An introduction to experimental methods for language researchers

Monica Gonzalez-Marquez, Raymond B. Becker
and James E. Cutting

1.   Introduction

Imagine yourself as a field researcher sent to Africa to study lion colonies. The savannah can be treacherous in its monotony. Inching forward through the tall grasses you suddenly discover yourself face to face with the snarling leader of a lion tribe. Would you say, "Pardon me, but I would like to know a little more about how you communicate with the other members of your family. For starters, how is it that when you, say, roar, that comes to mean 'an elephant is coming'?" Maybe you would have a flashback to your philosophical coursework, and begin to wonder about Wittgenstein's famous aphorism, "If a lion could talk, we could not understand him." Clearly, *asking* a lion how he communicates would be fruitless. What you would need to do is arrange controlled circumstances under which you could record his vocalizations in response to situations you construct and whose meaning you do understand. For somewhat similar reasons, there are serious epistemic weaknesses to simply asking a *human being* how she communicates. Although she may be able to reply in a language that you understand, the introspections that she shares with you will have undergone a great deal of perceptual, cognitive, and social alterations and filtrations. Her subjective report of how she uses language will only be a shadow of how her mind really uses it. To make matters worst, we know that people are not good at understanding why it is they behave in certain ways (Wilson & Nisbett). For cognitive linguists, this issue has particular implications, specifically those of embodiment theory. The following sections will flesh out how to go about asking similar questions, albeit about people, within the confines of the scientific method. They will also address practical concerns that should serve as a preliminary guide to doing experimental work.

2.   Experimenting with experimentation

How do you do an experiment? How do you figure out what to study, which questions to ask, which methods to use to ask them? Answering these and all of the other related

questions adequately, takes many years of training, false starts and successes. It all begins however, with learning to read a research paper. The approach can be thought of rather like reverse engineering, as each section of a research paper has a relationship to the research process. In the following sections we will deconstruct the research article with two goals in mind, to help you, as a new consumer of research literature, better understand published studies, and to help you, as a new experimentalist, design and develop your own experiments. As such, the material will first be presented for the benefit of a naïve reader followed by elaboration for the experimentalist. In the first section we recommend that you concentrate on the structure of the paper. Do not worry if you cannot follow the arguments well. What is important is to become familiar with the way a research story is told. Anything else that you manage to grasp will be icing on the cake. In the second section try to put yourself in the shoes of the researcher conducting the project by trying to understand why the researcher made the choices she did. Your ultimate goal, after all, is to understand the nature of those choices well enough to be able to conduct your own experiments. All of the information provided here should be general enough to apply to any type of behavioral research. To benefit most from this chapter, we recommend that you follow with a copy of a published experimental research article. We recommend those from a prestigious journal such as "Psychological Science" or "Cognitive Science."[1]

3.   Reading the research article

"Stephen King's new novel will be available this Saturday at Modern Times Bookstore." This announcement may sound appealing to a Stephen King fan or inconsequential to someone who is not. Intimidating, overwhelming and confusing are not words usually evoked at the thought of actually reading the book. Many people would use precisely these terms, however, to describing reading a research article. There are reasons for this. Most people know what to expect from a novel. A novel is the telling of a story, and since King writes popular fiction, it is expected that the plot will contain a given set of elements that will be linked to each other in predictable ways. What makes a research article so intimidating to some readers is that they do not know what to expect. The American Psychological Association publishes a handbook that describes the general structure of articles. Though the handbook is aimed at researchers writing up their findings, following is an adaptation intended to help new consumers of psychological research know what to expect from a scientific research article.

The analogy between reading a novel and reading a research article was intended to help you conceptualize an article as a story unfolding paragraph by paragraph. As in popular fiction, when we read a paper, what we want to know is, 1. What is the setting, 2. who did what to whom, where, why, how and with what, 3. what was the outcome, and 4. what

---

1.   Those readers who do not have easy access to scientific journals will find that the proceedings of the Cognitive Science Society conference are available free online at http://www.cognitivesciencesociety.org/cogsci.html

does the outcome mean for what we thought the story was. This information can usually be found in predictable places. Research articles, unlike (unfortunately!) most novels, have very strict space limitations. This means that every piece of information is significant. Everything from the title to the list of references will be relevant to understanding the piece of research as a whole.

3.1 Title

The title of an article serves two purposes, to inform the reader as briefly as possible about the subject matter of the paper, and to facilitate indexing in databases. Because authors have a very limited amount of space, they tend to use key terms in an eye-catching kind of way. For example, if a researcher were writing up an investigation into how bilingual children respond to image schemas while listening to stories in both the languages they speak, it is very likely that the title might be something like "Image Schemas and Narrative in Bilingual Children." This title makes it amply clear to readers interested in any of these three topics that this paper might be relevant to their needs. In addition some journals now allow inputting keywords for indexing.

3.2 Authors and affiliations

Authors listed are responsible for the research reported. The order of the authors reflects the agreed upon level of responsibility for the published work. The first author is most responsible and so forth. Affiliations typically reflect the institution that is the primary residence of each author. These might be academic institutions such as Cornell University, or they can be private research facilities such as the Santa Fe Institute. In addition, clinicians and others without institutional affiliation simply list their home city.

3.3 Place of publication

Different journals publish different types of research. Some are also more prestigious than others. Prestige is determined by different factors though the most important is consistent high quality research based on high selectivity. Examples of well respected journals that will be of interest to cognitive scientists in general include *Psychological Science*, *Trends in Cognitive Science*, and the various *Journal of Experimental Psychology* publications. As you read more research articles, it will become clear which journals are most relevant to your needs.

3.4 Year of publication

The year of publication is important to understanding general trends in research over time. Both within the context of an individual researcher's work as well as compared to other research, it will help you track the order in which findings emerged so as to better conceptualize the grounding of the findings.

As you become better acquainted with the literature of a certain field, it will become important for you to be able to reference different studies. These 4 pieces of information typically serve this purpose. It is considered good practice to commit to memory the authors, and the year, the publication outlet, and the title of a paper in decreasing order of importance. Google (either the general search engine or the specialized Google Scholar) will help you find the paper generally with the first and second pieces of information, and easily by including the third. Having this information readily available to memory will prove invaluable in activities ranging from preparing a study to having a conversation with a colleague on a research topic.

3.5 Abstract

An abstract at about 120 to 150 words is a concise summary of a research article. Though dense, it is generally well organized and is expected to contain specific types of information. The reader can expect to find what the main problem under investigation is, sometimes the characteristics of the subjects in the experiments, including sex and language abilities if relevant, etc., a description of the experimental method used, the findings, and the conclusions and implications. Like the title, the content of the abstract is also used for indexing in databases.

3.6 Introduction

The introduction is the first step into the paper itself. It is meant to set the stage for the rest of the paper by providing the reader with three pieces of information.

1.   What the research question is.
2.   What the literature has to say about it.
3.   What the authors have done and why.

The research statement will be described at the beginning, usually very briefly. The literature review will involve presenting brief summaries of key studies related to the research question. Care will usually be taken to only discuss studies that are directly relevant. In addition, only the relevant parts of the studies themselves will be discussed. This point is especially important to keep in mind if you wish to do research along similar lines since it means that in order to get a complete picture of a cited article you will have to read the original article and should not rely on a brief summary. The final part of the introduction will give the authors' hypotheses and their motivation for conducting the research. In other words, they will briefly describe why they think the research is necessary.

3.7 Methods

The methods section is in many ways the 'proof is in the pudding' section in that it gives enough information about the experiments conducted so that anyone else can also conduct them (more about why this is important below). Details about the experiments will fall into these categories:

1. subjects
2. apparatus
3. procedure

3.7.1 *Subjects*

The subjects section will describe the subjects' characteristics. Typically demographics will be given along with the number of subjects, and the method used to recruit them. In addition, authors are encouraged to include information such as sex/gender. ethnic/cultural/national origin, socio-economic status (SES), and linguistic abilities. The rationale is that though these details may not be the target variables in the current studies, they might be relevant to other researchers for future studies. For example, a paper might say "50 Spanish/English early bilinguals participated for course credit. Their average age was 20 years of age with a range between 18 and 26 years. 25 were male and 25 were female. They were recruited from introductory psychology courses at the University of Texas at Austin. All subjects were of Mexican origin. Their SES was working class." Someone conducting a study five years hence might be investigating the impact of socio-economic status on cognition. The fact that the researchers above reported this information will prove valuable to the later researcher. Another important bit of information found in the procedure section is attrition rate. Authors will usually explain why any subjects were excluded, and how the criteria were set.

3.7.2 *Apparatus*

The apparatus or equipment used in an experiment can vary from the quite sophisticated such as a functional magnetic resonance imaging machine (a machine that uses blood flow to take pictures of activity in the brain) to the mundane such as playing cards. Regardless of what is used, all details necessary to obtaining the same tools will be provided. This includes model numbers of equipment as well as the names and versions of specialty software. If the equipment is particularly complex or unusual, an appendix may be added containing the necessary elaborated descriptions.

3.7.3 *Procedure*

The procedure section contains the details of exactly how and where the experiment took place. The grouping of the subjects, the nature of the stimuli, and the general breakdown of how the trials varied step by step is stated here. Figures, which are helpful for adding context to the description, can be found here as well. In brief, this section covers the steps from the beginning instructions to the debriefing.

3.8 Results

The results section provides a summary of the data collected using both descriptive and inferential statistics (see Nuñez this volume for an elaboration). The results will be given first in plain English and then in their statistical form as means, p-values, F-tests, etc. All analyses deemed necessary to support arguments in the discussion section will be pre-

sented here. In addition, any tables and graphs necessary to explain the results will also be found here.

3.9 Discussion

The previous parts of the article set all of the cards on the table, so to speak. It is in the discussion section that the authors will evaluate their results in relation to their hypotheses. This last point is particularly important because much of the disagreement about papers is specifically about whether the conclusions drawn follow from the data (or whether they can be explained with alternative explanations). In general, they will address points dealing with how their research was useful in resolving the target problem as well as any implications or consequences to the theory they are working in. Although this is the only place where anything resembling an opinion on the part of the authors is appropriate, most researchers will tend to be cautious about their claims, instead adopting the more productive position of encouraging further inquiry.

Many articles are composed of more than one experiment. When that is the case, you should expect short preliminary discussions to follow each experiment, with a general conclusion at the end.

3.10 References

The references section will list, in APA format, all sources cited in the research article. Aside from giving credit where credit is due, it is also an excellent starting point for further readings.

4.   Other types of articles

The previous sections described the construction of an article reporting original research. There are two other main types of articles that you will encounter. The first is a review article and the second is a theoretical article. The review article is in many ways a tutorial of the subject matter. The author will review the available literature on a topic so as to evaluate its progress. You should expect to find summaries of key articles relating to the target question. A good review will also point out inconsistencies, contradictions, and gaps in the research, as well as recommendations for future research. They are especially useful for anyone considering research in a new area as well as for students rounding out their training.

A theoretical article shares many of the characteristics of a review article. The main difference is that where review articles tend to be geared toward seeing how well a research question has been addressed by the scientific community, a theoretical article will seek to support a theoretical model using extant research. Though most journals will publish both review and theoretical articles, especially good sources are *Behavioral and Brain Sciences*, *Perspectives on Psychological Science*, and especially *Psychological Review* for review articles

and *Psychological Bulletin and Review* for theoretical articles. For beginning researchers, *Current Directions in Psychological Science* will prove a particularly accessible publication.

At this point you should be quite familiar with the type of information contained in the different parts of a research article. Hopefully you will have read not one, but several papers using the guidelines set out above and your mind is filled with questions about how and why things were done the way they were. Later, we will present additional criteria for reading journal articles that should help you develop solid experiments. Before continuing with the next section, following are a few preliminary details that should help you better conceptualize the nature of the experimental process.

5. The scientific method

> The purpose of the scientific method is to make sure nature hasn't misled you into thinking you know something that you don't actually know.
>
> Robert M. Pirsig. Zen and the Art of Motorcycle Maintenance, 1974

The word 'science' comes from the Latin *scire* meaning 'to know.' Science presents one of many ways of knowing. Other examples include the method of tenacity, the method of authority and a priori method (Rosnow & Rosenthal 1993). The method of tenacity refers to the manner in which people will cling to an idea simply because it seems to be common sense. The method of authority refers to taking what an authority figure such as a priest or a teacher says as fact without question. The a priori method refers to the use of pure reason and logic to come to conclusions about the world. Though all three are commonly used, they share a characteristic. The only verification they require is that of the individual having the belief. What sets science apart is its dependence on intersubjective verification, the possibility that knowledge can be empirically tested by different researchers. The Scientific Method, in general terms, is an intellectual framework geared at generating the most reliable findings possible as well as at facilitating their verification.

There are many misconceptions about how the scientific method is actually used. A common one is that it describes the step-by-step procedures followed unvaryingly by all scientific disciplines. The reality is quite different. Researchers use different methods developed specifically to understand their chunk of the environment (Tattersall 2002; Solso & Johnson 1984; Rosnow & Rosenthal 1993). The procedures used to investigate lava flow do not necessarily apply to understanding how neurons communicate nor to understanding how to cure cancer. They are all scientific methods insofar as their motivation is to produce reliable and verifiable findings, though it is doubtful that they share too many other practical characteristics. It is actually more appropriate to refer to the scientific method as scientific methodology (Solso & Johnson 1984) so as to make amply clear that there is actually a compendium of methods, all with a common goal.

Two other common misconceptions are actually different sides of the same coin. In the general description of scientific methodology, the possibility that findings could be verified by different researchers was given as an important component. The consensus possible from multiple testing has sometimes been misinterpreted as agreement by con-

sensus. This type of agreement is also known as majority rule. Actually, since science is not a democracy, what it refers to is agreement by repeated recognition of low probability of falsehood (Tattersall 2002). Statistics is a central factor in reaching consensus (see Nuñez this volume). The general form that an accepted idea takes (and which will be further developed below) is that it has been shown to be very unlikely the result of chance. Specifically, that when different researchers went through the process of collecting data in comparable ways, and analyzed their results, the statistical analyses they conducted indicated a very low probability that the results were the product of random variation. Thus, there is consensus about the high probability that the phenomenon described by the results is not illusory, leading directly to the other side of the coin for this issue.

If you were to ask any person off the street what they thought was the goal of science, most would unfortunately respond that it was the pursuit of The Truth about the world. Not quite. Most scientists, in fact, believe that there is no way we will ever know the true nature of the world, a perspective formalized in the 20th century as a philosophical movement called Positivism. And that therefore, the best we can do is to develop possible explanations supported by empirical findings. The goal of science is then to find the best available explanation for a phenomenon, with 'best' qualified as being the least likely to be false. This is not the same as saying that it is most likely the correct explanation, but only that it is better than any other under consideration. It also means that it too can be dislodged the moment another explanation is found that better accounts for the target phenomenon.

Science as a human endeavor, prone to all of the subjectivities of the human experience, continues to be the topic of much debate. Though some would disagree, we propose that science is our attempt as a species to approach the closest thing to objectivity that we can achieve as subjective beings. Since our goal is to introduce the issues to the novice, we will limit discussion to just these few points. Readers interested in further elaboration are encouraged to consult Carnap's (1966) "The philosophy of science," Chalmers' (1976) "What is this thing called science?" Kuhn's (1962) "The structure of scientific revolutions," Lakatos' (1970) "Falsification and the methodology of scientific research programs," and Popper's (1959) 'The logic of scientific discovery."

6.   Conducting research

Aronson, E., Wilson, T. W., & Brewer, M. B. (1998) claimed that there was no better way to understand the research process than by the careful reading of research articles. At the beginning of the chapter, you were introduced to the structure of a paper and were instructed in your initial readings to only pay attention to the general context and not necessarily the content of what was written. Understanding the research process via journal articles obviously requires more than structure based readings. Besides knowing where to look for information, it is also necessary to have the criteria to evaluate the contents. The basis for which depends on background knowledge of the subject matter as well as on an understanding of the nature of the different components involved in doing research. The first element can only be acquired through the deep study of the choice subject, the method-

ological aspects of which will hopefully be facilitated by the contents of this volume. The latter, along with a short repertoire of questions formulated to facilitate evaluation of each research component, is discussed below.

The following sections describe scientific methodology as developed to investigate cognitive phenomena. As above, the different steps will be presented within the structural context of the research article. In other words, each step will be discussed in reference to the section of a research article where it appears. This does not mean that there is a one-to-one correspondence between research steps and the sections of a paper, only that all the steps are represented in some form throughout. Our goal is to help you recognize them so that you will better understand the research process. (Note: All papers described in this section marked with "CS" are available for download at http://www.cognitivesciencesociety.org/cogsci.html. They can be found in the 2003, and 2004 proceedings.)

Conducting research generally involves the following steps, roughly divided in 2 parts (boldface indicates where the information can generally be found in a research article.)

Part One:

Introduction –
1. Narrowing down the topic of interest
2. Conducting an exhaustive literature review
3. Deciding on a question
4. Formulating a hypothesis

Part Two
1. **Methods** – developing an experiment
2. **Results** – analyzing data
3. **Discussion** – interpreting results

6.1 Part one: The introduction – the thinking and reading steps

The first 4 steps of conducting research are typically addressed in the introduction. They involve the project's conceptualization and motivation. Though they are given as steps, they might be better described as components in that their relationship with each other is interdependent. The ultimate goal is to produce a question that can be addressed experimentally. The decisions made along the way determine the general architecture of the research, and as such, should be considered carefully.

6.1.1 *Narrowing down the topic of interest*
The first step in conducting a research project is to narrow down the topic of interest. If you are reading this book, chances are very high that you have at least a couple of ideas that you would like to test experimentally. Chances are also very high that your ideas are far too broad to be addressed by one project, i.e. "what is the relationship between the body and language;" 'how is spatial information used by language," etc. This is common to new researchers. In fact, a large part of doing good experimental research involves learning how to ask questions that can be addressed experimentally. A good way to think of this process

is as an inverted pyramid, the first step of which is determining your general area of interest, followed by deciding on a particular question and ending with the development of a specific hypothesis. Regardless, deciding which general phenomenon you want to study is the first step. An important consideration as a beginning researcher is that your topic have a history of being studied by other more experienced researchers. That is not to say that you should always shy away from topics that have not been investigated experimentally, only that in learning how to do research, working from well established models will prove infinitely useful. In short, you want to do as the experts have done.

Deciding what area you want to study begins with reading broadly. If you are interested in language and space, the first step is to introduce yourself generally to the issues. A good place to begin if you have no background might be with an undergraduate textbook. Sternberg's (2003), is a relatively up to date survey of cognitive psychology. The section on knowledge representation with its treatment of spatial cognition will prove quite enlightening if you are interested in the meaning structures underlying language. (Note to readers with a background in a different discipline: as you read the material, it will be very useful to suspend judgment about the perspective taken. You may not agree with the topics of study nor with the conclusions. Regardless, it will be key that you try to understand the perspective that led to the types of questions asked, a skill that will be pivotal to accurate critical assessment of all the research you encounter.) The next step might be more specialized books such as Bloom, Peterson, Nadel, and Garret's (1996) *Language and Space*, and Newcombe and Huttenlocher's (2000) *Making Space*. Books might be more appropriate as they tend to give more conceptual descriptions of research, and may be easier to digest for the research novice. Alternately, review articles by journals such as *Current Trends in Psychological Science* and *Psychonomics Bulletin & Review* will be useful. The goal is to find something that piques your curiosity.

6.1.2   *Conducting a literature review*

After deciding on a general topic, the next step is finding out what other researchers have already done, and that requires learning how to search. If you did begin your inquiry by reading a book or a review article, the bibliographies will be a very useful place to begin your background reading. Otherwise, the place to begin will be at the library using databases. Most libraries have online databases you can access either on site or via the internet. Each will have specific guidelines and tutorials for how to conduct searches. Walk yourself through these before beginning; it will save you much aggravation. Searching databases in general involves using keywords. Though your library will have guidelines to help you along, a place to start will be with target terms used in what you have read as general background in a textbook, etc. An alternate source involves using a cataloguing software such as Devonthink, which, among its many features, can provide you with frequency lists of the words used in an article. These often prove quite useful when conducting a search.

Besides library databases, other resources include http://scholar.google.com, a new search engine dedicated specifically to academic concerns. Doing a generalized 'google' search (http://www.google.com) can also be a good idea. In choosing what to read, however, avoid anything that has not been published in a reputable journal. Make sure, at

the very least, that the piece was written by someone whose work has been referenced elsewhere or that you know.

Your goals in actually sitting down and reading the masses of papers your search will have produced are 1) to find out what has been done, 2) why it was done and 3) how it was done. Going section by section, following are questions you should ask yourself as you read through the articles (adapted from Jordan & Zanna 1999). A last note before beginning to read, it is probably not a good idea to read the sections of articles in sequential order. Instead start off by reading the abstract then the introduction followed by the discussion/conclusion. When the answers to questions 1 and 2 are reasonably clear, continue with the rest of the paper. If the paper has multiple experiments, skip to the general discussion at the end of the paper, then read the inner sections

6.1.3 *Questions to help you understand a research article*

6.1.3.1 *Abstract*

– What is the paper about?
– How did the researchers set up their study? i.e., who were the subjects, what was the experiment?
– What did the experiment measure?
– What were the main results of the study?

6.1.3.2 *Introduction*

– What were the theoretical considerations underlying the research?
– Why was the particular topic chosen for study?
– Does the chosen topic have implications beyond itself?
– What are the authors hypotheses?
– What questions do the researchers hope to answer with the results of their study? (Note that this is a different question than what their hypotheses were.)
– How did the authors decide on their research strategy, i.e. did they develop an experiment or chose to do a correlational study?

6.1.3.3 *Method*

– How were the hypotheses turned into testable questions?
– How were the variables manipulated, i.e. how was the experiment done?
– Were appropriate controls used?
– Were the measures used appropriate to the question being asked, i.e. is income an acceptable measure of socio-economic status?

6.1.3.4 *Results*

– What are the main results of the study?
– Can the results be used to answer the research question?
– Can the results be generalized beyond the context of the study?

6.1.3.5 *Discussion*

–   What conclusions do the researchers draw from their results?
–   What questions were left unanswered by the study?

After finishing the paper, should you have found it particularly insightful and interesting, you might want to do a background search on the authors to find out what else they have published. Finding out where their work has been cited will also be helpful in developing a clearer picture of how the research fits the broader scope of work on your topic and on cognition in general. A database such as the Web of Science (found at your library's online database directory) as well as Google Scholar can provide you with this information.

6.1.4 *Developing a research question*

Reviewing the literature is closely tied with choosing a research question. You can think of it rather like spirals that feed into each other. You begin with a vague idea of what you want to study, i.e. language and space, then you read a few papers. It strikes you that most of what has been written has been on the English language, so you decide you would like to look at prepositions in French. You go and find everything you can, theoretically and experimentally, on French prepositions. After reading even more, you decide you are interested in locative prepositions, etc. By the same token, as you fine tune your question, persist with your literature search. A researcher who uses experimentation is first a researcher. This means that your first goal is to find the answer to your question. It is possible, if not likely, that in reviewing the literature you will find the answer to your original question, in which case you will have to decide whether to extend the findings somehow or to change topic.

Once you have done at least a preliminary literature review, the next step is deciding what the question itself should be. What do you want to know? Be as precise as possible. There are issues to keep in mind as you concretize your thoughts. The question you decide upon should be one you find deeply interesting. Doing research takes a lot of time and patience. A deep interest will be required to keep you motivated through the rough spots. The question should also be relevant. You should carefully think through why the research is timely and important. A good question will always go beyond itself to help explain a larger phenomenon. Though it will focus on a seemingly small problem, it should contribute to our understanding of language and cognition. Ideally, your question should also be novel, elucidating a previously unexplored issue. If it cannot be completely novel, at the very least, it should shed further, necessary, light on a well-established topic.

There are also practical considerations. The question must be testable. There is a famous Sidney Harris cartoon showing two bespectacled, lab-coat-arrayed scientists discussing their new elixir. One says to the other, "It may well bring about eternal life, but it will take forever to test." Time is not the only obstacle to testability, of course. Questions such as 'Can consciousness be transferred from one person to another?' may be deeply interesting but, to the authors' knowledge, are currently far beyond our means to explore.

Given that your question is testable, research questions vary in the scope, or the level, at which the investigation will address the problem. The scope of the research question

plays a balancing act with the practicality of finding an answer. As the scope gets wider, the more difficult it is to find a clear answer to a question. However, if the scope is too narrow, then the answer, although more easily found, might not be interesting.

Adequately addressing these issues will greatly facilitate the rest of the research process. A lack of clarity here will unavoidably be reflected in the finished product, i.e. the manuscript. You should only proceed with the more practical aspects of research after careful examination of the 'thinking' steps.

Ember and Ember (2001) lay out how to come up with a question that a researcher will be satisfied pursuing. Following are four types of research questions, using possible examples from cognitive linguistics.

> *1. Descriptive questions* are concerned with the prevalence of a particular phenomenon, i.e., What type of metaphors do people use to refer to temporal events in English versus Aymara? How is snow described in French as compared to Inuit?

> *2. Causal questions* are *why* types of questions, where the investigator wants to know the reason for the difference in languages, i.e., Why do Spanish speakers conflate manner of motion in the verb (e.g., *subir*), whereas English speakers use a satellite-frame construction (e.g., *go up*)? Why do Spanish speakers use a count noun in the plural to refer to popcorn (i.e., *palomitas*), where English speakers use a mass noun?

> *3. Consequence questions* ask about the effect of a particular grammar on other cognitive processes. For example, given that space is an integral component of the structuring of meaning in language, does being raised with two languages, each with their own way of representing spatial relationships, have any implications for the way that space is processed independently of language, as assessed by Shepard and Metzler's classic mental rotation task? One of the authors is currently engaged in addressing precisely this question.

> *4. Nondirectional relational* questions are typically used in correlational studies, where the investigator is not making conjectures about causality. Brain imaging techniques are often employed in an investigation of these questions, i.e., What parts of the brain lights up (are active) when people listen to a sentence describing active versus abstract sentences?

Each of these types of questions will yield qualitatively different answers. They each have a place in the investigation of any topic. Regardless of the type of question you choose, it must be translated into a research hypothesis.

6.1.5 *From research question to research hypothesis: scientific methodology*
Ayala (1994) describes the process of developing a research hypothesis as the most creative and imaginative element within research. The process begins, of course, with deciding on a research question, which is then interpreted as a research hypothesis. Interpreting your question as a scientific research hypothesis requires that you rethink it in terms of predicting what the relationship between two (or more) variables will be in a given context. Factors such as explanatory value and falsifiability also need to be integrated.

Earlier, scientific methodology was described as "an intellectual framework geared toward generating the most reliable findings possible as well as at facilitating their verification."

A more 'practical' definition of scientific methodology is:

> *Scientific methodology encompasses standardized methods of testing whether an idea, translated into a hypothesis, has explanatory value over a phenomenon in a setting that allows falsifiability.*

In other words, the next step is to frame your research question such that it makes predictions about the type of behaviors to expect whenever your variables (see below) interact in the ways you predicted such that your research question is addressed as directly as possible. In addition, your question must be framed such that the prediction can be shown to be wrong, meaning that it can be falsified.

A couple of examples should help clarify the process.

> *Example 1:* Casasanto and Boroditsky (2003, CS) are interested in how it is we understand concepts not directly available via direct sensory experience. This is their topic of interest and it is stated in the first few sentences of their paper. They also give the sources that started them thinking along those lines. Reading the work of researchers such as Gattis, Gibbs and Lakoff, etc. was instrumental in their choice of topic. They begin to narrow down their scope of interest by stating their interest in spatial metaphors for time. Notice that they give references to additional readings, i.e. Alverson, Boroditsky, etc. Their research question is whether the relationship between space and time, well-established in linguistic tasks, will also emerge in simple psychophysical tasks with nonlinguistic stimuli and responses. They re-interpret their research question as a research hypothesis centered on the asymmetry of the dependency between space and time appearing in language, postulating that it will also appear in psychophysical tasks.

> *Example 2:* For Bergen, Narayan and Feldman, the topic of interest is mirror neurons and their possible relationship with motion verbs. The sources that got them thinking about the issue include Gallese, Rizzolatti and Buccino. They are specifically interested in whether the comprehension of motions verbs involves some type of mental simulation, and whether this possible simulation might involve activation of the parts of the brain responsible for the behaviors described by the motion verbs. Further review of the literature led them to their research hypothesis. Namely, that information processed simultaneously by different modalities will occur at varying rates depending on its similarity.

Each of these sample papers begins with a broad statement of interest. The process resulting in a research hypothesis is well outlined by the literature they cite. Each step of the methodology is accompanied by supporting research. Though the whole process appears neatly packaged in the journal article, keep in mind that the authors spent a considerable amount of time reviewing different research, weeding out the useful from the irrelevant, all in an effort to develop as concise an image as possible of what the literature had to say about the topics they were interested in, as well as about what their research question should be and the form it should take as a research hypothesis. All the while keeping in mind the practicalities of actually finding out whether their idea could make a contribution or not.

6.2 Part two: The practical steps

6.2.1 *The methods section: Developing an experiment*
The methods section of a paper gives you the details of how a research hypothesis was tested. The authors will have arrived at this point only after having a very clear research hypothesis, as should you before beginning to consider the more practical steps. This is particularly important as you will next need to re-envision it within the context of the type of information an experiment can actually give you. Besides describing different research methods, this volume can also be thought of as a survey of the types of information produced by experiments. Common ones are response times, eye-movements, and differential activation in given areas of the brain. Deciding on a methodology involves deciding which type of information will best be able to address your research hypothesis. In order to do so, a clear understanding of research variables, experimental design, and differences between types of hypotheses will be necessary.

6.2.1.1 *Variables* Conducting a study, regardless of the type, is about attempting to understand how different components of a phenomenon, called variables, relate to each other. The most common types of relationships studied involve determining whether a variable is a part of a target phenomenon, or alternately, how it is relevant once it has been established that there is a connection.

6.2.1.2 *Variable classification* A variable can best be qualified as a set of events that can take on different values. Typical ones include sex, age, scores on an exam, number of milliseconds required to respond to a stimulus etc. They can generally be categorized in two different ways. Variables can be grouped according to their nature or according to how they are used in research.

Variables classified by their nature are behavioral, stimulus and subject (also sometimes called organismic). Behavioral variables comprise responses made by entities such as people or animals. How quickly a rat snatches a piece of chocolate is an example of a behavioral variable. Stimulus variables are those factors that precipitate a behavior. This might be a social context such as a party, or the reading of a sentence as a prompt. Subject variables are characteristics of objects of study that are used to classify them. They are generally not changeable. Sex and age are common examples.

Variables as classified by their use in research are Independent, Dependent, Extraneous and Constant. The independent variable is what is manipulated such that it will affect the dependent variable. Inversely, the dependent variable is what is affected by the independent variable. This relationship is unidirectional. For example, let us say that we were interested in finding out which social settings produced the most dancing. The independent variable would be the setting in which dancing took place, i.e. a birthday party, a wedding, an office meeting, a funeral. These are called levels of the independent variable, in this case 4 since there are 4 settings. The dependent variable would be the amount of dancing. The hypothesis is that the setting affects the dancing not the other way around, therefore, the independent variable is the setting and the dependent variable is the amount of dancing. It is important to keep in mind that there is no a priori reason why either set-

ting or dancing should be the independent or the dependent variable. Their classification as a given type of variable is entirely dependent on the research hypothesis.

Extraneous variables are factors that might affect the independent variable. In the dancing examples above, alcohol might be considered an extraneous variable in that it might dispose subjects to dance more so than a particular social setting. To avoid alcohol from having any effect you control for it by making sure that no alcohol is consumed during the experiment.

What if you decided that you were in fact interested in how predetermined amounts of alcohol, i.e. 0–4 drinks, might affect the amount of dancing in each of the 4 settings? Then there would be 2 independent variables, the settings and different amounts of alcohol. In formal terms, what you would be interested in is the possible interaction between setting and alcohol consumption quantity.

By the same token, you can decide that you are interested in any physical reaction that might be indicative of response to the music at each of these settings (assuming that the music remained constant.) You could then operationalize (see below) behaviors such as foot tapping or swaying as levels of dancing, along with full body motion on the dance floor. This would give you a design (discussed further below) that has 2 independent variables, one with 4 levels, i.e. the 4 different social settings, and one with 5, i.e. the 5 different quantities of alcohol, beginning with no alcohol at all up to 4 drinks, and 1 dependent variable with 3 levels.

The last type of variable is a constant variable. These are factors that remain unchanged across the different conditions of an experiment. For example, a constant here would be to have the same number of subjects in each social situation. A further constant would be to have the same number of males and females. Constants are kept so as to prevent unwanted variation. It could be that if the number of subjects were different in each of the different dance conditions that the amount of dancing would vary as a result, i.e. different ratios of males and females induce more or less dancing. A constant can be thought of as a possible extraneous variable that has been controlled for.

6.2.1.3 *Internal structure of a variable* Regardless of how they are categorized, variables have internal structure. They can be discrete, i.e. the number of times a particular image is chosen in a forced-choice task, someone's gender, etc. or they can be continuous, i.e. the amount of force a subject applies to a lever, someone's age, etc. The key difference is that discrete variables are not decomposable, i.e. a family cannot have 1.5 children, while continuous ones are, i.e. today you ran for 22.37 minutes while tomorrow, if feeling more energetic you might run 44.06 minutes.

A variable's internal structure can also be described as qualitative or quantitative. The former refers to kinds, someone can be female, Chinese and left-handed. These are all kinds of things that she is, each of which has no intrinsic numeric value. Quantitative variables predictably refer to variables than can occur in different amounts. How much money is (or mostly likely, is not) in a graduate student's bank account, the amount of time she vacillates before calling up her parents for a loan, etc. are both examples.

6.2.1.4 *Operationalizing variables* When variables are used in an experiment, all target ones need to be operationalized. This refers to transforming any concept you are interested in into an observable behavior that can be measured. In the dancing example, we can operationalize 'dancing' as (attempted!) rhythmic movement performed by a person not including swaying while sitting or leaning against a wall, or foot-tapping while sitting or standing. A video recording might be made such that you can measure how much time people spent 'dancing' only according to the operationalized criteria. This way the amount of dancing will have an objective measure instead of being left to subjective interpretation, i.e. what may count as 'a lot of dancing' to one person may seem like not much at all to another. In addition, should it be the case that another experimenter doubts your results and wants to see for himself, he will know exactly what measures to use to test the reliability of your results. Proper operationalization is also basic to testing the theoretical concept you want to test. e.g., if you want to see how accessible a word is, you can operationalize this question in terms of how quickly people say a certain word (naming), or how quickly they determine it is a word in English (lexical decision). There is much debate on what measures best test a theoretical notion.

6.2.2 *Experimental design*

In the dancing experiment example above, the assumption was that music in given social settings would make people dance, and that alcohol consumption in moderate but increasing quantities would make them want to dance more. As described, the experiment is lacking two elements to qualify as a valid experiment. It requires a control group and random assignment.

Returning to our assumption, the implication is that people require a social situation of some type to want to dance, a desire that could be augmented by alcohol, i.e. no social situation equals no dancing. In order to control for the possibility that people might still dance when not in a social situation, i.e. when alone, 5 more groups will have to be created. In one group, one person will be alone listening to music during a period of time matching that of the social situations. The other 4 groups will also consist of people alone listening to music though they will receive the different quantities of alcohol. These conditions will serve to control for the possibility that situations involving music, though not social settings, might also induce dancing. Our design now involves a control condition involving no social situation, 4 social settings, and 5 alcohol consumption quantities. Since 'no social setting' can be considered a type of social setting, i.e. a non-social setting, we group it with social settings meaning that we have a 5 × 5 design with 3 levels of the dependent variable.

The last element is subject randomization. All this means is that all subjects are randomly assigned to each group. This is to prevent a particular group of people, students coming from a dance class, for example, to all be inadvertently assigned to the same group, thus biasing the results.

There are two general variations for the way this experiment can be conducted. If different subjects are assigned to one and only one group, then it is called a between subjects design because it means that when the analysis is done you will be making comparisons between groups. A stronger though more time consuming design is a within subjects de-

sign, although this has the possible drawback of informing all the subjects of the complete experimental design, and thus would require counterbalancing of conditions. Here, every subject is assigned to every group once, though in different randomized orders, each in turn used the same number of times. What makes this a strong design is that by putting every subject in every situation we can rule out the possibility that differences between groups were the result of unintended and uncontrolled differences between the subjects themselves. For example, the subjects in the birthday party setting who had 4 drinks might have just happened to like to dance a lot, regardless of the social setting or the amount of alcohol consumed. Since we will have data for every subject in every condition we can be more assured that the amount of dancing observed is the direct result of the social situation and the alcohol consumed, should alcohol prove to be an important factor, and not an extraneous variable.

6.2.3 *Research hypotheses and experimental hypotheses*

Whereas a research hypothesis makes predictions about a phenomenon, an experimental hypothesis makes predictions about the type of information that can be used to support a research hypothesis, i.e. these are hypotheses involving the experiments themselves. The basis for this distinction is practical. Most of the time it is impossible to observe or measure directly whether your variables will interact as predicted by the research hypothesis. For example, if the research hypothesis took the form of 'if A then B' using capitals, then the experimental hypothesis would be 'if a then b' with small letters. The catch is to make reasonably sure that the relationship between 'a' and 'b' maps to the relationship between 'A' and 'B.' The discussion section of a paper will typically address the effectiveness of the proposed mapping in detail. A large part of learning to be a critical reader involves learning to assess whether a researcher's claims about the research hypothesis are substantiated by the claims about the experimental hypothesis.

The methodology of your experiment will ideally present you with observable and measurable data that can help to support or negate the research hypothesis. This book is intended to introduce you to the different types of methods used to investigate language from different perspectives. Regardless of the type of question you ask, it is likely that the methods you find here will help you decide on the experimental protocol best suited to your question.

6.2.4 *The experimental and the null hypotheses*

Earlier we mentioned falsifiability and prediction. These two ideas are directly concerned with the development of hypotheses. The two simplest and also most common that you will encounter are 1) that the chosen task will produce the predicted effect or 2) that it will not. In this case they would be the experimental hypothesis and the null hypothesis, respectively. Prediction is involved with the experimental hypothesis since your hypothesis 'predicts' that given a certain setting, the given variables will interact in a certain way. The null hypothesis deals with plausibility. In essence, you cannot have a true experimental hypothesis that does not assume the possibility of being incorrect (Popper 1959). It is this possibility, which we call falsifiability, that is captured in the null hypothesis, i.e. that the effect you predict from the interaction of the given variables does not exist.

In actually conducting your experiment, the goal will not be to prove the experimental hypothesis correct but instead to show that the null hypothesis can be rejected. In other words, to establish whether your results merit rejecting the null hypothesis. Recall that earlier we described the scientific method as an attempt to come up with the best possible explanation for a phenomenon. This is achieved by showing that other explanations cannot account for the data. The null hypothesis encapsulates the other explanation as the proposal that the relationship predicted by the experimental hypothesis is not meaningful. If the relationship is shown to be meaningful using inferences based on statistical probability, then the new hypothesis explaining the relationship becomes the best available explanation for the phenomenon in question, thus rejecting the null hypothesis. If however, the proposed relationship is not shown to be meaningful, the null hypothesis is not rejected, thus maintaining the original explanation for the target phenomenon as the best available.

6.2.5 *Following are two examples of how all of these pieces fit together*

> *Example 3a:* Lozano & Tversky (CS, 2004) were interested in whether gesture benefits communication. Their research hypothesis is loosely that gesturing while explaining something benefits the person gesturing by facilitating reasoning, problem solving, and other cognitive processes. As an experimental method, they needed a controlled situation where people could be observed gesturing while explaining something. They decided to use a video-taping environment in which one person assembles something, then using their experience, reassembles the same object while explaining how to do it. A television cart was the object chosen, likely because it is relatively easy to assemble and disassemble. Their experimental hypothesis given this paradigm is that subjects who speak and gesture will reassemble the TV cart more accurately and efficiently than those who only gesture and a control group who will do neither. The null hypothesis is that gesture will have no effect on how well reassembly is done. The researcher's independent variable is the type of communication during reassembly, and it has 3 levels, speech + gesture, gesture only, or control. The dependent variable is the number of errors made during reassembly. Since the researchers want to know how much gesture benefits communication, counting the number of errors made as a direct result of a communicative instance is a very good measure of communicative effectiveness.

> *Example 4a:* Wiemer-Hastings, Barnard & Faelner (CS, 2004) were interested in the possible structural differences in abstract versus concrete categories. Their research hypothesis was that abstract and concrete item categories are structurally different because abstract items are organized more by situational relations, whereas concrete items are organized more by taxonomic relations. In their own words (p. 1453) "The main hypothesis was that abstract and concrete item categories differ in the amount of constraint that they place on membership." They needed an experimental method that would allow the membership flexibility of abstract categories, if real, to become readily apparent without manipulating the rigidity supposedly associated with concrete categories. They chose a listing paradigm in which subjects would list exemplars for each of 24 categories, half of which were concrete and half abstract. Their experimental hypothesis was then "that significantly more subjects would list the same items for concrete categories than for abstract ones" (ibid.). The null hypothesis was that there would be no difference in the listing for the two types

of categories. Their independent variable was the use of either abstract or concrete items. Their dependent variables were the frequency of types produced and token/type ratio, i.e. a higher ratio is indicative of a category, which places strong constraints on cognitive processing.

There are two additional experiments in this paper. Take it as an exercise to break them down the way we have done with the others here.

6.2.6 *Reliability and validity*

Earlier we discussed the importance of learning to assess whether the research hypothesis and the experimental hypothesis mapped well to each other. There are two types of criteria instrumental in making this evaluation. They are reliability and validity.

Reliability refers to the consistency of the effect produced by your experimental method. If the test yields consistent results, it is considered reliable. Two common measures of the reliability of the experiment are test-retest and inter-observer. If an experiment is conducted on two or more separate occasions and the results are consistent, the experiment is considered to have met 'test-retest reliability'. In practice, most researchers will replicate, or conduct their experiments at least twice, to ensure that the results are trustworthy. A frequent factor interfering with test-retest reliability or even with reliability across subjects is inter-observer reliability. Reliability fails when different people performing different tasks for an experiment, i.e., collecting the data or coding it, do so differently across subjects or data sets. The goal of every investigator is to conduct every aspect of an experiment so that all tasks involved with it are done as consistently as possible so as to ensure that results are not made unreliable due to inconsistencies. This will be discussed further below in the section on data collection. Tests of reliability are assessed by the degree of correlation among measures (see Núñez for further discussion, this volume).

Validity primarily refers to how well your experiment measures what your research hypothesis says it measures. Validity can be assessed according to different criteria, and unfortunately, there are no statistical tests to measure it. Knowledge of the phenomenon in question is usually the only metric against which to assess the validity of the experiment. Common validity criteria are construct, ecological, internal, and external.

Construct validity is concerned with the operational definition of an abstract concept in an experiment. In our dancing example above, dancing was operationalized as full body rhythmic motion while on both feet, and not including foot tapping or swaying. Would you consider this a valid description of dancing? Can you think of a more exacting, less subjective metric?

A second kind of validity, ecological validity, involves how well an experimental design, and its results, fit with what people already know about their natural environment. It is sometimes difficult to see what a study on language as described in psychology journals really means for ordinary people in real life situations. Due to the necessity of a controlled environment, ecological validity is sometimes called into question. It is important that the investigator be cautious in understanding the trade-off between good experimental control, and doing work that is meaningful and informative outside of the laboratory.

The last two types are internal and external validity, and refer, unsurprisingly, to internal and external factors of an experiment. Internal validity concerns ensuring that

the outcome of an experiment is not due to factors other than those intended by the experimenter. These can include elements such an uncontrolled extraneous variable, or inconsistency in data coding. External validity deals specifically with how well an experiment can generalize to the population at large. A common criticism of most university run research is that the subjects are mostly university students. This is a rather specialized population in terms of education, and sometimes socioeconomic status, among other possible factors. Deciding that findings from this group are applicable to other groups can sometimes be questionable.

Actually choosing a methodology to conduct an experiment means satisfying all of the criteria we have described. Once you have a research hypothesis, you must design an experimental protocol to address it. Choosing from among the methods described in this volume, or perhaps from a more closely related paper, you must identify all of your variables and your experimental hypotheses, as well as consider validity and reliability issues. Only then can you proceed to actual data collection.

6.2.6.1 *Following are the reliability and validity assessments for examples 3a and 3b*

> *Example 3b:* Lozano & Tversky (CS, 2004)
>
> Reliability assessment: In this experiment reliability could be assessed by replication. That is, if another group of subjects came into the lab and they were randomly assigned to the three groups, they would show similar results.
>
> Validity assessment: The validity of this experiment rests on the assumption that reassembling a TV cart is a good measure of logical reasoning or problem solving.
>
> *Example 4b:* Wiemer-Hastings, Barnard & Faelner (CS, 2004)
>
> Reliability assessment: If another group of subjects were tested and showed similar results as the original group then it could be said that the measure is reliable. Do note that it will not matter if the abstract item categories themselves are variable. Reliability is dependent only on the consistent variability of the tokens for each category regardless of the group sampled.
>
> Validity assessment: Validity in this experiment is the suitability of self-reported lists as a measure of mental organization. Is mental organization what is really being measured by these frequency data?

6.2.7 *Data collection*

There are three general aspects to data collection, obtaining subjects, actually collecting data from them, and ensuring their informed consent and safety.

6.2.7.1 *Collecting data* Although it is time-consuming and, for lack of a better term, grunt work, collecting data is a crucial step in the scientific process. In order for things to run smoothly, care must be taken to ensure that consistency reins throughout data collection. A few things to keep in mind follow.

It is usually best if the subjects do not know what your hypotheses are so as to prevent their affecting the results. A cover story is usually given to subjects to account for the procedure without giving away critical information.

It is also best if the person in charge of executing the experimental procedure does not know the hypotheses of the experiment. Experimenter bias can be a problem. We communicate with each other using means other than language. It is possible for an experimenter who wants a certain result to unconsciously signal the subject in some way. In order to prevent this, it is usually best to have naïve research assistants collect the data. If this is not possible, then the experimenter can try to assign subjects to groups randomly, so that the experimenter herself does not know which group they are in. This is called a double-blind technique. In addition, a strict protocol must be followed where instructions are read in the same way to all subjects, and careful attention is paid to treat all subjects consistently. Deviation can result in extraneous influences on the results.

6.2.7.2 *Obtaining subjects* University students tend to be the most common source of data. Many schools have a system in place to facilitate contact between researchers who want data, and students who will give it to them in exchange for course credit or a small monetary reward. There are other options for collecting data, however. And given the arguably valid criticism that comes from trying to extend findings from university-educated 18 to 25 year olds to the general population, it may be desirable to do so. Subjects can be obtained through advertising in local newspapers, by contacting community organizations, etc. A factor that will contribute to how easily this is accomplished will be transportable research equipment, i.e. people tend to be more willing to participate if they do not have to go out of there way very much to do so, thus if you can bring your equipment to them, you will have better luck. Alternately, the use of financial rewards may be required. Some studies requiring one observation per subject simply go on buses and subways and ask people one question.

6.2.7.3 *Subject informed consent and safety* The early history of human experimentation is filled with many exciting discoveries. Less well known is the history of mistreatment of experimental subjects. Much has been done in recent years to prevent any future abuse. All universities in the United States where research is conducted have human subjects experimentation review panels whose duty it is to oversee the conduct of all experimenters on their respective campuses. These panels follow guidelines set by the federal government.

As a researcher, your first priority will always be to ensure the safety and well-being of all subjects as well as to safeguard their rights and privacy, regardless of where the research is conducted. All practicing researchers learn early in their careers to only design experiments that adhere to these constraints. We cannot stress enough that there is no negotiation possible on this point.[2] Guidelines have been developed by the American Psychological Association, known as the Code of Ethics (http://www.apa.org/ethics/code2002.html,

---

2. Basic research must adhere strictly to these constraints. There are slightly different guidelines for cognitive research in medicine. The reasoning is simple. Basic research involves knowledge for its own

1992.) to help you plan your research. Though the guidelines go into extensive detail, the general ideas can be summarized as follows:

> *Well-being:* The experiment should not compromise the participant's well-being. This refers to experiencing unreasonable physical or emotional discomfort.
>
> *Informed consent:* Subjects must receive enough information about the experiment to be able to make an educated choice to participate. This, of course, does not mean that you should give away your hypotheses. It only means that subjects must receive information describing the procedure in enough detail such that they know what will happen. Subjects then sign a document describing the extent of their participation, and must receive a copy for their records.
>
> *Confidentiality:* Each participant's confidentiality must be safeguarded. This means that the only documents that can contain any identifying information must be their signed consent form and any possible subject list for the day. Identifying information must never appear in relation to any of the data collected. Additional information regarding the use of subjects for gesture and Sign Language data can be found in Wilcox & Morford, this volume.

The usual procedure for a new research project is that a proposal for research is submitted to the review panel. The proposal will typically contain a brief summary of the experiments to be conducted focusing on the treatment of subjects. It will be authorized for execution based on the acceptability of its procedures for the treatment of human subjects. Regardless of how innocuous you may believe your experiments to be, it is imperative to follow the requisite guidelines or you risk losing the privilege of conducting research.

6.2.7.4 *Additional points Conducting research off campus:* The requirements to collect data from persons who are not students at your home institution should be similar to those for using student subjects. The only difference will be that your identified subject source will not be students. If you are uncertain, it is best to contact your university's officials.

Conducting experiments with children: In the U.S., all research conducted on persons younger than 18 years of age, or in a university setting that may include 17 year old freshman, must be authorized by their legal guardian. This follows regardless of where the research is conducted. In other words, if you are based in the United States but plan to collect children's data in Uruguay, you will need to obtain written consent from the legal guardian of every child the same as you would if you were in the U.S. In terms of differences in age constraints, follow the strictest ones. For example, if your home country requires legal guardian consent for persons under 18 but your target country considers persons over 16 to be adults, follow the constraints of your home institution as these are the strictest.

Conducting experiments at different universities: The rule of thumb in conducting research at different universities is to investigate what the ethical requirements are at every location, well before you begin research, and to follow their guidelines to the letter.

---

sake. As such, no sacrifice from subjects should ever be expected. Medical research usually involves an acknowledged risk/benefit exchange.

These guidelines must be followed in addition to those of your home institution. If the university where you will be conducting research does not have a similar review panel, follow the most careful guidelines you can find. Well-respected research universities are a good source.

Conducting research in foreign countries: When conducting research in a foreign country, if you will be associated with a university, contact their human subjects review board the same as you would at your home institution. Otherwise, contact government officials for information on how to proceed. In countries where there are no set guidelines, be sure to contact the local government officials to ensure that you have all the necessary authorizations. In addition, the general protocol is that you must satisfy the guidelines of your home university as well as those of your research location.

6.2.8 *Data analysis*

Data analysis involves two types of statistical analyses. The first are descriptive and the second inferential (see Núñez this volume, for extended discussion.)

Descriptive statistics describe what the data look like using means, medians, modes, standard deviations and variances. The mean is the average of a set of data. For example, if you have 6 data points, {2, 4, 5, 6, 6, 9} the average would be 32/6 = 5.33. The median is the value that falls at the center of the data points if they were to be ordered from smallest to largest. If the data array has an odd number of entries it is the middle entry; if it is even numbered it is the mean of the two middlemost entries, (5 + 6)/2 = 5.5. The mode is the largest cluster of similar values. In this case it would be 6, because there are two of them and only one of each other value.

Of these three, the most commonly used is the mean. The standard deviation and the variance depend on this number. They both measure the spread of your data. The standard deviation is (loosely!) the average value by which each data point deviates from the mean. It is calculated by subtracting the mean from each data point, producing that data point's deviation, then squaring it, then doing the same for all data points, then adding the sums of all of the squares of all of the deviations for all of the data points, dividing by the number of data points minus one, i.e. n-1, then taking the square root of that total. The chart below should make this much clearer. The formula is $s = \sqrt{\frac{\sum(x-m)^2}{n-1}}$

Data point (x)	Mean (m)	Deviation ($x - m$)	Squared deviation ($x - m$)^2
2	5.33	−3.33	11.0889
4	5.33	−1.33	1.7689
5	5.33	−0.33	0.1089
6	5.33	0.67	0.4489
6	5.33	0.67	0.4489
9	5.33	3.67	13.4689
	Sum of the squared deviations = $\sum(x - m)^2 =$		27.3334
	Divided by $n - 1(6 - 1) =$		5.46668
	Squared root =		2.34

Here it is 2.34. In general, the smaller the standard deviation, the more consistent your data set. Usually the standard deviation is transformed into a variance by squaring it. Given our standard deviation, the variance here is 5.48. The variance is preferred to the standard deviation as a measure of the variability in a data set because it is considered nicer from a mathematical perspective, however both are used in inferential statistics, discussed below. For our purposes, both measures are useful.

Once you have calculated the descriptive statistics for each of your groups, the next step is to begin comparing them. The key question while exploring your preliminary results is whether there are noticeable differences between the different sets. This is typically done by graphing them in different ways, i.e. using bar-charts, pie charts, box and whiskers plots etc. What you are looking for is visual evidence of differences between your groups, as well as patterns connecting the differences. Returning to the dancing example above, graphing the data might reveal that the experimental condition where the most dancing occurred was the birthday party where subjects each had 2 drinks. At least, in looking at a bar chart of the data, this bar seems to be that largest. Testing whether this finding is significant requires inferential statistics.

Inferential statistics are used to show whether the differences you see between the different groups are meaningful enough to reject the null hypothesis. Though there are different ways that this can be done, they all depend largely on probability theory, or the likelihood that your findings are the result of chance. The details of these procedures are complex enough that they have been given their own chapter. Núñez in the following chapter, will provide you with a detailed description of how to proceed.

6.2.9 *Interpretation*

The ideal situation is that your results turn out exactly the way you predicted. In this case, all that is required is a formalization of your findings in the form of a research article. Fortunately, human cognition is far too interesting and complex to make things that easy. In fact, the results of an experiment rarely turn out exactly as the design predicted. For example, what if a group of researchers wanted to know if there was a relationship between hair color and height. The researchers recruited 60 men and women, and assigned them to either the "tall" or "short" group. They find that there are significantly more brown-haired people in the short group than the tall group. Thus, it could be concluded that there is a relationship between brown hair and height. However, there is also a significant interaction of gender by hair color by height. What should be reported?

First, the researchers should report all significant interactions. However, if you have a 5-way design, then there are 20 interactions making this rule less than practical. Often people do not report the higher level interactions simply because they are uninterpretable. (However, there are cases where reviewers or readers will find significant effects interesting or even interpretable, despite the author's initial classification as irrelevant.) Though it is a bit of a gray area, the authors should report unusual, or unexpected, significant results within reason. It may then be necessary for them to go back and look for possible causes of the effect in their data, a flaw in how the data was recorded or a factor they had not originally considered. A careful check should be done of how people were assigned to groups to ensure that it was done without bias. In other cases, where a computer or other

recording apparatus was used, it should be checked for recording error as well. Once these checks are in place, the researchers must treat the effect as real, and offer an interpretation with their report. In this case, it may be that the tall men are mostly blond, whereas the number of blond women do not differ in either the tall or short group. It can be said that "the men carry the effect" in the resulting relationship between brown hair and height. And it is the reason why there is a significant interaction.

Such an explanation might seem obvious in this example, however, as we discussed earlier it is necessary at various stages of the scientific process to revisit the literature. At this point any explanation is post hoc, which means that the researchers did not predict this effect, but given its emergence, can make a logical argument for it. Post hoc explanations are not very convincing to scientists, who would be more convinced by a second experiment. There is a silver lining to unforeseen effects in that they will sometimes lead to further experimentation which can either be set up to have better controls, or modulate the effect of the newly considered variable.

7.   Then there is the rest of the planet…

Cognitive linguistics (CL), perhaps more so than any other discipline encompassed by cognitive science, has an implicit interest in cross-linguistic research. For CL, embodiment is a basic premise. Language is not a complete abstraction. Instead it is grounded in the experience of our bodies in all of the environments they inhabit, from the physical to the social, to everything in between. In other words, CL is based on the idea that language is an extension of our environments. If our environments differ, then likely so will our languages. To better understand how this process works, it is necessary to study and compare many languages in conjunction with the cognitive systems they are a part of.

The initial sections of this chapter present an experimental environment that more and more researchers recognize as an idealization. The possible effect of factors such as cultural or linguistic differences and abilities, have not usually been considered. Nonetheless, evidence keeps emerging supporting cognitive sensitivity to these influences.

A telling example from research on the cultural basis of cognition comes from Cohen and Gunz (2002). When the authors investigated differences in memory recollection between westerners and easterners, they found that when easterners recalled a memory of an event of which they had been the focus, they spoke about themselves in the third person, but when they had not been the focus, they spoke about themselves in the first person. This is exactly the opposite of what would be expected in the west where the opposite is the norm. According to the authors, these differences in perspective indicate deeper differences in the way that information is processed and in the way that memories are eventually encoded.

Other work explores the ways that specific languages mold cognition. Levinson's (1996) work comparing Tzeltal and Dutch showed that the system used to describe spatial relationships in a language had consequences for the way that space was understood in non-linguistic processes. Languages tend to use either a relative location (left/right, front/back) or absolute location system, akin in meaning to the English cardinal direction

system, i.e. north, south. Dutch uses a relational location system and Tzeltal an absolute one. Levinson developed a task that involved comparing arrows in different orientations set on two different tables. Subjects were asked to point to the arrow in the second table that matched the direction of the arrow in the first. Whereas Dutch speakers tended to choose the arrow that matched the original direction relative to themselves, Tzeltal speakers chose the one that matched the absolute direction of the first, a result which mapped to the spatial system used in the subjects' languages.

Much as we would like to follow with another item by item set of instructions for how to deal with cultural differences in experimentation, this is impractical given not only space considerations, but the sheer impossibility of preparing for every possible scenario. The best we can do is to help make you aware of the issues. They center on three foci, types of human groupings, linguistic variation and endemic bias in the interpretation of cognitive phenomena.

7.1 Types of human groupings

People tend to define themselves according to the groups they belong to as much as by those they do not. Common groups include 'culture,' 'race,' 'ethnicity,' 'nationality,' and 'socio-economic status.' If addressed directly, most people easily indicate those they belong to. The problem is that there are no a priori agreed upon constraints for who belongs to which group. Some people might choose to classify themselves as belonging to a certain group using completely different criteria than other people. In terms of race, for example, in some countries, having any African ancestry qualifies someone as belonging to the African race. In others, having any European ancestry designates them as white. The classification is thus entirely subjective and tied to a given set of societal norms. All human groupings are subject to such constraints. As a researcher, you will find yourself invariably trying to determine how to best select your subjects so as to minimize effects from possible external variables. This will be important whether you are interested in contrasting different groups or focusing on a phenomenon in as generic an environment as possible. There is no easy answer for how to control for grouping factors. The best you can do is to be aware of their nature so as to be better able to control for them, and to document them as completely as possible in your research report. The following descriptions of some of the main groupings are intended to introduce you to the issues involved in making useful classifications.

7.1.1 *Culture*

What is culture? A textbook definition is "the customs, behaviors, attitudes, and values, and the objects and implements that can be used to identify and characterize a population (Beins 2003: 309). Though this description might seem uncontroversial enough, following is a scenario that should help you better conceptualize the issues. Say you think of yourself as an American, for example, visiting a country different from your own, say Canada. Your 'culture' as an American is probably very similar to that of Canada. Or is it? Though there are countless similarities between the two cultures, there are differences. Canada has a

significant social safety net and much lower levels of violent crime, for example. Are these differences significant enough to, in turn, merit cognitive differences?

7.1.2 *Race*
From a scientific perspective, there is no such thing as different human races. We are one race. However, despite the ivory tower we may sometimes choose to find refuge in, we live in a deeply politicized world. Many governments assume racial distinctions based on differences in physical morphology when collecting census data. A difficult task complicated by the fact that many people are "mixed." Given the absolute lack of any biological evidence of anything but cosmetic differences between different people, this is not likely to be an attribute that will need to be dealt with directly in research. However, physical morphology is often correlated with other types of groupings. Perhaps the most common is ethnicity, described below.

7.1.3 *Ethnicity*
Ethnicity, as used in the United States, refers to cultural heritage (Shiraev & Levy 2001). Typically this includes religion, language, ancestral origin, traditions, diet, etc. Belonging to the same ethnicity does not entail having the same national origin. There are ethnic Hungarian communities in Slovakia, Austria and obviously in Hungary, among other countries, for example. By the same token, people that might be grouped as belonging to the same ethnicity by a given country's sociopolitical structure do not necessarily share more than language and possibly religion. Mexican and Puerto Rican people in the United States are a case in point. Though both original cultures were colonized by the Spanish, resulting in the adoption of the same language and religion, most similarities end there. Puerto Rican culture resulted from the mixing of surviving Taino, enslaved and imported Africans and colonizing Spanish. Mexican culture resulted mostly from the mixing of established though subjugated Native American societies and colonizing Spanish. The subsequent extant cultures are hybrids of the originals, and therefore, though superficially similar, are actually quite different.

7.1.4 *Nationality*
Nationality can be just as confusing as ethnicity. Though the dictionary definition describes a group of people who share history, language, geographical origin and exist under the protection of a recognized and formalized political entity, i.e. a country, people belonging to the same nationality can be quite diverse. Someone can have emigrated from Iran, and have been naturalized as a United States citizen. Another person might belong to a family that has lived in the United States for 8 generations, having originated from England. Yet a third might be a direct descendent of the people the Pilgrims encountered when they first arrived. Nationally, all of these people are considered Americans.

7.1.5 *Socio-economic status*
Socio-economic status usually involves a grouping based on family income, parental educational level, parental occupation, and social status in the community. Here again there are a number of inconsistencies that can arise due to intergenerational differences and em-

igration. A member of a family that is well off, could for many reasons, work in a fast food restaurant and barely make enough money to live. A person whose parents were subsistence farmers could receive a scholarship and become a well-educated high-paid engineer as an adult despite having been raised in poverty. An immigrant who was a teacher in his home country can find himself cleaning bathrooms upon arriving in his adopted country.

How do you decide which of these groupings to use as subject characteristics and how to use them effectively? Once again, there is no neat answer. It really does depend on the specifics of the phenomenon you are investigating. Further, perhaps you have noticed that all of the distinctions of types of groups are based on the constraints of the English-speaking world. There is no guarantee that the people you might be interested in studying would necessarily use them. At the risk of sounding like a dangerous stunt television show, a strong recommendation is to not attempt to do a study comparing these factors either as a first study or alone if you have no experience with studies in general. Science is a strongly collaborative endeavor. Enlist the help of someone with at least some experience. That said, the best you can do, once again, is to document as completely as possible the choices made, as well as to be as consistent as possible with the internal structure of your sample.

7.2 Linguistic ability

Bilingualism and multilingualism in general present another set of variables. If your direct goal is to investigate cognitive differences between groups with differing linguistic abilities, you should obviously develop criteria for inclusion and exclusion of subjects into your various study groups. Though the required guidance is beyond the scope of this volume, the "Handbook of Bilingualism" edited by Kroll and De Groot (2005) will prove a useful guide.

Chances are that your study will not focus on multilingualism and you may wonder how relevant the phenomenon might be to your research. Consider the following recent studies.

Marian and Spivey (2003a, b) conducted a series of eye-tracking studies on phonological interference. They tested Russian-English bilinguals to see if the phonological attributes of one language would interfere with those of the other. They found that when subjects were presented with an array of objects whose names overlapped phonetically with each other, albeit in different languages, i.e. "marker" in English versus 'marka' in Russian meaning stamp, and asked to point to an object, for example, the marker, that they also made eye-movements toward the stamp. According to the authors this indicated phonological interference from the second language.

In a study on short-term memory, Thorn and Gathercole (1999) showed that when bilingual, monolingual, and ESL students were tested on a word-recall task, the number of recalled words was a function of how well the subject knew the language the words belonged to, indicating that short-term memory is not a language-independent process.

Emerging evidence shows what appear to be differences in non-language related cognitive abilities in bilinguals. Bialystok (1992, 1994, 1999, 2005) has shown that bilinguals exhibit greater control of attention in the execution of cognitive tasks. In other words, that they are better at identifying and concentrating on the relevant elements of a task while ef-

fectively ignoring peripheral details. In terms of spatial abilities, Bialystok and Majumder (1998) showed that bilinguals were favored over monolinguals in spatial control tasks. Further, McLeay (2004) showed that bilinguals were both faster and more accurate in a mental rotation task involving knots.

These are all examples of ways that multilingualism produces cognitive differences. As a responsible researcher, the course of action to address the possibility that your results might be influenced by variations in linguistic ability is the same it would be with any other variable. Control for it. This means ensuring that your subjects all belong to the same linguistic group or else documenting any within-subjects variation. That way, if you do get results that are not quite what you expected, you can use statistical analyses to rule out those effects. This is precisely the same technique used for any other subject variable not the direct object of investigation. Sex, for example, is commonly documented for each subject. Researchers routinely check to see if there are any significant differences in the results attributable to sex. Linguistic variation should be treated in much the same way.

8. Endemic bias in the interpretation of cognitive phenomena

Earlier, in describing the scientific method, we talked about the inherently subjective nature of any human endeavor. Often, subjectivity takes the form of biased thinking of which people are largely unaware. Shiraev and Levy (2001: 55–70) provide an excellent description of common biases in conducting research between different groups. Following is a brief summary of the main points along with the antidotes for each bias proposed by the authors.

8.1 The evaluative bias of language

Words are meant to describe the entities in our environments. What we do not realize is that they also prescribe what they are. If I call a flat surface suspended on four legs a table, then it 'is' a table, if I call it a desk, it becomes a desk. In terms of the way that people are classified, consider these pairing, all of which are meant to refer to the same characteristic. *Old/mature; obsessed/committed; lunatic/visionary; dead/ontologically impaired.* These terms each carry value judgments which are impossible to escape, i.e. one term is considered more positive and the other more negative. Try to come up with a non-value laden description of *narcissistic/high self esteem*! Scientific language is not immune to biasing, whether it be in describing a target phenomenon or writing a consent form.

Antidotes:
1. Remember that descriptions, especially concerning personality characteristics can never be entirely objective.
2. Become aware of your own personal values and biases, and how they influence the language you use.
3. Avoid presenting your value judgments as objective reflections of truth.
4. Recognize how other people's use of language reveals their own values and biases.

8.2 Differentiating dichotomous variables from continuous ones

There are phenomena in our environment that occur as mutually exclusive or contradictory pairings. A woman can be pregnant or not, but she cannot be a little pregnant. Many other phenomena are treated as if they were also dichotomous, even though they are actually continuous. Examples are cooperative/competitive; introverted/extroverted; good/bad. This can be a problem, for example, while designing an experiment in that a variable can be inadvertently treated as if it were dichotomous, when in fact it is continuous.

Antidotes:
1. Learn to differentiate between variables that are dichotomous and those that are continuous.
2. Remember that most person-related phenomena – such as traits, attitudes, and beliefs – lie along a continuum.
3. When making cross-cultural comparisons, try to avoid artificial or false dichotomies.

8.3 The Barnum effect

The Barnum effect refers to statements that are generally true about most people. It references Barnum in that it has 'a little something for everyone.' Barnum statements treat the general as if limited to the specific. One could say, 'women are sensitive to sexism' which seems true though upon deeper thought is also likely to be true of men.

Antidotes:
1. Learn to differentiate Barnum statements from person and group specific descriptions and interpretations.
2. Be aware of the limited utility inherent in Barnum statements. Specifically, remember that although Barnum statements have validity about people in general, they fail to reveal anything distinctive about any given individual socio-cultural group.
3. Whenever feasible and appropriate, make it a point to reduce the Barnum effect by qualifying personality descriptions and interpretations in terms of their magnitude or degree.

8.4 The assimilation bias

We categorize everything in our environments as a matter of course. Typically, not much thought is involved when we use schemas or cognitive structures to organize our beliefs. Though they are useful to a point, i.e. you are a woman alone walking on a street when you see three men walking in your direction and so you immediately assume they might be dangerous and cross the street, they can be problematic when they function as unquestioned stereotypes, the three men were actually your colleagues from work whom you have now insulted by so obviously avoiding them. In short, what schemas do is help you make assumptions about a person or situation based on limited information. They are a problem when we do not see beyond them.

Antidotes:

1. In situations in which you are likely to utilize the representative heuristic, make a conscious effort to consider the possibility that the schema or prototype in question might be inaccurate, biased, or incomplete.
2. Take into account relevant statistical information, such as base rates, sample sizes, and chance probability.
3. Beware of the natural tendency to overestimate the degree of similarity between phenomena and categories.
4. Recognize that your personal attitudes about people and group prototypes can bias your comparison and subsequent judgments.

8.5 Fundamental attribution error

This error refers to the tendency to assume that 'cause' of someone's behavior is their character and not the situation. In essence, behavior is attributed to internal influences versus external ones. For example, you go to a crowded restaurant and are served by a waiter that responds brusquely and is late with your order. You may assume that the person is a bad waiter, and dismiss him at that. What you may not know is that the kitchen is in disarray, and that the cook just yelled at the waiter for no reason other than that he was the first person to walk through the door. If the waiter is actually usually quite courteous and timely, in categorizing him as a bad waiter you have just committed fundamental attribution error by not considering that the situation might also have a role to play in his behavior.

Antidotes:

1. Do not underestimate the power of external, situational determinants of behavior.
2. Remember that at any given time, how people behave depends both on what they bring to the situation as well as on the situation.
3. Keep in mind that this attributional error can become reversed, depending on the perceiver's point of view. Specifically, although people are prone to underestimate the impact of others' situations, they tend to overestimate the impact on their own situations.
4. Be sure to take into account both cognitive and motivational biases that are responsible for producing these attributional errors.

8.6 Correlation does not prove causation

This is actually one of the most common errors made in considering the relationship between two phenomena. We assume that because two events happen at approximately the same time, that somehow, one 'caused' the other. For example, there is some evidence that watching violent television programs is correlated with violent behavior. Some people have interpreted this to mean that watching violent television programs makes one violent. The possibility that violent people might like watching violent programs is not considered, thus reinterpreting the correlation not as causative but as symptomatic of otherwise violent behavior.

Antidotes:

1. Remember that correlation or coappearance is not, in itself, proof of causation.
2. Keep in mind that correlations enable us to make predictions from one event to another; they do not , however, provide explanations as to why the events are related.
3. When a correlation is observed, consider all possible pathways and directions of causation. For example, if Event A and Event B are correlated, does A cause B? Does B cause A? Do A and B cause each other? Does C cause A and B?

9. Conclusion

The goal of this chapter was to get you on your feet, so to speak, about experimentation. It would be naïve to imagine that all you need to get started can be found here. As social creatures we tend to learn best if we have a model of how to proceed. If at all possible, in lieu of going it alone, try to find someone who would be willing to serve as your mentor during this process. At the very least, seek out someone who will allow you to observe as they develop a project through to the point of writing up the research paper. The experience will provide you with a valuable concrete model of how to conduct your own work.

There are also excellent methods and statistics books available to help you further develop your knowledge of the material introduced here. We strongly recommend those referenced here as well as Geoffrey Keppel's "Design and Analysis: A Researcher's Handbook" in whatever edition you can find it in, as well as Julian Meltzoff's (1998) "Critical Thinking about Research: Psychology and Related Fields."

References

American Psychological Association (1992). Code of ethics (http://www.apa.org/ethics/code2002.html)

Aronson, E., Wilson, T. W., & Marilyn B. Brewer (1998). "Experimentation in social psychology." In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology, Volume 1, 4th/ed*. Oxford: Oxford University Press.

Ayala, F. J. (1994). On the scientific method, its practice and pitfalls. *History and Philosophy of the Life Sciences 16*, 205–240.

Beins, B. (2003). *Research methods: Tools for life.* Boston: Allyn & Bacon.

Bergen, B., Narayan, S., & Feldman, J. (2003) Embodied verbal semantics: Evidence from an image-verb matching task. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*.

Bialystok, E. (1992). Attentional control in children's metalinguistic performance and measures of field independence. *Developmental Psychology, 28*(4), 654–664.

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition, 16*(2), 157–168.

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development, 70*(3), 636–644.

Bialystok, E. (2005). The impact of bilingualism on language and literacy development. In Kroll & De Groot (Eds.), *Handbook of Bilingualism.* New York: Oxford University Press.

Bialystok, E. & Majumder, S. (1998). The relationship between bilingualism and the development of cognitive processes in problem solving. *Applied Psycholinguistics, 19*(1), 69–85.

Bloom, P., Peterson, M. A., Nadel, L., & Garrett, M. (1996). *Language and Space.* Cambridge: MIT Press.

Carnap, R. (1966). *The Philosophy of Science.* New York: Basic Books.

Casasanto, D. & Boroditsky, L. (2003). Do we think about time in terms of space? *Proceedings of the 25th Annual Meeting of the Cognitive Science Society.*

Chalmers, A. (1976). *What is this Thing Called Science?* Queensland: Queensland University Press and Open University Press.

Cohen, D. & Gunz, A. (2002). As seen by the other… Perspectives on the self in the memories and emotional perceptions of Easterners and Westerners. *Psychological Science, 13,* 55–59.

Ember, C. R. & Ember, M. (2001). *Cross-cultural research methods.* New York: Altamira Press.

Jordan & Zanna (1999). How to read a journal article in social psychology. In R. F. Baumeister (Ed.), *The Self in Social Psychology* (pp. 461–470). Philadelphia: Psychology Press.

Keppel, G. (1994). *Design and Analysis: A Researcher's Handbook.* Englewood Cliffs, NJ: Prentice Hall.

Kroll & De Groot (2005). *Handbook of Bilingualism.* New York: Oxford University Press.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programs. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the Growth of Knowledge.* Cambridge, MA: Cambridge University Press.

Levinson, S. (1996). Language and Space. *Annual Review of Anthropology*, *25,* 353–382.

Lozano, S. C. & Tversky, B. (2004). Communicative Gestures Benefit Communicators. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society.*

Marian, V. & Spivey, M. (2003a). Comparing bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics, 24*(2), 173–193.

Marian, V. & Spivey, M. (2003b). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition, 6,* 1–19.

McLeay, H. (2004) The relationship between bilingualism and spatial tasks. *Bilingual Education and Bilingualism, 6*(6), 423–438.

Julian Meltzoff's (1998) *Critical Thinking about Research: Psychology and Related Fields.* Washington: American Psychological Association.

Newcombe, N. S. & Huttenlocher, J. (2000) *Making Space: The Development of Spatial Representation and Reasoning.* Cambridge, MA: MIT Press.

Popper, K. R. (1959). *The logic of scientific discovery.* London: Hutchinson.

Rosnow & Rosenthal (1993). *Beginning behavioral research: A conceptual primer.* New York: Macmillan Publishing Company.

Shiraev, E. & Levy, D. (2001). Introduction to Cross-Cultural Psychology. Needham Heights: Allyn and Bacon.

Solso, R. L. & Johnson, H. H. (1984) An Introduction to Experimental Design. In *Psychology: A Case Approach* (3rd ed.). New York: Harper & Row.

Sternberg, R. (2003). *Cognitive psychology* (3rd ed.). Belmond, CA: Wadsworth.

Tattersall, I. (2002). The Monkey in the mirror: Essays on the science of what makes us human. New York: Harcourt.

Thorn, A. S. C. & Gathercole, S. E. (1999). Language-specific Knowledge and Short-term Memory in Bilingual and Non-bilingual Children. *The Quarterly Journal of Experimental Psychology A, 52*(2), 303–324.

Wiemar-Hastings, K., Barnard, K. K., & Faelnar, J. (2004). Structural differences in abstract and concrete item categories. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society.*