functions should have decreasing marginal returns, but nothing hinges on this difference – our conclusions also hold if we only consider the portions of the success functions that have decreasing marginal returns.

The second assumption that we made concerns the order in which agents make their decisions about which project to work on. Insofar as one can interpret MCR models as saying something about individuals, the procedure assumes that there is only one individual left to make a choice, and that everyone else has already allocated their labor appropriately. In our models, however, every agent has to decide which project to work on. This happens sequentially, but in a randomized order each time the simulation is run.

We made these first two assumptions in order to reinterpret MCR models into an agent-based framework, but a third assumption of our models introduces structure not found in Kitcher's and Strevens' models. In our models, agents are distributed in random locations on a torus of 35x35 units. This spatial dimension corresponds to communication distance, not physical proximity. We further defined a *radius of vision* within which agents can "see" the project choices of other agents. Agents inside the radius of vision of one another are within communication distance. When the radius of vision is greater than or equal to Ö578 units, the spatial structure is superfluous and all agents can see and communicate with one another, making our model essentially equivalent to MCR. Adjusting the radius of vision is the main tool by which we can relax the distribution assumption.

Our models are initialized with a specified number of agents randomly assigned to a project. In random order, each agent first determines what the other agents in its radius of vision are working on. It then calculates its marginal contribution to its current project as well as its potential marginal contribution to the other projects. On the basis of the payoff function, each agent then calculates its expected reward for each project and then chooses the project that maximizes its expected reward[8]. Since each agent follows this procedure in sequence, we repeat the procedure ten times to ensure that

[8] The expected payoff of working on a given project J with total payoff π is as follows: $EU_J = (P(J(n+1))^* \pi(J(n+1)) - (P(J(n))^* \pi(J(n))$, where J(n) is project J with n agents working on it. An agent chooses which project to work on by taking the MAX over all EU_J. For any social utility calculations, we assume that the payoffs combine additively.

the community of agents finds the equilibrium distribution[9].

When the radius is maximized and every agent has the same success function, we were able to replicate Kitcher and Strevens's results. For example, we examined how our model compares to Strevens' model when the agents chose between two projects that differed only in degree of difficulty (values for r). Using the *Marge* payoff function, we observed the following: With small numbers of agents, all chose to work on the easier project. As the number of agents was increased, an incentive was created for a minority of scientists to work on the harder project. When the number of agents was increased further, scientists allocated themselves to both projects, and eventually the number of scientists working on the harder project overtook the number working on the easier project. These are qualitatively[10] the same results one gets with Strevens' constrained maximization model, suggesting that we have successfully interpreted it without distortion. We next report what happens when the distribution and success function assumptions are relaxed.

The Distribution Assumption

The distribution assumption is the assumption that all the agents know the distribution of cognitive labor at all times. Because it seems extremely implausible that every scientist knows the full distribution of cognitive labor across all of science, we interpret this as the claim that all scientists in a particular research domain know what all the other scientists are doing in that domain. The projects in MCR models thus correspond to a single domain.

In some very small sub-fields, the distribution assumption might be approximately true,

---

9 This is an artifact of applying MCR to a concrete population of individuals: Ordering effects that are not relevant or possible at a high level of abstraction become possible with discrete decisions in a finite population. MCR models consider the decisions of the marginal agent, so the rest of the population has already made decisions. Several rounds are needed to ensure that each agent can be treated as the marginal agent. Without each agent performing this calculation, there is no way to see what the overall distributional effects of a given incentive system would be.

10 Qualitative match is the most we could aim for, as Strevens does not offer particular quantitative results to compare against. However, our results match the behavior he describes.

although in general we believe that it is an idealization.[11] It is unlikely that scientists know what all of the others in their subfield are working on because there are too many scientists, too many research programs, and too much physical distance between scientists to maintain the level of communication necessary for this knowledge. Scientists are much more likely to be informed of the work being carried out by colleagues in their own laboratories, colleagues in close proximity, and those with whom they have pre-established relationships.[12] Prima facie, this creates a problem because accurate information about how many scientists are working on each project is crucial for calculating the payoff of working on a project in the MCR scheme. To assess the extent of the problem, we report on a set of simulations where we systematically reduce the distribution information available to scientists.

In order to reduce the information each scientist possesses about the distribution of cognitive labor, we reran the *Marge* simulations described above, but systematically varied each agent's radius of vision from a full radius of vision (Ö578 or approximately 24.04 units) down to a radius of one unit, allowing them to "see" only those agents to which they are adjacent. To keep the analysis simple, we studied a 2-project model and fixed the number of total scientist-agents at 500. We kept the maximum probability of success and the total payoff equal for the two projects, but we made Project 2 harder than Project 1, meaning that more scientists are required to realize Project 2's probability of success.

[FIGURE 1 HERE]

Figure 1 summarizes the results of this study. As the radius of vision becomes smaller, meaning that each agent lacks information about what some of its colleagues are working on, the distribution of

---

[11] That is not to say that there are never instances in which this idealization holds true. While we affirm that Kitcher's motivating case of DNA was an instance of an appropriately small sub-field, and further agree that there are other similarly small and tight-knit research communities, we doubt the generality of these cases.

[12] For some empirical support of these claims, Mark Newman has done extensive study of scientific research networks. For example, see M. E. J. Newman, "The Structure of scientific collaboration networks." *Proceedings of the National Academy of Science*. *USA* 98, (2001): 404-409

cognitive labor begins to deviate from the optimal distribution, with increasing numbers of scientists working on the easier project. We observe a qualitative shift in the distribution of scientists when the radius is 12 units or smaller. At this point, a majority of scientists begin working on the easier project, which is a severe misallocation of cognitive labor given the total number of scientists that are available. As the radius of vision drops below 7 units, not a single scientist works on the harder project. Subsequent simulations reveal that these qualitative changes occurs under many different parameter sets, but the radius at which the qualitative shifts in distribution are observed depends on the population density. As the population of scientists is more densely packed on the torus, small samples become more representative of the whole population and it takes considerably smaller radii to affect qualitative shifts[13].

These simulations show that the first epistemic assumption—perfect knowledge of the division of cognitive labor—is crucial for generating the main result of MCR models. When communication between scientists is reduced and they are not in possession of perfect information about how many scientists are currently working on the projects, self-interested choices do not necessarily lead to an optimum division of cognitive labor. In fact, when this information is limited considerably, all scientists may chose to work on the easier project, or the project with higher payoff. Thus, a major result of MCR models is not robust to changes in the distribution assumption.

Before analyzing the mutual knowledge assumption, it is worth considering how MCR models might be altered to accommodate this lack of robustness. One possibility is that each agent could treat the distribution of cognitive labor that it sees as a representative sample of the larger population. This seems like a sensible enhancement of the model because it probably is how scientists actually do try to assess the current division of cognitive labor.

[13] This is only true so long as the initial distribution of projects to agents is random, with a uniform distribution. In all of our simulations this condition is maintained. If this condition were weakened to any given distribution, the problems for MCR become worse, as localized sampling would be biased.

Despite moving the model in the direction of greater realism, we do not think this will solve the problem we have discussed above. For one thing, any sample the agent sees is unlikely to be representative because local clusters of scientists such as lab groups and research units are usually composed of people who are working on the same or similar problems. Similarly, scientists tend to talk about research more often with those that have similar interests than with random members of the scientific community. For both of these reasons, any sample the scientist sees is likely to be biased.

Perhaps these problems could be overcome and a non-biased sample could be taken, but there is an even more significant problem for the sampling approach. The MCR approach requires that scientists know the actual number of scientists working on each project, not just the proportional distribution. To find this, scientists would have to take their sample and scale it up with reference to the total number of scientists in a research domain. But there seem to be no mechanisms by which an average scientist can determine the size of such a research community. While the membership in professional societies or numbers of conference attendees are occasionally reported, these are weak proxies for the actual number of active scientists in a given discipline. So not only is there no sufficient way for a scientist to take a non-biased sample of the larger population, such a sample couldn't be translated into the actual distribution. Thus, we see little hope for enhancing MCR models by adding more realistic scenarios for the assessment of the current division of cognitive labor.

Success Function Assumption

The second assumption of MCR we will discuss is the success function assumption, which says that every agent knows the true success function for each project. Although Kitcher and Strevens do not explicitly discuss this requirement, it is entailed by the structure of MCR models. Agents make decisions based on calculating the expected utility of joining a given research program, which requires knowing the actual success function for each project.[14] Without knowing the precise shape of this function, calculating one's marginal contribution to the project is impossible. But how plausible is this assumption? Are success functions the kind of things that could be agreed upon by everyone in a community?

Recall that success functions take the number of scientists working on a project as input, and output the probability that the project will be successful. To some extent, the plausibility of the success function assumption depends on how the probabilities that success functions output are interpreted. We do not believe that these probabilities can be frequencies of success because, presumably, the projects are novel and will only be worked on once. This leaves two possible interpretations of the functions: They tell us the objective intrinsic probabilities of success, or they are subjective assessments about probabilities of success.

Much of what Kitcher and Strevens say about success functions suggests they have in mind the first interpretation, where the probabilities reflect the intrinsic chances that the project will succeed[15]. A project with a 0.8 probability of success has some intrinsic features that make it very likely to succeed. But we are not sure that one can make sense of this interpretation. Say that the success function tells us that a project's intrinsic maximum probability of success if 0.25. Clearly this doesn't mean that if we repeatedly set up scientific communities, one quarter of them would be successful. The project either has faulty assumptions behind it, in which case it will fail, or good assumptions behind it, in which case it will be successful. Phlogiston chemists had no chance of ever understanding the real principles of chemical reactivity. The intrinsic probability for finding those principles from phlogiston-based chemistry must be zero. Proponents of DNA as the molecule of heredity were correct, so the intrinsic

---

14 Strevens (pp. 64-65)

15 Strevens refers to the "intrinsic potential" of projects repeatedly. See Strevens (pp. 61, 64, 69, 70, 72, 74). Kitcher refers to the objective probabilities of projects, and in his footnote 6 claims that his account is compatible with any common interpretation of probability. See Kitcher 1990 (pp. 5, 7, 9-12)

probability that their research program would succeed must have been one, despite skepticism and alternative views in the scientific community. Thus, if success functions have probabilities not equal to one or zero, they must be subjective; they are best understood as the bets the scientific community is willing to take.

Although we believe that subjective probabilities provide the best way to interpret the outputs of success functions, the use of such probabilities raises a new set of questions. While MCR models assume that success functions are uniform among scientists, if the probabilities really are subjective, this may often not be the case. As the number of scientists in a research community increases, the probability of all their priors agreeing almost certainly will decrease, unless there is some mechanism by which their priors can be coordinated. Since we do not see where such a mechanism could come from, we think a more plausible option is that empirical information learned in the course of scientific research leads to a convergence of posteriors. Empirical information gained from scientific investigation must somehow force scientists to have nearly identical success functions. So we should ask: Do scientists have access to enough information about the projects they are choosing to work on that would cause their posteriors to converge?

We doubt that this could happen very often. In the most extreme cases like Kuhnian revolutionary science, all the concepts and approaches are novel, so there is little information that could be used to construct success functions. In more everyday cases, success functions might be constructed by examining the successes and failures of past research, but it seems to us that this only gives qualitative support for constructing success functions. It is compatible with scientists holding quantitatively different success functions, even while all agreeing, for example, that Project 1 has a higher probability of success than Project 2.

It seems clear, then, that the success function assumption is another idealization. But do the main results of the MCR approach depend crucially on this assumption? We set out to test the status of this assumption by systematically weakening it in our models. Once again, we began with a two-project, agent-based version of Strevens' *Marge* model. However, instead of giving the agents uniform beliefs about the success functions for the two projects, we let these functions vary among agents. To keep the analysis simple, every agent knew that the success function was logistic and knew the function's true value for *K*. However, we allowed the agents to have different beliefs about the value of *r*, the easiness parameter, but fixed the mean of these beliefs to the "true" values: 0.02 for Project 1 and 0.03 for Project 2. When the simulation initialized, each agent was assigned a value for *r* that was drawn from a normal distribution whose mean was fixed to the true value. By changing the variance of the distribution, we could increase or decrease the degree of uniformity among agents' beliefs.

In our simulations, we began with populations of 500 agents that drew their values for *r* out of a normal distribution with zero variance. As predicted, this initial simulation yielded identical results to the ones generated by Kitcher's and Streven's models: Agents correctly allocated themselves to the two projects. We then increased the variance in the distribution of *r* values and observed misallocation in the direction of the easier project. This was done several additional times with increasing values for the variance.

With the smallest variance in *r* that we tested (2.5E-5), 1% of agents are misallocated when compared to the correct distribution. When the variance was increased to 1.0E-4, this increases to a misallocation of 2.5% of the agents. With the much larger variance of 4E-4, we found that 7% of the agents are misallocated. The percentages of misallocated agents are fairly small, but they show that considerable misallocation occurs even when the distribution of beliefs is symmetric and the success functions are similar, but not identical.

A more dramatic result happens when we increase the relative difficulty of one of the projects. By changing Project 2's true value of *r* to 0.06, we make the project considerably easier than it was originally, requiring almost 14% fewer agents to optimally maximize its probability of success. Looking again at our diagnostic variance levels, with variance 2.5E-5 the agents misallocate by more than 2%. A variance 1.0E-4, this increases to more than 4%, and at variance of 4E-4, there is a nearly 12% misallocation of agents to projects. In every case with 500 scientists, the direction of this misallocation is the same: Too many scientists are allocated to the easier project. However, the trend actually reverses with much smaller total population sizes. With only 50 agents to allocate between projects, there is about 17% misallocation from the easier project to the harder project at the smallest variance level.

We have now shown that weakening either the distribution or the success function assumption independently can lead to misallocation of scientists, and in dramatic cases, can allocate nearly all of the agents to the easier project. Our final analysis involves weakening both assumptions simultaneously. To do this, we assumed once again that 500 scientists needed to allocate themselves, Project 1 had a true value of 0.02 for *r*, and Project 2 had a true value of 0.03 for *r*. Following the procedure described above, we distributed the agents' *r* values normally with variance 1E-4, but this time, we additionally decreased the radius of vision in order to restrict information about the current distribution of cognitive labor.

As information about the distribution of cognitive labor was restricted, the misallocation of agents to projects increased. When agents had incomplete, but still considerable information about the distribution of labor (radii of vision between 24 and 9), they misallocated themselves to the easier project more severely than when either assumption was relaxed alone. However, when they were given even less information about the current distribution of cognitive labor (radii of vision between 9 and 5), the agents actually improved their allocation and came closer to the expected MCR result. This seemed to be the result of two kinds of distortion canceling each other out. Less information about the distribution of labor (small radii) encourages more scientists to work on the easier project because it seems to them that their marginal contribution to the project's success is larger than it actually is. But if views about the differential difficulty of the two projects (values of *r*) vary among agents, some of the agents will think the harder project is easier than it actually is. So this partially mitigates the effect of the small radius size, although not for an epistemically interesting reason.

We thus conclude that MCR models are not robust to weakenings of the success function assumption either alone or in conjunction with weakening the distribution assumption. In either case, when the success function assumption is weakened, there is a tendency for scientists to misallocate away from the harder project to the easier one.

Conclusions

Using MCR models, Kitcher and Strevens tried to explain how scientists could manage to more-or-less optimally divide their cognitive labor without a central planner. We have argued that Kitcher's and Strevens' explanations are problematic because of the way that their MCR models depend on non-robust idealizations. The distribution and success function assumptions will only rarely obtain in real scientific communities. Despite the rapid flow of information in scientific communities, scientists do not typically know the full distribution of cognitive labor, nor do they have uniform beliefs about which projects are more likely to be successful. Further, we have shown that when the distribution and success function assumptions are weakened, the resulting MCR models can divide cognitive labor far from optimally. So because these assumptions are neither robust nor are they likely to be instantiated in real communities, models based on them do not possess enough fidelity to explain the self-organized division of cognitive labor in the scientific community. In light of this analysis, we must re-evaluate the broader conclusions that Kitcher and Strevens draw from their models.

Like Kitcher, we believe that the scientific community is often right to hedge its bets by distributing cognitive labor, and also believe that self-interested motives of individual scientists can lead the community to better fulfill its epistemic goals. Since Kitcher's discussion of MCR was largely oriented toward building a basic model for the study of the social structure of science, he did not use the model to make a more specific epistemic or policy recommendation. However, Kitcher claims that the MCR approach he develops can be easily extended to include more realistic assumptions. The main challenge to this, he argues, is mathematical complication.[16] We believe that the results presented here challenge this assumption. Slightly more complex models show qualitatively different behavior, and this suggests that the basic models themselves need to be reconsidered.

Strevens puts MCR models to work in service of a more ambitious claim. He uses such models to argue that the Priority Rule, a payoff function that gives all the credit to whichever scientist successfully completes the project first, is the best incentive structure for promoting the optimal distribution of cognitive labor. The Priority Rule encourages scientists to hedge their bets by distributing cognitive labor, but biases the distribution in favor of the project most likely to succeed. In order to show this, Strevens compares the results of MCR models employing the *Marge* scheme to those employing the Priority rule. The latter encourages more agents to work on the project with the higher maximum probability of success, which is closer to the optimal distribution.

While we fully agree with Strevens' argument for MCR models making the distribution and success function assumptions, we do not believe that his analysis holds up if either of these assumptions is relaxed. We have shown that relaxing these assumptions misallocates cognitive labor toward the easier project. Similar results are obtained when the projects are equally easy but one has a higher maximum probability of success. In this case, agents are overallocated to the high-probability project. If we add the Priority Rule on top of this over-allocation, agents will be grossly over-allocated, perhaps resulting in no one working on the low-probability project at all. This is precisely what an incentive system inspired by social epistemology should be able to avoid.

The desirability of the Priority Rule in Kitcher's and Strevens' model scientific communities is hard to dispute. However, we do not believe that real scientific communities optimally distribute their

---

[16] Kitcher 1990 (pp. 7, 18, 22)

cognitive labor when this rule is in place. Indeed, the opposite may be true in a wide variety of cases. Like Kitcher, we think that it would be surprising if the current set of incentives in science are anywhere close to optimal.[17] These incentives were designed for communities of scientists that much more closely approximated the epistemic assumptions of MCR than the scientific community of the twenty-first century. As we have demonstrated, once the epistemic assumptions of MCR models are weakened, the direction of the allocation of cognitive resources can change. Adopting the Priority Rule exacerbates this problem, resulting in greater misallocation than then when the *Marge* payoff function is employed. As a result of this, we do not believe that Strevens' has established the desirability of the Priority Rule for real scientific inquiry.

In closing, we note that it might be possible to develop revised MCR models, either by weakening the idealizations as we have done or by finding ones that are more robust to perturbation. This work may turn out to be fruitful, although our own experience of modifying MCR models to test for robustness suggests that this approach generates extremely complex models. As an alternative, we propose the development of a new framework for studying the division of cognitive labor that makes fewer epistemic assumptions about scientists from the outset. In such a framework, scientists make decisions about what projects to work on "on-line," as information about their own and their community's success and failures become known.[18] Such a framework may advance the project of understanding the social nature of scientific epistemology championed by Kitcher and Strevens by providing a more fertile basis for continued research.

---

[17] (*op. cit.*, p. 22)

[18] In Weisberg, M. and R. Muldoon "Epistemic Landscapes and the Division of Cognitive Labor" (Manuscript, 2007), we argue for an alternative modeling approach inspired by ecological models that remains simple while also ensuring that the agents in the model are under similar epistemic constraints as actual scientists.