The 'synopsis': one size fits all?[1]

Frode Olav Haara and Kari Smith

Høgskulen i Sogn og Fjordane, Universitet i Bergen, Norway

In this essay, a doctoral candidate describes his initial searching process on how to write the synopsis for his dissertation and how this was influenced by the lack of a tradition as has been established for monographic dissertations. Most of the text is written by Frode Haara, but this note includes comments from Kari Smith, the supervisor, in italics, together forming a reflective dialogue. Comments from the translator Jan Vermaat are in footnotes.

Introduction

In September 2009, I had come to the stage of my doctorate job that it was time to start work on the synopsis. I had been a fellow student since 1 January 2007, with funding from my employer Høgskulen in Sogn og Fjordane (HiSF) and admission to the PhD educational program at the Faculty of Psychology at the University of Bergen (UiB). Here written guidance was available on writing an article-based doctoral dissertation. This work should consist of three articles for publication in peer-review-based journals, as well as a synopsis.

Through my affiliation with the Faculty of Psychology at UiB, I also joined the Western Norway Graduate School of Educational Research (WNGER). This research school was a joint project of UiB, Bergen University College (HiB), Norsk Lærerakademi (NLA), Stord / Haugesund University College (HH) and HiSF, with administrative and academic management at UiB. Within this research school, active participation of the PhDs in seminars was a key principle.

In September 2009, HiSF hosted such a seminar, and the participants were challenged to contribute a post related to our doctoral thesis. The status of my progress at that time was that I had one of my three articles accepted and two had just been submitted for peer review. I had begun to toy with the idea of starting work on the synopsis, but did not really have anything that I thought interesting for the research school seminar. I discussed this a little with my supervisor (coauthor Kari Smith) and she challenged me to remain close to what I was busy with at the moment. Among the ideas I presented to her was also a future-oriented topic: What is a synopsis in a doctoral dissertation and how should it be? In an attempt

[1] This is an 'actively revised' translation after a first round with Google Translate by Jan Vermaat from a paper in Norwegian, which appeared in UNIPED (2011) 34: 79-86. Interestingly, the software alternatingly translated 'kappe' literally as throat, jacket, capping, cape, cloak hood, or gown or similar. I stick to synopsis. Lovely examples of machine translations are: 'The cloak will lift and discuss ethical issues at work', or 'The cloak should be nothing more than a summary or a summary (actually: noe mer enn sammendrag eller oppsummering).' This accidental poetry made the translation actually fun to do. I have tried to stay as close as possible to Frodes text, although I found him pretty 'wordy' in places. Here I tried to restrain myself, but I have still repeatedly shortened the text. This is compensated by my foot notes.

to elucidate these two questions, we agreed that precisely this would be a valuable exercise for myself and my fellow PhDs and supervisors in the research school.

In Norway, the monograph has been the traditional form for a doctoral dissertation. The use of a system with a number of articles, and an associated synopsis, is relatively new. Using a synopsis - both in structure, guidance and priorities related to work with a dissertation - is therefore yet without a widely accepted template have been established. A professor with ample experience in guidance of PhD candidates (invited as lecturer at the seminar) told me: "We have no tradition of writing a synopsis and its guidance and as a supervisor I experience this now as challenging in relation to my own guidance. And I'm not alone." This means that both supervisors and doctoral candidates must make choices without having a tradition to fall back to.

In the following, I tried to compile my presentation into a coherent text, which actually shows how I tried to approach the task of writing a synopsis. In addition, my supervisor, Kari Smith, has submitted her comments (in italics). This also provides a supervisor's perspective on starting on the synopsis.

Synopsis - some attempts at definitions
A basically easy way to get an impression of what a synopsis is meant to be is to see what educational institutions in Norway that offer doctorate have written down as their interpretation. Here are some examples of definitions I found (all online material is retrieved in the period 5-14 September 2009).

Former School of Business and Administration in Bodø (HHB) (2008)
If the doctor chooses to an article-based thesis, a synopsis (kappe) must be written, which represents an independent effort, which will document the whole of the dissertation and summarize the issues and conclusions presented in the articles. The summary of the thesis should not only summarize but also synthesize the issues and conclusions presented in the articles in an integrating scientific perspective, thus documenting the coherence of the dissertation. A summary must be included of the thesis's contribution to the research field - both practical and theoretical. The synopsis will also explain the methodological choices used in the articles, in addition to what is explain in the separate papers.

University of Oslo (UiO) (2009)
The final work on the synopsis should be carried out towards the end of the doctoral thesis. However, the synopsis should also serve as an outline for the work so that an overall perspective is available early in the work process. At the same time, it is natural that the synopsis is modified along the way in relation to how the shape and content of the articles evolve.

1. The synopsis will summarize and synthesize the issues and conclusions presented in the separate papers so that the dissertation appears as a whole. The synopsis will present the results of the individual papers in such a way that their ‘internal coherence’ relationship is visualised. The synopsis can thus help shaping connections between individual findings and can invite to discussions at a more theoretical level. Complexity and nuance shall be discussed in the light of methodological, science-theoretical and theoretical issues[2].

[2] Personally, I would not ask my PhD student to seriously consider the full bearing of this sentence too much. What a job!

2. The synthesis must contain necessary theoretical and methodological assessments related to doctoral work. The synopsis will also explain where concepts or elaborations on different themes are found in the dissertation.
3. The synopsis will highlight and summarize the thesis's contribution to the relevant research field.
4. The synopsis will discuss ethical issues of relevance for the work.
5. The synopsis shall contain updates when necessary. This may be necessary when major changes have occurred after publication of one or more individual papers.
6. The synopsis must be written by the doctoral student alone.
7. The synopsis should normally have a range of between 60-70 pages.

Norwegian University of Science and Technology (NTNU) (2009)
The dissertation can be delivered as a major collective work (monograph) or as a collection of articles. If it consists of a collection of articles, it should normally be 3-5 papers in addition to a compilation (synopsis).

After having thought over these three examples, I still wondered what a synopsis was meant to be in practice. Is it supposed to be a summary? A compilation? A synthesis? An overview? A digest? Is there any difference between the terms used? Some quick dictionary checks[3] offered several explanations.

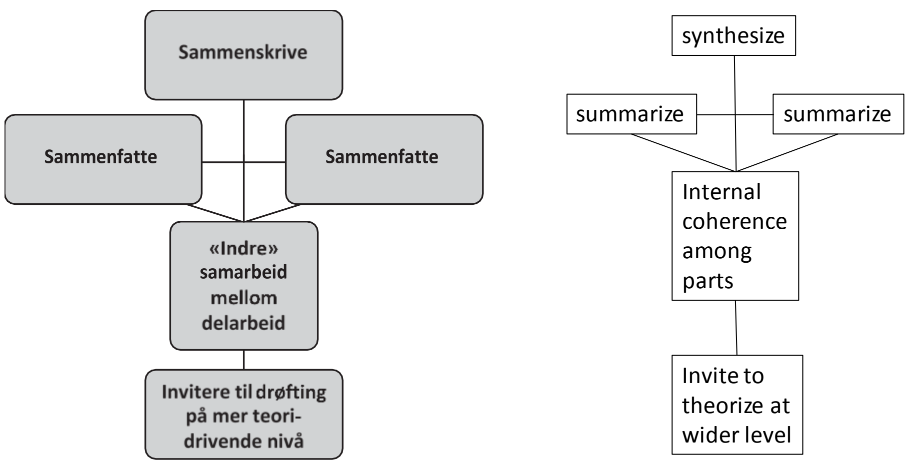


Figure 1. An attempt to visualise the form and function of the synopsis. Left the original Norwegian, right and English interpretation.

The function of the synopsis

Because I still lacked an unambiguous conception of the meaning of a synopsis, I changed focus from definition to function. I therefore recaptured the above seven concrete points in UiO's guidelines. By re-reading these with a functional view, both the shape and content become more concrete. The synopsis should be more than a summary. In a purely visual sense, therefore, the synopsis should be seen as a compilation of the project, where compilation can be interpreted to include summary, compilation and overview of the independent articles that are part of the dissertation project. The compilation requires the visualization of (possible) "internal" relationships among the articles, thus inviting discussion of the

[3] I do not repeat and try to translate the words and their explanation he reproduced in Norwegian here, it is too language-specific to be useful, and anyone can check a dictionary. It is not funny either.

research results against the theoretical basis of the dissertation at a level beyond the theoretical basis[4] of the individual article. The synopsis invites to discuss the independent research results in relation to each other and in context, but also in relation to the theoretical basis chosen for the entire dissertation project. This I have visualized in Figure 1. I then tried to project this conception of what a synopsis should be and how it should be constructed upon the structure of five paper-based theses that had already been approved. The five theses that I compared came from 4 different disciplines (natural sciences, sport, pedagogics and anthropology). I allow myself to re-order the suggestions from UIO into a structure for a synopsis.

1. Introduction to problem (s).
2. Review. Place your own work in the field of research and thus refine the focus on your own work.
3. Theoretical perspective and central concept. What's needed beyond what's in the articles?
4. Methodology. Method selection, selection.
5. Main findings.
6. Meta Reflection. Discussion.

I think it is necessary to see the role of the synopsis in relation to the doctoral thesis in its entirety, and, in the light of a comprehensive understanding, to plan the synopsis before we begin to discuss it as a separate part of the dissertation. A doctoral thesis based on three articles is different from a monograph, but it should still be genuine in its research within a chosen topic. Therefore, it is necessary to begin with the theme of the study, and the theme itself must be significant and embedded in one (or more) theory (s). The dissertation poses an overall problem, and in the course of the dissertation the candidate will find the answer to the overall question. Previous research on the subject has been presented, and the dissertation presents the methodological choices that have been made to investigate the overall question. This is common to all doctoral dissertations, as I see it. In an article-based dissertation, the overall question is based on three (or more) questionnaires that are highlighted in the three articles the thesis contains. But they are all directed at the overall question and help to highlight the thesis's theme. Each article is an independent unit within the thesis theme, but it is important to clarify how each article is part of the search for answers to the overall question. The dissertation requires a discussion of findings that highlight the overall problem statement and clarify the significance of the dissertation in developing new knowledge about the subject.

As a summary of my thoughts around the synopsis' structure, it is possible to list the words I have written in bold italics in the text and describe the dissertation:

- *theme*
- *theoretical basis*
- *overall research questions*
- *previous research*
- *Methodical choices*
- *sub-problems (articles)*

[4] Somewhat like I argued before in footnote 2, I see little difference between these two theoretical bases. Such a sentence sounds great, but is more easily written down than thought through in concreto when you start to project it unto the reality of your own thesis. The subsequent sentence largely is a repetition.

• Discussion related to overall problem
• significance as knowledge development
• The thesis's limitations and new research needs
More technical parts of any scientific work come in addition, and Frode will include them in the next section.

In my subsequent comparison of five approved synopses, I found the following structural elements:

• Preface / acknowledgments
• summary
• List of articles related to the dissertation (references)
• Introduction (point 1 above)
• Location of the research done (point 2 above)
• Theoretical approach that underlies the research done (point 3 above[5])
• Methodological approach, including ethical considerations (point 4 above)
• Introduction of articles that are summarized and synthesized in the cloak (point 5 above)
• General discussion and conclusions (point 6 above)
• Implications
• References
• Any attachments

This relatively primitive mapping process shows that - as I already suspected - there are some common features, both from the supervisor's and the candidate's point of view[6]. After all, presumably, several people have been involved in the design of UiO's guidelines, and all the five dissertations I compared had different supervisors.

The lists presented above may assist the candidate as checklist and suggest a structure that seems to be agreed upon among supervisors, and members of the review committee. I am afraid that such lists can quickly become too detailed and used technically so that there is no room for variation and creativity. Therefore, I am opposed to a centrally controlled template, while at the same time it is important for the candidate to have some guidelines. After all, the quality of the dissertation depends more on content than on structure.

Hi! You cannot just find yourself a synopsis!

With the acceptance of a commonality in the perception of what a synopsis should contain and how it should be built up, the following question came to my mind: Is it just okay to simply fill in the structure of an already existing synopsis from my field of study? This question can be met with both positive and negative arguments:
- Yes, you can because it is becoming a standard template for how a synopsis should be built up; Wholeness and coherence are covered this template.
- No, you cannot because the adjustments that necessary to cover your research work in the best possible way is not simply pasted from previous synopses. The specific form and content of your own

[5] Personally, I am not sure whether one should be rigorous about the sequence 'problem statement – previous research – theoretical background – research questions' Frode and Kari take different positions here. Neither of them uses the word hypotheses
[6] Should these differ, actually?

research work should be guiding the development of your synopsis.
- My own conclusion**: t**he candidate himself must make choices and customize the template. This may involve addition or removal of elements in the structure of the synopsis.

Final

After completing my mapping process, my supervisor and I agreed that it was worth inviting others in the form of a post for discussion at the seminar. I finished my presentation containing a "hardly" draft outline of my coat. It looked very similar to the template visualized above. This original draft outline did not stand during the writing process, instead it changed in line with the development of the content. But the challenges met and choices made through this process prompted me to establish a framework. Furthermore, the initial work on a draft synopsis opened up for good discussions with my supervisor, as well as clear advice on how to build the cloak, both by safeguarding my dissertation project as good as possible and by preventing a continuous challenging of the framework of the synopsis. I delivered my thesis for evaluation in December 2010. It was a long process from the first awkward attempt to set up a credible framework to the final delivery-ready version. I went through a number of reviews that included moving large sections and toning down some too outspoken headlines. My synopsis was turned upside down, torn apart and sand-papered until it finally felt as a coherent product which corresponded to the template, while I kept the necessary room for adaptations to my research work. In this process, the supervisor was an important support and discussion partner.

Currently, there appears a need to enhance such guidance skills, both for master and PhD level. This can be done through the establishment of a supervisor network, competence-building seminars and conferences. The UH network Vest (which consists of UiB, HiB, HSH, HiSF and Holdskuld i Volda (HVO)) has acknowledged this need. Also at the National Research School for Teacher Education (NAFOL) the content of guidance at Ph.D. level is a priority area. The discussion whether a synopsis should have an official standard structure should be part of such guidance training. If so, then the question will remain whether such a template really suits everyone.

The challenge for the supervisor is to assist the candidate in writing a dissertation that represents the knowledge the candidate has gained while working on the subject, whilst also ensuring that the work meets formal requirements in terms of quality and structure. It is a balance between breakthrough work, creativity and formal academic frameworks and requirements. This also applies to the synopsis and its structure. I would warn against having strict formal, often technical requirements that inhibit new forms of presentation of knowledge development and progress. It is the supervisor's professional judgment that must be used in fostering a new generation of researchers who dare to be pioneers in their field whilst at the same time respecting the profession's norms and frameworks.

References

HHB (2008). Rules for the degree Philosophiae Doctor (Ph.D) at the Business School in Bodø. Downloaded September 2009
Knowledge Agency (2005). Book goal dictionary, 3rd edition. Oslo: Knowledge Agency / Aschehoug.
NTNU (2009). Ph.D. in social work. Original downloaded version (downloaded September 2009) has been moved without change: http://www.ntnu.no/studier/phsarb
UiO (2009). Guidelines on the cloak of an article-based dissertation. Downloaded September 2009: http://www.uv.uio.no/forskning/doktorgrad-karriere/forskerutdanning/gjennomforing/Krav_til_kappen. html

Av Frode Olav Haara og Kari Smith

Kappen: «One size fits all»?

Frode Olav Haara
(Doktorgradskandidat)
Høgskulen i Sogn og Fjordane
frode.olav.haara@hisf.no

Kari Smith
(Professor og veileder)
Universitetet i Bergen
kari.smith@iuh.uib.no

Sammendrag
I artikkelen beskriver en doktorgradskandidat sin søkende prosess i den innledende fasen av arbeidet med sin avhandlingskappe og hvordan den ble influert av at det ennå ikke er etablert en like tydelig tradisjon knyttet til denne typen avhandlinger som det er knyttet til monografier. I teksten er det også lagt inn kommentarer fra doktorgradskandidatens veileder.

I september 2009 var jeg kommet til den fasen i mitt doktorgradsarbeid at det var på tide å begynne arbeidet med kappen. Jeg hadde på denne tiden vært stipendiat siden 1. januar 2007, med finansiering fra min arbeidsgiver Høgskulen i Sogn og Fjordane (HiSF) og opptak til obligatorisk forskerutdanning ved det psykologiske fakultet ved Universitetet i Bergen (UiB). Ved dette fakultetet lå det allerede på denne tiden en føring på at man helst burde skrive en artikkelbasert doktoravhandling. Dette arbeidet burde bestå av tre artikler for publisering i peer-review-baserte tidsskrift, samt en kappe.

Gjennom min tilknytning til det psykologiske fakultet ved UiB, ble jeg også tilknyttet forskerskolen Western Norway Graduate School of Educational Research (WNGER). Denne forskerskolen var et fellesprosjekt mellom UiB, Høgskolen i Bergen (HiB), Norsk Lærerakademi (NLA), Høgskolen Stord/Haugesund (HSH) og HiSF, med administrativ og faglig styring lagt til UiB. Innenfor denne forskerskolen var de deltakende stipendiaters aktive deltakelse et sentralt prinsipp knyttet til forskerskolesamlingene.

I september 2009 var min arbeidsgiverinstitusjon HiSF vert for en slik forskerskolesamling, og vi som skulle delta ble tradisjonen tro utfordret til å bidra med et innlegg knyttet til vårt doktorgradsprosjekt. For min egen del var statusen for min fremdrift på denne tiden at jeg hadde fått godkjent en av mine tre artikler og to artikler nettopp var sendt til peer review-vurdering. Jeg hadde så vidt begynt å sysle litt med tanken om å begynne arbeidet med avhandlingens kappe, men hadde egentlig ikke noe som jeg kjente på som interessant å legge frem for deltakere på forskerskoleseminaret med tanke på mulighet for peer-review-respons og -kritikk. Jeg diskuterte dette litt med min veileder, professor Kari Smith ved UiB, og hun utfordret meg nærmere på hva jeg var opptatt av i prosjektet mitt for tiden og hva andre deltakere på samlingen kunne ha interesse av at jeg fokuserte på i et seminarinnlegg. Blant ideene jeg presenterte for henne, var det ideer som var basert på emner jeg allerede hadde arbeidet med knyttet til prosjektet, men også ett fremtidsrettet emne: Hva er en kappe i en doktorgradsavhandling, og hvordan skal en kappe være? I et forsøk på å belyse disse to spørsmålene, ble vi enige om at det ville være en verdifull øvelse for meg selv med tanke på det påtroppende arbeidet og et tema med interesse for stipendiatkollegaer og veiledere innenfor forskerskolen.

I Norge har monografi vært den tradisjonelle formen ved levering av doktorgradsavhandling. Bruk av et system med et antall artikler, og en tilhørende kappe, er derimot en relativt ny form for avhandling, i alle fall etter norske forhold. Å basere seg på bruk av kappe – både i oppbygging, veiledning og prioriteringer knyttet til arbeid med en avhandling – er derfor en prosess hvor det ennå ikke er tråkket opp en bred sti der kutyme og en allment akseptert mal er etablert. En professor med lang fartstid og erfaring knyttet til veiledning av doktorgradskandidater (som ikke tilhørte WNGER, men var invitert som foredragsholder ved seminaret) sa til meg etter at seminarinnlegget mitt var ferdig: «Vi står uten en tradisjon knyttet til skriving av kappe og veiledning i forhold til det, og som veileder opplever jeg dette nå som utfordrende i forhold til min egen veiledning. Og det er jeg nok ikke alene om.» Det betyr at både veiledere og doktorgradskandidater må gjøre valg uten å ha en tradisjon å støtte seg til.

I det følgende har jeg forsøkt å skrive sammen presentasjonen jeg hadde til en sammenhengende tekst, som egentlig presenterer hvordan jeg forsøkte å nærme meg det å skulle skrive en egen kappe. I tillegg har min doktorgradsveileder, professor Kari Smith, lagt inn noen kommentarer i kursiv her og der. Det gir også et veilederperspektiv på den innledende fasen i tilnærmingen til kappearbeidet.

Kappe – noen forsøk på definisjoner

En i utgangspunktet enkel måte å skaffe seg et inntrykk av hva en kappe er ment å være i forbindelse med en doktorgradsavhandling, er kort og greit å se hva utdanningsinstitusjoner i Norge som tilbyr doktorgrad har nedfelt skriftlig som sin tolkning av hva en kappe skal være. Her er noen eksempler på definisjoner og avgrensninger jeg fant (*alle* nettreferanser som ble brukt i forberedelsene til seminarinnlegget 16. september 2009 ble hentet i perioden 5.–14. september 2009):

> ***Daværende Handelshøgskolen i Bodø (HHB) (2008)***
> *Om doktoranden velger å utarbeide mindre arbeider (artikler), skal det lages et sammendrag (kappe) som skal representere en selvstendig innsats, og som skal dokumentere helheten i avhandlingen gjennom en sammenfatting av problemstillinger og konklusjoner som fremlegges i artiklene. Sammendraget i avhandlingen skal ikke bare sammenfatte, men også sammenstille de problemstillinger og konklusjoner som legges frem i artiklene i et helhetlig vitenskapelig perspektiv, og på den måten dokumentere sammenhengen i avhandlingen. I dette ligger også en oppsummering av avhandlingens bidrag til forskningsfeltet – både praktisk og teoretisk. I sammendraget skal det også redegjøres for metodologiske og metodiske valg anvendt i artiklene, i tillegg til det som fremgår av delartiklene.*

Universitetet i Oslo (UiO) (2009)
Det endelige arbeidet med kappen bør tas mot slutten av doktorgradsarbeidet. Kappen skal imidlertid også fungere som en disposisjon for arbeidet slik at det helhetlige perspektivet kan klargjøres tidlig i arbeidsprosessen. Et slikt innledende arbeid med kappen kan virke bevisstgjørende for det helhetlige perspektivet som avhandlingen skal ha. Samtidig er det naturlig at kappen modifiseres underveis i arbeidet i forhold til hvordan form og innhold i artiklene utvikler seg.

- *Kappen skal sammenfatte og sammenstille de problemstillinger og konklusjoner som legges frem i delarbeidene slik at avhandlingen fremtrer som en helhet. Kappen skal presentere resultatene i de enkelte delarbeidene på en slik måte at den indre sammenhengen mellom dem synliggjøres. Kappen kan dermed bidra til å skape en sammenheng mellom funn som kan invitere til drøftinger på et mer teoridrivende nivå. Kompleksitet og nyanser i funn skal diskuteres i lys av metodiske, vitenskapsteoretiske og teoretiske problemstillinger.*
- *Kappen skal inneholde nødvendige teoretiske og metodiske vurderinger knyttet til doktorgradsarbeidet. Kappen skal også redegjøre for hvor begrepsavklaringer eller utdypinger om ulike tema finnes i avhandlingen.*
- *Kappen skal synliggjøre og oppsummere avhandlingens bidrag til det aktuelle forskningsfeltet.*
- *Kappen skal løfte frem og diskutere etiske forhold ved arbeidet.*
- *Kappen skal inneholde faglige ajourføringer når det er nødvendig ut fra artiklenes publiseringstidspunkt/tidspunkt for ferdigstillelse. Dette kan være nødvendig for at avhandlingen som helhet skal fremstå som faglig oppdatert.*
- *Kappen skal være utformet av doktoranden alene.*
- *Kappen skal normalt ha et omfang på mellom 60–70 sider.*

Norges teknisk-naturvitenskapelige universitet (NTNU) (2009)
Avhandlingen kan leveres som ett større samlet arbeid (monografi) eller som en samling av artikler. Dersom den består av en samling av artikler, bør det normalt være 3–5 arbeider i tillegg til sammenskrivning (kappe).

Etter å ha latt disse tre forsøkene på å få frem og presisere hva en kappe er ment å være synke noe inn, ble jeg sittende å lure på hva en kappe i praksis er ment å være i forbindelse med en doktorgradsavhandling. Er den ment å være et sammendrag? En sammenstilling? En sammenfatning? En sammenskrivning? Er det egentlig noe forskjell på begrepene som brukes?

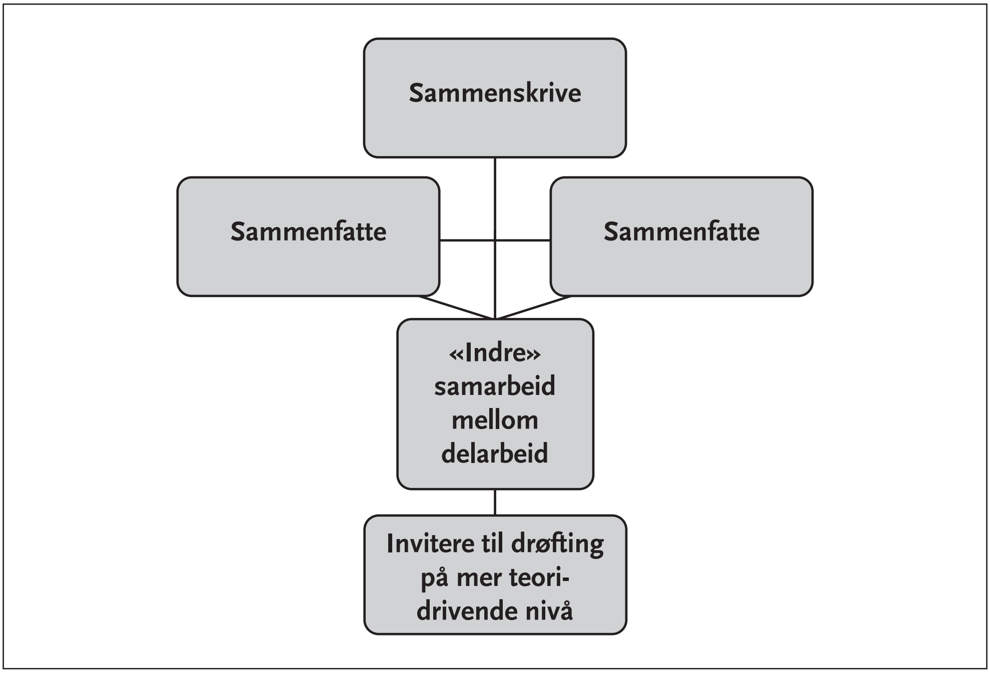
Noen raske ordbokoppslag (Kunnskapsforlaget, 2005) bød på følgende avklaringer:

Sammenheng:
1 *det at noe henger (logisk) sammen; tanke*
2 *forbindelse (med andre forhold)*
Sammendrag: *kortfattet gjengivelse av noe; resymé*
Sammenstille: *stille side ved side for betraktning, sammenligning*
Sammenfatte: *gjengi i hovedpunkter; resymere*
Sammenskrive: *(ikke definert i ordbok)*

Kappens funksjon

Uten å se en entydig definert oppfatning av hva en kappe er tolket til å ivareta gjennom institusjonenes definisjoner, skiftet jeg fokus fra definisjon av kappe til å se direkte på kappens funksjon. I den sammenheng hentet jeg frem igjen de syv konkretiseringspunktene i UiO sine veiledende retningslinjer for hva en kappe skal være (se over). Ved å lese disse punktene i forhold til hvilken funksjon kappen bør forstås til å ha, blir både kappens form og innhold noe mer konkretisert. Kappen skal være noe mer enn et sammendrag eller en oppsummering. Rent visuelt kan derfor kappen heller ses som en sammenskrivning av avhandlingsprosjektet, der sammenskrivning kan tolkes til å innbefatte både sammenfatning og sammenstilling av de selvstendige artiklene som er del av avhandlingsprosjektet. Sammenskrivningen krever synliggjøring av (eventuelle) «indre» sammenhenger mellom artiklene, og inviterer dermed til drøfting av forskingsresultatene opp mot det teoretiske grunnlaget for avhandlingen på et nivå som går utover teorigrunnlaget for den enkelte artikkel. Kappen inviterer til drøfting av de selvstendige forskningsresultatene i forhold til hverandre og i sammenheng, i forhold til teorigrunnlaget som velges for hele avhandlingsprosjektet. Dette kan visualiseres slik:

Figur 1: Et forsøk på å visualisere kappens form og funksjon.



Oppfatningen av hva kappen skal være og hvordan en kappe skal bygges opp, ble så forsøkt sett i forhold til hvordan fem artikkelbaserte avhandlinger som allerede hadde blitt godkjent for disputas var bygd opp. De fem kappene som ble sammenlignet med hverandre og min gryende oppfatning av kappens form og funksjon representerte fire forskjellige fagfelt (realfag, idrett, undervisning, antropologi). Tidligere i undersøkelsesprosessen ble det vist til UiO sine retningslinjer for forståelse av kappens innhold, og jeg tillater meg her å forsøke å liste dem opp i form av disposisjonspunkt for en kappe:

1. Introduksjon med problemstilling(er).
2. Review. Plassere eget arbeid i forskningsfeltet og dermed avgrense fokuset for eget arbeid.
3. Teoretiske perspektiv og sentrale begrep. Hva trengs utover det som er i artiklene?
4. Metodologi. Metodevalg, utvalg.
5. Hovedfunn.
6. Metarefleksjon. Diskusjon.

*Jeg tror det er nødvendig å se kappens rolle i relasjon til doktorgradsavhandlingen i sin helhet, og i lys av en helhetlig forståelse planlegge kappen før vi begynner å diskutere kappen som en egen del av avhandlingen. En doktorgradsavhandling som bygger på tre artikler er ulik en monografi, men den skal likevel være genuin i sin undersøkelse innenfor et valgt tema. Derfor er det nødvendig å begynne med **temaet** for handlingen, og temaet i seg selv må være betydningsfullt og nedfelt i **en (eller flere) teori(er)**. Avhandlingen stiller en **overordnet problemstilling**, og gjennom arbeidet med avhandlingen skal kandidaten finne svaret på det overordnede spørsmålet. **Tidligere forskning på temaet** er presentert, og avhandlingen presenterer de **metodologiske valg** som er gjort for å undersøke det overordnede spørsmålet. Dette er felles for alle doktorgradsavhandlinger, slik jeg ser det. I en artikkelbasert avhandling er det overordnede spørsmålet konkretisert gjennom **tre (eller flere) underspørsmål** som blir belyst gjennom de **tre artiklene** avhandlingen inneholder. Men de er alle rettet mot det overordnede spørsmålet og hjelper til å belyse avhandlingens tema. Hver artikkel er en selvstendig enhet innenfor avhandlingens tema, men det er viktig å klargjøre hvordan hver artikkel er en del av søken etter svar på det overordnede spørsmålet. Avhandlingen krever **en diskusjon av funn som belyser den overordnede problemstillingen** og klargjør **avhandlingens signifikans** i utvikling av ny kunnskap om temaet.*

Som oppsummering av mine tanker rundt kappens struktur, er det mulig å liste de ordene jeg har skrevet med fet kursiv skrift i teksten og som beskriver avhandlingens

- *tema*
- *teoretisk grunnlag*
- *overordnet forskningsspørsmål*
- *tidligere forskning*
- *metodiske valg*
- *underproblemstillinger (artikler)*
- *diskusjon relatert til overordnet problemstilling*
- *signifikans som kunnskapsutvikling*
- *avhandlingens begrensninger og nye forskningsbehov*

I tillegg kommer de mer tekniske delene av ethvert vitenskapelig arbeid som Frode tar med i sin opplisting i neste avsnitt.

I min etterfølgende sammenligning av fem godkjente kapper, ble følgende store og små disposisjonsfellestrekk identifisert for oppbyggingen av en kappe:

- Forord/takksigelser
- Sammendrag
- Liste over artikler knyttet til avhandlingen (referansene)
- Introduksjon (punkt 1 over)
- Plassering av forskningen som er gjort (punkt 2 over)
- Teoretisk tilnærming som omkranser forskningen som er gjort (punkt 3 over)
- Metodologisk tilnærming, inkludert etiske betraktninger (punkt 4 over)
- Introduksjon av artikler som sammenfattes og sammenskrives i kappen (punkt 5 over)
- Generell diskusjon og konklusjoner (punkt 6 over)
- Implikasjoner
- Referanser
- Eventuelle vedlegg

Denne relativt primitive kartleggingsprosessen viser at det – som jeg mistenkte allerede innledningsvis i min kartleggingsprosess – er noen fellestrekk i oppfatningen av hva en kappe bør inneholde og hvordan den bør bygges opp, både fra et veilederståsted og fra et kandidatståsted. Tross alt har antagelig flere vært involvert i utformingen av UiO sine retningslinjer, og alle de fem avhandlingene jeg sammenlignet med hverandre har hatt forskjellige veiledere.

Listene som er presentert ovenfor kan være til hjelp for kandidaten som sjekklister for å gi kappen en struktur som det ser ut til å være enighet om blant veiledere, og naturlig nok da også komitémedlemmer. Men jeg er redd for at slike lister fort kan bli for detaljerte og brukt på en teknisk måte slik at det ikke er rom for variasjon og kreativitet. Derfor er jeg imot en sentralt styrt mal for hvordan kappen skal være, samtidig som det er viktig for kandidaten å ha noen retningslinjer å holde seg til. Tross alt er avhandlingens kvalitet avhengig av innholdet og mindre av den tekniske strukturen.

Hei! Du kan jo ikke bare finne deg en kappe!

Med aksepten av at det finnes klare fellestrekk i oppfatningen av hva en kappe bør inneholde og hvordan den bør bygges opp, fant jeg det mulig å stille følgende spørsmål: Er det like greit å bare «fylle inn i disposisjonen til en allerede godkjent kappe som er knyttet til det fagfelt prosjektet er innenfor»? Dette spørsmålet kan det knyttes både aksepterende og kritisk argumentasjon til:

- Jaaa … det kan du, fordi:
  - Det er i ferd med å bli etablert en mal for hvordan en kappe bør være bygd opp.
  - Helhet og sammenheng ivaretas av malen.

- **Nei, det kan du ikke, fordi:**
  - De tilpasninger som det er rom for å gjøre og som skal til for at du skal få kappen til

å fremstille ditt forskningsarbeid på beste måte, er ikke mulig å finne i andre kapper. Valgmulighetene knyttet til (detaljer i) oppbyggingen av en kappe åpner for dette.
– Forskningsarbeidets form og innhold kan være styrende for kappens oppbygging.

Min konklusjon

Kandidaten selv må gjøre valg og tilpasse kappemalen. Dette kan innebære tillegg av punkt eller å droppe punkt i oppsettet av kappen.

Avslutning

Etter at jeg avsluttet min kartleggingsprosess, diskuterte prosessen med min veileder og ble enig med henne om at dette var en «blottleggingsreise» som var verdt å invitere andre med på i form av et innlegg til diskusjon på et forskerskoleseminar, avsluttet jeg min presentasjon på seminaret med å legge frem et «knapt» utkast til disposisjon av min kappe. Den lignet veldig på den standard som er synliggjort gjennom kartleggingsprosessen jeg gjorde. Dette utkastet til disposisjon ble nok ikke stående ettersom arbeidet med selve kappearbeidet gikk fremover. Disposisjonen endret seg i tråd med utviklingen av innholdet i kappeteksten. Men utfordringer og valg som jeg møtte gjennom denne prosessen, førte til at jeg både fikk etablert en ramme og en retning for hvordan kappen min burde være. Videre åpnet det innledende arbeidet med kappedisposisjonen til gode diskusjoner mellom meg og veileder samt klare råd til meg om hvordan kappen burde bygges opp, både i forhold til å ivareta avhandlingsprosjektet best mulig og å unngå å utfordre rammene for hvordan en kappe bør være bygd opp. Jeg leverte min kappebaserte avhandling til vurdering i desember 2010.

Det er en lang prosess fra det første vaklende forsøk på å sette opp en troverdig disposisjon for en kappe, til den leveringsklare versjonen av kappen. Jeg gikk gjennom en rekke revisjoner som innebar flytting av store innholdsdeler og demping av «frittalende» overskrifter. Kappen min ble høvlet, slipt og til slutt filt til et produkt som helhetlig og sammenhengende kunne møte en mal som ser ut til å være i ferd med å bli etablert, samtidig som jeg beholdt det nødvendige rom for tilpasninger til mitt forskningsarbeid. I denne prosessen var veileder en viktig støttespiller og diskusjonspartner.

For tiden er det fokus på behovet for å rekruttere veilederkrefter for å møte et økende behov for veiledning på både master- og Ph.D-nivå. Dette gjøres blant annet gjennom etablering av veiledernettverk og kompetansebyggende seminarer og konferanser. Nettverket UH-nett Vest (som består av UiB, HiB, HSH, HiSF og Høgskulen i Volda (HVO)) har for eksempel sett behovet for dette. Også ved nasjonale forskerskoler (for eksempel Nasjonal forskerskole for lærerutdanning (NAFOL)) er innholdet i veiledning på Ph.D-nivå et prioritert område. Diskusjon av om en kappe skal ha en offisiell form for mal eller om det i prinsippet skal stå doktorgradskandidaten fritt å utforme kappen, bør være en del av arbeidet som gjøres innenfor fokuset på veiledning fremover. Da melder i så fall også spørsmålet seg om det finnes en uoffisiell form for mal og om en slik kappemal egentlig passer for alle.

> *Utfordringen for veilederen er å hjelpe kandidaten til å skrive en avhandling som representerer den kunnskapen kandidaten gjennom arbeidet med temaet har ervervet seg, og samtidig sørge for at arbeidet møter formelle krav både med hensyn til kvalitet og struktur. Det er en balanse*

mellom nybrottsarbeid, kreativitet og formelle akademiske rammer og krav. Dette gjelder også for kappen og dens struktur. Jeg vil advare mot å ha for strenge formelle, ofte tekniske krav som hemmer nye former for presentasjon av kunnskapsutvikling og fremgang. Det er veilederens profesjonelle skjønn som må brukes i arbeidet med å fostre en ny generasjon forskere som våger å være pionerer innenfor sitt felt og samtidig har respekt for profesjonens normer og rammer.

Referanser

HHB (2008). *Reglement for graden Philosophiae Doctor (Ph.D) ved Handelshøgskolen i Bodø*. Lastet ned september 2009: http://webcache.googleusercontent.com/search?q=cache:tGY_SMAPOMIJ:www.hibo.no/index.php%3FID%3D4807+%22Om+doktoranden+velger+%C3%A5+utarbeide+mindre+arbeider+%22+Bod%C3%B8&cd=1&hl=en&ct=clnk&source=www.google.com

Kunnskapsforlaget (2005). *Bokmålsordboka*, 3. utgave. Oslo: Kunnskapsforlaget/Aschehoug.

NTNU (2009). *Ph.D. i sosialt arbeid*. Opprinnelig nedlastet versjon (nedlastet september 2009) er blitt videreført uten endring: http://www.ntnu.no/studier/phsarb

UiO (2009). *Veiledende retningslinjer om kappen ved en artikkelbasert avhandling*. Lastet ned september 2009: http://www.uv.uio.no/forskning/doktorgrad-karriere/forskerutdanning/gjennomforing/Krav_til_kappen.html

8 · Checklist of problems

Most of this book deals with weaknesses in the central constructs of ecology. In contrast, most ecological research addresses smaller, more tractable questions which relate only obliquely to the central tenets of the discipline. Many ecologists may feel that the faults discussed so far have limited bearing for their ecological specialties, and specialists may be tempted to ignore criticisms of general ecology. The same conclusion is encouraged by the exceptional character of the general constructs that are the focus of most ecological discussion and criticism. Lessons learned from the likes of Darwin, Elton, Hutchinson, MacArthur and May are hard to apply to the bulk of the scientific literature, simply because there are so few contributions of that stature, scope and form. Right or wrong, ecological syntheses rarely follow the standard format of a scientific paper; they rarely consider methods, and they rarely address the details of a particular case. In short, because seminal works in ecology are rare and exceptional, both their virtues and their faults misrepresent the field.

Dismissal of general ecological criticisms as irrelevant to the bulk of working ecologists would be unjustified. Intellectual leadership in a field is awarded by consensus of the field, and flaws in leadership reflect on the entire field. Those who read ecological journals and especially those who review manuscripts or research proposals know the problems of normal ecology and the many ways in which the frailty of the science is manifest in its primary literature. No critique of ecology would be complete if it confined itself to the problems of our leaders and ignored those of normal ecologists.

This chapter outlines common flaws in normal ecological articles, under the premises that the weakness of general ecology reflects weaknesses at its scientific base, and that the lack of an effective standard for scientific achievement has encouraged laxity at all levels. To a large extent, the chapter will cover familiar ground because the difficulties it addresses are those of writing or delivering a good scientific paper. This topic is covered by books, articles and the 'Instructions to Authors' found in most journals, and that large part of university education involving laboratory reports, term papers, and thesis evaluations that seeks to transfer acceptable standards of scientific reporting to the next generation.

Writing a scientific paper is not a so much literary exercise as a scientific one. The hallmarks of a good paper – logic, clarity, precision – are those of good science and the goal of scientific exposition, like that of research in general, is to provide a good piece of science. Like good science, good writing is straightforward, concrete, exact, rigorous, clear-headed and concise (Woodford 1968). When this is not so, more vigilant criticism is needed.

To facilitate application, this outline follows the classical format of a scientific paper – Introduction, Methods, Results, and Discussion – pointing out common deficiencies in each section like a checklist or field-guide. This device provides a structure for the chapter, but should not be taken rigidly, for ecological problems are rarely confined to single habitats. Since the weaknesses of larger ecological constructs are repeated in these smaller studies, there is necessarily some overlap between the criticisms of this chapter and those that went before, but this has been minimized.

A listing like this can never be complete, so the chapter is only a beginning for criticism of particular contributions. The reader is encouraged to expand this list, to modify it and to illustrate it with appropriate examples in the light of personal experience and particular application.

The Introduction

The Introduction of any scientific paper must do three things: It must establish the relevance of the topic for an informed reader; it must provide a context for the study by describing the status and shortcomings of present knowledge; and it must relate the study's goal to this context. For example, a paper introducing a model of the reacidification of lakes after neutralization with lime (Sverdrup & Warfvinge 1985) might begin by relating lake acidification and liming to the problem of conservation:

> Anthropogenic acid precipitation threatens the biota of many lakes in granitic basins, but symptomatic relief may be had by liming.

Next there should be a concise statement of the present state of knowledge about liming, specifying the lacuna in this knowledge that the present paper will address:

> Models exist to calculate the dose and form of calcium carbonate required to neutralize acidified lakes (Sverdrup 1983), but not the dose which optimizes the schedule for future treatments following the inevitable reacidification. Instead it is

recommended simply that excess carbonate be added, although reacidification must depend on annual acid load to the lake, the flushing of the neutralized water from the lake, and the solubilization of sediment carbonate.

Finally, the introduction must explain how the paper fills this lacuna:

This paper (a) describes a model of reacidification that is based on flushing time, acid load, and the amount and distribution of carbonate on the sediment and (b) tests this model against rates of reacidification observed in limed lakes.

In general terms, good introductions present the questions being asked as hypotheses under test and show where these questions are relevant. Such introductions are easy to write if the study has been directed by explicit, testable hypotheses identifying the goals of the research. These goals should have been established when the research was proposed, not after the research is complete. Given a sound proposal, writing a good introduction is relatively simple, because its elements have been clear since the beginning of the project.

Many ecological papers fail to meet the elementary requirement of a clear hypothesis. In review of 87 publications about optimal diet theory, Gray (1987) was unable to determine what was being tested in 30% of the cases. A decade earlier, Fretwell (1975) reported that the proportion of papers in *Ecology* which tested explicit or implicit hypotheses had risen from about 5% of the total in 1950 to 1955 to almost 50%. This great improvement should not obscure the fact that, in 1975, over half of the papers in *Ecology* apparently did not address hypotheses. Such papers are extremely hard to judge or read because their authors' intentions are never clear.

A clear introduction is the key to a successful paper because it defines the hypothesis under test and therefore the question on which the paper depends. When this is done, the methods can be judged as to whether they are adequate to the test, the relevance of the results is clear, and the direction of the discussion is determined. However, this is possible only if the proposed question is clear and testable, if the methods chosen to make the test were appropriate, and if the results provide a clear answer to the question. These are requirements for both good papers and good science, but they are routinely ignored in ecology where the vagueness of questions addressed in single research programs reflects a vagueness in the discipline. The uncertainty of ecology results in vague purpose and ultimately in vague papers. This is only compounded by the perceived personal need to publish at all costs so that even inconclusive tests now clog the literature.

Selection of hypotheses

The first problem for the scientist is to identify a relevant question or hypothesis to test as the goal of the research. This is likely also the most difficult problem, because it is the creative step. Once this step is taken, the rest of science involves deduction to determine what the hypothesis predicts, followed by the methodical application of relatively conventional intellectual and technical tools, like the appropriate instrumentation and statistics, to determine if the deduction holds. Most scientists consequently feel competent to test relevant hypotheses once these have been identified, but only a few are as confident of their continuing ability to form such hypotheses.

The challenge of selecting relevant topics for research explains the difficulty and frequent failure of impact assessments (Schindler 1976; Larkin 1984; Hecky *et al.* 1984). Valid assessment depends on what is assessed. Whether it is the agency requesting the assessment or the firm conducting it, someone must determine what environmental features are relevant in light of the proposed development for the site. Often this question is evaded with unthinking, routine observations about the presence and absence of different species and standard measurements of a few variables. However, because such variables are not always obviously pertinent to the problems posed by development, effective assessments require decisions about what, and in what way, observations are relevant.

The selection of variables to study requires fundamentally scientific decisions. Since identification of a single relevant variable tests the best minds of the science, it is scarcely surprising if ecologists in industry, who may lack the appropriate tools, surroundings and temperatment, sometimes find themselves stymied. As a result, poor impact assessments are so common that good ones stand out (Alexander & Van Cleve 1983; Berger 1977) as significant ecological achievements.

The difficulties in determining relevance were considered in the last chapter, so this section instead considers some of the ways ecologists have circumvented these problems of relevancy.

Justification by authority

Scientists idolize their heroes. The Newtons, Galileos and Eisteins are role models, even though very few researchers can achieve this stature (Glaser 1964). In ecology, this hero-worship is reflected in the zeal with which modern writers legitimize their position with quotes from

Darwin, Hutchinson or MacArthur. Some leading ecologists – Dayton (1979) cites Hutchinson, MacArthur, the Odums, and Watt – have had a Messianic appeal for their followers, so the views and interests of those men have defined much of modern ecology. Regrettably, the tentativeness and self-doubt that intellectual leaders commonly bear towards their own work are often lost when the approach is appropriated by acolytes whose admiration of the leader is so intense that they harden the great figure's ideas to an inflexible dogma.

One effect of charismatic authority is that the concerns of the master may become central to a field, without asking whether these are otherwise relevant or if they are scientifically tractable. Thus Hutchinson's famous question 'Why are there so many kinds of animals?' is still a central issue of ecology (Tilman 1982; May 1986), even though the form of the question begs an explanation which science can never provide. A similar situation arose when ecological opinion divided over the issue of whether the mechanism of population control was density-dependent or independent. This intractable question so fascinated leading ecologists, like Lack and Nicholson, that it overshadowed tractable problems associated with predicting population densities (Murdoch 1970; Strong 1983).

Ex cathedra statements

Ex cathedra statement represents a further aspect of authority often found in the introductions to ecological papers. Such assertions appear to be authoritative statements of fact, but are really only opinions that must be treated with circumspection. For example, the claim of Cates & Orians (1975) that slugs are generalist herbivores is central to their study of plant palatability, but the claim is not supported by rigorous studies of the diet of slugs relative to other organisms, but by unreferenced common knowledge that slugs eat a wide variety of plants. Similarly, the introduction of Macevicz & Oster (1976) declares that the social insects conform to a basic assumption in optimal foraging models that the systems be evolutionarily static. This claim is justified because 'the basic colony structure and strategies have remained largely unchanged for many millennia', but Macewicz & Oster (1976) cite no basis for such a seemingly substantive claim. Janzen & Martin (1982) base their argument for the coevolution of the trees and extinct megafauna of tropical America on anecdote, and Lehman (1986) suspects that 'applied ecology will be best served by continued basic inquiries into process and mechanism', but cites no evidence for that opinion. The critical reader should be suspicious of any unsupported assertions. Ex cathedra pronouncements may be an effective ploy in debate (e.g. Levins & Lewontin 1980), but unsupported opinion and hand-waving are treacherous foundations for intensive and expensive scientific research.

Emulation of authority

The attraction of authority has the further disadvantage that followers imitate the style of the great scientist as well. Thus MacArthur's iconoclasm freed his followers from stodgy rules of contemporary scholarship and breathed fresh air into the top ranks of American ecology (Fretwell 1975). This freedom is not wholly laudable, for MacArthur's style also seemed to sanction spotty and selective citation (Schoener 1972) and the evaluation of loosely formulated models with weak data (Fretwell 1975). Hutchinson's restrictive and idiosyncratic interpretation of the hypothetico-deductive method, whereby the hypothesis consists of an analogy between a syllogism and nature (Hutchinson 1978; Kingsland 1985), became the model for a generation of the brightest minds in the field, sometimes directing them to study syllogisms instead of nature.

In a field where general constructs are scientifically weak, the behaviour of those who built the constructs can be a poor example. Emulation tends to perpetuate and obscure the problems at the top by repeating them at the bottom. Thus mathematical, logical and verbal analogies or metaphors, inspirational writings based on vague and undefined concepts, why-questions about cause or mechanism, weak models, poor data, and unsupported opinion have become common elements at all levels of the ecological literature. However, the blame in this state of affairs does not belong to the leader, but to the followers, for science does not oblige us to follow anyone.

Faddism

The task of defining the hypotheses under test can be avoided by simply collecting data that are only 'related' to contemporary interests in some unspecified way. Unfortunately, refusal to specify relevant goals fosters research which is so unconnected to the rest of the field that it may wax and wane without affecting the science. Although such ephemeral research areas are easily identified as fads in retrospect, the failure of a research program to define its goals operationally and the circumlocutions that legitimize undefined research as 'providing insight' or 'furthering our understanding' are among the characteristics that should raise the critic's suspicions about faddism.

Dayton (1979) and Simberloff (1982) term 'competition' a fad, but

Abrahamson, Whitman & Price (1989) were unable to identify fads in the publication rate of papers within such large fields. If fads exist in science, they occur within narrow topics and the term should be reserved for lower elements in the hierarchy of ecological constructs. Thus competitive exclusion, limiting similarity, Hutchinsonian ratios, diffuse competition, and ecological divergence seem closer to the everyday use of the word 'fad' intended here, because interest in these areas is less enduring and more variable than the sum of all interests in the many aspects of competition, predation or life history.

Because fads are restricted in both the extent of their application and the duration of their interest, they are usually easier to identify within one's speciality. For example, a list of fads in freshwater biology (Peters 1988a) might include the measurement of primary production using $^{14}C - CO_2$ (Vollenweider 1971, and personal communication), nutrient turnover (Rigler 1975a), the size efficiency hypothesis (Hall *et al.* 1976), the sequence of studies of increasingly smaller organisms as we moved from studies of fish and zooplankton in the nineteenth centuries, to net plankton early in this century then on to nannoplankton, bacterioplankton, ultraplankton and picoplankton (Reynolds 1984; Stockner 1988). Other limnologists study the trophic cascade (Carpenter *et al.* 1985, 1987), nutrient responses and loadings (Peters 1986), top-down versus bottom-up control (McQueen, Post & Mills 1986), biomanipulation (Shapiro & Wright 1984) and food web manipulation (Benndorf 1987). In streams, the concepts of nutrient spiralling and the river continuum (Vannote *et al.* 1980; Statzner & Higler 1985; Minshall *et al.* 1985) are currently 'hot topics'.

These areas need not be weak. A great deal of good science can be done within them. Problems arise when researchers forget that their role is to test hypotheses and instead make observations that only 'pertain' to a topic in some unspecified way. Rigler (1982b) parodied this approach in an amusing model of the normal introduction to an ecological paper (Table 8.1). Such alternatives are especially tempting if it is easier and more enjoyable to make observations than to build and test hypotheses (Keddy 1987). In ecology, this choice is biased towards data collection by the pleasures of field work and away from hypothesis testing by the effort of intellectual rigour.

The trap of originality

Because many of the central issues of ecology are confused or untestable, ecologists are necessarily uncertain about what information the science

Table 8.1 *A substitute for hypothesis formation and a model for the Introduction of many ecological papers.*

X is a very
(*abundant, unusual, economically important*)
type of
(*species, community, ecosystem*)
Although other similar
(*species, communities, ecosystems*)
have been exhaustively studied, and some aspects of X have been studied
(references inserted here to demonstrate a knowledge of the literature)
X's
(*feeding, production, physiology, lead content, etc.*)
has not. Therefore I decided to study
(*one of the last set*).

Note:
Unimaginative researchers need only select specific terms from between the brackets to create the outline for a normal Introduction (from Rigler 1982b).

needs. Nevertheless, as good researchers, ecologists recognize clever and engaging research, independent of the problem under study or its place in science. When decisions are made about future research, funding or publication, such originality in research design is often rewarded at the expense of less technically brilliant and inventive research that might yield more relevant and useful information.

In science, ingenuity should be the handmaiden of wisdom. Wisdom leads to significant questions about nature, and ingenuity to the most effective solutions to those questions. Uncertainties about what is relevant to ecology have encouraged ecologists to skip the first step and to judge research in terms of an inadequate criterion of ingenuity alone. This substitution is a perversion of science.

Change and novelty have no scientific value in themselves. Theories and facts neither wear out with repeated use, nor should they become stale and unappetizing with age. Nevertheless, research fads and bandwagons, thoughtless reverence for new machines and techniques, disdain for the older literature, and evident preference for cunningly conceived research into trivial ecological hypotheses on the part of major journals, especially *Science* and *Nature*, show that contemporary ecology has placed too high a premium on novelty and originality. This is particularly problematic in America where government funding agencies prefer 'neat and nifty' ideas over painstaking scholarship and useful data, to the detriment of both science and society.

Verification and falsification

The imprecision of many ecological constructs makes them easy to confirm and difficult to falsify, because they are inconsistent with so few possible observations. As a result, it is easy to design research programs that will verify weak hypotheses, so verification must not be the goal of research (Murray 1986). As Dayton (1979) puts it, if ecologists are only interested in verifying a given thesis, this can be achieved most easily by designing experiments which maximize the chance of finding information that confirms the theory, and minimize any possibility of rejection. Such tests confirm the prejudices of the scientific community and are therefore likely to meet a sympathetic audience in both the reviewer and future readers. Introductions to such papers are pernicious because, if the research was undertaken to support an hypothesis, one purpose of the Introduction must be to obscure the intentional weaknesses and bias of the test.

Other introductory flaws

Logical lapses

The Introduction should form a logically ordered whole. Instead, relevance and context are sometimes 'demonstrated' by a bald list of other works in the area, with little attempt to describe the essence of this work, much less show how this precedent, thoughtfully considered, necessitates the present study. Sometimes, the relevance, context and lacunae described are vast, but the study is so narrow that the paper can only disappoint. Sometimes, the test is relevant to only part of the general hypothesis discussed in the Introduction so the reader feels cheated by the bit of information the paper holds. Sometimes the general hypothesis is so redefined by the test, that the test has little meaning for the hypothesis as it appears in the statement of relevance. The best defense against these non sequiturs is careful attention to the logical connections between each sentence and each paragraph (Woodford 1968).

Stock problems

Some pitfalls for Introductions to scientific papers need no discussion. The description of the paper's place in science should be brief, so that its point is not lost in a welter of references and detail. Since a research report is not a literature review, the point of the Introduction is to provide a logical context for the research and support for substantive claims by citing key references defining the hypothesis under test, rather than to prove one has read widely. This does not excuse a poor grasp of the literature. The lacuna in scientific knowledge identified by the paper must be real and the approach subsequently used to address that lacuna must be sound, not reflections of insufficient reading or selective interpretation.

Methods

The sections of a scientific paper are read in a different order from that in which they are presented (Woodford 1968), and doubtless with a different frequency, enthusiasm and impact. The actual reading order is Title, Abstract, Figures, Tables, Introduction, Discussion, Results, and Methods. This is understandable, but regrettable, for proper evaluation of a scientific paper requires evaluation of its techniques. Failure to read the Methods places more trust in the referees and authors than is extended to the remainder of the paper. This places a special onus on the authors as they select, describe and consider the methods.

The purpose of the Methods is to describe techniques in sufficient detail that an informed reader would be able to judge and repeat the work. The methods are more easily described if the researcher uses standard procedures, wherever these are appropriate, and is consistent in the application of whatever techniques are used. If modified or new methods are introduced, the novelties should be described, justified and tested. Explanation of the purpose of the major techniques, relating them briefly to the hypothesis under test, helps order and demystify methodological description. Many descriptions fall short of these standards.

Technical problems

Many of the shortcomings of the Methods involve highly technical and specific points that lie beyond the responsibilities of general criticism. One general weakness associated with these technical aspects involves the search for a 'technological fix' in the belief that better or just newer instrumentation by itself can resolve the problems of ecology. This can produce a succession of newer, larger, more expensive devices to monitor ecological systems at ever finer scales in space and time. It some areas, like physiological plant ecology, biogeochemical cycling, and remote sensing, this has led to a technological race, like the arms race, as different laboratories vie for the latest in analytical power.

Better analytical capacity is essential to test hypothesis that require better measurements, like those involving trace quantities of toxicants or minor nutrients. However more-sophisticated measurement must not become an end in itself, justified on the grounds that it somehow gives a truer picture of nature. Such a claim cannot be substantiated, because there are no independent estimates of what true pictures of nature might be. Our knowledge of the world consists of theories about how observations hang together (Hutchinson 1953) and the most sophisticated measurement has meaning only if incorporated in a theory relating that measurement to other variables of interest. The introduction of new instrumentation to a field should therefore illustrate the greater utility of the new measurements to theory.

New technology should not be applied for its own sake, and the pressure to justify the expense of new technologies by publication must be resisted. Papers that simply describe new methods seem to have little impact because they are among the least cited (Jumars 1987). This contrasts sharply with the success of papers that describe well evaluated methods with a clear role in the science; the latter are the most cited of all (Garfield 1988).

Citation analysis suggests that new technologies initially experience a phase of expansion, as they are examined and adopted by laboratories world-wide. Subsequently, citation rates may slow as doubts and second thoughts accumulate, and finally a steady, low level of citation reflects use of the technique in a few specific tasks (Peters 1989a). In aquatic sciences, this description seems to fit the evolutions of the ^{14}C-bicarbonate estimates of primary production in the phytoplankton (Vollenweider 1971) and of electronic particle size analyses in zooplankton feeding experiments (Peters 1984). No doubt technologies in many parts of ecology share a similar history.

'Mathematistry'

An almost identical situation occurs with respect to new mathematical and statistical tools. Periodically, the ecological literature is swept by claims for new calculating devices, like the logistic (Kingsland 1982, 1985), the simulation models of the 1970s (Patten 1975; Rigler 1976), catastrophe theory (Thom 1975), fractals (Mandelbrot 1982), and chaos (Gleick 1987). These may all be useful tools in certain situations, but initial propagandizing often oversells their importance and utility whereas initial 'applications' seem little better than self-justified abstractions (Hall & DeAngelis 1985) that may only repeat scientific commonplaces (Gray 1987). Box (1976) called this concern with mathematical abstractions 'mathematistry'. Critical readers in ecology should be wary of papers which serve only as vehicles for mathematical abstractions (Slobodkin 1975; Levin 1981b).

Statistical considerations

The most common weaknesses of the Methods are likely to be statistical. Proper tests invariably invoke statistical comparisons and these comparisons are generally stronger if the study is designed with a particular test in mind. Ecological statistics is too vast a field to be summarized here and much lies well beyond any claim of competence on my part. Statistics are better learned from direct applications of the statistics in the context of one's own research, supplemented whenever possible with appropriate readings, texts and courses. In the absence of such experience, Green's ten rules (Table 8.2) provide a concise summary of statistical advice for biological research, and will usually be a sound basis for critical assessments of the literature. Those rules related to the necessity of estimating the uncertainty in measurements merit the emphasis of reiteration here.

The difficulty of effective controls

Green's rules must be applied with the knowledge that real data only approximate his desiderata, and that too rigorous an interpretation may prove counterproductive. For example, it is unlikely that 'true' controls, those that differ only in one characteristic, are ever achieved in biology. This is particularly evident at the ecosystem level and Likens (1985) suggests that the term 'reference system' be used instead of 'control' for such comparisons.

Whatever term is used, the role of a reference or control is to determine the behaviour of systems not subject to treatment. This requires an estimate of both mean and variation associated with untreated systems which in turn requires replicated controls. Since the temporal variation within a single system can differ from spatial variation (Knowlton, Hoyer & Jones 1984; Likens 1985; Schindler 1987, 1988; Marshall, Morin & Peters 1988), reference systems must be either run parallel to the treatments or replicated in time. They must be replicated in space in any case. Without replication, the results will only support untested speculations, not valid conclusions. For example, the whole lake experiments of Carpenter *et al.* (1987) involve comparisons

Table 8.2 *Ten statistical principles for ecological research (Green 1979).*

(1) Be able to state concisely to someone else what question you are asking. Your results will be as coherent and as comprehensible as your initial conception of the problem.
(2) Take replicate samples within each combination of time, location, and any other controlled variable. Differences among can only be demonstrated by comparison to differences within.
(3) Take an equal number of randomly allocated replicate samples for each combination of controlled variables. Putting samples in 'representative' or 'typical' places is *not* random sampling.
(4) To test whether a condition has an effect, collect samples both where the condition is present and where the condition is absent but all else is the same. An effect can only be demonstrated by comparison with a control.
(5) Carry out some preliminary sampling to provide a basis for evaluation of sampling design and statistical analysis options. Those who skip this step because they do not have enough time usually end up losing time.
(6) Verify that your sampling device or method is sampling the population you think you are sampling, and with equal and adequate efficiency over the entire range of sampling conditions to be encountered. Variation in efficiency of sampling from area to area biases among-area comparisons.
(7) If the area to be sampled has a large-scale environmental pattern, break the area up into relatively homogeneous subareas and allocate samples to each in proportion to the size of the subarea. If it is an estimate of total abundance over the entire area that is desired, make the allocation proportional to the number of organisms in the subarea.
(8) Verify that your sample unit size is appropriate to the size, densities, and spatial distributions of the organisms you are sampling, then estimate the number of replicate samples required to obtain the precision you want.
(9) Test your data to determine whether the error variation is homogeneous, normally distributed, and independent of the mean. If it is not, as will be the case for most field data, then (a) appropriately transform the data, (b) use a distribution-free (non-parametric) procedure, (c) use an appropriate sequential sampling design, or (d) test against simulated H_0 data.
(10) Having chosen the best statistical method to test your hypothesis, stick with the result. An unexpected or undesired result is *not* a valid reason for rejecting the method and hunting for a 'better' one.

among three lakes representing two different treatments and one control; those of Schindler (1978) only one treatment and one control. Because the designs preclude any estimate of the variation associated with either treatment or control and because the number of degrees of freedom in any comparison is zero, no statistical comparison of the lakes or treatments is valid, and treatment effects cannot be distinguished from casual differences unassociated with the manipulations. Such distinctions can only be provided by lumping these treatment comparisons with other similar data from our experience outside the experiment, not from the limited evidence of the experiment alone.

Ineffective controls are not restricted to the ecosystem level where the scale of the study may put proper control beyond the budget of most ecological researchers. They are a common, but even less acceptable, feature of tests at all scales in ecology.

The necessity of replication

To determine if treatments have had a significant effect, ecologists often compare the means of different treatments or they correlate and regress responses against treatment levels. In all cases, effects can be distinguished only if they are large relative to the inherent variability of replicates and the number of samples that can be taken. Thus, variability is best estimated in a pilot study and can then be used to estimate the number of samples required to test for effects of a given magnitude (Elliott 1977). This is part of 'power analysis' (Toft & Shea 1983; Peterman 1990), for it indicates the statistical strength or power of the comparisons.

If the expected magnitude of the effects is so small that it is unlikely to be detected with realistic sampling effort, the study should be abandoned. If pursued, it will prove an inconclusive waste of valuable scientific resources. Insufficient sampling will incur a 'Type II' error in which no significant difference is found, even though such a difference may exist. For example, Emlen (1986) was unable to accept his own findings that Hawaiian bird densities were unrelated to site productivity, partly because his censuses were sufficiently erratic to obliterate any effect of productivity. The tests were invalid due to Type II error reflecting the poor design of the study.

The dangers of pseudo-replication

Hurlbert (1984) has identified a further pitfall associated with the analysis of controls and treatments: the biasing of the estimates by 'pseudo-replication'. Pseudo-replicates are repeated samples which misrepresent the range of possible values that treatments or controls might take because the replicates are not statistically independent and the number of independent measurements overestimated. For example, Fowler & Lawton (1985) criticize experiments purporting to show that herbivore damage to one tree induces a defense reaction in others, because of pseudo-replication. In this case, replicates did not involve the susceptibility of a series of different trees, but the susceptibility of one tree to a series of different larvae. Variations in herbivory estimated in such an experiment reflect variation in larval feeding rate, not variation in tree susceptibility, and any shift in the mean degree of herbivory on this tree might reflect any characteristics that distinguish this tree from the control, not just the effect of treatment. Even different trees in the same incubator may be pseudo-replicates because their response could as easily reflect an incubator effect as a treatment effect. A similar doubt hangs over the work of Hassell, Southwood & Reader (1987) who analyzed the population dynamics of white fly (*Aleurotrachelus jelinekii*) larvae on individual leaves of viburnum (*Viburnun tinus*). Since all the leaves were on a single bush, the supposedly independent populations could have instead been entrained and correlated by the behaviour of their single host.

Pseudo-replication is related to the older requirement that the sample represent the population. It differs because it does not emphasize the 'true' nature of the population being sampled, but the undesirable effects that interdependence of the samples can have for statistical description of the larger population. The result of this bias is that the parameters need not apply to the population as a whole or to other populations. Any statistical tests invoking parameters derived from pseudo-replicated studies could be invalid.

Problems of representative sampling

The problems of taking representative samples and of its converse, sampling bias, have had a long history in ecology (Southwood 1966; Elliott 1977; Downing 1979; Green 1979; Morin 1985). The problems endure because they are insoluble and they are insoluble because we can never know the reality beyond our samples. We are therefore limited to comparisons among different samplers and protocols to determine which comes closest to the device and technique we consider most reliable. Reliability is usually subjective.

A more effective criterion for the best sampling protocol would be the utility of the estimate, where utility refers to the function of the estimate

as a variable or boundary condition in one or more theories. The best sampling strategy results in estimates that lead to the best predictions within logistic constraints. Such estimates are 'representative', because they provide more information. Critical scrutiny of the methods should therefore consider the potential utility of the estimates in available theory and ensure that the methods represent the variables as defined in the theory under test.

It is usually assumed that the most reliable samples are also the most useful. That this is not always the case is demonstrated by the coexistence of pure and applied science (Cartwright 1983). Applied science exists as a separate field because pure science does not work well in application; presumably pure science exists because it is considered more reliable (i.e. realistic) notwithstanding. Comparisons of the utility and reliability of planktonic chlorophyll a concentrations estimated with sophisticated high pressure liquid chromatographic (HPLC) and with standard spectrophotometric analysis of crude pigments in an organic extractant provide an ecological illustration. HPLC estimates are surely more reliable, but crude extractions are far more informative, because they appear in predictive relationships for transparency (Carlson 1977), fish production (Jones & Hoyer 1982), zooplankton biomass (McCauley & Kalff 1981) and primary production (de Lafontaine & Peters 1986). The HPLC estimates cannot be introduced to such equations, because these reliable estimates bear no consistent relation to the crude extractions on which the informative regressions are based (Jacobsen 1975; Sartory 1985); HPLC estimates may be up to an order of magnitude below the crude estimates and would therefore underestimate the response variable in these equations. It is even doubtful that relations built on reliable HPLC chlorophyll a values would perform as well as the existing relations. The crude extracts likely reflect the range of biologically important properties of the phytoplankton community better than the sophisticated measurement. This situation will be resolved only when freshwater biologists give up the fiction that traditional techniques actually measure 'chlorophyll a' and adopt a different term for the object of the crude extractions.

Results

'The Results' describe the observations, establishing what differences were significant and what were not. Since the study should be directed towards an hypothesis, these findings are easier to grasp if they are related

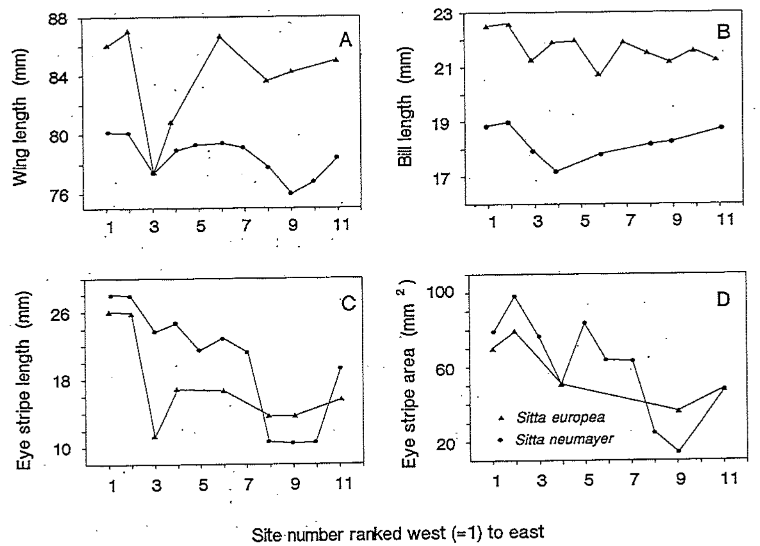


Fig. 8.1. Peak analysis of the covariance between four morphological characteristics of two sympatric species of rock nuthatches in 11 areas lying along a cline from Yugoslavia to Iran. Although the original author felt that these show 'good correspondence between the species', his own correlation analyses revealed significant correlations only in the case of eye stripe area and bill width. The latter relation was not plotted here because it was omitted in the original (From Grant 1975).

to the hypothesis under test, leaving comparison of the results with the literature for the Discussion.

Results and Discussion are frequently fused to avoid needless repetition. Although these papers may be easier to write and read, the merger may obscure the distinction between observation and conjecture. Writers who fuse Results and Discussion should therefore be especially careful to separate the two trains of thought with appropriate paragraph structure and verb tense.

Description of the Results

One of the most vexing problems in assessing papers in contemporary ecology is to discover just what the results were. High publication costs, editorial precepts, competition for journal space, and the awesome capacities of modern computers encourage extensive reduction of the data before publication. Although some reduction of the raw data is essential to presentation, this can be carried so far that the observations are obscured. Even when familiar data-reduction techniques are applied, the results are often so digested that no check on the accuracy of the author's calculations is possible (Gray 1987). This might be acceptable if ecologists made few calculation errors, but that premise would strain the credulity of the most casual critic. Full disclosure of the data is an elementary precaution against both honest mistakes and charlatans. In any case, the financial support provided by society is normally given so that the results can be used by the community and scientists should therefore feel obligated to make their results public.

The problems of data access could be easily circumvented if authors accepted that publication of the research entails the availability of the data to the community. Extensive data may be placed in inexpensive data banks and repositories dedicated to this purpose (Prothero 1986; Downing 1979), but most scientists seem unaware of such services. In my experience, direct requests for the data behind published reports are usually either refused or ignored. Many scientists are too busy or lazy to prepare data in accessible form, some are too jealous or defensive to relinquish their proprietary rights, but most simply cannot find the old data. In all cases, their behaviour is antithetical to the open commerce of knowledge upon which science depends.

Analysis of the results

Once the results are established, they must be analyzed to determine whether patterns predicted by the hypothesis under test occur or not. Attendant problems are normally associated with inappropriate techniques for pattern analysis. Once the analysis is complete, it should be a simple process to decide if the hypothesis has been supported or not.

Bias in peak analysis

A preliminary step in the analysis of some data arrays is to plot different variables along a single axis. The plotted arrays are then examined for co-occurring peaks and troughs in the data set, so the technique can be called peak analysis (Fig. 8.1). This approach is often used to compare biological response and environmental factors along geographical or temporal gradients, but it is equally effective in any ordered display of the data.

Peak analyses are only appropriate for casual explorations of the data, because they are biased. Outliers are highly influential in identifying peaks and troughs, whereas the bulk of the data receives little or no weight. In addition, the correspondence between peaks and troughs in different variables is often hard to judge because the data are presented in different panels. This uncertainty is sometimes abused by implicit or ad hoc time lags to allow for imperfect matches in the data. Data may be ordered in this way if the analysis examines the relation between the response variables (e.g., phenological patterns or geographical gradients in biological response) and the ordering variable (e.g. time, site). Such ordinations are also effective in isolating eccentric data. However, if the point of the analysis is to relate the biological response to some other correlate of space or time, the appropriate visual tool is a scatter diagram relating biological response to the environmental correlate. The appropriate statistical analysis is a regression or correlation, not a verbal summary of the fancied correspondence between high and low values in space and time.

Misapplication of statistical models

Parametric statistics depend on the fit of the data to an abstract statistical model, like the normal or binomial distribution. This is inevitably an approximation (Box 1976), but many statistical tests are robust to such abuse. Nevertheless, there is a real possibility that lack of fit between the data and the assumed statistical model will bias any statistical test. For example, the frequency distributions of many biological data are closer to log-normal than normal (Koch 1966, 1969). When this is the case, the data should be analyzed after logarithmic transformation, for parametric analyses of the raw data would be misleading. This is the contention of Eadie *et al.* (1987) who explain the regularity of 'Hutchinsonian ratios' as an artifact of comparing the arithmetic means of log-normally distributed data. Such confusions can often be avoided by scrutinizing the data to see that it agrees with the results of statistical tests. When this is not so, transformation or non-parametric analyses (Conover 1971) may be required.

Negative evidence

Statistical tests are always probabilistic and therefore they are never certain, even if applied well. The possibility that Type II error might obscure real differences (and therefore lead us to 'accept the null hypothesis of no difference as true when it is false') was considered briefly in discussing experimental design. Type II error is so common in ecology that ecologists are wary of any 'negative evidence' that purports to show

the absence of an effect. Such results are less interesting to consider and are generally hard to publish (Connell 1983; Toft & Shea 1983; Rotenbury & Wiens 1985; Peterman 1990). These difficulties can be overcome if the study includes a careful consideration of the possibility of Type II error and if the results of the test are in some sense unexpected. In this case, evidence that no difference was found is fully as informative as 'positive' evidence that differences exist.

Type I error

Type I error refers to the complementary problem of falsely identifying a significant difference where none exists. If the statistical models used are appropriate to the data, the probability of a Type I error is the level of significance (the α level) of the test. Thus when there is a 5% probability that a certain observation occurred by chance, the probability of Type I error is also 5%. Because the fit of the data to the model is always approximate, such probability levels should not be interpreted too strictly; nevertheless they are usually the best available indicators of the chance of a Type I error.

The likelihood of Type I error can be increased if an extensive data set is analyzed for many possible patterns. This might be the case in 'all subset regression' which regresses every variable against every other or in multiple pairwise comparisons. In 100 regressions or comparisons, one would expect to find that five of these were significant at the 5% level, on the basis of chance alone. An example is available in the key factor analysis of *Viburnum* white fly mortality by Hassell *et al.* (1987). This study considered 108 regressions between different life stages of the fly in different years and corresponding population densities, and identified 16 significant relations. Since about a third of these can be expected to represent Type I error, the significance of all 16 must be suspect. Because similar ambiguities result whenever extensive multiple comparisons are made, the critical level of significance must be lowered. A Bonferroni correction designates the new critical level to be $5/n$ % where n is the number of regressions considered (P. Legendre, personal communication). In such cases, significance at the typical levels of 5% or 1% can only be a tentative indication that a relation may exist.

Discussion

The Discussion should further relate the results of the study to the hypothesis, acknowledge weaknesses that may compromise this interpretation, point out the consistency of the results with existing knowledge, and suggest some of the implications of the new results for scientific theory. Like the Introduction, the Discussion is usually one of the most interesting sections to write because it allows the writer some scope. As a consequence, it is also likely to share debilities with the Introduction, including loss of focus with attendant confusion of the reader. Such difficulties frequently result because the author's views are so skewed by precept that the paper is distorted to support these views, rendering the order and selection of material idiosyncratic and illogical. Those who read Discussions, and those who write them, should be aware of the many problems that can be encountered there.

Flew's flaws and fallacies

Many of the problems in scientific discussions occur in most human discourse and are only specific examples of the general difficulty of clear thinking. These everyday problems have been described by Antony Flew (1975) following an old suggestion of the philosopher, Schopenhauer, to give each of many fallacies and logical errors a short and obvious name 'so that when a man used this or that particular trick, he could at once be reproached for it.' Although this chapter can no more do justice to applied logic than it can to statistics, some good comes from a simple list defining Flew's terms with ecological examples. If nothing else, this exercise may encourage others to read his short and witty book, but it should also serve to warn of some of the problems that may be found in the Discussion.

Denying the antecedent and affirming the consequent

In logic, propositions are given the form 'if P, then Q', but there is nothing intrinsic to logic in the form. Because scientific theories are also propositions, they fit the same model: If an animal is a 6-kg North American omnivorous mammal, then its population density will be between 0.2 and 100 km^{-2} (Fig. 2.1). Science must be concerned with both the truth (i.e. the predictive or factual value, which is usually addressed in Results) of the proposition and the logical validity of inferences drawn from the proposition (which is primarily the concern of the Introduction and Discussion). Logic, the topic of this section, deals only with logical validity.

Two valid inferences can be drawn from the proposition 'if P, then Q'. Assuming the proposition holds, then wherever P obtains, Q will

also obtain, and wherever Q does not obtain, then P will not obtain either. Thus, if the proposition holds, one can deduce that the density of a 6-kg mammalian omnivore lies between 0.2 and 100km^{-2} and that an animal whose density is 10,000 km^{-2} is unlikely to be a 6-kg omnivorous mammal. Two other inferences would be fallacious: if P does not obtain, there is no reason to assume that Q also does not obtain; and if Q is found, one cannot infer that P also occurred. In the concrete terms of the theory in Fig. 2.1, just because an animal is not a 6-kg North American omnivorous mammal, we cannot infer that its density is not between 0.2 and 100 km^{-2}; and a density between these limits is not enough to identify an animal as a 6-kg North American omnivore.

Reiterating the explanation in Chapter 7, the traditional name for the fallacy of asserting P when Q is observed is 'affirming the consequent', because one correctly affirms that the consequent (Q) was observed in order to assert fallaciously that the antecedent (P) also occurred. The second fallacy is called 'denying the antecendent' because it correctly states that the antecedent (P) was not found but fallaciously concludes that the consequent (Q) could not hold. Both fallacies occur widely in ecology. For example, it may be proposed that if two mice species are strong competitors, then they will be allopatric, but it would be fallacious to conclude from that proposition that allopatric mice are competitors, because this affirms the consequent (allopatry) in a vain ploy to affirm the antecendent (competition). The fallacious claim that mice should be sympatric simply because they are not competitors instead denies the antecendent (competition) to deny the consequent (allopatry).

Similar fallacies are often involved in mechanistic arguments. For example, it might be proposed that if a large composite process, like primary productivity, has a Q_{10} of 2, then the rates of its constituent processes double with each 10°C in temperature. This is a testable scientific proposition. However, one cannot affirm that the Q_{10} of the whole will be 2 because that value was determined by laboratory study of one or more constituent processes, nor can one presume that Q_{10} of the constituents is not 2 simply because the whole responds differently to temperature change. In ecology, these seeming inconsistencies may be explained away with multiple causality and scale effects, but as logical fallacies, they need no such effort.

The Un-American fallacy and the no-true-Scotsman move

The 'Un-American fallacy' is named to recall Senator Joseph McCarthy and the House Committee on Un-American Activities who used the fallacy to identify communists as all those who shared some belief, pacifism or atheism for example, with communists, even though the belief was in no way particularly or necessarily communist. The traditional name is the 'fallacy of the undistributed middle', and this fallacy imputes all characteristics of a given class to any object having one of these characteristics. As a result, any 'average' characteristic applies to all members of the class.

This fallacy can be used to smear opponents in ecology, too. For example, because Christian fundamentalists are opposed to the principle of evolution by natural selection, defenders of evolution may be tempted to protect their position by implying that all evolutionary critics are fundamentalists (Ruse 1982). Lehman (1986) associates the empirical work of his limnological opponents with the fraudulent, but putatively empirical, work of the psychologist, Cyril Burt. The fallacy may find less-emotional application in bolstering typologies: for example, if *K*-selected species tend to lay large eggs or live long, then it is tempting to class any organism that has large eggs or a long life as *K*-selected (Pianka 1970).

A second and related fallacy excludes undesirable cases as unrepresentative of the class. Thus a proud Scot might save his nationalist self-esteem simply by reclassifying some deplorably behaved highlander as 'no-true-Scotsman'. In ecology, one might accept both the evidence in Fig. 4.3 that animals which eat both plants and animals are common and Pimm's (1982) contention that omnivores are rare (Chapter 7), by defining omnivores as organisms which take at least 20% of their diet from non-adjacent trophic levels (Pimm 1980). In other words, the omnivores in Fig. 4.3 are 'no-true-omnivores'.

Redefinition is not fallacious or reprehensible in itself. What must be avoided are the logical inconsistencies that result if familiar words are used in unfamiliar senses without first warning the audience of this change in meaning, or if different definitions are confounded, for example, by proposing a new definition but then shifting ground to use the term in its older and more widely accepted sense. In other words, Pimm's assertion has no bearing on the scarcity of omnivores in the English sense of animals that eat both plants and animals.

The logically-black-is-white slide, the genetic fallacy, and the heaper

Definitional problems often arise where distinctions are so vague that a more or less arbitrary line is drawn between classes. 'The logically-black-is-white slide' exploits the resulting uncertainty to argue that all distinctions based on that division are in reality non-existent or unimportant. This argument denies significance to differences in degree and stresses the importance of quality over quantity. This fallacy encourages and is encouraged by widespread acceptance of types, categories and individuals as ecologically important, and by mistrust of the mathematical arguments that often accompany quantitative treatments, even though quantitative treatments lend themselves better to more rigorous and logically transparent analyses than typologies permit.

The arbitrary use of critical significance levels provides a concrete example of the problem. Because any critical level is arbitrary, one may be tempted to assume that the distinction between significance and non-significance is illusory and subject to reinterpretation. Such a refusal to play by the accepted rules and policies of science can represent a significant rupture of the implicit contract between the scientific expositor and his or her audience.

'The genetic fallacy' is a special case of the logically-black-is-white slide. The genetic fallacy holds that whatever developed or evolved from something is still really the same as its antecedent. A chicken is always an egg. Flew (1975) uses the example of Desmond Morris' (1968) contention that man is only a naked ape on the grounds that man evolved from the apes. Leaving aside the issue of which end of the evolutionary continuum should be more offended by this assertion, the naked ape remains an example of refusing to consider differences in degree as important.

The 'heaper' or 'sorites' is the related fallacy that no matter how many small sequential changes occur, there is never a point where one thing changes into another. Thus if a heaper (in Greek, *Sorites*) adds grains of sand to a small pile one-by-one, there is never a point at which this pile changes into a heap.

Three fallacious premises

The 'truth-is-always-in-the-middle damper' refers to doctrinal adherence to the principle of the golden mean. Although the truth may often lie between opposing viewpoints, this cannot be taken for granted and Flew demonstrates that, in practice, the position is incoherent and absurd. If the truth in any debate lies at the mid-point between the two debaters, then the truth will vary with the extremity of the positions taken. A Machiavellian opponent could coerce the truth to any position simply by taking extreme enough positions for the sake of argument.

This might work in bargaining for public opinion, but it ought not sway scientists for whom the value of a construct depends on its predictive value in dealing with nature, not on a gimmick of exposition.

It is regrettable that Flew redefines all of his fallacies in contemporary English, for the long title of 'the whatever-follows-must-be-the-consequence fallacy' is less euphonious than the traditional 'Post hoc ergo propter hoc'. Both English and Latin phrases refer to the fallacy of attributing causal force to whatever antecendent conditions may have been observed or remembered. In ecology, this fallacy frequently occurs as casual explanations used to explain away outliers with ad hoc devices, in the causal attribution following insufficiently replicated treatments, and in discussion of dichotomous natural experiments that relate observed changes to whatever variable the writer feels is important. The problem recurs in the broader stricture that correlation is not causation and reflects the general difficulties and ambiguities involved in causal attribution (Chapter 5).

The 'first maxim for Balliol men' is that even a truism may be true. Flew (1975) introduces this aphorism to counteract the insidious and fallacious premise that only new propositions need be entertained or that propositions may be dismissed simply because they are old, boring or hackneyed. This premise also encourages the pathological emphasis on originality noted in discussing the Introduction earlier in this chapter.

The subject/motive shift and four derivatives

One of the difficulties in writing or reading any scientific prose is to keep one's mind on the subject. All too easily, discussion can shift to related but different problems and a convincing argument can be elaborated which has nothing to do with the topic under scrutiny. The best defense against such shifts is to stress what substantive thing one wants to know and how that knowledge may be achieved, instead of stressing who 'wins' the argument. Flew (1975) discusses these problems in relation to a change from the subject of discussion to the motives behind the various positions: the 'subject/motive' shift.

'The but-you-can-understand-why evasion' shifts the debate from a consideration of whether a given position is scientifically correct because the theories it espouses are better predictors to a justification of the position in terms of scientifically extraneous criteria that explain why an erroneous position is held. Valid motives for maintaining a position are confused with valid positions. Thus we may understand if capitalists stress competition and socialists stress mutualisms as key ecological interactions, but the scientific strength of those positions is entirely independent of their supporters' politics.

The 'but-they-never-will-agree diversion' confounds adequacy of support for a position with the effectiveness of that position in convincing the opponent. Scientific questions must be decided on the merits of different views as evidenced by information value and not on the extent of opposition or approbation offered by others. The opinion of others may have no basis at all in fact, but depend entirely on precept and misinformation. Scientific debate must focus on which position has the strongest theoretical and factual support, not which has proven most appealing to other scientists.

The grounds for discussion can also be shifted by simply taking the subject of the discussion for granted, by 'begging the question'. Thus the debate between theoretical and empirical approaches to an ecological question may lead to the dismissal of each side by the other as 'too theoretical' or 'only empirical'. Such entrenchments beg the question, but do not resolve disagreement.

Begging the question is a defence similar to 'the fallacy of the pseudo-refuting description' whereby the opponents' positions are dismissed by the simple act of giving them a name. For example, 'reductionism' may be used falsely to refute mechanistic studies or 'naive falsificationism' can be used to decry the criterion of predictive ability. In these cases and many others, the critic must explain what is fallacious in those positions, and not simply tar them with a vaguely condemnatory name.

Two debating ploys

The 'pathetic fallacy' is better termed a misconception than a fallacy, for it is not a logical error but an unsupported proposition based solely on analogy. Nevertheless, Flew (1975) calls attribution of human emotions, motives, and desires to non-human entities 'the pathetic fallacy' because it requires us to experience a sense of pathos or compassion for the object of our research. Since there is rarely evidence for the feelings of such entities, this is scarcely a strong position, yet ecologists repeatedly employ the pathetic fallacy when they discuss 'ecological strategies'. This may be an acceptable shorthand, but the critic must be wary that the strength of the argument depends on the power of the theory to which the analogy leads and not the appeal of the analogy.

The 'fallacy of many questions' is also not a fallacy, but simply a trick to set the audience on a desired track. This is achieved by posing questions that build false assumptions into any answer. The stock example is 'When did you stop beating your wife?' A number of ecological examples were collected in Table 1.1.

Comparisons of data and hypothesis

In science, instruments fail, sampling schedules are disrupted, unexpected behaviours appear, statistical assumptions are only approximated, and unusual events occur, so the results of even the best-planned experiments are less clearcut than intended. In these cases, the empirical support for the hypothesis depends on interpretation of the data. In ecology, this is especially necessary because hypotheses are poorly phrased, methods uncertain, and data irrelevant. Consequently, discussions frequently begin by interpreting the data in light of the hypothesis and critical readers must be alert to the possibility of bias in this interpretation.

Superficial and partial tests

Superficial analyses take the agreement between some general aspect of the phenomenon and some implications of a mechanistically or hierarchically structured hypothesis as a confirmation of that hypothetical structure at all levels. This was discussed with reference to mechanism in Chapter 5 and as the fallacy of affirming the consequent in this chapter. Since different processes can induce the same gross phenomenon, such tests do not effectively discriminate among alternate mechanisms. Loehle (1987a) holds that this is a common error in the evaluation of computer models. Dayton (1973) addressed this problem in his complaint that models might be right for the wrong reason, because analyses of the hypothesis at lower hierarchical levels were not upheld. Belsky (1987) has termed this a confusion of scale. Kalff (1989–MS) details the problems associated with making ecological predictions from physiological studies.

Partial tests present a similar problem, because the support for one aspect of a complex hypothesis is held to confirm other aspects, but is distinguished from superficiality in that the other phenomena are at similar scales. Simberloff (1978) holds that this is the case in island biogeographic theory where the existence of the species–area curves is taken to confirm the existence of a stable equilibrium between unmeasured rates of colonization and extinction.

The confusion of correlation and causation is another example: simply establishing a correlation between two variables does not imply that

manipulation of one will induce variation in another. Manipulability must be established in separate experiments. For example, evidence from dogs suggests that the correlation between mammalian size and gestation time is not manipulable, because dogs of all sizes gestate in about 63 days. In contrast, a similar correlation between size and food intake is manipulable, for big dogs eat more than small ones. Both partial and superficial tests are sometimes called inferential because their confirmation is used to infer qualitatively different attributes of the system from those which were observed or are necessitated by the observations. Such inferences may make good hypotheses for future studies, but they are not deducible from the evidence available in partial or superficial tests.

Special pleading

Many studies generate eccentric values that contradict the hypothesis under examination, but are not considered falsifications of the hypothesis under examination. Such data are normally explained away by invoking bias, or artifact, or some other ad hoc explanatory device. This is entirely acceptable if it is not overused by claiming unusual circumstance more frequently than necessary (Weatherhead 1986). Nevertheless, the ploy protects the hypothesis from falsification and may hide particularly interesting data that do not fit conventional theories. The reader must be informed when data are excluded and told how much such exclusions influence the interpretation. Only then can the critical reader decide whether the exclusion was acceptable or not.

The quagmire of plausibility

In 1887, Stephen Forbes could legitimately claim that, given the virtual absence of appropriate ecological information and the effort required to collect such information, it might be more effective to use the 'a priori road' of speculation and plausible argument. A century later, the same argument has a hollow ring, but plausibility still plays a large role in contemporary ecology. Plausible arguments are an amalgam of clever writing, commonsense and dogma which should be regarded with considerable suspicion. As Howe (1985) writes in his critique of the view that tropical American plants coevolved with now extinct gomphotherian mammals as agents of dispersal: 'It is unsettling that the idea is most plausible when applied to fruits we know virtually nothing about'. The a priori road may be valid for erecting hypotheses, but it is a poor tool for judging them.

Literature comparisons

Any ecological measurement could be biased by factors which the author was unable to consider. As a result, discussions seek to reassure the reader that the data collected were appropriate and representative of the phenomenon under study. One way to do this is to cite other measurements from the literature which indicate that similar values have been observed elsewhere and that the results of this test are consistent with the expectations based on an accurate reading of the literature. If this is not the case, the difference between the literature and the observations should be explained with a testable hypothesis.

Selective comparison and self-delusion

Where the literature is extensive, it is easy to flatter one's point of view by unconsciously selecting references and information. This selection might be achieved by restricting comparisons to trends rather than absolute amounts. In the early 1970s, the literature on phosphorus excretion by zooplankton claimed agreement because all studies showed that specific excretion rate declined with increasing size. This agreement virtually excluded the further observation that published estimates of excretion for animals of the same size differed by more than 1000-fold (Rigler 1973) or that the rates of decline with increasing size differed.

Comparisons are therefore best made quantitatively and statistically, but this is effective only if the data are an unbiased sample from the literature. If this is not the case, comparisons with the literature may degrade to meaningless polemics (Schoener 1972) or simply to self-delusion whereby we support our biases by unconciously selecting data that are most similar to our own.

Over-interpretation

Given sufficient good will and a sufficiently loose hypothesis, aspects of many different phenomena will seem consistent with our hypothesis and beliefs. One flaw in discussion may therefore consist of an accumulation of diverse observations which are interpreted as consistent with the hypothesis, even though they are only incidental. For example, Howe (1985) sees many of the points raised to demonstrate the coevolution of gomphotheres and neotropical trees (Janzen & Martin 1982) as too subjective, qualitative, and irrelevant to test the hypothesis adequately. As a result, using these points to support the hypothesis is over-interpretation. Such over-interpretations are easy to make when the

hypothesis they appear to support is particularly engaging (as in the theory of megafaunal dispersal), or particularly widely accepted (as in the case of competition theory; Simberloff 1982) or particularly loose (as in the case of adaptationist arguments; Gould & Lewontin 1979; Ollason 1987). One litmus test against misinterpretations is to ask if the hypothesis would have been falsified by observing the converse of what was actually observed. When this is not the case, the evidence has been over-interpreted.

Over-interpretation is a common vice in ecology, for many ecologists see that their purpose is to organize the diversity of ecological observations, rather than to predict then. This clogs the literature with intellectual jig-saws that exclude and overshadow testable hypotheses.

Significance and magnitude

If the sample size is large enough, even very small differences are highly significant. It is therefore appropriate to ask if very small, but statistically significant, effects are also biologically important. Fowler & Lawton (1985) adopted the position that small differences obtained in laboratory tests may not be biologically important in the field because the highly defined, invariant conditions of laboratory testing may prove unrepresentative on extrapolation. For this reason, isolated statements of probability are insufficient; the magnitude of the difference and its associated variation must also be considered.

Vacuous contrast

A frequent trick in favorably comparing one's results to the literature is to make a vacuous contrast, claiming that the hypothesis of choice has some special virtue not contained in an unspecified, and actually non-existent, alternative. For example, Janzen & Martin (1982) cite the rotting of seeds and fruit under the parent canopy of certain South American trees as evidence in favour of their hypothesis that the extinct megaherbivores once ate these fruits and so dispersed the seed; This would carry weight only if the seeds of other trees, whose dispersal agents are not extinct, did not rot under trees. Since they do (Howe 1985), this is a case of vacuous contrast.

Critics often use vacuous contrasts to create and criticize straw men. This step may be necessary in ecology because many ecological constructs have to be defined before they can be addressed. However, this opens the criticism to the charge of irrelevance, because some aspect of the ecological concept is almost inevitably excluded by the process of definition (see Chapter 4), and vacuous contrast, because the model being criticized is a figment of the critic.

Since it seems unlikely that any scientific paper is flawless, this discussion of all the ways a scientific paper could be improved may also seem a vacuous contrast. This is not so because the argument is not that ecology should consist of perfect papers (thereby invoking a non-existent perfect science), but rather that the science will progress faster if existing flaws are recognized. The contrast between the literature and perfection would be vacuous, that between better and worse elements in the existing literature is not.

Extensions and hypotheses

Perhaps the most important function of a scientific paper is to describe where we go from here as scientists: the Discussion is therefore important because it draws out the implications of the hypothesis. It explains what we now can predict that might not have been possible before and how we can build on this with future research. This should be one of the most exciting parts of the paper, but often it falls short because the authors are so unaware of the limitations of their work that they feel all questions have been resolved, or because they do not recognize the importance of new hypotheses, or because they have become so confused by their own rapportage that the hypotheses they offer are either unrelated to the paper or untestable as stated. The critical reader should be aware of these difficulties and be alert to any hypotheses which the author has made, but not recognized, or recognized but not made.

Tunnel vision

Most scientists work with a limited set of ideas and these limitations restrict their breadth of view on their work. As a result, the self-evaluations represented by discussions are likely to favour prevailing dogma and personal beliefs, and alternative approaches are less likely to be considered thoroughly.

For example, fisheries biologists are so firmly wedded to the concept of density-dependent population control that they find it difficult to conceive of alternatives, even though the evidence for density-dependence in fish stocks is virtually non-existent (Beverton *et al.* 1984). Beverton *et al* (1984) claimed that 'it is difficult, if not impossible, to conceive that the upper extreme of population size is not limited

ultimately by density dependent processes, which is one manifestation of self-regulation (homeostasis) . . . ' and Rothschild (1986) can protest that, despite all the evidence to the contrary, 'there must be some relation of recruitment to stock, otherwise stocks would not be persistent . . . '

The historical record of any persistent stock should show a tendency to decline from high populations and to rise from low ones. With enough data, this pattern, which is inherent in the definition of an historical mean population size, will generate negative relationships between the size and growth rate of the population and will result in regression coefficients of less than unity when the logarithm of population at time t is regressed against that of the population at time $t+1$. Neither of these common tests for density-dependence entails a density-dependent, causal process (Eberhardt 1970; Maelzer 1970; St Amant 1970). Murdoch (1970) recommends that the words be used simply to indicate relationships between density and the change in numbers, rather than some underlying homeostasis. Fisheries biologists have mistaken precepts and logical models for the data, even though these precepts and models do not apply to the data in hand. It would be surprising if this confusion led to particularly effective fisheries management policy.

In the context of the Discussion, this bias is likely to lead to one-sided arguments in which the preferences of the author usually carry the day. For example, supporters of optimal foraging theory find that it predicts effectively (Stephens & Krebs 1986; Stearns & Schmid-Hempel 1987), but its critics conclude just the opposite (Gray 1987; Pierce & Ollason 1987). Similarly, Gilpin & Diamond (1984b) are frustrated at the inability of Connor & Simberloff (1984a,b) to see the virtues of competition theory. No doubt, Connor & Simberloff (1984b) are just as puzzled that the vices of competition theory are not equally apparent to Gilpin & Diamond. It is important therefore to recognize the existence of scientific bias. Writers of Discussions should strive to free themselves from this handicap by self-criticism; but since they are likely to fail, critical readers should be alert to the possibility of bias.

Gould (1981) has demonstrated the vulnerability of science to bias in his analysis of the evidence supporting sexist and racist preconceptions of human intellegence. Gould (1986) divides these into three groups: the frauds, like Cyril Burt, who manufacture evidence; the finaglers, like Morton, who subtly and perhaps unconsciously bias their results by selective application of otherwise acceptable corrections; and those, like Broca, who suffer such disabling bias that their interpretations of the data are warped and untrustworthy.

Ad hoc hypotheses
Many papers introduce ad hoc hypotheses to explain away eccentric data from the literature or from their own experiments. This is a normal process and is not too damaging if the ad hoc hypothesis is testable (Fretwell 1987). Ideally, this test is performed in the context of the paper, but if this is not possible, it should be tested in some future study. If this is not done, the test offered in the current paper is so compromised that it is meaningless, because whatever conclusion was reached rests on a crucial, untested assumption. In practice many ad hoc assumptions are not intended to be tested and therefore should be discounted as special pleading.

Field relevance
Theories in ecology are relevant only if they can be applied in the field. Consequently, one of the most interesting elements in the Discussions of ecological papers describes the implications of the paper for organisms in nature. To the extent that a study has been removed from nature, this consideration is pivotal, and the reader should expect to find some evidence that such an extension is warranted. Alternatively, there should be warnings that no effective field testing has been done, specifying the sorts of information which would constitute such a test. This need not be extensive, but it is a useful protection against unwarranted claims for field relevance.

Extraneous material
Of all the parts of a paper, the Discussion is most likely to contain material that is not essential to the evaluation of the hypothesis under test, and the inclusion of such extraneous material is one of the most confusing aspects of many Discussions. For example, ecologists often confuse the precise use of words to present data and theories with their vague use to motivate and suggest ideas (MacArthur 1972b). Since only the former use is appropriate in the Discussion of a scientific report, this confusion can bewilder the reader. The same result is achieved when the Discussion is used to introduce arguments, facts and theories into the literature, not because they are germane to the hypothesis under test but because they concern the author. Woodford's dry advice to the author who wishes to write a good Discussion is 'avoid megalomania'. No reader wants to know all of the author's thoughts about something, so the author should decide which points are most germane, make these points and keep the rest to himself or herself.

Minims
Mentis (1988) identified a special problem of megalomania in ecology as the production of trivial hypothesis or minims. This term is an analogue of a maxim and is defined as a statement of proverbial form having no general application or practical use. Any test of the proposed hypothesis would require so much effort and yield so little information that no one would wish to do so. It is an aphorism which appears to offer wisdom but actually says very little. The competitive exclusion principle (Hardin 1961) might be one example.

Extrapolation
Extrapolation in its broadest sense refers to the extension of an hypothesis beyond the original domain over which the hypothesis was built and tested. In one sense, this is inevitable. Because every situation is different in some way, each application of the hypothesis represents an extension of the original domain of application. Such extrapolations present no new problem for science.

The charge of extrapolation is usually limited to cases where one of the boundary conditions or one of the predictor variables takes a value outside the domain of application specified in the theory. The commonest case is likely the extension of a regression line beyond the range of the data set (Fig. 8.2). Such extrapolations are considered risky because they presume that the effect of a variable is unaffected by scale. Sometimes this is so: general allometric relations for rates of metabolism seem to extrapolate relatively well to very large organisms. But this is not always the case: relations relating speed to size overestimates the maximum speeds of large animals (Bonner 1965).

One widely discussed extrapolation involves the growth of the human population. Von Foerster, Mora & Amiot (1960) fit a set of past estimates for the human population of the Earth as a function of date (in years AD) such that:

$$\text{Population} = 1.79 \times 10^{11}/(2026.87 - \text{Date})^{0.99}$$

This leads to the expectation that world population will reach infinity on Friday, 13 November, 2026 – Doomsday. This relation has done surprisingly well at predicting human growth (Caswell 1976), although it now underestimates the world population (Umpleby 1987). However, since we can be certain that the population will never reach infinity, we know that it cannot extrapolate to doomsday. Indeed, even before doomsday, the equation must prove false for it would require doubling

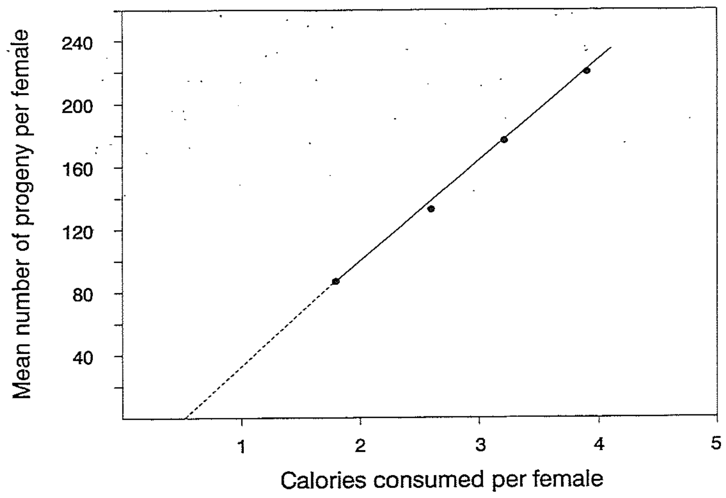


Fig. 8.2. An extrapolation beyond the domain of the data on which the regression, describing reproductive output as a function of ration for *Daphnia*, is based. (Modified from Hassell 1978.)

rates of the human population that are beyond the biological capability of our reproductive system (Deevey 1987).

Extrapolation does not always introduce a fatal flaw. The extrapolation in Fig 8.2 is particularly interesting, because it does not accord with the expectations of most zooplankton ecologists (Porter, Orcutt & Gerritsen 1983; Frost 1985; Geller 1985; Stemberger & Gilbert 1985) who prefer to fit such data as a concave downward plot. However, the available data suggest that in this case linear extrapolation is more likely to be consistent with the evidence than a more complex curvilinear relation (Condrey 1982; Condrey & Fuller 1985).

Summary – The challenge of good science

Science is best when it is straightforward, logical and concise, but achieving these simple goals is not easy. Science requires a judicious blend of wisdom and knowledge in choosing an appropriate question, single-minded dedication to that question in selecting and pursuing the appropriate techniques, followed by a perceptive and unbiased judgement about the importance and relevance of the findings. This development requires sound ethics and broad knowledge to select a relevant,

soluble, unsolved problem, a practical grasp of epistemology to address the problem so that it will yield meaningful results, appropriate statistical and technical tools so that meaningful results can be properly evaluated, and basic but powerful logic to pursue the implications of the analyses, all bound together in varying proportions by effective writing. Strong science is difficult to achieve, but that is no excuse to try for less.