**Why Mathematics is Boring**

*John C. Baez*

VERY ROUGH DRAFT

Department of Mathematics
University of California
Riverside CA 92521
email: baez@math.ucr.edu

June 28, 2007

**Abstract**

Storytellers have many strategies for luring in their audience and keeping them interested. These include standardized narrative structures, vivid characters, breaking down long stories into episodes, and subtle methods of reminding the readers of facts they may have forgotten. The typical style of writing mathematics systematically avoids these strategies, since the explicit goal is "proving a fact" rather than "telling a story". Readers are left to provide their own narrative framework, which they do privately, in conversations, or in colloquium talks. As a result, even expert mathematicians find papers — especially those outside their own field — boring and difficult to understand. This impedes the development of mathematics.

## Introduction

In their research papers, mathematicians usually eschew narrative techniques designed to keep readers interested, since their main goal is not to "entertain" or even explain, but present logical arguments as efficiently as possible. While this makes a certain sense, it neglects the human dimension of mathematics. It neglects the fact that a piece of mathematics is almost useless if almost nobody understands it. But, before anyone can understand a piece of mathematics, they must first become *interested* in it. So, for a mathematician who wants to fully develop a piece of mathematics, discovery and proof are only the first steps on a longer road. The next step is getting people interested.

Unfortunately, mathematicians are not trained in this art. Indeed, their writing is famous for being "dry". There are exceptions, and these exceptions are worth studying. But it also makes sense to look to people whose whole *business* is getting people interested: story-tellers.

Everyone enjoys a good story. We have been telling and listening to stories for untold millennia. Stories are one of our basic ways of understanding the world. I believe that when we read a piece of mathematics, part of us is reading it as a highly refined and sublimated sort of story, with characters and a plot, conflict and resolution.

If this is true, maybe we should consider some tips for short story writers, and see how they can be applied — in transmuted form — to the writing of mathematics. These tips may sound a bit crass to mathematicians, or even readers of "serious" fiction. But they go straight to the heart of what gets people interested, and what keeps them interested, in a piece of writing.

Here are ten tips for short story writers, taken from a typical online guide [2], but listed in a somewhat different order:

- Write a Catchy First Paragraph
- Choose a Point of View
- Use Setting and Context
- Develop Your Characters
- Write Meaningful Dialogue

- Set up the Plot
- Create Conflict and Tension
- Build to a Crisis or a Climax
- Find a Resolution

Let us gauge a typical mathematics paper according to these guidelines, and see what we learn.

## Write a Catchy First Paragraph

We are constantly encountering texts; we don't bother reading all the way through most of them. Once texts were rare and precious. Now, in the era of the world-wide web, there is always too much to read. We must efficiently cull out most of the material vying for our attention. Often we base our decision on the first sentence or two.

Since most writers of short stories succeed largely on their sheer number of readers, and few people read stories because they need to, writers of short stories learn the importance of quickly grabbing the reader's attention. In a catchy story, each sentence makes the reader want to read the next. The first few sentences bear the brunt of this responsibility.

Mathematicians operate in a more forgiving environment, with guaranteed permanent employment for many. They can succeed with only few people reading their work. Consider two of the most famous mathematicians of recent years: Andrew Wiles and Grigori Perelman. How many of us have really read Wiles' proof of Fermat's last theorem, or Perelman's sketched proof of the Poincaré conjecture? Even among professional mathematicians, most are satisfied to know that a few experts vouch for these proofs' validity. So, instead of broadening their readership, mathematicians are mostly concerned with impressing other experts in their field.

Most mathematics papers begin according to a strict format. First comes a paragraph-long "abstract". This summarizes the main results of the paper, usually in language understandable only by specialists in the given field. Here is a typical example [1], randomly chosen from the main electronic database of math papers, the arXiv:

> First and second fundamental theorems are given for polynomial invariants of a class of pseudo-reflection groups (including the Weyl groups of type $B_n$), under the assumption that the order of the group is invertible in the base field. Special case of the result is a finite presentation of the algebra of multisymmetric polynomials. Reducedness of the invariant commuting scheme is proved as a by-product. The algebra of multisymmetric polynomials over an arbitrary base ring is revisited.

Few people would call this an attention-grabbing start. The passive voice, the ungrammatical second sentence ("broken English is the international language of science"), and most of all the density of technical terminology conspire to filter down the potential audience of this paper to the few experts who are eager to know more about polynomial invariants of pseudo-reflection groups.

An abstract like this is enough to make most non-mathematicians want to run away screaming. More important here is how trained mathematicians react. As a widely read dilettante, perhaps my personal reaction is worth noting. Different branches of mathematics have their own "fundamental theorems". I don't know which "first and second fundamental theorems" were being alluded to here. This makese me curious, but also intimidated. If the author eventually deigns to explain these theorems, I'll be happy, but if he doesn't I'll be frustrated. Experience has taught me that the odds are about 50-50.

I also don't know what a "pseudo-reflection group" is, but I can guess it is a generalization of something I already know and love: a "reflection group". Reflections groups describe the symmetries of regular polygons, Platonic solids, and their higher-dimensional kin. In particular, the Weyl group of type $B_n$ is the symmetry group of an $n$-dimensional cube. I know a bit about this group, and would like to know more.

I also know a bit about "polynomial invariants" of reflection groups. For the symmetry group of the $n$-dimensional cube, these are just polynomial functions on an $n$-dimensional cube that don't change when you rotate or reflect the cube. So, they are functions of $n$ variables, say $x_1, \ldots, x_n$, that don't change when you permute these variables or replace any variable with its negative. Here are two examples when $n = 2$:

$$f(x_1, x_2) = x_1^2 + x_2^2$$

and

$$g(x_1, x_2) = x_1^2 x_2^2$$

You can see that if you switch $x_1$ and $x_2$, or replace either $x_1$ or $x_2$ by its negative, these functions don't change. That's what "invariant" means.

I first became interested in polynomial invariants of reflection groups when I learned they have applications to geometry, topology and physics. The tale behind this is intricate and fascinating. Though reflection groups describe *discrete* symmetries, each one is closely linked to to a specific *continuous* group of symmetries. For example, the Weyl group of type $B_n$ is linked to the group of rotations in $(2n + 1)$ dimensions. Through a chain of reasoning too clever to explain here, this winds up implying that any polynomial invariant for the Weyl group of type $B_n$ gives a recipe for computing information about the topology of $(2n + 1)$-dimensional spaces! These recipes, called "characteristic classes" are also important for theoretical physicists studying fields in $(2n + 1)$-dimensional spacetimes.

Given all this, I would happily learn more about polynomial invariants of reflection groups — and even "pseudo-reflection groups", whatever those are. But, I know that polynomial invariants of reflection groups have been studied for a long time; my understanding must lag decades behind the state of the art. So, I can guess that the Weyl groups of type $B_n$ are mentioned merely because the author considers them an elementary example of the more general "class of pseudo-reflection groups" he is really interested in. This discourages me from reading this paper to indulge my interest.

The phrase "the order of the group is invertible in the base field" puts me on notice that the author is studying polynomial invariants that take values not in the rational, real, or complex numbers — the cases I'm most familiar with — but in significantly different number systems. Again this fills me with curiosity but also some trepidation. I don't know what "multisymmetric polynomials" are, and I am not sure I want to. Finally, the phrase "reducedness of the invariant commuting scheme" lets me know that to enjoy this paper, it would help if I knew more algebraic geometry than I do.

Overall, the main effect the abstract has on me is to inspire curiosity, but also hint that reading this paper will not satisfy that curiosity. Nothing suggests the author will bend over backwards to help out an amateur like me. So, I feel like not reading this paper.

But beware: if you think I'm trying to criticize this particular paper, you have completely missed my point. I think my reaction to this paper is similar to how *most* mathematicians react to *most* mathematics papers. For any given paper, I bet at least 95% of the mathematicians who see the abstract would decide it's not worth going on to read further. (It would be interesting to compile statistics on this. With the facilities of the arXiv, it should be possible to see what fraction of the people who look at a given abstract go on to download the paper. But, even looking at the abstract indicates a higher-than-average level of interest in the subject in question.)

In short, "attention-grabbing" is precisely what the typical beginning of a math paper is *not*.

## Use Setting and Context

In a short story, the reader is usually "located" as an observer to some scene of action, with a definite point of view — perhaps in a room somewhere, perhaps in some character's mind, or whatever. The story should quickly and unobtrusively establish this context.

In a typical math paper, setting the scene is usually done in the "introduction", which comes after the abstract. This section, usually between 1 and 10 pages long, explains the main results in a bit more detail than the abstract, and put these results in their historical and mathematical context.

It's very hard to appreciate the virtues of piece of mathematics without the necessary background. At the very simplest level, this requires understanding all the *words*: mathematics bristles with technical terminology. So, the introduction to a good math paper should set the scene as simply as possible, with a minimum of fancy vocabulary. Often this requires "watering down" the results being described — stating corollaries or special cases instead of the full theorems in maximal generality. Sometimes one even needs to leave out technical conditions required for the results to really be true. In this case, one should warn the reader that one is doing so.

While most mathematicians could do better at explaining terminology, they are aware of the issues I just mentioned. Reviewing definitions is common practice, and phrases like "under mild assumptions" abound in the introductions to math papers. Where mathematicians really fall down is in explaining the *context* of their work. New results rely on previous ones not only for their proofs, but for their *interest*.

Here is the introduction to the paper whose abstract I already analyzed. I'll interrupt it with a few small remarks, and then comment in more detail.

> Fix natural numbers $n$ and $q$, and a field $K$. Apart from Theorem 2.7 and Section 5, we shall assume that $n!q$ is invertible in $K$, and assume that $K$ contains a primitive $q$th root of 1. Denote by $G = G(n, q)$ the subgroup of $GL(n, K)$ consisting of the monomial matrices whose non-zero entries are $q$th roots of 1. The order of $G$ is $n!q^n$, and as an abstract group, $G$ is isomorphic to the wreath product of the cyclic group $C_q$ of order $q$ and the symmetric group $S_n$; that is, $G$ is isomorphic to a semi-direct product $(C_q \times \cdots \times C_q) \rtimes S_n$.

This first paragraph is on the dry side even for mathematics papers, but it's not unusual. It is not welcoming; it assumes the reader *already wants to know* the results this paper contains, and dives right in, defining a class of groups $G = G(n, q)$. With careful attention, a non-expert like myself could guess that these are the "class of pseudo-reflection groups" mentioned in the introduction. This is confirmed in the next two sentences:

> Consider the natural action of $G$ on $V = K^n$. Since $G$ is generated by pseudo-reflections, [....]

However, the author does not expend much energy to make this clear. It would have been easy to do so, simply by starting the introduction with a sentence like "In this paper we study a class of pseudo-reflection groups $G(n, q)$ defined as follows". Without a cue like this, the non-expert can easily miss what's going on.

The sentence beginning "Since $G$ is generated by pseudo-reflections..." also informs me that I'm unlikely to learn what a pseudo-reflection is here: the time for defining them has arrived, but no definition is given. So, my curiosity about these will be frustrated.

Continuing:

> Since $G$ is generated by pseudo-reflections, by the Shephard-Todd Theorem [Shephard-Todd] (see [Chevalley] for a uniform proof in characteristic zero, and [Smith] for the case when $\mathrm{char}(K)$ is positive and co-prime to the order of $G$) the algebra $K[V]^G$ of polynomial invariants is generated by algebraically independent elements.

While some mathematicians would find this sentence forbidding, it actually comes as a great relief to me, since it is the first one that explicitly mentions a result I know. Let me explain what the Shephard–Todd theorem says in a special case. Suppose $G$ is the Weyl group of type $B_2$ — that is, the symmetry group of the square. The algebra $K[V]^G$ mentioned above is just another notation for the polynomial invariants of $G$. In the case at hand, these are polynomial functions on a square that don't change when you rotate or reflect the square. I mentioned two examples:

$$f(x_1, x_2) = x_1^2 + x_2^2$$

and

$$g(x_1, x_2) = x_1^2 x_2^2$$

The Shephard–Todd theorem says you can get all the rest from these two. More precisely: if you have *any* polynomial in the variables $x_1$ and $x_2$ that doesn't change when you permute the variables or replace a variable by its negative, you can express it as a linear combination of functions of the form $f^n g^m$. Even better, the Shephard–Todd theorem says there's just just one way to express it in this manner! — this is what "generated by algebraically independent elements" means in the above passage.

I had thought that Chevalley proved this result; I didn't know it was called the Shephard–Todd theorem. So, I just learned something from reading this introduction. By glancing at the bibliography, I can see that Chevalley's paper appeared in 1955, while Shephard and Todd's paper dates back to 1954. And, I can see that in 1985, someone named Smith generalized this result to number systems significantly different from the rational, real and complex numbers — so-called "fields of positive characteristic". Smith's paper is entitled "On the invariant theory of finite pseudoreflection groups." So, if ever want to know what a "pseudo-reflection group" is, I now have a reference to try.

These are facts worth knowing if I ever get around to studying this subject further. More importantly, they make the author's work into an episode of a larger ongoing story. This is not primarily a story about mathematicians: it's a story about mathematical entities, and how they are gradually becoming better understood.

However, all this information is packed into a single sentence, which can only be appreciated by readers already familiar with polynomial invariants of reflection groups!

Next:

> Now consider the diagonal action of $G$ on $V^m = V \oplus \cdots \oplus V$, the direct sum of $m$ copies of $V$. The algebra $K[V^m]^G$ is no longer a polynomial ring if $m \geq 2$ (and $G$ is not the trivial group). In the present paper we show a very short and simple argument that yields simultaneously the generators of $K[V^m]^G$ (first fundamental theorem) and the relations among these generators (second fundamental theorem). Our main result is Theorem 3.2, which provides an explicit finite presentation of $K[V^m]^{G(n,q)}$ in terms of generators and relations. In the proof we apply Derksen's degree bound on syzygies [Derksen] and ideas of Wallach and Garsia [Haiman].

Since I understand the notation and terminology here, I am happy: I now know what the author means by "first fundamental theorem" and "second fundamental theorem"! You may recall that when I described a special case of the Shephard–Todd theorem, my description consisted of two parts:

- Every polynomial invariant for the symmetry group of the square can be built up from two functions $f$ and $g$. In this situation, the process of building up fancy functions like $f^2 g + 3fg^3$ is usually called "generating", and the building blocks $f$ and $g$ are called "generators".
- Every polynomial invariant can be built up in a unique way from the functions $f$ and $g$. In other words, no equations like $2f^2 g = g^2 + f^4$ are true. In this situation, we say there are no "relations" among the generators.

From what the author writes above, I can tell that the first part, which deals with generators, is what he is calling the "first fundamental theorem". The second part, which deals with relations, is what he is calling the "second fundamental theorem".

> In the special case $q = 1$ we have $G = S_n$, and $K[V^m]^{S_n}$ is the algebra of multisymmetric functions, which received much attention in the literature (see the references in Remark 2.6 (ii)).

This sentence also makes me happy, because it explains the phrase "multisymmetric polynomials," which appeared earlier in the abstract. I will not bother to explain the term here.

If we were to compare this introduction to a short story, the best analogy might be a tricky detective story. Clues from later in the text shed light on secrets hidden in earlier portions. To understand what is really happening, the reader must read back and forth until the meaning gradually oozes forth. This is quite typical of mathematics papers.

The introduction concludes with, some helpful hints for readers wanting to know how the paper fits into the context of earlier work:

> Our approach gives new insight even in this special case, especially by its simplicity and transparency compared to the other approaches in the literature. In addition we gain some technical improvements in the known results. We mention also that no finite presentation of the algebra of multisymmetric functions appeared in prior work (apart from the case of $\mathrm{char}(K) = 2$ studied in [Feshbach]).
>
> Another interesting special case is when $q = 2$, and the group $G$ is the Weyl group of type $B_n$. The generators of $K[V^m]^{G(n,2)}$ were determined in [Jerjen] and [Hunziker]. To the best of our knowledge, the relations have never been considered in the literature when $q > 1$.
>
> The fundamental relation appearing here can be deduced from the theory of trace identities of matrices. This observation leads to the corollary that the $GL(n, \mathbb{C})$-invariant commuting scheme is reduced, see Theorem 4.1.
>
> The present paper joins the content of our preprints [Domokos1] and [Domokos2] (some digressions from the preprints have been omitted). In addition, in Section 5 we adjust our method for arbitrary base rings and clarify and strengthen the known results in this case. In particular, in Theorem 5.5 we give a new characteristic free (infinite) presentation of the ring of multisymmetric polynomials.

Note, however, that no *reason* for generalizing the original Shephard–Todd theorem is ever given. The original theorem is part of a fascinating web of ideas involving geometry, topology and physics. Does this web extend to include the generalizations? I can't tell.

To summarize: the introduction of a typical mathematics paper does indeed "set the scene," but in a way that's utterly incomprehensible to non-mathematicians, and — more importantly — only understandable by non-specialists if they exert themselves and know enough to pick up the clues scattered here and there. The broader context, which answers the question *why does any of this matter?*, is often left unspoken.

I already mentioned my guess that at least 95% of the mathematicians who see the abstract of a given paper will not read further, because the abstract does not grab their attention. I suspect that of those who go on to read the introduction, at least 90% will stop there, because the scene has not been set in a way they can understand. To the mathematician who understands its context, a paper may be very exciting. Without enough context, it will be boring.

## Develop your Characters

If a mathematics paper is a kind of story, the "characters" must be the mathematical entities involved in this paper. Some of these characters are more important than others; there are usually just a few heroes (and sometimes villians). For the paper to be enjoyable, the main characters must be introduced in a way that marks them as special and highlights their already known properties: their "personality".

In the paper discussed above, the hero is a certain "class of pseudo-reflection groups". It makes its entrance already in the first line of the abstract:

> First and second fundamental theorems are given for polynomial invariants of a class of pseudo-reflection groups (including the Weyl groups of type $B_n$), under the assumption that the order of the group is invertible in the base field.

Note however that it enters masked: there is no way from the abstract to tell precisely *which* class of pseudo-reflection groups will be discussed. We must commit to reading the introduction to find that out. The hero lets his mask slip in the first paragraph:

> Fix natural numbers $n$ and $q$, and a field $K$. Apart from Theorem 2.7 and Section 5, we shall assume that $n!q$ is invertible in $K$, and assume that $K$ contains a primitive $q$th root of 1. Denote by $G = G(n, q)$ the subgroup of $GL(n, K)$ consisting of the monomial matrices whose non-zero entries are $q$th roots of 1. The order of $G$ is $n!q^n$, and as an abstract group, $G$ is isomorphic to the wreath product of the cyclic group $C_q$ of order $q$ and the symmetric group $S_n$; that is, $G$ is isomorphic to a semi-direct product $(C_q \times \cdots \times C_q) \rtimes S_n$.

To an insufficiently expert reader, this way of starting the paper will inevitably seem dry as dust — an unmotivated barrage of assumptions and notation. As mentioned earlier, part of the problem is that the author fails to mention what must seem obvious to him: the groups $G(n, q)$ are none other than the "class of pseudo-reflection groups" the paper is all about!

After penetrating this ruse, we can note something very important: the *tendency of mathematics to singularize the plural*. What is really a *class* of groups $G(n, q)$, depending on the numbers $n$ and $q$, is for most of this paper treated as an *individual* group $G$. The shift in perspective begins in the first sentence, where the author tells us to "fix" natural numbers $n$ and $q$ — that is, pick particular ones, without saying which. So, by the time we reach the sentence "Denote by $G = G(n, q)$ the subgroup...," we are discussing a *particular* group $G$, a typical representative of the class of groups under discussion.

While this sort of move has been thoroughly analyzed by logicians, and there is nothing mysterious about it from a purely logical point of view, its relation to the art of story-telling may have been neglected. It is harder to imagine or sympathize with a "class" of entities than a particular representative of that class. When listening to a story, we prefer to imagine one person doing something, not a mob of similar people doing similar things. Even authors of the crudest sort of politically engaged fiction, seeking to depict the "plight of the working class", know enough to tell their story about a particular member, not the the whole class all at once.

## Create Conflict and Tension

The "conflict" in a mathematics paper is usually the struggle to know — often manifested in the struggle to prove something. As Piet Hein noted,

Problems worthy of attack
prove their worth by fighting back.

The best known epics in mathematics, like the long saga of Fermat's Last Theorem, gain their interest from the way *truths can resist being known.*

### Find a Resolution

The conclusion of a math paper should set our feelings at rest by assuring us those problems that have been solved have indeed been solved, while reminding us of those that have *not* yet been solved.

## References


[1] M. Domokos, Vector invariants of a class of pseudo-reflection groups and multisymmetric syzigies, available as arxiv:0706:2154.

[2] K. Kennedy, Short stories: 10 tips for novice creative writers, available at `http://jerz.setonhill.edu/writing/creative/shortstory/`.