

HARKing: Hypothesizing After the Results are Known

Norbert L. Kerr

Department of Psychology
Michigan State University

This article considers a practice in scientific communication termed HARKing (Hypothesizing After the Results are Known). HARKing is defined as presenting a post hoc hypothesis (i.e., one based on or informed by one's results) in one's research report as if it were, in fact, an a priori hypotheses. Several forms of HARKing are identified and survey data are presented that suggests that at least some forms of HARKing are widely practiced and widely seen as inappropriate. I identify several reasons why scientists might HARK. Then I discuss several reasons why scientists ought not to HARK. It is conceded that the question of whether HARKing's costs exceed its benefits is a complex one that ought to be addressed through research, open discussion, and debate. To help stimulate such discussion (and for those such as myself who suspect that HARKing's costs do exceed its benefits), I conclude the article with some suggestions for deterring HARKing.

*A reader quick, keen, and leery
Did wonder, ponder, and query
When results clean and tight
Fit predictions just right
If the data preceded the theory
Anonymous*

Practically every modern textbook on scientific research methods teaches some version of logical empiricism's hypothetico-deductive (HD) approach (e.g., Hempel, 1966), particularly for experimental research. That approach prescribes deducing or deriving one or more explicit and testable hypotheses from some plausible theory(ies) about the phenomena of interest prior to designing one's research. The whole process of research design is fundamentally guided and constrained by these hypotheses: What one chooses to look at and how one chooses to look depend vitally upon what

question(s) one chooses to pose. And, until fairly recently, nearly all textbooks that offered instruction on how to write up one's research for publication prescribed organizing the introduction of one's research report around the a priori hypotheses. Such hypotheses serve important functions, both justifying and giving meaning to the empirical research to be reported.

However, in the last few years, several textbooks and articles have suggested a departure from this traditional approach (Baumeister, 1993; Bem, 1987, 1991; Dane, 1990; Sternberg, 1992). One of the earliest and probably the clearest statement of this nontraditional approach can be found in Bem (1987):

There are two possible articles you can write: (1) the article you planned to write when you designed your study or (2) the article that makes the most sense now that you have seen the results. They are rarely the same, and the correct answer is (2) ... the best journal articles are informed by the actual empirical findings from the opening sentence. (p. 172)

Bem (1987) acknowledges that the traditional HD strategy should be used in some situations when writing up one's results, requiring an author to explicitly report findings contrary to one's a priori hypotheses (i.e., negative results): "If your study was genuinely designed to test hypotheses that derive from a formal theory or are of wide general interest for some other reason, then [negative results] should remain the focus of your article" (p. 173).

But Bem (1987) goes on to argue that if one's study is actually exploratory or initiated by "speculations of the 'I-wonder-if ...' variety, ... nobody cares if you were wrong" (p. 173). How is one to judge whether

Editor Note: This article was submitted and reviewed after Norbert Kerr had completed his term as associate editor of *Personality and Social Psychology Review*.

Support during the period in which the initial draft of this article was written was generously provided by the Department of Social and Organizational Psychology, University of Leiden.

The ideas in this article have been challenged and refined by the comments and suggestions of dozens of colleagues. Particularly helpful were the comments of John Adamopoulos, Steve Fuller, Tony Greenwald, Sara Kiesler, Pat Laughlin, Dick McGlynn, and Arjan Wit. The final draft of the article was improved substantially by suggestions of the editor and an anonymous reviewer.

Of course, I must take personal responsibility for the various ways I have employed (and, perhaps, at times, distorted) within this article the provocative suggestions that so many have offered.

Requests for reprints should be sent to Norbert L. Kerr, Department of Psychology, 433 Baker Hall, Michigan State University, East Lansing, MI 48824-1117. E-mail: kerr@pilot.msu.edu.

one's negative findings merit reporting as such? Bem (1987) appears to argue that it is largely a matter of how clearly and strongly (in one's own judgement) the current data support one's post hoc insights.

The data may be strong enough to justify recentering your article around the new findings and subordinating or even ignoring your original hypotheses. ... If your results suggest a compelling framework for their presentation, adopt it and make the most instructive findings your centerpiece. (p. 173)

This strategy of report writing can lead to what I call *HARKing* (or, Hypothesizing After the Results are Known). When speaking of *HARKing*, I am not referring to traditional scientific induction. Induction, inferring some relation or principle post hoc from a pattern of data, is also literally an instance of *HARKing*. But when I refer to *HARKing*, I mean something more specific—presenting post hoc hypotheses in a research report as if they were, in fact, a priori hypotheses.

This practice raises many questions, including:

1. What does and does not constitute *HARKing*?
2. How widespread is this practice?
3. Do scientists approve of the practice?
4. What are the incentives or reasons for *HARKing*?
5. What, if any, are the costs to science of *HARKing*?
6. Do the costs exceed the benefits of *HARKing*?

The immediate purpose of this article is to explore these questions. It may be well to admit, a priori, what will later become evident to the reader, that, in my judgment, the costs of this practice do indeed exceed its benefits. Thus, I also examine the question:

7. How may *HARKing* be deterred?

The ultimate goal of the article, however, is not to moralize about *HARKing*; too many complex arguments exist on each side to make “the evils of *HARKing*” the theme of a compelling sermon. Rather, the ultimate goal of this article is to stimulate discussion, debate, and research on an issue that routinely arises in scientific work but that is rarely discussed.

What Does and Does Not Constitute *HARKing*?

HARKing may be defined as presenting a post hoc hypothesis in the introduction of a research report as if it were an a priori hypothesis. A useful baseline against which to contrast various forms of *HARKing* is the classic HD approach. This model of writing research report introductions (as well as many alternative models that involve some degree of *HARKing*) can be de-

Table 1. *Cross-Classification of Hypotheses by A Priori and Post Hoc Status*

Status of Hypothesis Before the Study Is Run	Status of Hypothesis After the Results Are Known	
	Plausible	Implausible
Anticipated & Plausible	a	b
Anticipated & Implausible	c	d
Unanticipated	e	f

scribed using the 3 × 2 contingency table presented in Table 1. The rows of this table offer a three-level categorization of the a priori plausibility of all hypotheses that one might possibly advance in the introduction of one's research report. The first category (i.e., the first row) includes all those hypotheses that were both known to the investigator and judged to be plausible prior to the data collection. The second row includes all hypotheses that were also known or anticipated by the investigators but dismissed as implausible, a priori. Finally, the third row includes all hypotheses that were unanticipated by the investigators a priori.

The columns of the table dichotomize all possible hypotheses post hoc after the study is completed and the results are known. The first column of the table includes all those hypotheses that are plausible in light of the results in hand. The second column of the table includes all those hypotheses that are implausible post hoc (i.e., that are clearly contradicted by the results in hand). Of course, because either theories or research findings can be equivocal, certain hypotheses can be neither indisputably consistent nor inconsistent with the results in hand. But for the present purposes, let us assume a simpler, more definitive world.

The cells of the 3 × 2 contingency table represent the conjunction of these a priori and post hoc classifications. For example, Cell *a* would include that subset of all anticipated and plausible hypotheses that are also plausible post hoc.

In discussing the classic HD approach and various *HARKing* alternatives to it, I write about “hypotheses advanced in the introduction to one's research report.” By this, I mean hypotheses presented as a priori rationales for the research to be reported; that is, hypotheses that ostensibly guided the design of that research and for which the data to be described provide an independent empirical test. I do not mean any hypothesis, whether presented in the article's introduction or elsewhere, that is explicitly acknowledged as post hoc—that is, as having been suggested by the data to be presented or whose ostensive a priori plausibility has been revised in the light of those data (relative to the plausibility actually perceived prior to knowledge of those data).

In terms of Table 1, the classic HD approach to research design and reporting requires that any hy-

pothesis presented in the introduction to a research report be drawn exclusively from the first row (i.e., *a + b*) and without any consideration of its post hoc plausibility. That is, the only hypothesis(es) that one could advance in the introduction must have been anticipated as plausible a priori. Within this general constraint many variations exist. Only one such hypothesis may exist, or more than one may exist. In the latter instance, identifying empirical questions for which the alternative plausible hypotheses make different or even opposite predictions is desirable (Platt, 1964).

Essentially, HARKing involves taking the post hoc plausibility of hypotheses into account in deciding what hypotheses to advance in the report's introduction. There are nearly as many variations as combinations of cells in Table 1. I restrict my attention, however, to a representative subset of these.

Version 1: Pure HARKing

Bem's (1987) advice to adopt the most compelling post hoc framework(s) for presenting results implies that one would choose that hypothesis from the first column of the table (i.e., Cells *a*, *c*, and *e*), which is most consistent with known results, and most particularly with the results of the present research report. Alternatively one could advance more than one such hypothesis. With this approach, whether the hypothesis was actually anticipated or seen as plausible prior to the execution of the study (i.e., which row the hypothesis falls in) is of no concern.

Version 2: Pure HARKing + Straw Man

Another approach augments pure HARKing by presenting (although, probably personally dissociating oneself from) one or more hypotheses known post hoc to be contradicted by the data (i.e., from the second column of the table). This approach could create an illusion of competitive hypothesis testing (Grant, 1962; Platt, 1964).

Version 3: Suppress Loser Hypotheses

Another approach restricts one's choice of hypotheses to be advanced to those seen as plausible both a priori and post hoc; that is, to Cell *a* of the table. This approach is a variation on the classic HD approach. Its primary difference is that any plausible a priori hypothesis(es) that was (were) contradicted by the data (i.e., those in Cell *b*) would not be advanced. In this version of HARKing, an author would suppress initially plausible hypotheses that failed the immediate empirical test.

Version 4: Post Hoc Plausibility + Necessity of Anticipation

In yet another approach, one might expand the set of hypotheses that might be advanced under the preceding (i.e., suppress loser) approach by including Cell *c* as well. In this approach, a hypothesis would have to be anticipated a priori, but would not have to be judged plausible a priori.

Version 5: Empirical Inspiration

Yet another approach augments the classic HD approach with a bit of pure HARKing. One is required to advance the hypothesis(es) seen as plausible a priori (Cells *a* and *b*), but one can also include any hypothesis that seems plausible post hoc (i.e., in the first column). We might call this approach *empirical inspiration* because it can involve reassessing as plausible a hypothesis originally seen as quite implausible (i.e., entries in Cell *c*) or advancing new hypotheses wholly unanticipated prior to the study (i.e., entries in Cell *e*) in light of the current empirical evidence.

There are other approaches to research report writing that might also be termed *HARKing* and can be defined using Table 1. Their common feature is expanding or altering the set of hypotheses that it is permissible to advance from those prescribed by the classic HD approach, namely, Cells *a + b*. And, some approaches that also might be termed HARKing cannot be defined in terms of the table. For example, in an approach we might term *empirically inspired scholarship*, empirical disconfirmation of one's original hypotheses might inspire a (selective?) search for useful theoretical ideas that may have existed (e.g., have been published or presented somewhere; have been suggested by a reviewer or editor), which were not known to one prior to the current study, and which (had they been known) would have been viewed as equally or even more plausible than one's genuinely a priori hypothesis(es). Or, in an approach we might term *reevaluation of theory*, empirical disconfirmation of your original hypothesis might prompt reexamination of your original theoretical logic (i.e., a review of the soundness of one's original deductions; a check to see if all explicit or implicit theoretical or measurement assumptions were met in this study), with the result that you conclude the theory really should have led you to a hypothesis rather different from the one originally conceived and more consistent with the current data.

Intuitively it seems reasonable to suggest that HARKing is not a simple, "either-one-is-or-one-isn't" distinction. The various approaches sometimes differ in the degree to which knowledge of the data influences the content of one's introduction. For example, Version 1 (pure HARKing) would disregard both whether a

hypothesis had been anticipated and whether it had actually been deemed plausible a priori, whereas Version 4 (post hoc plausibility + the necessity of anticipation) would only disregard the latter. So Version 1, in this regard at least, might be considered to be a more extreme form of HARKing than Version 4. Similarly, Version 5 (empirical inspiration) might be viewed as a less extreme form of HARKing than Version 3 (suppress loser hypothesis) because the latter would expurgate disconfirmations of hypotheses seen as plausible a priori, whereas the former would not. And some approaches (or variations on some approaches) can fall into a grey region of HARKing where the influence of knowledge of results is unclear or subtle. For example, is it HARKing if one advances a new and useful theory that comes to one's attention independently of but after one's knowledge of the results of this study?

Summary

My purpose here is not to propose a definition for HARKing that can be unambiguously applied in any particular instance. Nor do I suggest (or believe) that the various forms HARKing might take are equivalent (e.g., in terms of their costs). My immediate purpose is to define conceptually and illustrate a (fuzzy) set of approaches to advancing hypotheses in research reports that share a common feature—namely, knowledge of the results of the study alters the set of hypotheses advanced in the report's introduction.

How Widespread Is the Practice of HARKing?

Any objections to HARKing would be fairly moot unless (or until) some number of scientists actually do HARK. Although admittedly inconclusive, some evidence suggests that some do.

Circumstantial Evidence

Every editor, reviewer, and active reader of scientific reports has probably encountered such evidence. I propose the following as diagnostic symptoms of HARKing (with tongue only slightly in cheek).

Symptom 1: The too-convenient qualifier. This refers to a qualification of some otherwise-sensible prediction that (a) does not arise in any clear or direct way from otherwise-sensible theory, and (b) is confirmed by the results of the study. For example, an author may follow a straightforward theoretical derivation with a qualification such as, "however, we expect

this only to occur for males (or older participants or participants in the larger laboratory room or seniors who belong to fraternities or ...) because" The accompanying rationale may be plausible, but no more or less so than any of many other such conceivable boundary or moderating (cf. R. Baron & Kenny, 1986) conditions. Such a presentation suggests that (a) the effects of primary interest in the study were not as robust or general as expected, (b) that some variable(s) were discovered post hoc upon which those effects depended, and (c) the original theory was supplemented in a post hoc (and, perhaps, ad hoc) fashion (see the section "HARKing and the Relative Value of Accommodation Versus Prediction" later in this article) to accommodate the observed facts.

Symptom 2: The just-too-good-to-be-true theory.

This is similar to the preceding symptom, except that there is no obvious "rabbit from a hat" production of a useful qualifier. Rather, the theoretical argument seems reasonable, but a palatable mismatch exists between the plausibility, coherence, and apparent power of the theory and the complexity or expectedness of the (inevitably) confirming findings. One can easily imagine plausible variations in the author's theoretical assumptions or parameters that would lead to markedly different, perhaps completely opposite, predictions.

Symptom 3: The too-glaring methodological gaff.

If an author HARKs, the chances increase for a poor fit between the study's design and its apparent goal simply because the study was not originally designed to test the HARKed hypothesis. So, for example, one may be surprised at the nonoptimal way in which an experimental treatment or a measure was operationalized, the absence of an obviously informative control condition, or the author's failure to measure a variable central to the purported mediating process (R. Baron & Kenny, 1986).

Unfortunately, although each of these (and other) symptoms is more likely to be seen when HARKing has occurred, as with most disease symptoms, none of them is perfectly diagnostic. Sometimes theories that look too-good-to-be-true turn out to be true (at least in some context of interest); sometimes investigators simply make methodological gaffs when testing genuinely a priori theories. Thus, no matter how frequently these symptoms appear in a literature, knowing what, if any, degree of HARKing they reflect is difficult.

More Direct Evidence

Informal interviews and observation suggested that, indeed, HARKing sometimes occurred (at least within

my own and my acquaintances' disciplines). But such preliminary, undocumented, and anecdotal reports did not really provide persuasive evidence. In order to provide some better evidence, Susan Harris and I (Kerr & Harris, 1998) surveyed 156 behavioral scientists from three subdisciplines: social psychology, clinical/community psychology, and sociology. Respondents were asked to report the frequency with which they had personally observed scholars in their discipline use a version of the classic HD approach as well as three forms of HARKing (viz., pure or "simple" HARKing, suppress/omit loser hypotheses, and empirical inspiration; see Kerr & Harris, 1998, for details on the survey's methods). Although the HD approach was reported as occurring most frequently, two of the HARKing approaches (viz., pure HARKing and empirical inspiration) were also observed at a frequency statistically indistinguishable from the HD approach. These data suggest that HARKing not only occurs, but that certain forms are as common as the most widely

prescribed traditional approach. (See Figure 1 for a plot of survey results.)

Respondents were also asked to estimate the frequency they suspected that each of the four strategies was used in their disciplines. In this question, they were allowed to draw on any indirect, circumstantial evidence as well as their direct observations. Respondents suspected that the HD approach tended to be used less and all three forms of HARKing tended to be used more than they had personally observed. The most dramatic such difference between observation and suspicion occurred for pure HARKing. In fact, in terms of respondents' suspicions, pure HARKing was the modal strategy and was believed to occur significantly more often than the HD approach (as was empirical inspiration).

Although dismissing such data as reflecting practices of particular disciplines might be comforting, there were no differences between disciplines in the observed or the suspected popularity of the various strategies. It

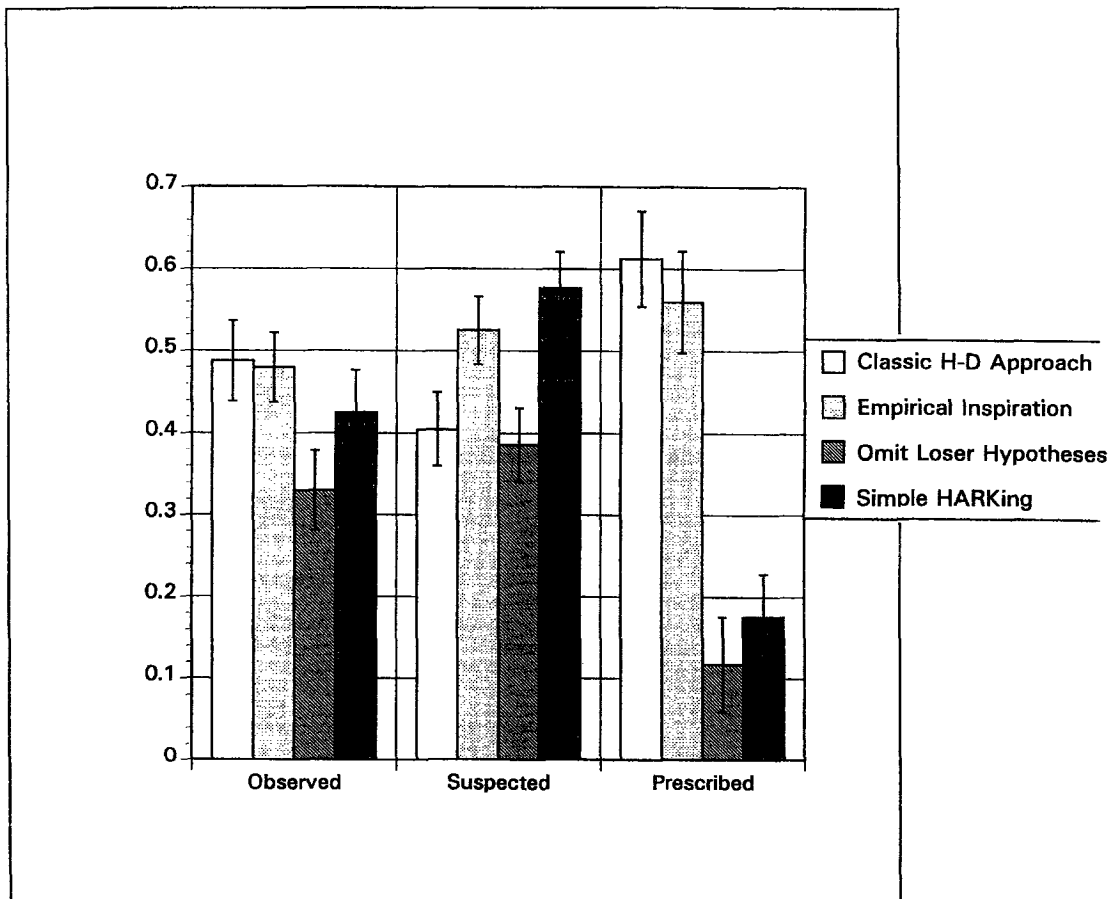


Figure 1. Frequencies that four strategies of advancing hypotheses are observed, suspected, and prescribed.

is, of course, possible that different results would emerge from surveys of scientists from other disciplines. Particularly interesting would be exploring such questions including representative samples of scientists from diverse natural, medical, and physical sciences. Also worth noting at this point is that none of the forces encouraging HARKing (to be discussed shortly) is in any way unique to the particular behavioral science subdisciplines included in Kerr & Harris's (1998) initial survey.

Do Scientists Approve of the Practice of HARKing?

Kerr & Harris (1998) also asked their survey respondents to indicate how frequently each of the four approaches ought to be used. Respondents were asked to take on the perspectives of an authority in their scientific discipline (e.g., an editor of a major journal, an author of a research methods textbook). The classic HD approach was the modal prescribed approach. Two forms of HARKing (viz., pure HARKing and suppress loser hypotheses) were clearly proscribed, but the remaining form, empirical inspiration, was recommended as strongly as the traditional HD approach. Kerr & Harris (1998) suggested that the prescriptive data were consistent with the following rule: It's perfectly acceptable to augment one's list of plausible hypotheses by adding ones suggested by the data, as long as one also sets forth any genuinely a priori hypothesis(es). The respondents seem to object to suppressing disconfirmations of genuinely a priori hypotheses, but not to advancing genuinely post hoc hypotheses.

Again, Kerr & Harris (1998) found no disciplinary differences in these prescriptive data. There were, however, interesting patterns in the data implying that HARKing is becoming more widely accepted and practiced. When respondents who obtained their doctorate degrees more recently (i.e., within the past 17 years) were contrasted with those who obtained them earlier, we found that younger scientists observed the traditional HD approach less and prescribed various forms of HARKing more than the older ones. In this vein, the only endorsements of HARKing Harris and I have been able to find in methods textbooks appear in relatively recent ones (late 1980s and 1990s).¹ None of the older textbooks we examined advocate HARKing (nor, for that matter, do most of the newer ones), and most (explicitly or implicitly) prescribe the classic HD approach.

¹Interestingly, even the recent texts that prescribe or condone HARKing when writing up findings (e.g., Dane, 1990; Judd, Smith, & Kidder, 1991) still advocate the classic HD approach to research conception and design.

What Are the Incentives or Reasons for HARKing?

Several factors encourage HARKing, and of these, I suspect that the paramount incentives lie within the publication process.

Publication Desideratum 1: The Ideal of Confirmation of a Priori Theory

A priori theory/hypotheses as an ideal. I have been struck in my experience on theses and dissertation committees by a recurring event. At some point in the proceedings, typically in reaction to a first draft of a research proposal, one or more of the committee members ask, "What are your hypotheses?" The question is rarely, "Do you have any hypotheses?" The presumption is that any worthwhile study always advances explicit, testable hypotheses. And, in my experience, this presumption often takes little notice of the current state of knowledge on the topic. It is usually not modified just because (a) no prior theory may exist on the topic, (b) the "theory" upon which such hypotheses are based is sketchy or incoherent (e.g., nothing more than intuition or the fact that someone, somewhere has observed something that may be somewhat relevant to the current question), (c) little or nothing is known about the phenomena in question, or (d) the objectives of the study are clearly exploratory. This point of view turns Sherlock Holmes's oft-repeated dictum, "it is a capital mistake to theorize in advance of the facts" (Conan Doyle, 1953, p. 770) on its head, becoming "it is a capital mistake not to theorize, regardless of the knowledge, understanding, or even existence of the facts."

In our (Kerr & Harris, 1998) survey of behavioral scientists we asked respondents to estimate how often, in the opinion of most journal editors and reviewers, a publishable research report should state an explicit hypothesis. Overall, they said that editors and reviewers believed that publishable articles should nearly always do so ($M = 83\%$ of the time). Moreover, when respondents were asked their opinions on this question, they were very nearly as insistent on explicit hypotheses ($M = 76\%$ of the time).

A couple of reasons for such a belief can be suggested. First, it is nearly axiomatic in the version of logical empiricism many scientists are taught. Functionally, most research methods textbooks only offer instruction on how to proceed when one is in the *context of justification* (Herschel, 1987); that is, when the scientist's task is the evaluation of an existing, testable idea or theory. Very few texts offer more than a few words of advice for how to proceed in the *context of discovery* (Herschel, 1987) when the scientist's task is coming up with such ideas. (See McGuire, 1983b, for a notable

exception.) Discovery is seen as a mysterious, chancy, creative process (which, in many cases, it is), but as essentially a preliminary step to the systematic, rigorous, “real” science of testing the theories and hypotheses already in hand. The textbooks’ science “script” (Schank & Abelson, 1977) always presents hypotheses prominently in the first scenes.

One can get to the same place by a very different route. A running theme in calls for qualitative methods in the behavioral sciences (e.g., Richardson, 1996) is that no wholly “theory-free” observations exist. Even if one does not begin research with a coherent or formal theory, one does begin with expectations, assumptions, or hunches, hunches that affect both where we look and what we see. Thus, being as explicit as possible about what our assumptions and expectations are is important. In the young behavioral sciences, such hunches are often the only theory we have. The prescription always to be explicit about one’s expectations can thus be read as a prescription always to advance one’s hypotheses prior to making any observations.

The ideal of confirmation. Popper (e.g., 1959) argued forcefully that disconfirmation of a theory is more informative than confirmation. A clear refutation can go far to eliminate a contending theory. An empirical confirmation merely permits the theory to survive for further study, but it does not narrow the range of possible explanations for a phenomenon.

Nevertheless, indications exist that scientists, like nonscientists, tend to find confirmations more informative and valuable than disconfirmations. Mahoney (1976) asked a group of scientists to indicate valid ways to test a simple hypothesis using Wason’s (1966) selection task. Nearly all (89%) recognized a confirmatory test as informative, but relatively few (39%) considered a disconfirmatory test as informative. Mahoney also found that scientists from many disciplines reported low probabilities that they would bother to write up, follow up, or successfully publish negative results (i.e., results that disconfirmed clear expectations).

The preference for confirmatory tests may reflect the simple belief that disconfirmed theory is “bad” theory (of no direct utility in understanding and controlling the phenomena of interest) and hence that disconfirmations are less informative than confirmations (Popper’s arguments notwithstanding). But it also probably reflects a more general psychological bias—the so-called “confirmation bias.” Considerable evidence now exists for such a bias in lay inference (see J. Baron, 1988; Evans, 1989; or Klayman, 1995, for reviews), which J. Baron (1988) summarizes as follows: “To test a hypothesis, think of a result that would be found if the hypothesis were true and then look for that result” (p. 231). If this is a familiar and comfortable approach to everyday inference, it seems plausible that scientists would both

follow and expect a similar approach in scientific inference.

The combined ideals: Confirmation of a priori theory. The preceding arguments suggest that, all other things being equal, research that advances one or more hypotheses and presents evidence confirming (at least some of) those hypotheses is more likely to be evaluated favorably for publication (and the many rewards contingent on publication) than research that either advances no hypothesis(es) or fails to confirm its hypothesis(es). Authors who HARK, of course, can satisfy both criteria: By knowing the pattern of results before writing the research report, they can ensure that any hypotheses to be advanced (and certainly any such hypothesis that they would personally embrace) in the report’s introduction is nicely confirmed by their data.²

The foregoing suggests that, all else being equal, when an author’s original hunches are contradicted by his or her results, the chances of publication are lower if he or she writes an article that suggests an explicitly post hoc explanation/hypothesis consistent with a study’s data (let’s call this *Draft 2*) than if he or she HARKs and prepares a draft that advances that same hypothesis as if it were an a priori hypothesis (let’s call the latter *Draft 1*). The respondents in our survey (Kerr & Harris, 1998) were asked to judge which of these two drafts was more likely to be published (all other things being equal). Ninety-six percent of the respondents reported that *Draft 1* had the better prospects for publication. When asked what objection editors and reviewers would raise to *Draft 2*, the modal response was that editors and reviewers would require a substantial new expenditure of resources (including time, that most precious resource for a junior scholar for whom the tenure clock is ticking)—namely to replicate the empirical finding.

If an author could write either draft, and *Draft 1* is much more likely to be published and *Draft 2* is more likely to receive a “go back to the drawing board”

²McGuire (1983a) and Greenwald, Pratkanis, Leippe, and Baumgardner (1986) describe another means to the same end. If one actually has a genuine a priori theory, one can search diligently for some set of conditions (e.g., operationalizations, research contexts, participant populations) under which the predictions of that theory are confirmed. If, as McGuire’s (e.g., 1983a) contextualist philosophy of science assumes, all theories are valid in some contexts, with diligence one can discover those contexts and then present research confirming one’s theory in them (neglecting, perhaps, to alert the reader to all the contexts within which the theory had not been confirmed). However, less diligence is required when one HARKs: One fits one’s theory to the facts rather than finding facts that fit one’s theory. (Today’s HARKer, however, might need to switch to the strategy described by McGuire and Greenwald et al. if he or she unambiguously embraces his or her HARKed hypothesis and it is disconfirmed by subsequent observations.)

evaluation, the author has a clear incentive to HARK. But it is not only through such indirect or implicit incentives that HARKing is encouraged within the publication process. Editors and reviewers may insist on it. Survey respondents in Kerr & Harris (1998) were asked if they had ever seen suggestions from editors or journal reviewers to rewrite the introductions to their article so that the article's revised theoretical framework would account for the article's results better than the original draft's. The majority of respondents (52%) reported that they had.³

Publication Desideratum 2: The Ideal of Efficient Scientific Communication

One can argue that journal space and readers' time and attention are too valuable to be wasted reviewing what turned out to be invalid approaches. For example, Bem (1991) writes:

Your report should not be a personal history of your stillborn thoughts. Scientific integrity does not require you to lead your readers through all your wrongheaded hunches only to show—*voilà!*—they were wrongheaded. Your overriding purpose is to tell the world what you think you have learned from your study about human behavior. That may or may not be the same as telling the world about what you used to think about human behavior when you began this investigation. (p. 455)

HARKing is just a variation on this theme. It brings the bottom line, one's final best understanding of one's data, to the top of one's report.

Clearly journal space and reader attention are limited and valuable resources, and research reports neither can nor should be detailed laboratory diaries. Research report writing must necessarily be selective. The difficulty, of course, is distinguishing between what is peripheral and uninformative and what is central and informative. Later I return to the suggestion that an author's "stillborn thoughts" and "wrongheaded hunches" are typically uninformative; that is, whether alerting the reader to hypotheses that were seen as more

plausible before rather than after one's research is worth the space required.

However, a call for concise, efficient writing does not seem to justify those forms of HARKing that portray genuinely post hoc hypotheses as a priori hypotheses. Little journal space is required to acknowledge the post hoc status of a hypothesis. For example, the phrase "this hypothesis did not guide the design of this study but rather was suggested by its results" adds only 17 words. Even the most space-conscious editor or reader would not grudge this much additional space to an author.

Publication Desideratum 3: The Ideal of Engaging Scientific Communication—Telling a Good Story

Earlier I argued that training most scientists instills a "good science" script in which explicit hypotheses are empirically confirmed. But considerable research now exists on narrative production and comprehension (e.g., Pennington & Hastie, 1993; Trabasso & van den Broek, 1985), which suggests a more general "good story script." Good (e.g., comprehensible, memorable) stories tend to have a certain structure. They begin with some initiating events, some "once upon a time" circumstances that trigger and direct the flow of the narrative (e.g., "Little Red Riding Hood wants to visit Grandma, but she lives on the other side of the woods"). Those circumstances initiate a sequence of purposeful, goal-directed actions by the protagonists of the story (e.g., "A wolf learns of Little Red's plan and gets to Grandma's house first, devours Grandma, waits in disguise for our heroine, and then ..."). Eventually, the various threads of the narrative lead to a clear consequence (e.g., "The woodsman cuts open the wolf to save Grandma and rescue Little Red Riding Hood, and they all lived happily ever after").

Another reason why HARKed research reports may fare better in the review and publication process is that they not only provide a better fit to a specific good science script, they may also provide a better fit to the more general good story script. Positing a theory serves as an effective "initiating event." It gives certain events significance and justifies the investigators' subsequent purposeful activities directed at the goal of testing the hypotheses. And, when one HARKs, a "happy ending" (i.e., confirmation) is guaranteed.

Contrast such a narrative with the following, non-HARKed story. "We began believing that H_1 was true. We carried out the following study to test H_1 and found little support for it. However, the results of the study suggests one or more new hypotheses, H_i , which seem plausible." Such a story, although it might be a more accurate depiction of what the scientist knows and how

³One colleague observed that nearly every behavioral scientist has his or her own preferred, pet theory. Such theories can be seen as relevant not just to one's own empirical questions, but to many others as well. So, when a reviewer or editor sees a novel result, one common response is to suggest (or insist) that the author acknowledge the ability of the pet theory to account for the finding. Although I have seen many such suggestions, I have never seen a reviewer insist that the author explicitly identify the reviewer's pet theory as having been identified post hoc (which, as noted earlier, tends to undermine the perceived interest value of the explanation).

and when he or she came to know it, lacks (for want of a better word) the drama of the HARKed story.⁴

If everyone agrees on the importance of a particular result, either because it resolves an important theoretical question, is of clear applied utility, or (as with many classic studies in social psychology) is highly counter-intuitive, then readers may not care too much how skillfully or engagingly one communicates that result. An article reporting a cure for AIDS or the discovery of cold fusion or the like is likely to be published and widely read no matter how well or poorly written it might be. However, when less agreement exists about what the important questions are (because theories are primitive and unrelated or because the utility for application is less clear), and when there is tremendous competition for journal space and for readers' attention, there will be a real premium on how well an author can convince the reader that his or hers is one of those important questions; that is, on how good a story an author can tell. I suspect that a greater premium is placed on telling a good story in relatively newer (and hence, immature) sciences (a testable conjecture), precisely because there is much less consensus on what the key questions are.

Probably everyone would agree that, all other things being equal, "good" (i.e., clear, coherent, engaging, exciting) scientific writing is better than "bad" (i.e., incoherent, unclear, turgid, unengaging) scientific writing. But everyone would also probably acknowledge that an author of a scientific report has constraints on what he or she can write under the guise of telling a good story. Scientific reports are not fiction, and a scientist operates under different constraints than the fiction writer. No matter how much the addition might improve the story, the scientist cannot fabricate or distort empirical results. The ultimate question is whether any such constraints should apply to the fictional aspects of HARKing (e.g., inaccurately representing certain hypotheses as those hypotheses that guided the design of the study).

Hindsight Bias

Hindsight bias refers to "a tendency for individuals with outcome knowledge (hindsight) to claim that they

would have estimated a probability of occurrence for the reported outcome that is higher than they would have estimated in foresight (without outcome information)" (Hawkins & Hastie, 1990, p. 311). In a typical experimental demonstration, participants are asked to judge the relative probabilities of several events occurring (e.g., Fischhoff, 1975) or of several answers to a factual question being true (e.g., Fischhoff, 1977) with and without knowledge of which event actually occurred or which answer is actually true. Those with such knowledge tend to believe that they "knew it all along" (i.e., tend to give higher probability for the known outcome even when instructed to ignore the outcome information). An experimental demonstration of special relevance to HARKing was Slovic and Fischhoff's (1977) study. In their 'foresight' condition, an experiment with two possible outcomes was described. For both possible outcomes, participants were asked to estimate the probability that outcome would be obtained in a specified number of additional replication attempts if the outcome were obtained on the first trial. Participants in the "hindsight" condition were told that the first trial had been completed and had resulted in a particular outcome, and then asked to make the same estimates that the result would be replicated as the participants in the foresight condition supplied. In three separate experiments, they found that participants believed that events known to have occurred were relatively more likely to reoccur than events that could, but had not yet occurred—reported findings seemed less surprising in hindsight than in foresight.

In many research projects, especially in new areas of research, we may consider (or might have considered) many possibilities prior to executing our studies. Also, with few exceptions (e.g., dissertation proposals), we do not permanently record the existence and our evaluation of all these possibilities prior to the study. After we have the results in hand and with the benefit of hindsight, it may be easy to misrecall that we had really "known it all along," that what turned out to be the best post hoc explanation had also been our preferred a priori explanation. That is, scientists may be vulnerable to the same bias that Slovic and Fischhoff's (1977) lay participants exhibited.

The hindsight bias has been replicated many times and appears to be a fairly robust phenomenon (Hawkins & Hastie, 1990). And, although several studies have identified ways of attenuating the bias (Arkes, Faust, Guilmette, & Hart, 1988; Davies, 1987; Hasher, Attig, & Alba, 1981; Hoch & Loewenstein, 1989; Koriat, Lichtenstein, & Fischhoff, 1980; Slovic & Fischhoff, 1977), it is not an easy bias to eliminate. Fischhoff (1977) found that the effect was equally strong whether participants were simply asked to disregard outcome information or told that people "exaggerate how much they would have known without being told the answer. ... Please do everything you can to avoid this bias" (p.

⁴This is not to say that the most engaging narratives are those with no surprises. We would probably enjoy the Little Red Riding Hood's story less if her mother reminded her at the beginning of the story that a friendly woodsman is usually within earshot in times of trouble. (In that vein, if HARKing were rife, little suspense or drama would be found in scientific reports because what "we predict" in an introduction would be tantamount to "we found." One wonders how much a norm of HARKing may thereby contribute to complaints about the low interest value of current scientific writing). Often, the most exciting discoveries are those that were unanticipated (Davis, 1971; van Andel, 1998). But I suspect that such surprises (such as those in a good mystery story) are ones that help to resolve questions posed early in the story, not surprises that fail to do so.

354). Such findings have led Fischhoff (e.g., 1975) to suggest that the bias may occur rather automatically and without effective conscious control. Certainly we have little reason to believe that scientists are immune from this particular bias (and a fair amount of evidence suggests that other experts are not; e.g., Tufte & Sun, 1975; Arkes, Wortmann, Saville, & Harkness, 1981). Thus, even without any conscious or deliberate desire to tell a good story, the best story may also be the most easily recalled story.

Summary of Incentives to HARK

I argue here that evaluative preferences by editors, reviewers, and (ultimately) readers implicitly reward HARKing. Furthermore, professional authorities sometimes sanction or even insist upon HARKing. Editors and reviewers will sometimes direct authors to HARK. Several colleagues have told me that their mentors urged or modeled HARKing behavior (cf. Tesser's 1994 video *On Becoming a Social Psychologist*). And, increasingly, authorities who train and advise scholars (such as methods texts) are sanctioning HARKing (e.g., Bem, 1987, 1991; Dane, 1990; Sternberg, 1992). And even if one were willing to ignore the incentives and counsel to HARK, certain cognitive biases such as the confirmation bias and the hindsight bias might lead one to inadvertently HARK.

In light of these many factors that encourage HARKing, we should not be surprised if one were to find that HARKing is commonplace as our survey (Kerr & Harris, 1998) suggests. But that survey also suggested that certain forms of HARKing were seen as inappropriate by scientists, which brings me to my next question, which is also the central question of this article.

What, if Any, Are the Costs to Science of HARKing?

HARKing Can Translate Type I Errors into Theory

Standard inferential statistics must accept a certain Type I error rate—a probability of falsely concluding that an effect is genuine. When such a Type I error is followed by HARKing, then “theory” is constructed to account for what is, in fact, an illusory effect. Two other biases of the publication process compound this problem. The first is a bias against reports of replication studies (e.g., Neuliep & Crandall, 1990, 1993), stemming from their lack of originality. A second is a bias against the null hypothesis (i.e., against studies that do not obtain significant effects), stemming from several considerations (many methodological flaws such as low statistical power, unreliable measures, and weak manipulations all tend to encourage Type II errors; the recognition that few null hypotheses are literally true;

Greenwald, 1975). In order to publish new work that documents that the original effect was, indeed, a Type I error, follow-up studies must overcome both of these biases: They must convince reviewers and editors to publish a replication study obtaining null results. If the original (but illusory) effect is surprising and the HARKed theory intriguing, they may be considerably easier to propound than to discredit.

Of course, Type I errors are always a risk. Are the costs of making such an error any greater when an author has HARKed than under other circumstances (viz., the hypotheses advanced in the introduction are genuinely a priori; all currently advocated hypotheses are clearly identified as post hoc)? I suspect so. A genuinely a priori hypothesis has some theoretical or empirical foundation that is independent of the current result. Even if that current result is a (apparently confirming) Type I error, there may be sufficient good reasons within that preexisting foundation to continue to entertain that hypotheses (perhaps in situations similar but not identical to the current experimental situation, where our assumption of a Type I error means that the predicted relationship does not really hold). An explicitly post hoc hypothesis implicitly acknowledges its dependence upon the result in hand as a cornerstone (or perhaps, the entirety) of its foundation, and thereby sensitizes the reader to the vulnerability of the hypothesis to the risks of an immediate Type I error. When HARKing occurs, a hypothesis that one ought to approach with the caution due a post hoc hypothesis masquerades as a hypothesis that is due the presumptions granted to a genuine, well-founded a priori hypothesis. (I'll return to this theme later.)

This suggests that any practices that inflate Type I error rates likewise increase the costs of widespread HARKing. In this regard, it is worth noting that certain statistical practices and techniques are more likely to have precisely the former effect. One example is data snooping in large and complex data sets (e.g., with multiple measures) without appropriate experimentwise Type I error rate protections. Another set of approaches worthy of mention here are linear causal models (e.g., path analysis; structural equation modeling; LISREL). Such models are powerful and useful tools, but when they are used not to test the plausibility of a priori models but in an exploratory mode (permitting any and all combinations of paths; cf. Spirtes, Glymour, & Scheines, 1993) and without a cross-validation sample (to ensure that a revealed model is robust), they are unusually vulnerable to Type I errors (Cliff, 1989; MacCallum, Roznowski, & Necowitz, 1992).

HARKed Hypotheses Fail Popper's Criterion of Disconfirmability

Popper (1959) argued that any useful scientific theory or hypothesis must be disconfirmable. Any hy-

pothesis that could never fail (to be confirmed) can never succeed (as a scientifically testable explanation). A HARKed hypothesis fails this criterion, at least in a narrow, temporal sense. When the investigator knows the results of the study in advance and HARKs a hypothesis consistent with those results, no immediate possibility of disconfirmation exists; a gambler would say, “the fix is in.” Of course, this objection may be only immediate and temporary—such a HARKed hypothesis could still be falsified through other observations (e.g., a new study).

HARKing and the Relative Value of Accommodation Versus Prediction

Several scholars (e.g., Brinberg, Lynch, & Sawyer, 1992; Horwich, 1982; Howson, 1982; Simon, 1955; Sternthal, Tybout, & Calder, 1987) have observed a widespread belief that an explanation crafted to fit results post hoc (a process Horwich, 1982, terms *accommodation*) is less likely to be correct than one that successfully (and genuinely) predicts those same results. Results of the survey Harris and I conducted (1998) confirm this informal observation. If this belief is well founded, one could argue that HARKing constitutes a harmful form of misrepresentation—disguising a less scientifically useful, post hoc hypothesis in the clothing of a more scientifically useful, a priori hypothesis. Our immediate goal in this section is to examine whether this belief is well founded.

A good starting point might be to note some of the reasons that have been suggested for this widespread belief. Brinberg et al. (1992) noted two such reasons. First, prediction applies the HD method, which is widely presumed to be more powerful than the inductive method upon which accommodation is based. The second reason is one developed earlier: In the short run, accommodation is invulnerable to falsification. Third, others (viz., Brinberg et al., 1992; Simon, 1955) suggested that post hoc explanations are usually generated to account for a specific new set of observations. The pool of plausible explanations for a specific and narrow set of such observations is likely to be much larger than the pool of hypotheses that are both considered plausible a priori (based on a larger and more diffuse set of observations) and also turn out to account successfully for the new observations. Thus, all else being equal, the “correct” explanation is less likely to be included in the former, larger set than the latter, smaller set. (I return to this argument shortly.) Fourth, Simon (1955) offered another, related argument, namely that the search for a plausible post hoc hypothesis will continue only until one is found (a possibility discussed later) and we have little reason to expect that the first hypothesis one thinks of that fits the data will ultimately turn out to be the correct one. Fifth, Horwich (1982) suggested when

facts fit a theory, we ask ourselves why. One comforting explanation is that the theory is correct. But another explanation is that the “theory was specifically tailored to fit [the facts]” (p. 111). The latter explanation is not possible for genuine prediction (one does not yet have the new facts), but remains a reasonable explanation for successful accommodation. Sixth, a final argument that I have heard (but not been able to find developed in print) goes as follows: One set of data can only accomplish so much. It is asking too much of one set of data both to suggest a new hypothesis to an investigator and simultaneously to provide an “independent” empirical confirmation of that hypothesis. Thus, until a replication of the findings is in hand, any post hoc explanation lacks empirical confirmation (see the discussion in the next section for more on the role of replication in HARKing).

I do not consider each of these arguments in detail here; several are addressed elsewhere in this article. Rather, in this section, I consider not why people believe that prediction is superior to accommodation, but rather whether, in principle, it actually is. Most prior attempts to consider this question have employed Bayesian models. Some (see Howson’s, 1982, discussion) have taken a rather extreme view—that evidence used in the construction of a hypothesis could never be used to support that hypothesis (a variation on the sixth reason in the preceding paragraph). However, Howson has shown that nothing in Bayesian reasoning, in principle, prohibits using new data as evidence for an even newer, post hoc hypothesis. Several other Bayesians have gone further to argue that, in principle, no reason exists to prefer an a priori to a post hoc explanation (see Brinberg et al., 1992; Howson, 1982; and especially Horwich, 1982, chap. 5).⁵ They all suggest that if one’s new results suggest a new hypothesis, one could, in principle, counterfactually estimate the prior probability of that hypothesis being true given knowledge of all evidence available except for those new results, and then use Bayes’s theorem to estimate how much belief now to place in the new hypothesis in light of the new results (i.e., the posterior probability). Under this argument, accommodation is no worse (or better) than prediction (at least in terms of how much faith we can put in their resulting hypotheses⁶). If we have no reason to believe a post hoc hypothesis any less likely to be true than an a priori one, one could also conclude that HARKing does no harm.

This argument is probably valid under idealized assumptions about human inference and scientific practice. However, under more realistic assumptions, I think

⁵See Sternthal et al. (1987) for a non-Bayesian argument with the same conclusion.

⁶Of course, prediction might still be preferred for efficiency of research design (Sternthal et al., 1987; and Symptom 3, p. 199 of this article, *supra*).

it breaks down. The equivalence of accommodation and prediction rests on one (of several; see Fischhoff & Beth-Marom, 1983) key assumption: Hypotheses generated in either way are evaluated in the same way and against the same body of empirical evidence. If an investigator with a new result also considered and weighed all available, relevant old results, then it really would not matter whether any successful hypothesis (i.e., one that accounts for those new results) had been conceived before or after the new results were known. If, on the other hand, knowledge of the new results biases the selection of relevant evidence, evaluation of relevant evidence, or both, this equivalence breaks down. In actual scientific practice, I suspect for a number of reasons that accommodation tends to bias attention on the results currently in hand (see also Fischhoff & Beth-Marom, 1983). One reason is hindsight bias (discussed earlier)—that is, reconstructing unbiased prior probabilities after one knows the result may be difficult or even impossible. Another is that after completing some research, one's attention is usually focused on understanding and explaining the results of that most recent work, not explaining the much larger and more diffuse body of older work (of one's own or, even less so, of others). Furthermore, when communicating and justifying one's recent research, the immediate challenge is to convince the reader (and one's peers and the funding agency) of the importance of that research, which tends to give it special prominence in one's thinking relative to older work. Moreover, which other results besides the most recent one need to be accommodated with it is often not clear (especially when theories are assumed to have narrow ranges of applicability). In summary, I argue that when one practices accommodation (i.e., develops a hypothesis post hoc to account for a finding), one is more likely to tailor a hypothesis to fit a narrow empirical base, dominated by the current result, rather than when one practices genuine prediction, where the in-hand result cannot engulf the attentional field.

If the actual practice of accommodation does bias the evaluation of evidence in this way, then prediction should (all else being equal) be preferred to accommodation. Simon's (1955) argument described earlier offered one reason—that there are many more possible explanations of a small set of facts, and all other things being equal, the probability of any particular one being the correct one is smaller than with the smaller set of facts. Horwich (1982) provided another reason. He drew a distinction between post hoc and ad hoc explanations. The defining aspect of a *post hoc explanation* is temporal: It is developed after facts it tries to explain. The defining aspect of an *ad hoc explanation* (at least for Horwich) is its narrow focus: A theory is ad hoc when its sole purpose is to explain certain specific facts. As we noted earlier, Horwich has argued that nothing

is inherently wrong with a post hoc explanation. But he also argued that truly ad hoc explanations are inherently implausible. They must have a low prior probability (i.e., prior to or exclusive of those specific facts) because if they did not, there would be more than those facts to justify them (contrary to the definition of ad hoc). Essentially, I am arguing that accommodation not only produces post hoc explanation, but also tends—because one focuses too exclusively on particular, recent facts—to produce more ad hoc explanations as well.

This (admittedly speculative) argument suggests that prediction is generally to be preferred to accommodation because accommodated hypotheses are more likely to try only to explain the current results, whereas successful, genuinely a priori hypotheses will usually strive to explain a broader set of findings. To the degree this is true, we should bring greater doubt and skepticism to a post hoc hypothesis than to an a priori hypothesis. HARKing masquerades a post hoc explanation as an a priori one, and thereby prevents us from exercising proper caution.

Replication as a Remedy for HARKing

Before enumerating additional potential costs of HARKing, a point needs to be made. A fairly straightforward remedy may exist for the dangers of HARKing discussed so far: empirical replication. Successful replication makes it very unlikely that one's original result was a Type I error, and hence, at least ensures that one is not HARKing to explain an illusion. If the HARKed hypothesis is sufficiently well developed to permit an empirical disconfirmation, it could fail in a replication attempt, thereby meeting Popper's disconfirmability criterion. Also, at least some of the objections to accommodation (e.g., the circularity of achieving discovery and confirmation with the same data set) are relieved by successful replication.

If these were the only costs attendant to HARKing, then it would be simple in principle (although not always in practice) to minimize them—simply insist on thorough replication of any phenomenon before any explanation is seriously entertained. In addition, of course, this is excellent advice for several reasons besides HARKing considerations. However, it would be a mistake to conclude that replication solves all HARKing problems. The problems discussed next can arise even when the effect of interest has been well and carefully replicated.

Valuable Information About One's Original Hypothesis May Be Lost Through HARKing

Bem (1991) suggested that unless one's original (i.e., actual, a priori) hypothesis is “derived from a

formal theory or ... of wide general interest for some other reason" (p. 455), nobody is likely to care if it turns out to be wrong. Some varieties of HARKing (e.g., suppress loser hypothesis; pure HARKing) result in ignoring such failed hypotheses in one's research report. But if one's original hypotheses had a sufficient rationale (theoretical, empirical, or even intuitive) to recommend itself to one researcher, why would it not also occur to others? Is not the information that a particular approach did not work very valuable? At the very least, it may help prevent others from traveling down paths that have been discovered to be dead ends. And if one's own tests of these "wrongheaded hunches" or "stillborn thoughts" are rigorous, they may serve the vital function of narrowing the range of plausible hypotheses (Grant, 1962; Platt, 1964). Furthermore, methodological choices that seem suboptimal or even inexplicable in the light of a HARKed hypothesis might be comprehensible and quite appropriate in the light of one's genuine original hypotheses (cf. the discussion of Symptom 3, p. 199 of this article).

It might be argued that it is not necessary for an author to explicitly bring his or her false leads to the attention of the reader. The interested reader who might entertain those same false leads can see for himself or herself that they are contradicted by the results of the article. This argument assumes that an implicit disconfirmation (when the disconfirmed hypothesis is not even mentioned) has the same impact as an explicit disconfirmation (where the empirical research was explicitly designed to provide a test of the disconfirmed hypothesis). This assumption is implausible, particularly in research domains where (a) the pool of conceivable hypotheses is very large and not widely shared, (b) the accompanying theory is not well developed (e.g., with well-understood boundary conditions), or (c) few standardized research methods or paradigms exist (so that comparability of findings across laboratories is potentially low).

HARKing Can Promote Statistical Abuses

Several situations exist in which investigators are granted some license in their statistical analyses. By this I mean that the investigators are allowed to relax or suspend usual and fairly conservative statistical standards. The most familiar illustration is the use of a directional or one-tailed test of a statistical hypothesis. This functionally enlarges the test's critical region and makes concluding that an observed difference is a statistically significant one easier. The rationale and prerequisite for such a practice is that a preexisting and well-established theory exists that justifies a change in the most plausible null hypothesis (from one of no difference between conditions to one of differences in the direction opposite to that theory's predictions). Other, similar practices include (a) performing specific

a priori contrasts that focus statistical power on comparisons of theoretical interest, and (b) disregarding particular effects that, although statistically significant, are of no real theoretical interest.

The justification for all of these statistical practices rests on the existence of a clear and genuine a priori theory. HARKing can involve post hoc theory masquerading as such genuine a priori theory. When HARKing is accompanied by use of any of these practices, the justification is lost. HARKers' use of such practices constitutes taking unjustified statistical license.

HARKing Presents an Inaccurate, Distorted Model of Science

Scientific publications do not simply communicate ideas and observations. They also present a model of how science is routinely conducted to the public and to students. It is HARKing's effect on students that concerns me most (I return later to the effect of HARKing on public perceptions of and attitude toward science). HARKed articles present a rosy picture—a prescient scientist anticipates and correctly predicts a complex pattern of results. When the student begins work, he or she is likely to discover that nature is only rarely so cooperative: Partial successes and undeniable failures are commonplace. Students react to this discrepancy in many ways. Those who make a situational attribution—that actual science is a lot more difficult and unpredictable than published science—are likely to persist and persevere (Seligman, 1991). Good mentoring can be important in bringing students to this understanding. On the other hand, those who make a personal, dispositional attribution—that they lack the imagination or talent to do publishable science—are more likely to give up. Sometimes they are right to do so; not everyone is suited to be a scientist. But sometimes (too often, I think) they are wrong. The problem is not that they lack the qualities necessary to do productive science, but rather that they have an inaccurate picture of how productive science normally proceeds. HARKing contributes to such a too-rosy picture.

This argument has many counterarguments. One can argue that publications primarily serve an archival function. Those on the cutting edge of scientific work communicate directly (at meetings, with preprints, and increasingly, via e-mail) and not through formal publications. In that same vein, one can argue that students shape their model of science more from their experience than from their reading. There is something to such arguments, but most students and scientists do rely upon standard publication outlets as a primary means of scientific communication. In this regard, it is instructive to view the interviews of prominent scholars in my own discipline, social psychology, in Tesser's (1994) video *On Becoming a Social Psychologist*.

When asked what advice they would offer to aspiring social psychologists, many recommended intensive reading of the literature of one's research question.

Is HARKing Unethical?

None of the existing ethical guidelines I have been able to find explicitly prohibit HARKing. The practice is not really considered in the guidelines that apply to my own discipline, *American Psychological Association Ethical Principles and Codes of Conduct* (e.g., Canter, Bennett, Jones, & Nagy, 1994). Nor does it appear explicitly in the National Academy of Science's blue-ribbon committee's list of ethical felonies (which they term *misconduct in science*, consisting of fabrication, falsification, or plagiarism in proposing, performing, or reporting research; Committee on Science, Engineering, and Public Policy, 1992) or ethical misdemeanors (termed *questionable research practices*, including failing to preserve or share data, inappropriate statistical analyses, inadequate supervision or exploitation of subordinates, or misrepresenting speculation as fact; Committee on Science, Engineering, and Public Policy, 1992). So, in terms of current ethical standards, no explicit objection appears to exist toward HARKing.

However, I think a case can be made that HARKing violates a fundamental ethical principle of science: the obligation to communicate one's work honestly and completely. Albert Einstein states this principle well: "The right to search for truth implies also a duty; one must not conceal any part of what one has recognized to be true" (cited in Committee on Science, Engineering, and Public Policy, 1992, p. iii). HARKing can entail concealment. The question then becomes whether what is concealed in HARKing can be a useful part of the "truth" (e.g., see the section on p. 120 of this article regarding information about one's original hypothesis lost through HARKing) or is instead basically uninformative (and may, therefore, be safely ignored at an author's discretion).

Such assessments are difficult and clearly depend upon the particular circumstances (e.g., how much good evidence was found and how much interest exists in one's original hypotheses; Bem, 1987, 1991). Unambiguous ethical standards can provide useful benchmarks. Probably the least ambiguous standard is the "capital offense" of science, fabrication of results. The primary reason that data fabrication is so serious is that it disregards the most serious constraint of science: to be empiricists and to let observation rather than preference or assumption be the final arbiter between contending ways of describing and understanding nature. But why else is this such a serious ethical breach and if so, do similar concerns arise for HARKing? Hull (1988) suggested that fabricating data is particularly noxious, in part, because it adversely affects more people and their

work than other forms of misconduct (e.g., plagiarism). Similarly, those adverse effects that have here been attributed to HARKing can affect any and all consumers of the HARKed reports. Data fabrication is also serious because it undermines mutual trust without which normal science would be nearly impossible. Of course, scientific communication is not just based on trust—we must be able to replicate one another's findings before we put too much trust in them. We are usually willing to attribute consistent failures to replicate to vagaries of chance (e.g., the original findings were Type II errors), to the importance of subtle contextual features (e.g., there were small but important differences in methods), or to some flaw in the original study (e.g., there was a contaminant; a mistake was made in analysis). But when we have to begin worrying about deceit or dissembling as well as such factors, scientific communication begins to break down. The content of what is concealed or misrepresented in HARKing is undoubtedly less crucial than what is misrepresented when results are fabricated, but the damage done by widespread and recognized HARKing to mutual trust among scientists may be qualitatively the same.

HARKing Breeds Cynicism and "Fudging"

Some might argue that discovering and accepting the gaps between scientific ideal and scientific practice is simply a part of one's education as a scientist. Sternberg's (1992) observations are instructive:

Journal articles are usually written in a manner that bears little resemblance to the way the research was actually conducted. This difference is not dishonesty: professionals simply know how the system works. I first learned this fact when I was in graduate school ... I'd just received an editor's letter. (p. 13)

Sternberg is right—"The system" can encourage or even require authors to tell a different story than the actual story, and HARKing may be an instance of this. But problems can result when a considerable gap occurs between what the system encourages and what we are taught and accept as ideal scientific practice. Such gaps are especially likely to breed cynicism if the reasons for the system's practices are unexamined (e.g., "That's just the way things work."), weakly justified (e.g., "No one is interested in what actually happened."); "The journals just don't have enough space to describe what actually happened."), or justified on grounds of expediency or self-interest (e.g., "If you want your work published, this is what the editors and reviewers prefer.").

HARKing may fall into that grey area of "questionable scientific practices" where the line between clearly appropriate and clearly inappropriate behavior is indis-

tinct. Codes of scientific conduct, principled graduate training, peer review, and reputational concerns all serve as forces to keep scientists operating on what the scientific community identifies as the appropriate side of these lines. Empirical results indicate (Kerr & Harris, 1998) that with regard to HARKing, quite a few may be operating on the wrong side of many people's line. Furthermore, in the absence of reasoned discussion to justify a clear redrawing of the line, such a situation encourages ignoring or blurring other such lines when convenient. As a member of the Vietnam War generation, I am generally suspicious of such "domino theories" or "slippery slope" arguments. A real difference I see in the current instance is that professional authorities (editors, reviewers, textbook writers) seem to be accepting, condoning, or even advocating a practice that is counternormative without explaining precisely and persuasively what is wrong with the traditional norm. This is not just recognizing that scientists have to operate on a slippery slope; it is greasing the slippery slope.

HARKing's Causes May Adversely Affect Scientific Discovery

This is less a consequence of HARKing per se than of one of its primary causes: insistence on confirmation of a priori hypotheses as the only route to new knowledge. The same motive that can lead to HARKing following the clear failure of one's original hypotheses can also lead to selective attention and forced interpretation when the hypotheses have not so clearly failed. Baumeister (1993) made this point:

Occasionally one sees researchers who are strongly committed to a particular point, and the work seems to suffer from this rigidity. Some such researchers ignore contrary findings, stretch interpretations to fit preconceived notions, run multiple studies or multiple analyses until they find what they want.⁷ (p. 10)

An investigator who is rigid in this sense is unlikely to notice or give credence to unexpected or serendipitous findings. And if, as some have argued (Merton, 1968; van Andel, 1998), such serendipity is not a rare event but an extremely common aspect of scientific discovery, the costs of such rigidity can be high. Van Andel (1998) documented the role of serendipity in dozens of important scientific discoveries. One Nobel laureate after another recounts how some unexpected, unpredicted observation turned out to be crucial in their work. Frequently, the importance of this unexpected observation was not evident when it was first made. Rather, the scientists' receptivity to that surprising ob-

servation, their unwillingness to ignore it as some uninteresting anomaly, made their ultimate achievements possible.

Adverse Effects of HARKing on Theory Development

HARKing promotes narrow (i.e., context and paradigm bound) new theory. A primary goal in science is to develop powerful, general theory with cross-situational utility. We are more likely to do so if the process of theory development is forced to confront, accommodate, and integrate findings from many diverse situations or contexts. However, as noted earlier, HARKing may foster preoccupation with explaining an effect in hand with an attendant lack of attention to the broader set of potentially relevant prior findings. If HARKing were commonplace, much of our new theory would have such a genesis—being developed to explain some specific effect observed in some specific experimental context and paradigm. And the research that built on such HARKed hypotheses would likewise tend to stick to that proven context and paradigm rather than wandering into new contexts or new paradigms where those hypotheses might fail.

HARKing promotes loose, disconfirmable old theory. Suppose one obtains not only an unpredicted result, but one for which the author can think of no good explanation (e.g., an unpredicted, many-factor interaction effect). In a scientific world in which unpredicted effects are seen as uninteresting (at least until they have been replicated), this poses a problem for a HARKer. One solution is to employ theory with which one can predict nearly any pattern of results in nearly any context. Such theory is likely to be vague in its concepts, to incorporate multiple (sometimes opposing) processes, and to have many auxiliary assumptions that are not firmly fixed and can be used in an ad hoc fashion to accommodate nearly any observation. In Lakatos's (1978) terminology, these are theories with small or even nonexistent cores and wide "protective belts." Such theories are the last resort of the dedicated but creative HARKer in extremis (i.e., when no other good explanation can be conceived), and, perhaps, the first resort of dedicated HARKers who are less creative. In a world where HARKing is commonplace, such scientifically bankrupt theories will always have willing proponents.

HARKing may inhibit the generation of alternative hypotheses. Previously I touched on a related point: the effect of HARKing (or HARKing's

⁷See Footnote 2, *supra*.

causes) upon receptivity to unexpected findings. But HARKing itself may also tend to have an adverse effect on another kind of receptivity—to alternative explanations.

That scientific progress is retarded by less than thorough consideration of plausible alternative explanations is not disputed. Failure to consider alternative explanations is a recipe for inefficient research design (Platt, 1964) and preoccupation with theories that turn out to be invalid or of limited utility.

J. Baron's (1988) review of research on the confirmation bias led him to suggest a two-part confirmation heuristic. The first part has been the focus on much hypothesis testing research: "To test a hypothesis, think of a result that would be found if the hypothesis were true and then look for that result" (p. 231). However, the second, concluding part of Baron's statement (on which relatively little research has been done, although see Wason, 1960; Gettys & Fisher, 1979) is of greater interest to us: namely, "... and do not worry about other hypotheses that might yield the same result" (J. Baron, p. 231). Some of the work reviewed by Baron suggests that people often use a "win stay—lose change" strategy for hypothesis development and that as long as one's current hypothesis is not clearly contradicted by available evidence, one does not need or attempt to consider alternative, plausible hypotheses (also see Shaklee & Fischhoff, 1982). One way of expressing this notion is that people tend to be satisficers rather than optimizers when generating hypotheses (Simon, 1955).

HARKing guarantees the provision of one or more hypotheses sufficient to account for a result in hand. If readers or the investigator himself or herself acted like such satisficers, they would not need to consider alternative explanations. And, assuming that the result is not a Type I error and that investigators do not wander too far from the original paradigm, subsequent replication of the result is likely to lead to high confidence in that hypothesis, much as Wason's (1960) participants tended to grow overconfident when their overly narrow hypotheses kept passing confirmatory tests.

Unlike many of the other objections to HARKing discussed herein, this possibility suggests tractable empirical questions. In a recent laboratory simulation of scientific inference (Kerr, Garst, Harris, & Sheppard, 1998), my colleagues and I explored one such question: "Will the existence of a sufficient hypothesis lead to less thorough generation of plausible alternative hypotheses than when one has no such sufficient hypothesis?" The results of three experiments all suggested that the answer is (a qualified)⁸ "yes."

⁸In our paradigm we did not find a global and undifferentiated inhibition of plausible hypothesis identification, but rather a more selective one. The simplest, most parsimonious alternatives were the most likely to be overlooked when our participants had a sufficient hypotheses already in hand.

I should also add that this logic and our experimental results do not differentiate between sufficing hypotheses that are genuinely a priori, overtly post hoc, or HARKed (i.e., covertly post hoc). We suspect—but have not yet been able to show empirically—that people (including scientists) are likely to be more complacent when the sufficing hypothesis has predicted results (i.e., the hypothesis is HARKed or genuinely a priori) than when it has accommodated them (i.e., is post hoc). And, as argued earlier, we also suspect that such complacency is more risky when the prediction is only apparent (i.e., HARKing is occurring) rather than real (i.e., the hypothesis was actually developed prior to seeing the results).

Summary of Costs of HARKing

Several potential costs of HARKing have been identified, including

1. Translating Type I errors into hard-to-eradicate theory.
2. Propounding theories that cannot (pending replication) pass Popper's disconfirmability test.
3. Disguising post hoc explanations as a priori explanations (when the former tend also be more ad hoc, and consequently, less useful).
4. Not communicating valuable information about what did not work.
5. Taking unjustified statistical licence.
6. Presenting an inaccurate model of science to students.
7. Encouraging "fudging" in other grey areas.
8. Making us less receptive to serendipitous findings.
9. Encouraging adoption of narrow, context-bound new theory.
10. Encouraging retention of too-broad, disconfirmable old theory.
11. Inhibiting identification of plausible alternative hypotheses.

And last, but not least

12. Implicitly violating basic ethical principles.

Do the Costs of HARKing Exceed Its Benefits?

Identifying potential risks of HARKing is much easier than showing that these risks justify proscribing or discouraging the practice. I have no compelling arguments on this question, but to stimulate discussion of it, I pose and discuss a few questions.

Can We Recognize and Discount HARKed Hypotheses?

The potential costs of HARKing discussed herein might not pose much of a problem if editors, reviewers, and readers are well aware of these risks and discount “too-good” fits between prediction and observation accordingly. This requires, however, that one can tell the difference between HARKing and genuine prediction. This is not easy. First of all, detecting HARKing is not like detecting other forms of questionable behavior where the relevant evidence is fairly straightforward (as with plagiarism) or where empirical means of checking one’s suspicions exist (e.g., such as detecting data fabrication). A judgment of HARKing involves challenging what somebody claims he or she was thinking in the past, and (as the hindsight bias suggests) may currently believe he or she knew all along. Clearly, challenging such a claim is not easy.

One could simply mistrust any and all claims of a priori hypotheses, but this is unsatisfactory on at least two grounds. First, it is not reasonable to discount every apparently successful prediction, including the real thing, because some unknown number of people practice HARKing. Second, to the degree that HARKing is undesirable precisely because it can breed mistrust among scientists, it seems poorly “solved” by indiscriminate and general mistrust.

One solution I have heard is to see if there is a decent theoretical rationale for any suspect hypotheses. When one does not exist, we would have good reason to suspect HARKing and could discount accordingly. This suggests, conversely, that we can be fairly confident that the author is not HARKing, and therefore, need not do any discounting, whenever he or she can muster persuasive theoretical or empirical arguments for his or her theory. But in many areas, especially in the newer and softer sciences, a lot of theory and research exists, often with very contradictory implications. I suspect that an able and motivated (and, quite possibly, unwitting or reluctant) HARKer can nearly always find a serviceable rationale for nearly any hypothesis. Perhaps the best defense against such selective scholarship is for those who evaluate the work to know the relevant theoretical and empirical literatures well. Editors and reviewers who are sufficiently knowledgeable should be able to distinguish a biased and selective appeal to the literature from a balanced and comprehensive one.

How Permanent Are the Costs of HARKing?

A great strength of science is that it is self-correcting. As long as inquiry is open (i.e., we avoid the kind of political or ideological constraints that people such as

Lysenko imposed and scientists such as Galileo had to contend with) and everyone agrees on the ground rules (e.g., that we are going to be empiricists), science eventually sorts out the good ideas from the bad ones. And many of the potential costs of HARKing (e.g., turning Type I errors into theory; advancing hypotheses that are not immediately disconfirmable; inhibiting immediate alternative-hypothesis generation; capitalizing on statistical rule bending) seem to be short-term ones that can only misdirect us temporarily. This suggests that although HARKing may be a source of inefficiency in science, it cannot do any permanent damage.

This is a good point, but I can see a couple of objections. The first is that not all of the costs of HARKing may be rectified by the normal course of science. In some cases, the harm could be cumulative, rather than temporary. It is not clear how normal science corrects the cynicism, the mistrust, and the erosion of ethical standards that HARKing may breed. The other argument is that we should not use the resiliency of science as an excuse for practices that harm science, even temporarily. For example, if someone fabricates data, eventually this will be discovered (assuming that other scientists find the result interesting enough to pursue): The damage is not permanent. But the widespread harm that such fabrication does (to misdirecting and wasting resources, to undermining mutual trust among scientists and public trust of science, and to the integrity of the whole scientific enterprise) is still bad enough to make such fabrication abhorrent. I am not arguing that HARKing has the adverse impact of data fabrication, but we do not excuse data fabrication simply because its damage may only be temporary. We should not use this as an excuse for HARKing either.

What Are the Transaction Costs of Doing Something About HARKing?

Of course, the answer to this question depends a great deal on what that “something” might be (see the section on how HARKing might be deterred on p. 213 of this article). But some costs may attend any kind of restriction on how scientists operate. Some (e.g., Feyrabend, 1993) have argued for a laissez-faire brand of science that holds that regulation and constraint only interfere with scientific progress. This viewpoint argues that science, particularly in the context of discovery, cannot function well under tight constraints that permit only certain acceptable ways of developing useful knowledge. And who cares if someone is HARKing as long as the idea turns out to be useful?

Of course, from the perspective of scientific progress, what matters most is the birth of useful new ideas, not the method of delivery. And we never know, in the short run, at least, which ideas will turn out to be useful

and which will turn out to be dead ends. But this seems to beg the question. I do agree, in principle, that it is possible to inhibit original, creative work by imposing constraints on one's conceptual or methodological strategies. But an ideal of minimizing restrictions on scientists is not the same thing as an ideal of eliminating all restrictions. Some restrictions (e.g., peer review to minimize risks to human participants) serve important functions. The real question, then is not whether constraints should be placed on scientific work, but rather which constraints are necessary. And, in our current discussion, should HARKing be among the scientific practices that we proscribe, or should it just be ignored?

The costs and difficulties of enforcing a proscription on HARKing may not be the only costs of opposing the practice. Suggesting that HARKing is both common and harmful in science could end up hurting science in general, the disciplines of those who make such suggestions in particular, and especially the individual scientists who do so. Anytime somebody suggests that something objectionable exists in scientific practice, he or she is bound to raise other scientists' hackles. Taking a look at what usually happens to whistle-blowers in science (e.g., LaFollette, 1992) is instructive. Whistle-blowers are much more likely to be criticized (or worse) than to be thanked for their trouble. Scientists are not generally thanked for rocking the boat (even if—as with HARKing—there may be no good reasons to assume that such problems are narrow personal or disciplinary ones, and excellent reasons to suspect that they are not). Raising issues such as HARKing may be likened to washing dirty linen in public. Science is vitally dependent upon public support. Those who oppose strong support for science would be only too happy to use such criticism of science (regardless of its intent) as ammunition against such support. Sometimes the damage done in raising and trying to solve a problem is greater (especially to those who are most concerned about solving it) than the damage the problem itself causes. I have heard colleagues argue that HARKing is such a problem.

It is hard to tell just when this line of argument would or would not justify ignoring any problems in actual scientific practice, and thereby foster their continuation. But ultimately, public support for science is not grounded in a public illusion that actual scientific practice is flawless. Rather, it is based on science's demonstrated ability to answer important questions and solve important problems. And one of the important self-correcting mechanisms of science that I mentioned earlier is a reliance on open discussion and inquiry, even when the object of that inquiry is science itself.

How Might HARKing Be Deterred?

I want to conclude with a discussion of how HARKing might be deterred. This section will probably be of

greater interest for those who believe that the costs of this practice exceed its benefits (plus the benefits of simply ignoring the question). But even those who are not now convinced that HARKing is much of a problem might be willing to consider what might be done if and when they were so convinced.

The incentive structure for HARKing may represent a social dilemma (Dawes, 1980; Messick & Brewer, 1983), a situation where individual interests conflict with the interests of the group as a whole. Individual HARKers are more likely to be rewarded (with publication and recognition) than individual non-HARKers, all other things being equal. In this sense, HARKing is a rational choice, but the net effect of everyone making this rational choice might be much worse for science and for the scientific community than if all made the personally "irrational" choice of not HARKing.

I believe that most authors who HARK are responding primarily to strong external incentives: to the way the system works. Other, internal incentives are at work to be sure (e.g., the hindsight bias; enacting one's own science script or story script), but I doubt if they would prevail if external incentives were neutral or discouraged HARKing. I raise this point to emphasize my conviction that the problem of HARKing is really not one of individual scientists making personally rewarding choices that they know to be costly to the scientific enterprise. For one thing, little open discussion about HARKing has taken place, and what has taken place has generally focused on its benefits (e.g., to facilitate communication; to tell a good story). In addition, for many young scholars the need to publish is less one of personal reward than of professional survival. Given the several (generally undiscussed) arguments that can be raised against HARKing, I suspect that most authors would, to the best of their ability, gladly refrain from the practice if the system did not encourage it.

Messick and Brewer (1983) distinguished between individual and structural solutions to social dilemmas. The former strive to alter individual choice behavior. The latter strive to change the structure of the dilemma itself. I also suspect that individual solutions to the HARKing dilemma cannot succeed without corresponding structural solutions. During a recent presentation on HARKing, a junior colleague asked: "I've just received an action letter from an editor that instructs me to move post hoc hypotheses up to the introduction. What am I supposed to do, refuse to do so?" I could not advise her to do so. Individual scientists have to work and survive in the system as it exists. Without systemic, structural changes, individual, principled choices not to HARK may be futile and professionally destructive. Thus, most of the suggestions that follow propose structural solutions that must be implemented by scientific

communities or their leaders and not by isolated, individual scientists.⁹

Education

Standard graduate training in research methods and philosophy of science should be supplemented with discussion not just of HARKing, but of all the grey areas of research practice that raise issues of scientific misconduct. As the National Academy's committee observed (Committee on Science, Engineering, and Public Policy, 1992), there are many such practices for which scientists have honest disagreements. At the very least, students should be exposed to the various arguments raised for and against such practices so they can both reach their own decisions on what scientific integrity requires and participate in the ongoing efforts to reach some consensus.

Undertake Research on HARKing

At least some of the purported risks of HARKing rest on assumptions that are empirical questions, including questions such as "How common is HARKing in various scientific disciplines?"; "What are scientists' attitudes toward HARKing in various disciplines?"; "Do we have direct evidence that authors who HARK improve their chances of publication?"; "Are HARKed narratives more interesting and memorable than alternative narratives?"; "Is HARKing associated with greater degrees of cynicism and various questionable scientific practices?"; and "Does HARKing tend to promote suboptimal generation of plausible alternative hypotheses by scientists?" Providing data on such questions could contribute to reaching consensus about the real risks of HARKing.

Address HARKing in Professional Codes of Conduct and Ethics

If a clear consensus of scientists could be reached in a discipline, formally incorporating that consensus into the professional codes of conduct could also be helpful. And, if the costs of clear violation of such provisions were perceived as sufficiently serious, appropriate penalties for such violation should also be provided for.

⁹Of course, structural solutions may also require individual action, particularly by the leadership of disciplines (e.g., journal editors, textbook writers, review panels). There are also purely individual remedies that could be helpful. For example, falling victim to the hindsight bias is more difficult if we routinely kept written journals recording the temporal development of our theoretical ideas, hunches, and so forth.

Alter the Incentives for HARKing Within the Research Publication Process

I speculated earlier that the primary forces encouraging HARKing stem from evaluative criteria and practices in the research publication process. If this is so, changes would need to be made in those criteria and practices to deter HARKing. Some of these changes could be implemented from the bottom up by altering the attitudes and behaviors of reviewers and readers. However, I suspect that such changes are implemented more quickly and effectively in a top-down process, with editors and editorial boards explicitly proposing and implementing changes such as the following.

Promote/insist on replication. As noted earlier, although replication is not a panacea, it can prevent some of the costs of HARKing.¹⁰ This could be achieved in several ways. Established journals featuring empirical reports could attempt to temper the usual bias against replication studies (Neuliep & Crandall, 1990, 1993). In addition, new journals might be established that feature reports of replication attempts (such as the short-lived *Replication Studies in Social Psychology* in my own discipline). The opportunities created by desktop and electronic publication make the latter option increasingly feasible.

Make HARKing a basis for rejection of an article. One can find all sorts of attitudes toward the traditional HD model of science among scientists and philosophers of science today. Among the former, I believe, the modal attitude is quite positive, so positive, in fact, that HARKing is widely rewarded. Many critics of this traditional approach feel that alternative approaches (e.g., qualitative methods, discourse analysis, case studies) should sometimes, usually, or even always be used.

This discussion of the risks of HARKing might be viewed as yet another criticism of the HD approach and another reason to curtail or abandon its use. I think that would be a mistake; anyway, it is certainly contrary to my intentions. Nothing presented here diminishes in any way all of the excellent reasons for using the HD approach developed by the logical empiricists. Many of the greatest achievements of science have resulted from scientists making and testing bold theoretical conjectures. The value of theory (for organizing what we know, for extending what we know, for guiding new research) is undeniable. For all these reasons, we should

¹⁰Insisting on replication, either within or between research reports, also carries certain drawbacks (cf. Wegner, 1992). For example, it may discourage work on problems that are difficult or expensive to replicate (e.g., long-term longitudinal research, group research).

continue to give special approbation to those who develop coherent theory, deduce its implications, and provide informative empirical tests. And, for all the reasons discussed herein, I think we should not only withhold such approbation from those who merely simulate this process but also penalize them (e.g., make obvious HARKing a basis for rejection of an article).

Of course, one objection to this recommendation is that determining the difference is often very difficult. HARKing is not a practice that can easily be detected. Most articles are written wholly after the data have been analyzed. An insistence that one be able to provide some theoretical or empirical rationale for one's hypothesis is certainly no guarantee that HARKing is not occurring—skilled or motivated scholars can find some such rationale for nearly any hypothesis.¹¹

Although it is difficult to detect HARKing, it is not impossible (at least to make educated inferences). The HARKing symptoms I described earlier offer useful (if inconclusive) clues. Trial by logical/scholarly plausibility is another method. HARKers are more likely to make a biased and selective appeal to prior theory and research to justify their hypotheses than non-HARKers.

Another criterion ought to enter into such assessments: the quality of the theory in question. The HD approach presumes that one's hypotheses can be deduced from one's theory. Deducing (with confidence) any specific hypothesis from poor theory is often very difficult. By poor theory, I mean theory that is internally inconsistent, incoherent, loosely reasoned (e.g., by analogy or association rather than through formal or mathematical logic) with unspecified boundary conditions and whose core concepts have only vague conceptual and operational definitions. In Lakatos's (1978) terminology, they are theories with wide protective belts, offering multiple handy escape hatches for apparent disconfirmations and a nearly empty core. This is not to say that even the poorest theory cannot be useful—even if its core is nearly empty, what is there could be valid or stimulate research that leads to better theory. Rather, I argue that tests of hypotheses derived from poor theories cannot be more valid than their derivation, and that because one can derive many, sometimes conflicting hypotheses from poor theory, an apparent empirical confirmation of one such hypothesis should not be viewed as a successful application of the HD model (Fischhoff & Beth-Marom, 1983). This is, I suspect, a

key confusion. We trivialize the term *theory* and overcredit apparent confirmations when we draw no distinction between hypotheses derived from poor theory and those from better theory. Perhaps we should withhold that special approbation reserved for genuine successful application of the HD method from hypotheses deduced from poor theory, regardless of whether that theory is (or alleges to be) a priori theory. (It might even be useful to require that a theory pass some reasonable quality threshold¹² before one could speak of one's hypotheses; perhaps we should reserve other terms—*conjecture?*, *prediction?*—for hypotheses derived from patently poor theories.)

Legitimize candid identification of research as exploratory. In those instances where little or no good theory exists, an author should be able to acknowledge frankly the exploratory nature of his or her research without thereby ensuring the article's near-automatic rejection. And, if the question is interesting and the results of the research are interesting, even our most selective, premier journals should make room for such work. It is noteworthy that the American Psychological Association (1996) Task Force on Statistical Inference has made a similar recommendation (see <http://www.apa.org/science/tfsi.html>).

This is not a call for atheoretical research or an endorsement of dust-bowl empiricism. As stressed in the previous section, clear and cogent a priori theory can be tremendously valuable, and we should always be pushing authors (and ourselves) to develop more and better theory in our work. Additionally, as noted earlier, making explicit even our most primitive theoretical notions has real value. But we gain little by insisting that authors pretend to have good, a priori theory when they, in fact, do not. When such insistence promotes HARKing, we actually stand to lose.

Destigmatize candid identification of hypotheses as post hoc. We should be evaluating alternative theories and their hypotheses by their scientific utility (i.e., the degree to which they help us describe, understand, predict, and control phenomena of interest). A research report that introduces a theory that clearly has (or well could have) such scientific utility makes a valuable contribution to knowledge. This is true whether that theory guided the design of reported empirical research or was even known to the investigator

¹¹This provides a counterargument to Wallach and Wallach's (1994) article in which they suggested that the findings of many (if not most) social psychological articles are "nearly tautological" because they can be deduced from premises that are almost certainly true. They show that one can, post hoc, construct such deductions for many articles from leading journals. I maintain that a motivated scholar could have constructed equally persuasive deductions for findings that were exactly opposite to those observed in these (or in as many such) articles.

¹²This assumes that shared and applicable minimum standards for evaluation of theory quality can be developed. But even if such standards are imprecise and idiosyncratic, editors, reviewers, and readers can and should still apply those standards when assessing the informativeness and interest value of an apparent confirmation (and potential instance of HARKing).

prior to designing and executing the research. In such cases, authors ought not to be punished (e.g., court immediate rejection and reflexive calls for replication) for frankly acknowledging when those theoretical ideas are post hoc.¹³

Again, this in no way minimizes the desirability of having theoretical ideas in hand a priori to research design. All else being equal, designs guided by a theory are more likely to yield results informative for that theory than designs not so guided. And, as I argued in an earlier section, focusing on a sufficing hypothesis prior to seeing the data may have undesirable effects (e.g., undue emphasis and attention on the immediate set of findings and corresponding neglect of relevant prior theory and research; suboptimal identification of alternative hypotheses). On both counts, post hoc theorizing is vulnerable to more risks than a priori theorizing. But such considerations ought not to lead us automatically to discount the interest value or scientific utility of an idea simply because it is post hoc.

Concluding Thoughts

By now it will come as no surprise to the reader that my assessment of the cost–benefit ratio of HARKing is greater than 1.0. For the various reasons presented here, I have convinced myself that HARKing adds nothing to a hypothesis' scientific utility. A genuinely post hoc explanation is no more valuable for having been falsely presented as an a priori one and may even be less valuable (e.g., if it inhibits the search for alternatives or overly narrows our attention). It is an illusion to believe that we gain something of value by insisting that authors pretend to have “known it all along.”

The issues raised by HARKing are complex. I would not be at all surprised if those who disagree with me about the cost–benefit ratio can come up with more and better arguments for their position than those few I have touched on here. If this article helps encourage discussion, debate, and research on HARKing, it will have fulfilled its purpose.

References

- American Psychological Association. (December, 1996). Task force on statistical inference initial report (draft) [posted on the World Wide Web]. Washington, DC: Author. Retrieved January 15, 1998 from the World Wide Web: <http://www.apa.org/scence/tfsi.html>
- ¹³Mindful of the advantages (for efficient communication; for narrative appeal) of introducing readers to the central ideas of a research report as early as possible, this might well take the form of describing post hoc hypotheses in the introduction to one's article along with a brief but explicit acknowledgment of the role that the soon-to-be-reported results played in the identification of those hypotheses.
- Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology, 66*, 305–307.
- Arkes, H. R., Wortmann, R. L., Saville, P. D., & Harkness, A. R. (1981). Hindsight bias among physicians weighing the likelihood of diagnosis. *Journal of Applied Psychology, 66*, 252–254.
- Baron, J. (1988). *Thinking and deciding*. Cambridge, England: Cambridge University Press.
- Baron, R., & Kenny, D. (1986). The moderator–mediator distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.
- Baumeister, R. (1993). Revising the ideal. *Dialogue: Newsletter of the Society for Personality and Social Psychology, 8*, 10–11.
- Bem, D. J. (1987). Writing the empirical journal article. In M. Zanna & J. Darley (Eds.), *The compleat academic: A practical guide for the beginning social scientist* (pp. 171–201). New York: Random House.
- Bem, D. J. (1991). Writing the research report. In C. Judd, E. Smith, & L. Kidder (Eds.), *Research methods in social relations* (6th ed., pp. 453–476). Fort Worth, TX: Holt.
- Brinberg, D., Lynch, J. G., & Sawyer, A. G. (1992). Hypothesized and confounded explanations in theory tests: A Bayesian analysis. *Journal of Consumer Research, 19*, 139–154.
- Canter, M., Bennett, B., Jones, S., & Nagy, T. (1994). *Ethics for psychologists: Commentary on the APA Ethics Code*. Washington, DC: American Psychological Association.
- Cliff, N. (1989). Strong inference and weak data: Covariance structure analysis and its use. In J. A. Keats, R. Taft, R. Heath, & S. Lovibond (Eds.), *Mathematical and theoretical systems* (pp. 69–77). Amsterdam: North Holland.
- Committee on Science, Engineering, and Public Policy. (1992). *Responsible science: Ensuring the integrity of the research process*. Washington, DC: National Academy Press.
- Dane, F. C. (1990). *Research methods*. Pacific Grove, CA: Brooks/Cole.
- Davies, M. (1987). Reduction of hindsight bias by restoration of foresight perspective: Effectiveness of foresight-encoding and hindsight-retrieval strategies. *Organizational Behavior and Human Decision Processes, 40*, 50–68.
- Davis, M. S. (1971). That's interesting: Toward a phenomenology of sociology and a sociology of phenomenology. *Philosophy of Social Science, 1*, 309–344.
- Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology, 31*, 169–193.
- Doyle, A. C. (1953). *The Complete Sherlock Holmes* (Vol. II). New York: Doubleday.
- Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Feyerabend, P. K. (1993) *Against method* (3rd ed.) London: Verso.
- Fischhoff, B. (1975). Hindsight foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 288–299.
- Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 349–358.
- Gettys, C. F., & Fisher, S. D. (1979). Hypothesis plausibility and hypothesis generation. *Organizational Behavior and Human Performance, 24*, 93–110.
- Grant, D. A. (1962). Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychological Review, 69*, 54–61.
- Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin, 82*, 1–20.

- Greenwald, A. G., Pratkanis, A. R., Leippe, M., & Baumgardner, M. R. (1986). Under what conditions does theory obstruct research progress? *Psychological Review*, *93*, 216–229.
- Hasher, L., Attig, M., & Alba, J. (1981). I knew it all along: Or, did I? *Journal of Verbal Learning and Verbal Behavior*, *20*, 86–96.
- Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, *107*, 311–327.
- Hempel, C. G. (1966). *Philosophy of natural science*. Englewood Cliffs, NJ: Prentice Hall.
- Herschel, J. F. W. (1987). *A preliminary discourse on the study of natural philosophy*. Chicago: University of Chicago Press.
- Hoch, S., & Loewenstein, G. (1989). Outcome feedback: Hindsight and information. *Organizational Behavior and Human Decision Processes*, *15*, 605–619.
- Horwich, P. (1982). *Probability and evidence*. Cambridge, England: Cambridge University Press.
- Howson, C. (1982). Bayesianism and support by novel facts. *British Journal for the Philosophy of Science*, *35*, 245–251.
- Hull, D. A. (1988). *Science as a process*. Chicago: University of Chicago Press.
- Jones, J. H. (1993). *Bad blood: The Tuskegee syphilis experiment*. New York: Free Press.
- Judd, C., Smith, E., & Kidder, L. (1991). *Research methods in social relations* (6th ed.). Fort Worth, TX: Holt.
- Kerr, N. L., & Harris, S. E. (1998). *HARKing—hypothesizing after the results are known: Views from three disciplines*. Unpublished manuscript, Michigan State University, East Lansing.
- Kerr, N. L., Garst, J., Harris, S. E., & Sheppard, L. (1998). *The effect of having a sufficient hypothesis on generation of alternative hypotheses*. Unpublished manuscript, Michigan State University, East Lansing.
- Klayman, J. (1995). Varieties of confirmation bias. In D. L. Medin, J. R. Busemeyer, & R. Hastie (Eds.), *The psychology of learning and motivation: Decision making from the perspective of cognitive psychology*. New York: Academic.
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 107–118.
- LaFollette, M. C. (1992). *Stealing into print*. Berkeley: University of California Press.
- Lakatos, I. (1978). *The methodology of scientific research programmes*. Cambridge, England: Cambridge University Press.
- MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*, 490–504.
- Mahoney, M. J. (1976). *Scientist as subject: The psychological imperative*. Cambridge, MA: Ballinger.
- McGuire, W. J. (1983a). A contextualist theory of knowledge: Its implications for innovation and reform in psychological research. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 16, pp. 1–47). New York: Academic.
- McGuire, W. J. (1983b). Toward social psychology's second century. In S. Koch & D. E. Leary (Eds.), *A century of psychology as science* (pp. 558–590). New York: McGraw-Hill.
- Merton, R. K. (1968). *Social theory and social structure*. New York: Free Press.
- Messick, D. M., & Brewer, M. B. (1983). Solving social dilemmas: A review. In L. Wheeler & P. Shaver (Eds.), *Annual review of personality and social psychology* (Vol. 3, pp. 11–44). Beverly Hills, CA: Sage.
- Neuliep, J. W., & Crandall, R. (1990). Editorial bias against replication research. *Journal of Social Behavior and Personality*, *5*, 85–90.
- Neuliep, J. W., & Crandall, R. (1993). Reviewer bias against replication research. *Journal of Social Behavior and Personality*, *8*, 21–29.
- Pennington, N., & Hastie, R. (1993). The story model for juror decision making. In R. Hastie (Ed.), *Inside the juror* (pp. 192–221). New York: Cambridge University Press.
- Platt, J. R. (1964). Strong inference. *Science*, *146*, 347–353.
- Popper, K. (1959). *The logic of scientific discovery*. New York: Basic Books.
- Richardson, J. T. E. (1996). *Handbook of qualitative research methods for psychology and the social sciences*. Leicester, England: British Psychological Association.
- Schank, R. C., & Ableson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Seligman, M. E. P. (1991). *Learned optimism*. New York: Knopf.
- Shaklee, H., & Fischhoff, B. (1982). Strategies of information search in causal analysis. *Memory and Cognition*, *10*, 520–530.
- Simon, H. A. (1955). Prediction and hindsight as confirmatory science. *Philosophy of Science*, *22*, 227–230.
- Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 544–551.
- Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search*. New York: Springer-Verlag.
- Sternberg, R. J. (1992). How to win acceptances by psychology journals: 21 tips for better writing. *The APS Observer*, *5*, 12–13, 18.
- Sternthal, B., Tybout, A. M., & Calder, B. J. (1987). Confirmatory versus comparative approaches to judging theory tests. *Journal of Consumer Research*, *14*, 114–125.
- Tesser, A. (Producer). (1994). *On becoming a social psychologist* [Video]. Athens: University of Georgia.
- Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*, 612–630.
- Tufte, E., & Sun, R. (1975). Are there bellwether electoral districts? *Public Opinion Quarterly*, *39*, 1–18.
- van Andel, P. (1998). *An anatomy of serendipity*. Unpublished manuscript, University of Groningen, Netherlands.
- Wallach, L., & Wallach, M. (1994). Gergen versus the mainstream: Are hypotheses in social psychology subject to empirical test? *Journal of Personality and Social Psychology*, *67*, 233–242.
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129–140.
- Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology* (Vol. 1, pp. 135–151). Harmondsworth, England: Penguin.
- Wegner, D. M. (1992). The premature demise of the solo experiment. *Personality and Social Psychology Bulletin*, *18*, 504–508.

