The basic logic behind this talk is:

PREMISES:

1. The purpose of academic research is to produce generalizations.
2. Meta-analysis is the process of generalizing across different studies/results by establishing a base result (i.e., average) and systematic differences.

CONCLUSION:
The purpose of academic research is to prepare for (and occasionally perform) meta-analysis.

While generally thought of as a technique for summarizing quantitative results, the basic thought process of meta-analysis applies to qualitative work as well. Any single study, no matter how well executed, has an infinite number of covariates that could potentially explain the results. By contrast, generalizations can emerge only from a collection of studies. Taking this point of view, a number of observations seem to follow:

1. THEORY, TYRANNY, AND EMPIRICAL GENERALIZATION

In the “it would be nice to have” category for advancing knowledge, two things seem particularly desirable. First is the identification of a repeatable phenomenon, A.K.A. an empirical regularity. That is, part of knowledge development involves establishing patterns that are likely to recur in the future. The limitation to much case and qualitative work is that it focuses on the particular situation in detail (which is good) without much concern for what other situations are similar or where the same patterns might recur (which is bad if your goal is general knowledge).

The second major desideratum is an explanation for what happened (note how anyone with a dictionary can use big words). While understanding why something occurs (the causal mechanism) is desirable, a simple descriptive story often provides value. Notice that rather than the value laden term theory (as in, don’t submit a paper without it), I use the term “story.” For all the homage we pay to the concept of theory, theory is basically a story that describes how, and where possible why, a phenomenon works. The current operational definition of a theory seems to be a story someone else managed to get published.

The appropriate goal of academic research is to develop empirically supported theory (stories). I prefer theory that is as specific and as quantitative as possible (i.e., a formula). Debates about which comes first, data or theory, are basically silly, having ended with Adam and Eve or maybe Socrates and Plato or chickens and eggs. Similarly debates about the inherent superiority of theory or data are akin to arguing whether the skin or the mind is more important; without one, the other cannot function. We should get on with the task of improving both rather than belabor the inadequacies of either. Most important, it is foolish to require theory before a result can be examined. Most humans construct theory to explain the world and rejecting data with no strong prior theory holds back progress (e.g., without Brahe, Kepler would have produced no laws). Theory is the appropriate end goal but not the only means.
2. OVERJUSTIFICATION (A.K.A. BIBLIOGRAPHIC OVERKILL)

Have you noticed the gradual increase in the length of literature reviews and bibliographies? Now it is important to place research in a context and for review articles the literature review is obviously crucial. However, it is not important, useful, or an efficient use of journal pages for every paper to completely review the field, especially when the review essentially lists past work with no original insights. Given access to computerized literature searches, there is even less need for extensive literature reviews now than there once was.

It has been rumored that some pad bibliographies and literature reviews to (a) appeal to the egos of cited authors who might be reviewers or (b) subtly suggest who the reviewers should be. This is not a good way to advance knowledge and may not even be effective at increasing the acceptance probability for a paper. Long bibliographies make missed cites that much more painful and increase the chance of mis-interpreting someone else’s work, which quickly alienates many reviewers.

Ask yourself questions like, “Does it make sense for a 30-page empirical paper to have 15 pages of literature review?” or “Do you really read these (unless you plan to use them to help you write your own literature review)?” Or, in a slightly different vein, “Why do job talks spend so much time on the literature and so little on what the dissertation will do?” Since at some level what is done must stand on its own, shouldn’t what was done, and not a preamble, be the focal point (both in emphasis and length) of a paper?

3. DERIVATIVE INCREMENTALISM

Most papers are remarkably unremarkable; that is, they fit neatly into established paradigms. In some ways this is related to the over-emphasis on literature reviews. Having spent considerable effort mastering (or at least citing) the literature, it becomes more difficult to think creatively. As Pope suggests, “Behold the bookful blockhead, ignorantly read, with loads of learned lumber in his head.”

Near-replications increase certainty about results, and form inputs to meta-analysis, which makes meta-analysts like me happy. Further, many/most researchers are better suited to, and make a real contribution to knowledge by, engaging in this type of work.

Still, two thoughts emerge. First, why do we insist on presenting these incremental papers as though they were earth shattering? You don’t have to be wildly creative nor apologize for not being so to contribute to knowledge.

Second, why not encourage more “discontinuous innovation” in our work? We owe a lot to those less timid souls who take a chance and view the world differently. “Far better it is to dare mighty things, to win glorious triumphs, even though checked by failure, than to take rank with those poor spirits who neither enjoy much nor suffer much, because they live in grey twilight that knows not victory or defeat.” (T. Roosevelt)

I’m not opposed to student samples; they are useful for many purposes. What isn’t very appealing, however, is observing behavior in situations that bear no relevance to the world consumers face. As part of your investigation of a theory, a study on real subjects or using existing data is helpful. Sure it’s messy but since the goal is generalization, triangulation by different methods is a plus, not a weakness.

Now this lack of realism is not confined to experimental researchers. Economic models are famous for this. Markets of 1 or 2 producers, 1 or 2 consumers, perfect information, etc. tend to exist only in our journals, usually accompanied by such comforting phrases as “without loss of generality, we assume ...” and “it is easily shown that ....” Ever wonder why, if it is so easy to show, that they don’t bother to show it? There is a procedure for showing a result holds for any number of competitors called proof by induction where you show (a) it is true for a monopoly and (b) given it is true for K firms, then it is true for K+1 firms. How often have you seen it used? Or why are fixed and variable costs considered important in operations research but assumed to be zero in so many models? The answer is mathematical tractability (a.k.a., convenience).

There was a time a generation ago when closed form solutions were needed. Given current computer power, however, it is quite feasible to numerically analyze complex functions that incorporate more realistic assumptions and then to describe the solutions with a simple formula or graph. That this approach has not been more widely adopted is a nice example of resistance to innovation.

5. HYPOTHESIS

H0: Construct A, also known as __________, will, when combined with B, produce result C under condition D. However, if B is combined with E or condition D’ occurs, then result F will appear unless it is Tuesday.

Our field suffers from a suffocation of hypotheses. To a reader of our work, it would appear that nothing is ever uncovered that is not hypothesized. Brian Sternthal and Alice Tybout have made the case for why it doesn’t matter when a hypothesis arises as long as you rule out alternative explanations. I agree. However, I want to go further and ask why we need so many formal hypotheses.

Hypotheses are implicit in what we measure and analyze. It is certainly instructive to briefly communicate why you think certain constructs are worthy of study. Still the fact you measured them eloquently communicates that you (or some outside party like a reviewer) thought they might have an impact. Why is this not adequate? Further, since often the expected relations can be communicated by a flowchart or series of equations, why repeat each link as a hypothesis?

Many of our hypotheses are basically checks on whether subjects are paying attention. For example:

H1: A positive reaction to __________ will increase attitude toward __________.

H2: As price increases, sales will decrease.

While there are situations where H1 and H2 may not obtain (e.g., when attitudes are strongly held or price is a major signal of quality), in most situations these results “better” obtain. Why give them the same stature as interesting hypotheses? Perhaps a category of manipulation check hypotheses should be created, especially for mathematical model based simulations where the assumptions clearly drive the conclusions.
Finally, avoid the tendency to create hypotheses where the null effect is zero. This is disingenuous at best and essentially dishonest, can only be true due to limited statistical power (no effect is ever exactly zero), and is contrary to the goal of cumulative learning.

6. POLYSYLLABIC Slobbering: An Ivory Tower of Babel

A number of research approaches are discussed and debated, including positivist, post-positivism, and post modern (which sounds like an oxymoron to me; is it really the future and if so, how do we know what it is?). All provide different views (or frames) on a phenomenon and are inherently both useful and incomplete. Yet the debate that has dragged on for several years seems intent on establishing the general superiority of a method by criticizing others. The debate reminds me of stories about people living in glass houses. Since at best we can establish method superiority for a given purpose, the debate seems off-base, a bit self-serving, and quite tiresome. The phrases “give it a rest” or “give it up” seem appropriate here.

As a discipline matures, it is natural for various sub-areas to form. Specialized language plays an important role in communicating among members of sub-areas. Unfortunately, that same language makes conversations across sub-areas difficult, especially since much of it either is not used in ordinary conversation (e.g., hermeneutic) or is used in ways at variance with its normal English (or other language) meaning (e.g., counter factual reasoning). Reading our work you might never guess hierarchical regression is the same as nested model testing. These language barriers work against generalizing knowledge. While the first ACR I attended in 1970 had behavioral researchers and quantitative modelers happily attending the same session and talking with each other, recent conferences often resemble a collection of mini-conferences except for cocktail parties and less-than-welcome lunch speeches like this one.

Of course barriers can be surmounted if people want to surmount them. Sadly, however, many people do not. The comfort of being in a sub-group is reinforced by the knowing smiles of approval when the particular paradigms/passwords (e.g., collinearity, demand effects, Stackelberg competition) are uttered. Why go to another track’s session when all your friends, role models, mentors, etc. are in your own area’s session? It will be uncomfortable and frustrating; while you know your group’s passwords, you probably won’t know theirs. Further you will probably marvel at either how trivial or how sloppy their work seems (without, of course, casting the same level of scrutiny on your own). Thus if you venture outside your area, Skinner’s operant conditioning and Bentham’s pursuit of pleasure will drive you home.

Why bother venturing outside the comfortable world of your own? There is no incentive if you are satisfied with being one of many fish in a small pond and incremental change. On the other hand, if you want to make major contributions/innovations, you have to. And even if you don’t, for the sake of your area consider discussing your work in such a way that outsiders can at least comprehend what you are doing. Try the following test: could you explain what you are doing to neighbors who are plumbers, your mother, or to a high school class?

7. CONTINGENCIES VS. CROSSOVER INTERACTIONS

Finding the boundaries in which a theory or result holds is an important goal of research. For example, where does the average price elasticity of Tellis’ meta-analysis (-1.76) provide a good estimate and where does it not? Trying a theory out in disparate areas of application both makes sense and is more likely to lead to different results (i.e., what is true for paper towels is more likely to be true for paper napkins than for oil well drilling equipment).

Now of course there is a certain excitement in finding a cross-over interaction or a category where the price elasticity is positive. However, to enhance general knowledge, finding an unexplored domain where the theory (or parameter) applies is every bit as useful as finding one where the result is reversed. Adding information about the new domain is far more important for knowledge development than whether the result matches or contradicts past patterns. Perhaps we should worry more about the domain extension rather than finding contradictions per se.

8. STATISTICAL STERILITY

The use of classical statistics, as typically practiced, has provided welcome rigor to our thinking. On the other hand, the blind use of cookbook statistics has produced a rigor mortis in our thought process.

Basically we approach problems with some notion about a phenomenon. That notion (which some call theory) is drawn from past experience, analogies, and formal learning. We then, subject to all the biases and imperfections in human judgment, alter the notion or our behavior to the extent that current information requires us to do so. This learning is essentially Bayesian, requiring gradual updates, or in a stickier form, related to control charts where at some level of discrepancy between current evidence and theory we dramatically alter our theory.

Now contrast this process of adaptive learning with the cookbook use of classical statistics. In standard statistics (i.e., those t, F, and X² values that appear in computer programs), we test the “null hypothesis” of nothing. That is, we examine the straw man of no effect or no relation.

First, statistical significance is pretty arbitrary. Why is a p-value of .09 so much different than one of .11? (Hint: they are not generally statistically distinguishable.) Answer: because we arbitrarily set .10 or .05 or .01 as a cut-off. As a consequence, we struggle to get “significant results” by hook (e.g., increasing the sample size or hyping the manipulation/signal or the attention payed to it) or by crook (e.g., only discussing hypotheses that lead to significant results.)

Second, consider the goal of empirical generalization. The primary substantive result is the impact of a variable (i.e., the mean difference due to a treatment or the size of a regression coefficient). If we want to weight results based on their reliability, then standard errors are needed. All the p-values, F ratios, and log-likelihood ratios provide is an indirect measure of standard error. Yet many articles report p-values, etc. and never directly present the size of the effect. Besides being frustrating to meta-analysts, this works against knowledge accumulation. (As an aside, the term “effect size” is unfortunate since it often refers to the explanatory power of a variable which depends on both the effect consistency and the average size of the effect.)

In summary, the real issue is not if or whether or when something has an effect (everything does, even if a very small
one), but rather how much effect it has in different circumstances. Focus statistical reasoning on this issue, recognizing past results are important information for current estimation. The best estimate for the effect in your study is some combination of what your data shows and what has been found in the past. If you must have a null hypothesis, then it should be based on past research and you should, for the sake of establishing generalization, hope to fail to reject it.

9. SOPHISTICATED ≠ COMPLICATED

We often mask simple results in our complicated calculations and confuse complication with sophistication. As an example, reconsider the figure used to highlight a 3-way interaction (Observation 7). While an ANOVA might find a significant interaction, examining the means could tell a different story. Seven of the eight conditions produce a mean of about 8 and are not (for reasonable sample sizes) statistically different. Only in the low, low, A cell is the mean different. Doesn't it make more sense to report this even if it isn't standard output or very elegant?

As another example, consider LISREL. Now it is an elegant and useful tool in the hands of a skilled user but likely to cause harm in the hands of a novice. For example, if we want to generalize, we may prefer to measure a construct the same way across studies. Yet as a one-step procedure that creates measures of constructs simultaneously with estimating structural relations among the constructs, the LISREL measure weights, and hence the operationalization of the construct, will differ across studies even if the same measures are used. Why not just use indexes that are simple averages and at the same time avoid capitalizing on chance variation?

Finally, consider the goals of (a) communicating to a broad audience and (b) providing input to meta-analyses. In general, simpler is better.

10. MEASURE VS. MANIPULATION

A fascinating tradition involves the choice to use an experimental manipulation of a construct to represent the construct instead of the measure of the construct. When a measure is taken (often for use in a manipulation check), why not use the measure of the construct directly in the analysis?

Other than tradition, the implicit reason has to do with error. Basically the true value of the construct depends on (a) the manipulation, (b) other influences on the construct (e.g., personal characteristics), and (c) a random component (e1). Similarly the measured value depends on (a) the true value and (b) a random component including measurement error (e2). Using the manipulated (generally binary) level assumes that the impact of other variables and the random component e1 is smaller than e2. For constructs such as involvement and mood and relatively weak manipulations such as “you may get your chosen snack” or a happy vs. sad story, one suspects personal factors play a large role. Assuming that the measured value has greater error associated with it seems a bit of a stretch, especially when multiple measures are used. Even if you can’t draw the typical 2 x 2 plot as easily, consider using the measured value.

11. SOCIAL APPROVAL/TENURE

Most of us like peer approval and if we want tenure, actually need it. This leads to the dreaded research strategy whereby you (a) develop a strategy to be an expert in some often-obscure area and (b) network yourself. Now focusing some of your efforts makes sense and being nice to people, especially smart ones, is both common sense and a way to learn something. Carrying these to extreme, however, is both unproductive (leading to derivative, non-interesting work) and unappealing. Can you think of many businesses that make strategies and don’t alter them for seven years?

Two questions arise. First, does game playing really contribute to general knowledge? Second, having been positively rewarded for one type of behavior for seven years, are you likely to change to different behavior even if you know intellectually it is superior? And if you change will it involve burn-out rather than increased commitment to scholarship?

I was fortunate to get in this business when tenure was easier to obtain but I still think I would rather be a roofer or run a chain saw than endure a seven-year mental and social makeover.

A PARSIMONIOUS EXPLANATION

I have observed some empirical regularities, but not offered an explanation. My story (theory) of the reason is complex but with a simple focal element: insecurity. Basically we tend to consider ourselves inferior to natural scientists or real psychologists or economists. We seize on their trappings without questioning and hide from practitioners in stilted prose.

SUGGESTIONS

What suggestions do I have? Here are my top ten:

1. Consider yourself first a student of consumer behavior and only secondarily an information processor or scanner data modeler. Work on disseminating as well as creating knowledge in your own sub-area.

2. Stop feeling inferior. Business schools exist because various “classic” disciplines such as economics, sociology, and psychology failed to see the opportunity that business in general and consumer behavior in particular provided for studying something that is both impactful and provides an important lens into the behavior of people in general. We aren’t going to win many Nobel prizes but neither are most doctors or economists. The input quality in marketing PhD programs has been high for 20 years which means we are up to competing on raw IQ. points.

3. Give back to the other disciplines. There is no need to continually defer to other disciplines. Neither is it helpful to stand by and knowingly criticize them (i.e., “look at the ridiculous assumptions economists make about consumers”). Why let economists eventually incorporate our results? Wouldn’t it be more interesting to incorporate them ourselves, in essence exporting finished product rather than just raw material in the form of assumptions?

4. Recognize we’re not curing cancer. I firmly believe market economies benefit consumers and by understanding them we can increase the benefits. However, we’re not providing dramatic life-saving new drugs or procedures. We chose to work in this area, many in business schools, because the combination of intellectual stimulation and monetary reward appealed to us. It does no good to feel bad about the choice; either change the choice or accept the trade-offs.

5. Recognize even our best work is inevitably wrong. In spite of constant reminders from reviewers, we often behave as though our work should somehow be correct. Philosophical
discussions about what is truth aside, in an active field subsequent work will alter, modify, or even invalidate the best that has gone before. Notice how few citations go back more than 15 years. Basically, 100 years ago they didn’t know you were coming and 100 years from now they won’t know you were here. This doesn’t mean you should take your work less seriously; since you chose to work on it, it must be the most important thing in the world to you at the time and should be treated accordingly. It does suggest you might be less sensitive to criticism and less strident in your criticism of others. “Whoever thinks a faultless piece to see, thinks what ne’er was, nor is, nor e’er shall be.” (A. Pope)

6. **Make the literature review an appendix.** I found during my time at MSI that literature reviews are one of the reasons practitioners don’t like to read academic papers. They are also one of the reasons I and many of you don’t like reading them, especially when they are bloated with references to please reviewers.

   One view of a paper is a conversation with a reader. Thought of this way a paper would have four sections:

   1. **An introduction that basically states what you are interested in.**
   2. **What you did, basically an abstract of the method section in plain English.**
   3. **What you found, restricted to means, cross-tabs, correlations, and OLS regression or ANOVA generated mean differences plus graphs.**
   4. **What next, describing future research.**

   Appendices would then be presented for:

   1. The traditional literature review.
   2. Details on the method.
   3. More extensive analyses.
   4. Detailed limitations and directions for future research.

   Notice this would make papers shorter and more accessible to members of other sub-areas of consumer behavior as well as practitioners and government employees, potentially leading to a more important role in policy making. (Notice how when questions are asked in the public arena about consumers, lawyers and economists are prominent, while marketers and consumer behavior researchers are rarely involved.)

7. **Structure your work so it aids future generalization studies.** Meta-analysis suffers from a preponderance of studies of a single type and a large fraction of empty cells. To contribute to knowledge development, fill the empty cells (which means method and paradigm pluralism is a virtue, not something to be wiped out by the stronger group). Put differently, exact replication isn’t of much value but Ehrenberg’s notion of differentiated replication/extension is. And whatever you do, report results so that they can be incorporated in a subsequent meta-analysis.

8. **Don’t test the null hypothesis of zero effect and report the size of the effect as the primary result.** You don’t believe there is no effect, though in some cases you may hope some are small enough to be ignored. The magnitude of effects are useful for accumulating knowledge; p-values are for juries deciding guilt and innocence.

9. **Recognize your contribution to the field will be hard to track.** Maximizing citations or awards may be mildly satisfying but chances are your biggest contributions may go unlauded. While awards may reflect short run impact, long run impact typically occurs through people and the subtle impact we have on them.

10. **Do what you want.** The advice you just received was at best free and at worst gratuitous. Research should be fun. Remember “Fanaticism consists in redoubling your efforts when you have forgotten your aim.” (G. Santayana)

I see my time is up. Thanks to those of you who stayed for your patience. I wish you and ACR well.
The following figures were prepared by Dr. Lehmann and were presented as slides during his Presidential Address. They are being published in the Proceedings as a supplement to his Address.
KNOWLEDGE GENERALIZATION AND THE CONVENTIONS OF CONSUMER RESEARCH: A STUDY IN INCONSISTENCY
UNDERLYING LOGIC

PREMISES:

1. The purpose of academic research is to produce generalizations.

2. Meta-Analysis is the process of generalizing across different studies/results.

CONCLUSION:

The purpose of academic research is to prepare for (and occasionally perform) meta-analysis.
GOALS OF META-ANALYSIS

1. Establish the base level (e.g., overall average)

2. Assess systematic differences

3. Establish the range of the results

4. Use past results to make predictions for other situations.
OBSERVATION 1: THEORY, TYRANNY, AND EMPIRICAL GENERALIZATION
OBSERVATION 2: OVERJUSTIFICATION

(A.K.A. BIBLIOGRAPHIC OVERKILL)

LIBRARY OF CONGRESS

NEW YORK PUBLIC LIBRARY

SOCIAL SCIENCE CITATION INDEX

ANY PAPER WRITTEN BY ANY POSSIBLE REVIEWER

"METHINKS THE AUTHOR DOTH PROTEST TOO MUCH"
OBSERVATION 3:
DERIVATIVE INCREMENTALIZATION

CURRENT THEORY/METHOD

REVISED THEORY/METHOD

THESIS

ANTITHESIS

CONSUMER BEHAVIOR VS. HEGEL

"The bookful blockhead, ignorantly read, with loads of learned lumber in his head." (A. Pope)

"Far better it is to dare mighty things, to win glorious triumphs, even though checkered by failure, than to take rank with those poor spirits who neither enjoy much nor suffer much, because they live in grey twilight that knows not victory or defeat."

(T. Roosevelt)
OBSERVATION 4: UNREALISM

(OR LIFE IN NEVER-NEVER LAND)
OBSERVATION 5: HYPOTHESES

H₀: Construct A, also known as ________, will, when combined with B, produce result C under condition D. However, if B is combined with E or condition D’ occurs, then result F will appear unless it is Tuesday.
OBSERVATION 6:

POLYSYLLABIC SLOBBERING –
AN IVORY TOWER OF BABEL
OBSERVATION 7: CONTINGENCIES VS. (CROSSOVER) INTERACTIONS

"MY KINGDOM FOR A 3-WAY INTERACTION"
OBSERVATION 8: STATISTICAL STERILITY

I HAVE A NEAT RESULT.

DON'T WORRY. WE'LL FIX IT RIGHT AWAY.

GENERALLY-ACCEPTED STATISTICAL PROCEDURES:
- t, F, p value
- log likelihood
OBSERVATION 9:

SOPHISTICATED ≠ COMPLICATED

\[ 7p = \exp - 13 \pi + \ldots \]

\[ F = MA \]

\[ \ldots \quad \text{QED} \]

OR

\[ \frac{A}{B} \]
OBSERVATION 10:

MEASURE VS. MANIPULATION

(WEAK) EXPERIMENTAL MANIPULATION

OTHER INFLUENCES

CONSTRUCT

MEASURE

OTHER INFLUENCES
OBSERVATION 11:

SOCIAL APPROVAL / TENURE

I'D GIVE ANYTHING FOR RECOGNITION

RESEARCHER

FAUST
TEN SUGGESTIONS

1. Consider yourself first a student of consumer behavior and only secondarily an information processor or scanner data modeler.

2. Stop feeling inferior.

3. Give back to the other disciplines.

4. Recognize we're not curing cancer.

5. Recognize even our best work is inevitably wrong.

6. Make the literature review an appendix.

7. Structure your work so it aids future generalization studies.

8. Don't test the null hypothesis of zero effect; report effect sizes as the primary result.

9. Recognize your contribution to the field will be hard to track.

10. Do what you want.
"Fanaticism consists in redoubling your efforts when you have forgotten your aim."

(G. Santayana)
THAT'S ALL, FOLKS

chain saw

Converse sneakers