Cultural adaptation in mathematics and physics

Frank Quinn

Abstract The thesis is that the organization, attitudes, and customs of a scientific discipline are strongly influenced by the nature of the subject matter. A case study is presented, comparing theoretical physics and pure mathematics. These share a great deal, but differences in goals and subject have led to striking cultural differences and a long history of culture shock at the interface. An analysis of this sort can be useful in developing science policy and managing change. We conclude for example that mathematics and physics have different needs in ethical standards, grant support, and electronic publication.

Some cultural adaptations in science are so obvious they cannot be missed: for example equipment forces a lot of herd activity in some areas ("big science") while others remain more solitary ("small science"). Close inspection reveals a multitude of ways in which even narrow specialties have cultural aspects micro-adapted to efficient development of their subject. Most scientists, if they can be brought to think about culture at all, find this point self-evident and conclude that culture will take care of itself. Social scientists studying science think a lot about culture. Unfortunately many of them assert that it can be studied without understanding content, or even that culture is primary and determines content. Neither point of view has led to a fruitful understanding.

In tranquil times there is not much need to understand cultural adaptation: the benefits of adaptation push research communities in the right directions without conscious effort. Unfortunately these are not tranquil times. Support structures, societal expectations, and even the publication and communication infrastructure are all on the verge of radical change. These changes will sweep away many slowly and painfully acquired cultural nuances. An awareness of culture and adaptation may make it possible reduce the damage.

We present a case study, contrasting pure mathematics and theoretical physics to reveal adaptation in both areas. This is a relatively simple case becase a wide spectrum of differences traces back to a single root cause: reliability. In other cases, particularly in experimental areas, the roots of culture will be more complex.

The setting By "physics" we will mean theoretical work in areas like high-energy particles, superconductivity, or quantum mechanics: areas where sophisticated mathematical apparatus is needed even to organize or interpret real data. By "mathematics" we mean the "pure" areas in which abstract mathematical apparatus is developed. The point of excluding experimental physics and applied mathematics is to obtain relatively homogeneous groups. They are still richly diverse in their subcultures, but the commonalities are strong enough permit useful conclusions.

There is a lot of traffic between these areas. Physicists rely on mathematical work, and many mathematical structures are inspired by physics. The shared use of mathematical apparatus and long history of interaction has led to a great deal of shared language and strong superficial similarities.

Reliability and logic The key to the differences between these fields seems to be reliability. Mathematics, through logical rigor, can achieve essentially complete reliability. Information in physics may be excellent but is never perfect. We remark that the reliability of logic is an "experimental" fact. The Incompleteness Theorem of Godel showed that we cannot *prove* that correct logic yields completely reliable conclusions, even if we set aside concerns about the circularity of proving something about proof, or about human fallibility.

However as a hypothesis based on experience, the reliability of logic has been extremely well tested over several thousand years.

To illustrate the significance of reliability, consider the use of proof by contradiction. In mathematics it is a standard technique to begin with a dubious assertion and build an elaborate logical structure on it. At the end something emerges which is known to be false. The conclusion is that the initial assertion must also be false. The usefulness of this method depends heavily on the complete reliability of logic and of the other information used in the demonstration. Any leakage will cause the method to fail. The false conclusion at the end may be a consequence of a flaw in the argument or an ingredient, rather than the falsehood of the target assertion.

Mathematical customs have adapted to this difference between perfect and imperfect information. The conclusion of a mathematical plausibility argument is traditionally called a "conjecture", while the result of a rigorous argument is called a "theorem." Theorems can be used without fear in a contradiction argument; conjectures are a possible source of error. Physics does not work this way: elaborate logical structures tend to magnify errors, so are suspect. A direct plausibility argument is usually more robust than an elaborate logical "proof." Accordingly the physics culture places a premum on short insightful arguments (supported by calculation: see the next section), even if wrong in detail. The "theorem-conjecture" distinction is not particularly useful.

These differences lead to culture shock. In physics conclusions from intuition and plausible argument have first-class status. Mathematicians tend to describe these conclusions as "conjectures" still needing proofs, or dismiss them as hopelessly imprecise. Physicists resent this. Conversely physicists tend to be disdainful of mathematical rigor as being excessively compulsive about detail, and mathematicians resent this. But there are good reasons for the values held in both disciplines. The problems come from customs adapted to the subject, not xenophobia or a power contest.

The effects of these mutual ill-adaptations are not symmetric. Mathematical practices used in physics are inefficient or irrelevant, but not harmful. Physical practices (no distinction between conjecture and theorem) used in mathematics can cause harm: it jeopardizes standard techniques such as proof by contradiction. There is widespread feeling among mathematicians that violating these "truth in advertising" customs should be considered misconduct [2]. This is a behavior which is productive in one field and misconduct in another, not because of contingent historical development of "standard practice", but because the subjects are different.

Process versus Outcome The decisive criterion for correctness in theoretical physics is agreement with experimental observation. This focuses attention on outcomes, not process. This point of view permeates even internal efforts in theoretical development that do not make direct contact with experiment. When a model is developed it is checked against others believed to be relevant: special cases, the "classical limit", etc.

In pure mathematics the primary criterion is internal. The reliability of logic can be rephrased as: *if an argument produces a false conclusion, then it contains either a logical flaw or an erroneous hypothesis*. Mathematicians have gone to some lengths to ensure their hypotheses are reliable, so the absence of logical flaws is a criterion for correctness. In practice it has been very effective. Attention is focused on the process (avoiding or detecting flaws), rather than the outcome.

This difference provides further opportunities for friction. A mathematician can offer work of great technical power to the physics community and be dismissed as having no connection to "reality": no testable outcomes. A physicist can offer work which reaches a desired goal by a magnificent leap of intuition, and be criticized for being sloppy.

There are many other ramifications of this difference in focus. For example mathematicians are more tolerant of apparently pointless exploration, as long as it conforms to internal standards of rigor. Physicists tend to be more relaxed about precision and more judgmental about significance. These differences cause problems in grant and paper reviews in border areas.

Efficiency Customs well-adapted to the subject should maximize return on resource investment. This means approaches seriously out of step with local customs may be counterproductive in some way. Alternatively, these customs may reflect adaptation to some influence other than the subject matter.

As an illustration we consider different levels of rigor expected in the two fields. Years often pass between an understanding satisfactory to physicists and a mathematical demonstration. Is the insistence on rigor a consequence of being sheltered from the demands of the real world? Or is it more efficient in some way? The history of mathematics reveals a lot of backsliding, but the predominent trend is toward greater rigor. Explaining this begins with another fundamental aspect of mathematics: since it is (usually) right the first time, it is not discarded. Over time it may become uninteresting or insignificant, but it does not become incorrect. As a result mathematics is an accretive activity.

In physics (and most of the rest of science) material must be checked and refined rather than simply accreting, and customs have evolved to support this. Duplication is tolerated or even encouraged. a great deal of material is discarded, and there is a strong secondary literature to record the outcome of the process. These activities use resources. Mathematics lacks many of these mechanisms: the payoff for working slowly and getting it right the first time is savings in the refinement process. In principle the same payoff is available to physics: if complete reliability were possible then the most efficient approach would be to seek it even at great sacrifice of "local" speed. But complete reliability is not possible, and an attempt to import this attitude into physics would be ill-adapted and counterproductive.

This adaptation has produced a vulnerability in mathematics. A group or individual can disregard the customary standards and seem to make rapid progress by working on a more intuitive level. But the output is unreliable. Mathematics largely lacks the mechanisms needed to deal with such material so this causes problems ranging from areas frozen up for decades, to unemployable students, to the outright collapse of entire schools of study [3].

As a second illustration of efficiency we consider the "fad" phenomenon in (theoretical) physics. It sometimes happens that an area becomes fashionable. There is a flurry of publication, with a lot of duplication. Then most of the participants drop it and go off to the next hot area. Physics has been criticized for this short attention span. But this behavior is probably adapted to the subject matter. First, the goal is development of intuition and understanding, and this is an effective group activity. Duplication in publication is like replication of experiments: several intuitions leading to the same conclusion increase the likelihood that the conclusion is correct. And after a period the useful limits of speculation are reached, and it is a better use of resources to move on than to try to squeeze out a bit more. However if all activity ceased after a fad then eventually all of physics would become unsuitable for further development. Different activities continue: experimentalists test the testable parts. Mathematicians clean up the logical parts. A few physicists remain to distill the material into review and survey articles. And after a period of solidification the area is ready for another round of theoretical development.

Mathematics has occasional fads, but for the most part it is a long-term solitary activity. The reviewing journal divides mathematics into roughly 5,000 subtopics, most sparsely populated. Mathematicians tend to be less mobile between specialties for many reasons: a greater technical investment is needed for progress; big groups are seldom more efficient; and duplication is unnecessary and usually discouraged. These factors tend to drive mathematicians apart. In consequence the community lacks the customs evolved in physics to deal with the aftermath of fads. If mathematicians desert an area no one comes in afterwards to clean up. There is less tradition of review articles: since the material is already right there is less sifting to do, and less compression is possible. Shifts of fashion may be an efficient behavior in physics, but they are not a good model for mathematics.

These considerations also suggest ways grant programs might be fine-tuned to mesh with cultural nuances. The physics group activity is often focused at conferences, while mathematical conferences are more oriented to communicating results. This suggests that mathematical conferences should (on average) be shorter and more frequent, while physics would benefit more from extended "summer institute" formats. Mathematical program officers might watch active areas for quality control problems, and sponsor physics-style review and consolidation activity. This might be a more effective use of resources than supporting the presentation of the newest results.

Publication Papers in pure math and theoretical physics look similar, treat similar subjects, and often reside in the same library. However publication customs and uses of the literature are quite different. Physicists tend not to use the published primary literature. They work from current information (preprints, personal contacts), and the secondary literature (review articles, textbooks). The citation half-life of physics papers is short, and there are jokes about "write-only" journals that no one reads. Duplication and rediscovery of previously published material are common. In contrast many mathematicians make extensive use of the literature, and in classical areas it is common to find citations of very old papers.

There are differences in the construction of the literature as well. In mathematics the refereeing process is usually taken seriously. Errors tend to get caught, and detailed comments often lead to helpful revision of the paper. In physics the peer-review process has low credibility. Reviewers are uninterested, and their reports do not carry much weight with either authors or editors. Published papers are almost always identical to the preprint version.

One view of these differences is that the mathematical primary literature is user-oriented: genuinely useful to readers. In physics it is more author-oriented, serving largely to record the accomplishments of writers.

These differences again reflect differences in the subject matter. The theoretical physics primary literature is not reliable enough to make searching it very fruitful. It records the knowledge development process rather than the end result. If material is incorporated into the secondary literature or some shared tradition then it is reasonably accessible, but it is often more efficient to rediscover something than to sift the primary literature. A consequence is that there is not much benefit in careful editing or refereeing. This leads to journals that are, in the words of one mathematician, "like a blackboard that must periodically be erased." In contrast, the mathematical literature is reliable enough to be a valuable asset to users.

The differences in literatures have led to differences in social structure. As noted above, mathematics has many sparsely populated specialties. More accurately these could be described as larger communities distributed in time and communicating through the literature. This works even though the communication is one-way, because the material is reliable. Less reliable material requires two-way give and take. As a consequence working groups in physics are more constricted in time, and appear larger because they are all visible at once. This also works the other way: a large working group with a lot of real-time interaction weakens the benefits of reliability, and in fact larger groups in mathematics often do become more casual about quality control. This in turn leads to a curious problem in the mathematical infrastructure. The leadership in the professional societies and top journals tends to come from larger and more active areas. As a result they tend to underestimate the importance of quality control to the community as a whole.

This analysis has applications to the structuring of electronic communications. Both mathematicians and physicists have become heavy users of electronic mail and preprint databases. Theoretical high-energy physics is particularly advanced due to the leadership of Paul Ginsparg at Los Alamos, and in that area the current published literature has become nearly irrelevant. If paper journals perish as a result, readers will lose the quality control, and authors will lose some credit mechanisms. In this area the quality control is marginal, and seems a small price to pay for the greatly increased speed and functionality. Authors may be briefly inconvenienced but new recognition mechanisms are already developing.

The needs of mathematics are different. If the physics example were followed too closely, and led to a significant decline in reliability, it would yield a literature seriously out of step with the needs of the subject. Sociological symptoms might include the demise of sparsely populated areas and an increase in size of working groups. New quality-control mechanisms would eventually evolve, but these would take time and are likely to be less satisfactory than the present literature-wide control. Re-adaptation of the social structure might take quite a long time. A mathematics-specific electronic publication model with greater emphasis on quality control seems to be called for [4].

Societal Influences There are strong outside influences on science. Some are accidental byproducts of other circumstances. For example the current underrepresentation of some racial groups and genders surely results from social forces unrelated to science. A more subtle example is given by Harwood [5], who argues that the old German ideal of the "universal scholar" led to a larger proportion of "mandarins" to "pragmatists" in early 20th century German genetics, as compared to the United States.

The more interesting influences are ones purposefully directed at science. There are inappropriate and obviously counterproductive examples like Russian genetics in the Lysenko era, Ayrian science in Nazi Germany, or church-controlled astronomy in Galileo's time. Some influences are appropriate in principle: society can reasonably expect some return on the investment, and is entitled to push science in productive directions. However these influences can still interfere with adaptation to the subject, and can be counterproductive. For instance Montgomery [6] suggests that plasma physics has been harmed by the forced march toward fusion. For a more subtle problem we note that it has been NSF policy for some years to encourage mathematicians to use computers. This is straightforward in applied work. It is harder to organize computation to provide the reliability crucial to pure mathematics. As a result quite a few mathematicians who wanted to use machines for more than e-mail and word processing have moved to applied areas. The machine/pencil dichotomy also seems to attract students to applied work. This shift of the entire field was probably not an intended consequence of the original program.

Many areas of science have been unusually free of societal pressure in the last fifty years: in the U. S. Vannevar Bush's "social contract" led to uncritical support of science as an abstract public good. In the Soviet Union it was often regarded as "production" and therefore intrinsically good. This era is ending [7]. As science policy becomes more demanding there is an increasing urgency to design it to mesh with the cultural structures that have proved effective in exploring nature.

Conclusion We have presented an analysis of the adaptation of culture and custom to subject matter in two scientific fields. Understanding the differences leads to conclusions about interdisciplinary work, professional ethics, electronic communications, science policy, and other "infrastructure" issues.

There are several cautions. First, this should be seen as explanations for observed cultural differences, not "proofs" that they must exist. Second, not all cultural differences are related to subject matter. Differences can come from societal influences, as discussed above, or from things like "founder effects" where personalities or circumstances of the formation of the field have left a lingering imprint. A final caution concerns the drawing of boundaries. Micro-adaptation produces cultural variation on fine scales. For example in experimental biology there are adaptations of researcher and organism to each another that drive diversity on a very small scale [8]. Consequently the strongest conclusions are limited to small scales and larger scale comparisons are limited to commonalties that transcend local variations. Pure mathematics and theoretical physics for instance, have significantly different commonalities. Applied mathematics is closer in spirit to physics, and -- curiously -- experimental physics has similarities to mathematics. Lumping together all of mathematics, and all of physics, would have given groups with enough internal diversity to overwhelm differences between the groups.

There are conclusions at the largest scale --- all of science --- just from the fact that there are cultural differences. Interdisciplinary workers should respect other cultures; there can be no detailed uniform code of ethics; science policy should mesh with cultural adaptations; area-specific nuances of publication should be preserved in the transition to electronic formats. Generally, a one-size-fits-all approach to any infrastructure issue will be suboptimal.

Finally, there are many other cultural divides in science resulting from differences in subject matter. There is the large science -- small science division mentioned in the introduction. Some areas (the genome project, x-ray crystallography) are primarily oriented to the production and analysis of data, while others (mathematics, theoretical physics) consist almost entirely of discursive argument. Subjects that require sophisticated use of statistics can be expected to differ from those that do not. Subjects like evolutionary biology and astronomy are oriented toward explanatory stories that organize observation, while the classical laboratory sciences emphasize testable prediction. Purely academic subjects differ from ones with significant commercial or national security interest. A great deal of worthwhile information should result from analysis of cultural adaptations to these differences.

Notes

[1] Background for this article appears in: A. Jaffe and F. Quinn, *Theoretical mathematics: a cultural synthesis of mathematics and theoretical Physics*, Bulletin of the American Math. Soc. Vol. 29 (1993) pp. 1--13. http://www.ams.org/bull/pre-1996-data/199329-1/Jaffe

[2] Not all mathematicians agree. For example W. Thurston (*On proof and progress in mathematics*, Bull. Am. Math. Soc. 30 (1994) pp. 161--177) suggests the primary goal of mathematicians should be *human understanding* of mathematics, rather than the production of reliable knowledge. On this basis he argues for heuristic argument as a basis for mathematical knowledge. This is, however, very much a minority view.

[3] See *Theoretical mathematics*, note 1.

[4] This issue is explored in more detail in F. Quinn, *Roadkill on the electronic highway: The threat to the mathematical literature* , Notices of the Amer. Math Soc. 42 (1995) 53-

56, http://www.ams.org/notices/199501/forum-quinn.html, and *A digital archive for mathematics* http://www.math.vt.edu/people/quinn/archive.html.
[5] Jonathan Harwood *Styles of Scientific Thought: the German Genetics Community 1900--1933* , University of Chicago Press, 1993.
[6] David Montgomery, letter to **Science**, vol. 269 (1995) 1328.
[7] R. Byerly and R. A. Pielke, *The changing ecology of United States Science*, **Science** 269 (1995), pp. 1531--2.
[8] See papers in the collection "The right organism for the job", J. History of Biology 26 (1993) 233--368.

Department of Mathematics,Virginia Tech
Blacksburg VA 24061-0123
September 1996
quinn@math.vt.edu