

# A KUHNIAN RATIONALIZATION FOR CREDENCE CONSTRAINT IN BAYESIAN CONFIRMATION THEORY

BRANDON LIVERENCE

*Stanford University*

## **Abstract:**

Bayesian Confirmation Theory (BCT) is the predominant approach to theoretical confirmation. Given one's credences, or subjective likelihoods, for a set of competing hypotheses, BCT provides a simple mathematical formula for determining how much these likelihoods should change given observation of evidence. It allows for subjectivity and disagreement exactly where one might expect to find it- early on in the process of theory testing, before evidence is considered. But some have argued that BCT may be too subjective- scientists with different subjective probabilities may actually disagree about whether a piece of evidence confirms or disconfirms a hypothesis. Convergence has been suggested as a solution to this problem- as more evidence is considered, every scientist's credences will start to look alike, and only true and nearly true hypotheses will get confirmed. But given any amount of evidence, there is some degree of bias that can prevent confirmation of true hypotheses. Because timely consensus is a goal of the scientific enterprise, it may be desirable to constrict one's credences in a way that aids the process of convergence. General principles for constraint have been suggested, but it may be impractical, as well as suboptimal, to impose the same constraints on all scientists. Constraints result in faster confirmation only when unexpected hypotheses, those for which most scientists would otherwise set low or zero subjective probabilities, turn out to be true, what Kuhn described as a crisis period. As I will show, Kuhn's theory of the social structure of science provides the framework for an optimal system of prior constraint- one in which paradigms are challenged by young scientists who apply pro-convergence constraints to their credences, whereas normal science is best advanced by older scientists who don't constrain their credences in the same way. The implications of this Kuhnian interpretation of BCT are considered.

## **The Fundamentals of Bayesian Confirmation Theory:**

In Bayesian Confirmation Theory (BCT), each competing hypothesis assigns a physical probability between 0 and 1 to a piece of evidence. If hypothesis  $h$  assigns a value greater than 0.5 to the physical probability of some evidence  $e$ , written  $Ph(e)$ , then it predicts that  $e$  is more likely than not to be observed. Upon observation of  $e$ , you should be more likely to believe hypothesis  $h$  than before. Conversely, if some competing theory assigns  $Ph(e) < 0.5$ , then observing  $e$  should make you less likely to believe hypothesis  $h$ . BCT incorporates these intuitions into a simple but powerful model of statistical learning. Given a set of hypotheses describing the same physical phenomenon, BCT will tell a scientist exactly how much to change his likelihood to believe each

hypothesis upon observation of physical evidence of the phenomenon.

These likelihoods of belief take the form of subjective probabilities, or credences, written  $C(h)$ , where  $h$  is some hypothesis. Credences, like physical probabilities, equal a number between 0 and 1. But unlike physical probabilities, credences are psychological phenomena. They do not make statements about the physical world the way that physical probabilities of the form  $Ph(e)$  do. A credence  $C(h)$  describes how likely some scientist is to believe that a given hypothesis  $h$  is true.  $C(h) = 0$  corresponds to being 0% certain that a given hypothesis is true. One consequence of the Bayesian formula is that a hypothesis with a credence of 0 can never be confirmed, no matter how much supporting evidence is observed. A credence of 1 corresponds to 100% certainty in a given hypothesis, meaning that no amount of contradicting evidence can convince you that the theory is not true. Most of the time, credence values fall somewhere in between. When there are multiple hypotheses describing the same phenomenon, the sum of all their credences must equal 1 at all times. It is an axiom of probability theory that you may not be more than 100% certain of any set of alternatives.

Upon observation of evidence  $e$ , credences are updated by substituting the values of the prior credences of all the competing hypotheses  $h_1, h_2, \dots, h_n$  (such that  $C(h_1) + C(h_2) + \dots + C(h_n) = 1$ ) into the Bayesian formula in order to calculate the posterior credence. This updating process is called conditionalization upon evidence. To calculate  $C+(h_1)$ , the posterior credence for hypothesis  $h_1$ , the Bayesian formula is written:

$$C+(h_1) = \frac{Ph_1(e)C(h_1)}{Ph_1(e)C(h_1) + Ph_2(e)C(h_2) + \dots + Ph_n(e)C(h_n)}$$

There are four key theoretical elements that, in combination, yield the Bayesian formula. First is Bayes' theorem:  $P(e | f) = P(f | e) \times P(e) / P(f)$ . Less formally, it states that the probability of some event  $e$  conditional on the occurrence of some other event  $f$  equals the probability of the event  $f$  conditional on  $e$ 's occurrence, multiplied by the probability of  $e$ , and divided by the probability of  $f$ . It is enough to know that the theorem follows from the definition of a conditional probability.

Second is Bayes' rule for conditionalization:  $C+(h) = C(h | e)$ . Bayes' rule states that upon observation of a piece of evidence, one should set his credence for some hypothesis that explains the evidence equal to his subjective probability for the likelihood of that hypothesis conditional on the occurrence of that evidence. Using Bayes' rule to conditionalize upon the evidence allows one to update prior credences, and thus relates credences at two different times. Bayes' rule makes two fundamental assumptions: first, that credences possess mathematical properties like probability assignments, and second, that there exists

some a priori justification for Bayes' rule, which does not follow from the axioms of probability because it relates probabilities at two different times.

The third is the probability coordination principle (PCP). Credences are psychological properties. They correspond to a human's belief about the likelihood of a given hypothesis. They are not statements about the physical world. It is only by way of PCP, which states that  $C(e | h) = Ph(e)$ , that they acquire the mathematical properties required for probabilistic formulation.

Lastly, the total probability theorem states that for some outcome  $e$ , and the set of mutually exhaustive (at least one must occur) and mutually exclusive (at most one can occur) outcomes  $f_1, f_2, \dots, f_n$ ,  $P(e) = P(e | f_1)P(f_1) + P(e | f_2)P(f_2) + \dots + P(e | f_n)P(f_n)$ . The right side term is a kind of weighted average of the probability assigned to outcome  $e$  by all of the outcomes in the set  $f_1, f_2, \dots, f_n$ . In the Bayesian formula, the class of hypotheses  $h_1, h_1, \dots, h_n$  will take the place of these outcomes. The total probability theorem follows from the axioms and the definition of conditional probability.

Combining the aforementioned principles, the Bayesian formula for conditionalization on evidence  $e$  for some hypothesis  $h_x$  (where  $x$  is any value between 1 and  $n$ ) can be written as follows:

$$C+(h_x) = \frac{C(e | h_x)C(h_x)}{C(e)} = \frac{Ph_x(e)C(h_x)}{Ph_1(e)C(h_1) + Ph_2(e)C(h_2) + \dots + Ph_n(e)C(h_n)}$$

The value  $Ph_x(e) / (Ph_1(e)C(h_1) + Ph_2(e)C(h_2) + \dots + Ph_n(e)C(h_n))$  is referred to as the Bayesian multiplier, because it determines by how much the prior credence will change upon conditionalization. The posterior credence simply equals the prior credence times the Bayesian multiplier, so for example, if the multiplier equals 0.5, then the posterior credence will be half as large as the prior credence. The numerator of the Bayesian multiplier,  $Ph_x(e)$ , represents the physical probability assigned to the evidence  $e$  by the hypothesis  $h_x$ . If  $h_1$  is the hypothesis that all ravens are black, and if  $e$  is a black raven, then  $Ph_1(e)$  is 1. If  $h_2$  is the hypothesis that 50% of ravens are black, then  $Ph_2(e)$  is 0.5. The denominator of the Bayesian multiplier represents the sum of the physical probabilities each hypothesis assigns to the evidence, each multiplied by the prior credence for that hypothesis. It is a kind of weighted average physical probability assigned to the evidence by the set of all hypotheses. If the numerator is greater than the denominator, or more exactly, if the physical probability  $h_x$  assigns to the evidence  $e$  is greater than the weighted average, then the value of the Bayesian multiplier will be greater than 1, and the credence for  $h_x$  will increase upon conditionalization. If the multiplier is less than one, then our likelihood to believe  $h_x$  will decrease.

To illustrate conditionalization, let's consider a simplified case with only three competing hypotheses:  $h_1$  (All ravens are black),  $h_2$  (50% of all ravens are black), and  $h_3$  (no ravens are black). Where  $e$  is observation of a black raven,  $Ph_1(e) = 1$ ,  $Ph_2(e) = 0.5$ , and  $Ph_3(e) = 0$ . Some scientist assigns credences  $C(h_1) = .25$ ,  $C(h_2) = .25$ , and  $C(h_3) = 0.5$ . The calculation for  $h_2$  follows:

$$C+(h_x) = \frac{Ph_1(e)C(h_x)}{Ph_1(e)C(h_1) + Ph_2(e)C(h_2) + \dots + Ph_n(e)C(h_n)}$$

$$C+(h_2) = \frac{0.25 \times 0.5}{(.25 \times 1) + (.25 \times 0.5) + (0.5 \times 0)}$$

$$C+(h_2) = \frac{0.125}{0.375}$$

$$C+(h_2) = 0.333$$

Hypothesis  $h_2$  gets confirmed because the Bayesian multiplier,  $0.5/0.375 = 1.333$ , which is greater than one, meaning that the weighted average physical probability for  $e$  was lower than that assigned by hypothesis  $h_2$ . The prior credence  $C(h_2)$  was 0.25, and upon conditionalization,  $C+(h_2)$  increases to 0.33. This process can be repeated for the remaining hypotheses, resulting in  $C+(h_1) = 0.666$  and  $C+(h_3) = 0$ , thus completing one cycle of conditionalization. When the next piece of evidence is observed, conditionalization can be repeated by entering the credences calculated from the first cycle into the right side of the formula to calculate new posterior credences.

When there are an infinite number of hypotheses to consider, for instance the set of hypotheses of the form "X percent of ravens are black," where X is any number between 0 and 100, then the set may be described mathematically as a probability density function  $f(X)$  and substituted into the Bayesian formula in place of credences for individual hypotheses. In such a function, the area beneath the curve between two points equals one's credence that the true hypothesis is some value of X between those points.

### Subjectivity and Convergence:

Subjectivity enters BCT in the form of each scientist's psychologically determined credences. How much a hypothesis is confirmed or disconfirmed by evidence is determined both by the claims that the hypothesis and its competitors make about the evidence, and by the individual scientist's credences for each hypothesis. This aspect of Bayesianism is often viewed as its greatest advantage, because it accommodates the behavior of real scientists. A scientist may initially favor one hypothesis above another equally probable one because hypotheses like

the first one have been successfully confirmed in the past, or perhaps because the first hypothesis more closely matches the dictates of that scientist's own previous work. Thus, BCT's subjective element can accommodate inductive reasoning, as well as the influence of a current paradigm.

Based on what we know about the relative stability of the natural world, it doesn't seem credible to assign an equal probability to preposterous hypotheses such as "all ravens are blight," where blight means black until the year 3000, then white afterwards, as to simpler hypotheses such as "all ravens are black." But without a mechanism to incorporate these subjective, common-sense beliefs into our Bayesian learning framework, we would be stuck believing both hypotheses equally until the year 3000. And even at that point, when the blight hypothesis is disconfirmed, there will always be variants that predict the color change in year 4000, or 5000, etc.

Despite all of the benefits of incorporating subjectivity into a theory of confirmation, subjectivity also leads to what is probably BCT's most severe problem. Two scientists with different credences will have different values for  $C(e)$ . If one scientist's value for  $C(e)$  is smaller than  $Ph(e)$ , then the Bayesian multiplier will be larger than one and the credence will go up upon conditionalization. If another scientist instead has a value of  $C(e)$  larger than  $Ph(e)$ , then his Bayesian multiplier will be less than one, and the hypothesis will be disconfirmed. In effect, the same piece of evidence can either confirm or disconfirm a given hypothesis depending on one's credences. Consider the example from before in which there are three competing hypotheses to explain raven color:  $h_1$  (All ravens are black,  $Ph_1(e) = 1$ ),  $h_2$  (50% of all ravens are black,  $Ph_2(e) = 0.5$ ), and  $h_3$  (No ravens are black,  $Ph_3(e) = 0$ ). Scientist 1, as before, assigns credences  $C(h_1) = .25$ ,  $C(h_2) = .25$ , and  $C(h_3) = 0.5$ . Upon observation of one piece of evidence (a black raven),  $C(e) = ((.25 \times 1) + (.25 \times 0.5) + (0.5 \times 0)) = .375$ . For  $h_2$ , the Bayesian multiplier is  $0.5 / .375 = 1.333$ , which is greater than one, so  $h_2$  is confirmed. Scientist 2 instead assigns credences as follows:  $C(h_1) = 0.5$ ,  $C(h_2) = .25$ , and  $C(h_3) = .25$ .  $Ph_2(e)$  is still 0.5, but now upon observation of the black raven,  $C(e) = ((0.5 \times 1) + (.25 \times 0.5) + (.25 \times 0)) = .625$ . Now the Bayesian multiplier is  $0.5 / .625 = .8$ , which is less than one, so  $h_2$  is disconfirmed. This result is often viewed as a particularly troubling problem for BCT, since it casts doubt on the ability of BCT to accurately model the scientific process. If scientists can't even reach a consensus on whether a piece of evidence supports or refutes a given theory, then it seems hard to believe that the scientific community could ever come to accept a single hypothesis or a single restricted class of hypotheses as being the most accurate.

Convergence may offer a solution to this problem. While prior credences for all of the competing hypotheses are needed in order to proceed with conditionalization in the first place, those prior credences

matter less and less as the amount of evidence being conditionalized upon increases. One consequence of repeated conditionalization on evidence is that with very high physical probability, the probability density functions of all scientists will gradually begin to converge as density is heaped around the same true hypothesis for all of them, and approaches 0 for all other hypotheses. The more evidence that is conditionalized upon, the higher the likelihood that everyone's credences will converge. Convergence results in a gradual washing out of the initial biases of the scientists. Let's consider an example: three scientists with radically different prior credences for three hypotheses concerning raven color. As before, the hypotheses are  $h_1$  (All ravens are black,  $Ph_1(e) = 1$ ),  $h_2$  (50% of all ravens are black,  $Ph_2(e) = 0.5$ ), and  $h_3$  (No ravens are black,  $Ph_3(e) = 0$ ). Prior credences are as follows:

Scientist 1:	$C(h_1) = 0.7$	$C(h_2) = 0.2$	$C(h_3) =$	$0.1$
Scientist 2:	$C(h_1) = 0.2$	$C(h_2) = 0.1$	$C(h_3) =$	$0.7$
Scientist 3:	$C(h_1) = 0.1$	$C(h_2) = 0.7$	$C(h_3) =$	$0.2$

Conditionalizing upon  $n$  examples of evidence  $e$  (black raven) produces the following results:

$n$	Scientist 1	Scientist 2	Scientist 3
1	$h_1 = .875$ $h_2 = .125$	$h_1 = .8$ $h_2 = .2$	$h_1 = .222$ $h_2 = .778$
3	$h_1 = .965$ $h_2 = .035$	$h_1 = .941$ $h_2 = .059$	$h_1 = .533$ $h_2 = .467$
5	$h_1 = .991$ $h_2 = .009$	$h_1 = .985$ $h_2 = .015$	$h_1 = .820$ $h_2 = .180$
10	$h_1 = >.999$ $h_2 = <.001$	$h_1 = >.999$ $h_2 = <.001$	$h_1 = .994$ $h_2 = .006$

In all three examples,  $C(h_3)$  immediately plummets to 0 because it assigns a physical probability of 0 to the evidence. The fascinating result is that despite such radically different prior credences, conditionalization upon only 10 pieces of evidence is enough to bring all three scientists' credence values for  $h_1$  to within 0.5%.

So, while a scientist's credences may initially govern how particular hypotheses are interpreted in light of the evidence, as more evidence comes in, that initial bias will tend to be washed out until the scientist is nearly in accord with (although never in exact agreement with) his colleagues. This seems like a reasonable account of consensus. We might believe there to be some disagreement at the beginning of the theory testing process, but once a large amount of evidence has been analyzed, trends in the evidence

should take precedence over the initial subjective beliefs of scientists.

The most illuminating challenge to convergence as a solution to the problem of subjectivity is that depending on one's credences, some given amount of evidence may never be enough to convince a scientist of the truth of a true theory. Even with a large amount of supporting evidence (say, 1000 black ravens in a row), a scientist may remain unconvinced that the hypothesis "all ravens are black" is more likely to be true than a hypothesis such as "half of all ravens are white" if he set his prior credence for the all black hypothesis sufficiently close to zero and for the half black hypothesis sufficiently close to one. Even worse, if a scientist set his prior credence for the most accurate hypothesis to 0, representing either the case where the scientist made an active judgment about the impossibility of such a claim, or the case where the scientist simply failed to even consider it, then confirmation will never be possible.

#### Credence Constraint:

Returning to the example presented above, it seems obvious that most scientists wouldn't give much weight to hypotheses that challenge the most fundamental laws and assumptions of their respective fields. Scientists are trained to accept as true particular assumptions about the physical world, largely because they have never been violated in the past. In biology, an event has never been observed where all individuals of a species spontaneously change color, so the stability of species color from moment to moment can be considered an assumption of the field. Biologists who accept the theory of evolution by natural selection to be true do believe that particular features of a species may change over time, but never spontaneously. One could well imagine a scenario where initially all ravens start out as black, and some genetic mutation leads to the birth of a white raven. Eventually, raven blackness is selected against, perhaps by some opportunistic predator, and white ravens come to dominate until they constitute the only remaining type of raven. This hypothesis would fit with the accepted notions of evolutionary biologists. Scientists comparing hypotheses of raven blackness by way of BCT might tend to assign a low, but non-zero, credence to hypotheses of the form of the evolutionary example. But scientists who accept the assumptions and principles of biology will assign a value of zero to hypotheses of the blight form.

Let's consider for a moment what would happen if the blight hypothesis turned out to be true. In 3000 AD, all ravens suddenly change color from black to white, due to some previously unobserved physical phenomenon. As our scientist begins to conditionalize upon evidence of white ravens, his credence for "all ravens are black" will plummet. The evolutionary hypothesis will get confirmed, even though it's not quite right, as will the "all ravens are white" hypothesis, even though its obviously

wrong. But the most accurate hypothesis will never get confirmed because the basic assumptions of his field prevented the scientist from assigning a nonzero value to radical hypotheses such as “all ravens are blight.”

This example illustrates that if BCT is ever to attain universal applicability, it may need to include principles for assigning prior credences that aid the process of convergence. One such principle, called open-mindedness, dictates that a scientist should always assign nonzero credences to every possible hypothesis. Open-mindedness is no light order: not only must a scientist suspend his disbelief about hypotheses that defy even the most fundamental assumptions of the field, he must also think of every possible hypothesis before he begins to conditionalize on the evidence. It's easy to be open-minded about a class of hypotheses that vary across only a single parameter, such as those of the class “X% of ravens are black;” this merely requires establishing a simple 2-D probability density function that is never equal to zero at any value of X. What blight-like hypotheses expose are more degrees of freedom: more parameters across which competing hypotheses may vary, such as time of change and color of change. Just considering those three parameters requires a scientist to formulate a 4-dimensional probability density function. In reality, there are an infinite number of parameters that could vary, and no way for a scientist to consider them all. Scientists are just humans after all. It is best to think of open-mindedness as an ideal behavior rather than an absolute prescription, a principle that should guide scientists to consider new parameters along which hypotheses might vary, rather than an expectation that scientists will always be able to think of the true hypothesis before they look at the evidence. Scientists who tend to be open-minded will be more likely to get confirmation of the most accurate hypotheses if the evidence happens to vary across a parameter that has not shown variation in the past.

A second principle for credence constraint dictates that one should be fair-minded and assign credences such that it won't take a prohibitively large amount of evidence to sufficiently confirm any given hypothesis. Fair-mindedness guarantees that the probability density function will quickly heap around the most accurate hypothesis, no matter how unlikely it might have seemed before conditionalizing upon evidence. A fair-minded scientist may still favor particular hypotheses by giving them a modest boost in their priors, as long as competing hypotheses are assigned only slightly lower priors. Whereas the probability density function of a highly biased scientist tends to resemble a sharp peak centered above the favored hypothesis, the fair-minded scientist's probability density function tends to look more like a hump. One consequence is that if the initially favored hypothesis turns out to be the most accurate, a fair-minded scientist will have to conditionalize upon slightly more evidence than a non fair-minded scientist in order to

reach the same level of confirmation. But the benefit is that that the fair-minded scientist won't have to spend his entire career conditionalizing upon evidence if his initially favored hypothesis turns out not to be true.

Credence constraint according to the principles of open- and fair-mindedness seems to represent ideal behavior on the part of a completely rational scientist. Perfect adherence to these principles guarantees timely convergence onto the most accurate hypotheses by ensuring that scientists can't set their priors for any hypothesis too low. But given the limitations of the human mind, no scientist can be expected to consider every possible hypothesis. Furthermore, no scientist simply ignores the history of scientific discovery when formulating new hypotheses. Instead, scientists will tend to assume that entire classes of hypotheses, such as the blight-like hypotheses with regard to raven color, are false based on existing scientific evidence. Many other scientists may simply be so biased towards their own hypotheses that they will purposely ignore the principles of open- and fair-mindedness by assigning priors that give their own models an unfair advantage. It may often be true that such behavior is irrational and counterintuitive to the goal of scientific progress, but might there be times when it's actually advantageous to ignore the principles of open- and fair-mindedness, and if so, what criteria govern when and by whom they should be applied?

#### **A Kuhnian Rationalization for Multiple levels of Credence Constraint:**

Thomas Kuhn believed that the history of science was characterized by alternations between long periods of what he called normal science, when the evidence supports the dominant beliefs and assumptions of the field, and extraordinary science, when the evidence begins to contradict the dominant beliefs. These dominant beliefs characterize what Kuhn called a paradigm- a broad consensus among scientists in the field with regard to fundamental questions and research methods.

During a period of normal science, the field is advanced by strict adherence to the overarching methods and assumptions of the paradigm. In Bayesian terms, normal science corresponds to a period when the hypotheses most scientists favor by assigning them high prior credences turn out to be the best ones. If the paradigm is that raven color (or more generally, the physical characteristics of all animals) will conform to stable proportions across the species, then scientists adhering to the paradigm will assign high credences to hypotheses of the form “X% of ravens are black.” Scientists may even have reasons to favor a particular value of X over another, say, if the paradigm also includes the assumption that all or most members of a species are the same color. In this case, they would assign a high prior credence to hypotheses such as “all ravens are black,” or “X% of all ravens are black,” where X is some fairly high proportion or range of

proportions. If the most accurate hypothesis turns out to be one of those initially favored by the assumptions of the current paradigm and given an initial boost in its prior credence, then confirmation to a high degree will be achieved very quickly. Furthermore, scientists who didn't constrain their credences will achieve any given level of confirmation more quickly than scientists who were open- and fair-minded. Credence constraint would entail distributing some probability density over hypotheses that violate the paradigm, such as blight-like hypotheses, at the expense of the ultimately more accurate hypotheses favored by the dominant paradigm.

During periods of extraordinary science, evidence tends to contradict the dominant paradigm, or simply fails to be explained by it. Without an influx of new theories, science will at best slow, and at worse crash to a halt during such periods. As shown above, being open-minded guarantees that if the true hypothesis falls outside of the dominant paradigm, it will still be assigned a non-zero prior credence, and will have the chance to eventually get confirmed. Being fair-minded simply speeds the process along.

If we accept Kuhn's model to be correct, then we must accept that during periods of normal science, confirmation happens more quickly when credence constraints are not applied, and during periods of extraordinary science, confirmation happens more quickly, and is sometimes only possible, when scientists constrain their priors according to the principles of open- and fair-mindedness. These principles are critical during periods of extraordinary science to allow radical hypotheses enough probability density to eventually get confirmed. Thus, neither system of credence determination has universal utility. If the same principles for setting credences were applied to every scientist, confirmation of accurate hypotheses would be slower during either normal science or extraordinary science as compared with a system where scientists use different credence constraints.

How might two such different forms of scientific behavior coexist, in order to produce the most efficient system of confirmation possible? Kuhn observed that it is generally younger scientists who advance radical theories characteristic of new paradigms during crisis periods, while it is generally older, more established scientists who oppose them and require overwhelming evidence to finally become convinced of their accuracy, if they ever do. Under a Bayesian framework, the behavior of younger scientists is characteristic of assigning higher prior credences to radical theories, and thus constraining credences according to the principles of open- and fair-mindedness. Conversely, the behavior of older scientists is characteristic of assigning low or zero credences to radical theories and higher credences to theories within the paradigm.

The social structure of science offers insight as to why scientific status and experience may determine credence setting behavior. It is often in the best interests of young scientists to be open-minded because

of the enormous pressures they face to make breakthrough discoveries in order to receive tenure. Furthermore, by being fair-minded, they leave themselves a way out should their radical theories prove to be wrong and the dominant paradigm prove to be successful once again. That way, they won't have to spend years conditionalizing on evidence to reach the same levels of certainty as their older, less fair- and open-minded colleagues.

It is important to stress that the principles of credence constraint are ideals that no human may really achieve, especially the principle of open-mindedness. The claim is not that every young scientist will be so open minded as to always assign some probability density to every radical hypotheses, simply that they are far more likely to do so than older scientists.

The circumstances are quite different for older scientists. Because defending the reigning paradigm is often inextricably linked to defending the relevance of their life's work, these scientists will more often than not oppose radical new theories that call their own into question by assigning them prohibitively low credences. Additionally, older scientists may be so deeply entrenched in the dominant paradigm that they simply fail to consider radical theories as readily as their younger colleagues, thereby assigning them credences of zero. Insofar as established scientists tend to fail to apply the credence constraints of open- and fair-mindedness, they will fail to reach high degrees of certainty in radical but accurate hypotheses as quickly as their younger colleagues.

In the workings of science, there is a place for both types of scientists. During a period of normal science, the unconstrained biases of older scientists help to cut through all of the implausible competing theories and direct effort towards achieving progress within the paradigm. It would simply be a waste of time to consider too many competing theories if the current paradigm continues to be supported by the data. During crisis periods, when the evidence begins to contradict the current paradigm, then the younger, more open- and fair-minded scientists make their biggest contributions by taking risks on new theories.

### Consensus and The Paradigm Shift:

By adopting a Kuhnian rationalization for multiple levels of credence constraint, we must abandon the requirement of absolute consensus for scientific progress. Insofar as scientific progress during a period of extraordinary science fits Kuhn's conception of a paradigm shift, such a proposal shouldn't be problematic, since no paradigm shift occurs with absolute consensus.

Rather than being polarized into two basic groups, scientists likely fall along a spectrum: at one end are the youngest, most open-minded and fair-minded scientists, and at the other end are the staunchest of emeritus professors, those who are most deeply entrenched in the

dominant paradigm and are consequently the most biased against radical hypotheses. Those scientists who are both open- and fair-minded will achieve the most rapid confirmation of radical hypotheses during a crisis period. This subset of scientists will constitute the earliest proponents of the new paradigm. Those scientists who fell at various points along the spectrum may tend to be open-minded but not fair-minded; they will eventually achieve the same level of confirmation of the radical hypothesis, but will require more evidence than their younger, less established colleagues. This represents the kind of opposition that Kuhn envisioned new paradigms would face from the scientific establishment. Lastly, those scientists who were neither open- nor fair-minded will simply never get confirmation of true but radical hypotheses. These scientists, the most established and deeply entrenched in the paradigm, may resist the paradigm shift for the rest of their lives. Max Planck wrote in his *Scientific Autobiography*: "A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it."

I have argued for a system of credence constraint in which there are no absolutes. It avoids the practical difficulties of forcing all scientists to set their credences in exactly the same way, while achieving confirmation of accurate hypotheses more rapidly during both periods of normal science and extraordinary science. But my system as stated remains vulnerable to an attack that has been leveled against open-mindedness in the past, namely, that it is a highly unreasonable demand to make of human scientists. But without perfect open-mindedness throughout the scientific community, it is hard to see how a broad consensus could ever be reached about the truth of radical or brand-new hypotheses. Any scientist that failed to consider a hypothesis before beginning to conditionalize upon evidence, and thus failed to assign a non-zero credence to it, could never achieve confirmation of the hypothesis. In order for other scientists who assigned credences of zero to the accurate hypothesis to become convinced of its accuracy, what may be needed is a non-Bayesian system to communicate those results.

Even in the extreme case, where only one scientist was open-minded enough to have actually assigned a nonzero prior credence to the radical but true hypothesis, all hope may not be lost for a paradigm shift. Once the scientist has conditionalized upon some evidence and achieved rapid confirmation of his radical theory, he may be able to convince other scientists of his accomplishment through methods of scientific communication such as journal articles, seminars, lectures, etc. These are not enough to convince the scientific community of the truth of the radical theory, since that would require everyone to reformulate their prior credences after evidence has already been conditionalized upon. Instead, those scientists (perhaps still a relatively small subset)

who are impressed with the rapid confirmation of this young innovator's hypothesis will assign non-zero prior credences to other testable hypotheses predicted by this new paradigm. They have effectively been convinced to become open-minded, and perhaps even fair-minded, in the future.

For example, if a scientist predicted a cosmic phenomenon that would occur at the beginning of the year 3000, which would radically change the structure of the pigment molecules in raven's feathers such that they would reflect all visible light instead of absorbing it, he would achieve rapid confirmation of the hypothesis "all ravens are blight." To the extent that such phenomena have never been described or observed in the past, this represents a paradigm shift. In the future, fair-minded and open-minded scientists may assign non-zero priors to hypotheses that fit the paradigm and explain other, related phenomena (perhaps the color changes of other species with the same pigment, or the effect another instance of the cosmic phenomena might have on other pigment molecules).

If the paradigm turns out to be a better fit to the data than the old one, then all of those scientists will subsequently achieve rapid confirmation of the blight-like hypothesis, due to its excellent fit to the data, and the paradigm shift will have begun.

## REFERENCES

- Hoyningen-Huene, P. (1998). "Kuhn, Thomas Samuel." In E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge. Retrieved January 13, 2004, from <http://www.rep.routledge.com/article/Q055SECT2>
- Kuhn, T. S. (1996). *The Structure of Scientific Revolutions*. Third Edition. University of Chicago Press, Chicago, IL.
- Planck, M. (1950) *Scientific Autobiography, and Other Papers*. Tr. by Frank Gaynor. Williams & Norgate, London.
- Strevens, M. (2002) *Notes on Bayesian Confirmation Theory*. Unpublished.