THE PROSPECTUS, AND HOW TO WRITE ONE

FRED BOOKSTEIN

A prospectus [auf Deutsch, *exposé*, but the English word "exposé" does not mean the same thing] is a document written for a small readership: your adviser(s), your department chair, and sometimes (especially for dissertations) the committee that has been set up to guide you and to read the manuscript that you eventually intend to write.

The prospectus has the status of a draft contract between you and your department. You agree to try hard to do the work as you have set it out in the prospectus; they agree to approve that work for your degree no matter what the facts turn out to be (i.e. whether or not your hypothesis was tenable) as long as you carried out the project competently and wrote about it in passable English.

A baccalaureate thesis or a master's thesis should try to do only one thing that is new. If you are trying out a new machine, don't try at the same time to invent a new statistical method to handle the data it gives you; if you are trying to develop a new statistical method, don't try to establish some pathbreaking scientific hypothesis with its aid. Following this advice tends to lower the risk of abject failure.

The prospectus for a diploma thesis should be about four pages; for a master's thesis, likewise about four pages; for a doctoral dissertation, no more than ten pages.

The prospectus is a scientific communication. It can benefit from diagrams, from photographs, and from any other teaching tool that you can think of. Some of these might well involve copyright violations that would not be legal in a real publication, but a prospectus is not a publication, and copyright violation is to be encouraged, especially when the hypothesis itself is not your own.

Basic structure of a prospectus

In its basic outline the prospectus resembles an abstract, a poster, or any other conventional scientific communication, but it differs epistemologically in that a systematic part of its text is in the future tense or the subjunctive voice.

For readers who haven't taken my course on Science Writing in English, here are the parts of an abstract (and thus, at least within the confines of this memo, the parts of a prospectus as well).

Introduction: Background, importance, reason for narrowing of focus Methods and materials: Sample(s) or subsample(s), machines involved in data gathering or standardization, measurements used (and not used), statistical methods, other choices

Results: Means, contrasts, trends, patterns; null-hypothesis significance tests (if you must); confirmations; surprises, agreements, disagreements.

Discussion: Strong inferences, if any; the Introduction in reverse—implications at small and large scales.

This outline also serves to guide you if you are writing the sort of paper Thomas Kuhn would call "normal science," a project that conveys absolutely no element of surprise: one that has a hypothesis that arises naturally out of some existing literature, that measures competently, and that finds the expected findings. No prospectus should promise surprise: that would be too risky. You can *hope* for

surprise, but you cannot say you expect it: for in that case it would not be a surprise. Your project must appear to grow organically out of things that "we" (meaning the community of people who are interested in the same kind of question) already agree on. You are a student, and would not be permitted to move the center of your field no matter how brilliant you turn out to be.

Remember, in short, that the thesis is not primarily a contribution to knowledge. That contribution is mainly the task of the papers, if any, that are derived from it. The thesis is mainly a part of your *training*, designed to let your evaluators pass you without anticipating any embarrassment. A good prospectus helps them do their job as much as it helps you do yours.

The modifications to this outline that apply to the prospectus arise from the epistemological problem mentioned above. You are required, in effect, to write the Introduction, the Materials, and the Discussion of a scientific publication without having in hand either the Materials or the Results.

Introduction

The purpose of the introduction, broadly speaking, is to explain to the reader how s/he should think about the hypothesis at the core of your proposal. We are all aware that you are a student at this point, not a colleague, and that from that student status a particular problem usually (but not invariably) arises. Realistically speaking, you are unlikely to be adept at weighing the importance of a hypothesis, or for that matter at phrasing one in the first place. In fact, probably the hypothesis of your thesis isn't actually yours, but came from somewhere else.

My courses here teach a good deal of the classic literature of hypothesis-formation, from C. S. Peirce's abductive syllogism ("The surprising fact C is observed./But if A were true, C would follow as a matter of course./Hence there is reason to believe that A may be true") through John Platt's "strong inferences" (experiments that are capable of supporting any of at least two contrasting hypotheses, the selection to depend only on the facts that have not yet been gathered) right down to the bottom of the barrel, which is null-hypothesis statistical significance testing. Your hypothesis is almost certainly not your own in any important way. It may have been your mentor's hypothesis all along (especially if that mentor is also your source of financial support); it may arise from somebody's puzzlement at the contradiction between two published papers; it may be a logical "next step" in a research programme known only to its programmer; or it may instead represent idle curiosity, speculation about a pattern observed in one place as it might apply, or not apply, in another.

The introduction to a prospectus can admit the truth of none of these possibilities. It must instead make your choice of topic appear to be interesting to a broader readership than just your mentor. It will, for instance, help determine your next academic position (whether graduate student or newly fledged post-doc); it will, if you are lucky, supply the Introduction to the paper(s) you pull out of the thesis or dissertation, papers that will be reviewed by a much more variable sample of "peers"; it must cover you and protect you regardless of whether the hypothesis you were checking is in fact true or false, whether the methods you chose (or were chosen for you) are sound or unsound, whether the promised research material or study samples did or did not materialize.

Thus the best introductions, in my view, are the ones that begin most generally, along the lines of "Modern Europeans are endlessly fascinated by the evidence for recent human evolution." (This prototype, of course, is from anthropology.) Or, "We are all aware of the variability of human attractiveness." Or, "Here it is already 2008, and still nobody knows what species to assign the strange fossils from Flores."

The Introduction then goes on to embrace a bunch of citations to earlier publications, going as far back in time as you can, a century or more if possible. By "a bunch" mean approximately ten, and the order in which they are cited needs to be calculated, beginning with the most widely-known and ending with the most relevant (which may be a separate, non-overlapping sublist). Our field has made progress in some aspects (e.g., coherent quantification of the output of machines, like surface scanners or spectrographs) and very little progress in others (e.g., unequivocal assignment of paleospecies). If at all possible, try to show that there is a problem accessible of comprehension to a broad variety of readers, not only those who share the special skills, interests, or ideology of your research group.

At this point, roughly midway through the Introduction, you change rhetorics. You now decide that your reader is likely to be specialized in a particular jargon, and you make free with that jargon in order to enumerate the virtues of the few (two to five or so) most immediate sources of inspiration for your work. Each of these sources should be described in a pithy sentence or two explaining what the author(s) think he/she/they accomplished and what you are planning to do with this knowledge in the context of 2008 and the claims of the other few seminal papers or books in this little review.

You need to have thought these matters through carefully. It is OK if your main sources disagree—in fact, it is better for you in the end, since then you have more chances to agree with one of them—but you need to have mastered them well enough to narrate their claims and their disagreements cogently enough to pass review by an expert in *precisely the same papers*. Also, as your reasoning will need to draw specifically on concepts and contents of each of these references in the course of the Results text below, they will need to be built into a scaffolding of some sort, one that won't fall down. For instance, either they need to be talking about the same thing, or you need to be able to tell how their topics (or operationalizations) differ. Either they show progress, or instead confusion; you need to know which. Your work will need to articulate back to these sources, so you need to respect them, rely on them as models, and generally consider them wise.

You are now at roughly the bottom of page 1 (for diploma thesis and master's thesis prospectuses) or page 2 (for the dissertation prospectus). It is time to turn to the second part, Methods and Materials.

Methods; Materials

In general, theses in the "inexact natural sciences," like medicine or anthropology, fall into two piles. In one pile, the Methods are familiar (Procrustes shape coordinates, etc.) but the Materials are not yet accrued (the people in a sample survey, or the next fifty patients needing maxillary restorative surgery at the Dental School). In the other pile, the materials are known (the set of Avar skulls from the NHM collection) but the methods are uncertain. It is rare to have a thesis for which both the materials and the methods are problematic—at least, for which both are known in advance to be problematic—and even more rare for both to be under adequate control: in this case there would not be a thesis, only a homework assignment. I cast the following notes as if the Methods are standard but the materials uncertain or problematic.

If the methods you are going to use are standard, say so. "I expect to gather landmarks specimen by specimen using the Microscribe in Room 2.024, analyze these locations via their Procrustes shape coordinates together with Centroid Size, examine their relative warp scores for evidence of clustering and then assess the dependence of shape on latitude, longitude, and δO^{18}." This is not the place to show that you can transcribe accurately—please do **not** attempt to review the

foundations of morphometrics or multivariate statistical analysis here. (There will be a place for those expositions in the finished document.) For this paragraph, assume a reader competent in the methods you yourself intend to use. To help the other readers, feel free to cite one authoritative source per method, preferably a textbook. There is no point in "reviewing the literature" here, only in reassuring the reader that you know how your hypotheses will eventually be confronted with your data.

With standard methods, far more interesting is the Materials part of this section. You may be selecting from an existing collection, or you may be interviewing drunks in Vienna bars, or you may be intending to go around the world measuring one or more fossils in every capital city. Unlike the Methods, this is the part of your prospectus that your readers do *not* already know. What they know, instead, is a long lugubrious list of all the ways that well-meant plans for this stage of research can go awry. You can't get random samples of Viennese; or, your machine is known to work only in broad daylight; or, you don't yet have permission to subject the NHM samples to destructive testing; or, people sometimes lie when you ask them questions; or, ... You need to assure the reader of your general competence and skepticism at this data-gathering phase of the work. It is here that statements are useful along the lines of "every twentieth sample [or: subject] will be remeasured [or: asked a supplemental series of questions]," or "obviously damaged femurs will be omitted" or, best of all, "quality control at this stage of investigation will follow the principles of [citation here], who reported a root-mean-square test-retest error of 5.3 on the scale we are using." If the task of the Methods section was to confirm that you have taken the necessary courses, the task of the Materials section is to indicate that you know what you're getting into, more or less, and that you have enough general intelligence to work your way around the haphazard appearances of a range of ordinary logistical problems standing between you and your results.

A particular aspect of this reassurance task is that you show us (the readers) that you understand how to set out the protocols that govern a study (the written rules of sample selection, inclusions and exclusions, and the like) and that your understanding of these appear realistic. You don't have the space to show these protocols in detail, but you would do well to say something like "Questions will be administered in a randomized order" or "A rigorous sequence of digitizing operations will be posted at every digitizing station." You are attempting to convince your reviewers that you are shrewd enough to be permitted to proceed; this persuasion does not require many words, but it does involve a good deal of careful thought prior to the writing.

You are now at roughly the bottom of page 2 (for diploma thesis and master's thesis prospectuses) or page 5 (for the dissertation prospectus).

Results

Of all the sections of a prospectus, the Results section is the farthest from the contents of a conventional scientific paper as you have grown to know them. You have to write about the results as you *expect* to find them, and you have to justify those expectations. Nobody will hold you to Zoroastrian standards of prophetic accuracy, but you need to show that, to put it bluntly, you are capable of recognizing when data has met the Interocular Trauma Test—that when a signal hits you squarely between the eyes, you will notice it.

In this part of the prospectus, then, you go *back* to the small list of main sources you summarized in the introduction, and now divulge a few more crucial details about each: namely, the way in which these worthy forebears of yours managed to *ground* the empirical inferences that, in the event, actually flowed from their

data (along with reminders to your reader of how your data design does and does not accord with theirs). For instance, "Bookstein's corpus callosum paper of 2002 showed the correlation of structure and function by examining the subsamples of his subjects that were extreme on a singular warp, and showed that they differed systematically on long lists of psychological measurements; if I find a significant relation between structure and function in my data, I will do the same to tease out its meaning in ordinary univariate form." Or, "Smith et al. (1906) found a regression coefficient of +0.30 for males and -0.25 for females. I will explore various combinations of mediating variables and paths in an effort to confirm her findings; if the change over the intervening century seems irremediable by statistical magic, I will search hard in contemporary records to ascertain additional factors that might account for this difference." Or, "The principal competing hypothesis to mine is Johnson's, dealing with age of the woman at first birth. I will compute a formal likelihood ratio test of the support of the data for his model over mine."

In this section, as in the Methods section, there are standard approaches to exceptions that can be invoked merely by mentioning them. Missing data will be noted and treated as follows: ... . Outliers appearing in scatterplots will be remeasured before the main analysis goes forward. Questionnaires with responses of all-0 or all-10 will be set aside.

Perhaps the most powerful single component of this section is what is called, in other contexts, a *pilot study*, a demonstration in the small of what you are proposing to do in a larger arena. You can show an image capture of one fossil, along with landmarks (hand-placed if necessary) on its surface; or, you can show a photograph of one subject (your mother) sitting in front of a computer screen bearing a reasonably relevant-looking mockup of a data form half-filled-in; or you can demonstrate that sitting at Ruberl you encounter at least three people per hour who would meet your selection criteria, or that the AKH reports 1,000 new patients with disease X per year, of whom you will need 1 per month for your sample.

You have arrived at the bottom of page 3 (for diploma thesis and master's thesis prospectuses) or page 8 for dissertation prospectuses. There is one section left, the Discussion.

Discussion

While this section is important, it is in many ways the easiest to write, especially if you have followed my hints over the preceding three sections. In an ordinary paper or poster, the Discussion section walks "backward up the ladder" that you walked down in the Introduction: first turning to a critique of your own prior work, modestly emphasizing its inadequacy and uncertainty; then to a confrontation of what you (expect to) find with the literature out of which your project grew; finally, back at the level of the larger scholarly community or even the general educated public, a statement of the implications of your findings if they turn out the way you expect.

You are not responsible for the truth of any of this, of course, as you have not measured (much) data; but that should not stop you from emphasizing the uses to which your findings might be put. "It is reasonable to expect that we could contribute considerable insight to the problem of the familial origin of attractiveness rating behaviors, which has been left in some confusion by the disagreement between Grammer and ... about ... . If our methods support the findings of Grammer in this somewhat different study design, it must be because ... . On the other hand, if we support Freud et al., it would be natural to follow up immediately with a measurement of the other jaw." "Depending on the dominant few terms in the discriminant function, future projects by the members of this work

group might emphasize new models and measures of the exocranium, or instead the endocranium, where prior work is somewhat less persuasive and the coverage more spotty, as I noted on page 1." "If the factor analysis I have sketched confirms the Big Five in this novel context, we will turn to the simplifications that shorten the questionnaire while preserving construct validity. If, as seems more likely, our factors turn out to be oblique to Costa's, we will extract the areas of greatest disagreement as the domain within which to go back to the raw video records for additional subjective coding. Such reworking of the archives, however, falls outside of the particular thesis work being proposed here."

In these ways, you have returned to the concerns I sketched for you several pages above, in the advice connected with the Introduction. The end of the discussion is not for the experts in your jargon and your laboratory protocols and your favorite arguments at colloquia; it is a reversion to the broader intellectual context of your department and your university, in a search for relevance of the specific project to the larger enterprise in which we are all engaged, to wit, the modern internet-based academic tradition. If you are skilled and fortunate, some of your promises here in the prospective Discussion will actually come true, and will turn into new projects, or new jobs, once you have finished the thesis whose pre-approval is the goal of all this arguing.