

When do stories work? Evidence and illustration in the social sciences¹

Andrew Gelman² and Thomas Basbøll³

22 Jan 2014

Abstract. Storytelling has long been recognized as central to human cognition and communication. Here we explore a more active role of stories in social science research, not merely to illustrate concepts but also to develop new ideas and evaluate hypotheses, for example in deciding that a research method is effective. We see stories as central to engagement with the development and evaluation of theories, and we argue that for a story to be useful in this way, it should be *anomalous* (representing aspects of life that are not well explained by existing models) and *immutable* (with details that are well-enough established that they have the potential to indicate problems with a new model). We develop these ideas through considering two well-known examples from the work of Karl Weick and Robert Axelrod, and we discuss why transparent sourcing (in the case of Axelrod) makes a story a more effective research tool, whereas improper sourcing (in the case of Weick) interferes with the key useful roles of stories in the scientific process.

I. Introduction

Stories have increasingly been recognized as important in social science. Cognitive psychologists have suggested that we understand the world in terms of storylike causal relations (Sloman, 2005), computer scientists have suggested that stories should be used in modeling human cognition (Schank and Abelson, 1977), and management researchers have identified storytelling as a key element in what makes ideas memorable (Heath and Heath, 2007). There has been much discussion in history and sociology on methods for integration of narrative into social science theory (see, for example, Abbott, 1983, 1995). The purpose of the present paper is not to review this literature but rather specifically to consider the ways in which stories inform the perspective of researchers in social science and statistical methodology. We hope our parochial view can yield some useful general insights, in the sense that every researcher is in some sense a methodologist. One always has some need to develop a unique approach and unique tactics when seriously studying any applied problem in social science.

Stories are sometimes used merely to illustrate or explain a model that has already been rigorously demonstrated, but stories can also play a more active role in the development and evaluation of hypotheses. Stories are not used merely as tools in communication but as steps in social science research. Strangely, it is sometimes argued that for developing social understanding, “any old story will do” (Czarniawska 2005): that is, the act of storytelling is more important than the story’s content, not

¹ To appear in *Sociological Methods and Research*. We thank Stephen Stigler, Susan Gelman, Howard Wainer, Matt Salganik, Joanne Gowa, Reinhold Martin, Chris Winship, Peter Bearman, and two anonymous reviewers for helpful comments.

² Department of Statistics and Department of Political Science, Columbia University, New York.

³ Department of Management, Politics and Philosophy, Copenhagen Business School.

only in practical situations, but for theory development. Sometimes scholars are even encouraged to “affirm” rather than to critique stories that circulate in social life (Weick 1982).

In this paper, we argue that “any old story” *won't* actually do. Rather, certain aspects of stories make them work as developing and evaluating research ideas, and we suggest two criteria for assessing the epistemic value of storytelling in social science. First, stories should be *anomalous*—that is, representing aspects of life that are not well explained by existing models—and, second, stories should be *immutable*: their details and contexts should be well-enough established that they have the potential to indicate problems with a new model. Our claims are both descriptive and normative, in that we argue that good stories have these desirable properties.

To make our argument, we examine two prominent examples of storytelling in the social sciences. First, we consider organizational theorist Karl Weick’s use of a story about a group of soldiers in the Alps who use a map of the Pyrenees to find their way back to camp. As Basbøll and Graham (2006) discovered, Weick’s story was transcribed from an uncredited source and subtly distorted in retellings.

Our concern here is not whether the story was believed by Weick’s audiences to be a true historical account, but rather the way in which the flexibility of the story—its ability in different versions to imply various contradictory messages for organizational behavior—could arise from its unsourced nature. Had the story been more clearly cited, we expect it would have been just as effective rhetorically, but it could not have been so easily used to make different, even contradictory, points in different tellings. Being unsourced gave the story a flexibility that made it convenient for the conveying of parables of management theory but much less valuable for the purpose of gaining insight into organizational behavior.

Second, we discuss political scientist Robert Axelrod’s highly influential game-theoretic model of cooperation in the trenches during the First World War. As Gowa (1986) showed, Axelrod’s analysis can be disputed using details that can be found both in Axelrod’s careful retelling of the story and his scrupulously referenced sources.

Both our engagements are critical rather than affirmative but they lead to radically different conclusions about the quality of the storytelling involved. This leads us to two general claims. Our positive claim is that, as social scientists and research methodologists, we do learn from stories, and more than in the simple sense that stories grab our attention and stick in our memory. Rather, we see stories as central to engagement with the development and evaluation of theories. Our negative claim is that plagiarism (or, more generally, improper sourcing) interferes with the key useful roles of stories in the scientific process.

When we suggest that stories be anomalous and immutable, this may seem no different than requiring them to be newsworthy or interesting while also requiring them to withstand the scrutiny of scholarly criticism. What is new in the current paper is the placement of these very general ideas of discourse into the framework of scientific evidence. Stories are universally recognized as important modes of communication but are often viewed more as techniques for illustrating existing ideas,

rather than as evidence in themselves. By exploring what we as social scientists learn from stories, we hope to gain insight into how and in what settings a story can aid in the evaluation of empirical models.

II. Background

We are all familiar with the idea that people think in terms of stories, and that stories help us to remember and organize our thoughts. Here we want to go further and consider some aspects of stories that make them useful in research and understanding.

Discussion often comes from the perspective that stories are good for conveying packaged ideas for others, not as research tools in themselves. Hence advice such as: When teaching (or when presenting research), present the ideas as stories. Stories are sticky, people remember stories, people think in terms of stories, and so forth. That's all good advice and true, but in this article we're making a different point, which is that stories are how *we* (researchers, both qualitative and quantitative) form our own conclusions.

We build confidence in new statistical methods not just from mathematical derivations but also from stories of how the methods have solved real-world problems. Similarly, models in economics, political science, and sociology often get their impetus from a stories in which they convincingly explain or reformulate real-world phenomena. Consider Adam Smith's story of the pin factory or Erving Goffman's stories of social interactions. These are not mere illustrations of self-contained theories; rather, they serve to both develop and justify their theoretical frameworks. For a slightly different sort of example, consider the theories of Sigmund Freud, which were formed through clinical experiences, with many of the motivating stories remaining controversial. Stories remain relevant for more recent work as well, on topics ranging from the mortgage crisis to prison policy to social networks.

This motivates the current article, in which we carefully explore what makes stories work, and we don't just mean "work" in terms of selling an idea to others, but also "work" in providing useful information for real-world inference.

There is an analogy here to the use of graphics in statistics. For many years, serious statisticians tended to think of graphs as frills or, at best, as ways of conveying, to students or other audiences, ideas that we, the experts, already understood. Only relatively recently has it been accepted within statistics that graphs help us understand, that graphical methods are statistical methods (see Tukey, 1977, and Gelman, 2004, for two different perspectives on the integration of graphics in statistical practice).

We think stories have a similar status. We (social scientists) have to admit that stories are important, not just as teaching tools, but that they're central to how we decide what we believe. If we take stories more seriously and think about what makes them work, then maybe they can serve use better. Abell (2004) considers stories as informal causal graphs, analogous to the structural models that are often used for causal inference in sociology or evolutionary sequences in biology. In the present

paper we consider a different perspective in which a story does not represent a model; rather, it is a form of data, a set of facts which can be used to refute existing models or suggest new ones. The more complex the story, the more useful it can be in model criticism; hence we see a connection between the qualitative idea of “thick description” and the contribution of stories to more formal model development.

III. When can stories be useful in the development of social science?

The point of this paper to understand the role of stories in social-science investigation. We claim that the most useful stories represent anomalies that cannot be handled by previously-existing theories, ideally examples with enough internal complexity that a serious engagement with their details can shed light on general ideas.

It would be tempting to claim that stories are a mere sugarcoating applied on top of serious work, just there to make our papers and books more readable. But no. Given our own experiences and much work by others that we respect, we believe that stories can serve an important function beyond mere motivation and illustration. As noted earlier, the key to our argument is that good stories tend to have two properties: they are *anomalous* and they are *immutable*.

Good stories are anomalous. As the saying goes, “Dog bites man is not news. Man bites dog is news.” When we use atypical stories to develop general understanding, this presents both a risk and an opportunity. The risk is obvious: By definition, atypical events do not capture most of life; thus we must be careful not to think that a strategy that “works” in an interesting setting will necessarily apply to mundane everyday reality. The opportunity arises because atypical stories can be those that are not easily explained by existing theories. In that sense, such stories play the role of the experimental anomalies that have such an important role in the philosophies of science of Popper, Kuhn, and Lakatos. Just as the progress of science is stimulated by unexplained phenomena, social science can move forward through serious engagement with puzzling sequences of events. This is one reason we believe that stories are central to so many important works of social science.⁴ As Bearman and Stovel (2000) put it, “people construct stories to account for non-canonical events that cannot otherwise be accounted for by culturally agreed upon narrative expectations.”

Good stories are immutable. Much can be learned from a true anecdote. The rough edges—the places where the anecdote does not fit your thesis—these are where you learn. There is a saying in statistics that God is in every leaf of every tree. What this means is that if you study any problem carefully and seriously enough, you will come to interesting open research problems. This is related to the concept of “thick description” in anthropology. Details matter; this is the difference between character and caricature. In statistics, this relates to the idea that assumptions can be checked

⁴ As discussed earlier, stories are believed to be central to human reasoning more generally, and in ways similar to that discussed in the present article. For example, in a study of children’s cognition, Legare, Gelman, and Wellman (2010) report that anomalous events stimulate theory-building: “inconsistent outcomes are an especially powerful trigger for children’s explanations and that the explanations children provide for inconsistent outcomes refer to internal causal properties, overriding perceptual appearances.”

by comparing data to simulated replications from the fitted model (Gelman and Shalizi, 2012). Whether the comparison is qualitative or quantitative, the point remains: the story or data used as a comparison must not be so pliable that it can be interpreted as consistent with any model (this was Karl Popper's famous criticism of the Freudian and Marxian paradigms).

Bad stories can do damage by giving a theorist a false sense of security. For the purposes of our discussion here, bad stories are those that masquerade as good stories but are not anomalous and immutable. A bad story can purport to demonstrate an anomaly even while actually being easily explained using existing theories, or a bad story can be contrived or manipulated so as to lose its ability to reject a model.

It has been through some recent engagement with controversial social-science stories that we have been motivated to think harder about the benefits of good stories, to attempt to move beyond the simple idea of stories-as-decoration and develop a theory of the value of stories in research.⁵

Consider the following anecdote (Gelman and Roell, 2011). The first author of this article was at a panel evaluating grant applications for a government-funded research program. One of the proposals had to do with the study of the effect of water-pipe smoking, the hookah. There was a discussion around the table. Not many people in the United States really smoke hookahs; so should we be funding this? Someone said that the hookah is becoming more popular among the young. And if younger people smoke it, they have a longer lifetime exposure, and apparently there is some evidence that the dose you get of carcinogens from hookah smoking might be something like twenty times the dose of smoking a cigarette. So even if not many people do it, if you multiply the risk, you get a lot of lung cancer.

Then someone at the table—and we could hardly believe this—said, “My uncle smoked a hookah pipe all his life, and he lived until he was 90 years old.” And we had a sudden flash of insight, which was this. Suppose you have something that actually kills half the people. Even if you're a heavy smoker, your chance of dying of lung cancer is not 50%, so therefore, even with something as extreme as smoking and lung cancer, you still have lots of cases where people don't die of the disease. The evidence is certainly all around you pointing in the wrong direction—if you're willing to accept anecdotal evidence—there's always going to be an unlimited amount of evidence that won't tell you anything.

The above anecdote actually happened. Would it work as well if it were an unsourced story or a joke? We think not. That the “my uncle lived to be 90” argument was used in a meeting to determine funding for medical research—this demonstrated to us the persistence of the anecdotal-argument fallacy, and it motivated us to think harder about how such arguments arise.

⁵ The importance of anomalous events in stories has been noted in other contexts as well. For example, writing about the role of stories in inspiring social activism, Ganz (2009) writes, “All stories have three parts: a plot, a protagonist, and a moral. What makes a plot a plot? What gets you interested? Tension. An anomaly. The unexpected.” This “tension” is similar to the “puzzles” that motivate academic social science, as well as the counterexamples that spur the development or replacement of scientific theories, as described by Popper, Kuhn, and Lakatos.

IV. Stories as central to learning and theorizing about the world

In the standard paradigm of research methodology, ideas are demonstrated or proved using more-or-less rigorous scientific reasoning, and then stories are added to make the message more palatable to the intended audiences. A statistics book, for example, will come from some coherent philosophy (hypothesis testing, or Bayesian inference, or some other approach) and provide a series of methods illustrated with appropriate examples (biology examples for the statistics-in-biology course, education and psychology examples for general undergraduates, business examples for business students, and so on). The examples do not mean anything on their own, and it is not considered particularly important to get the details right. To argue about survey nonresponse, for example, in a simple illustration of binomial proportions would make about as much sense as asking Einstein whether the conductor would get sick on that hypothetical train traveling at half the speed of light. Similarly, it should not matter whether the sentences and paragraphs in a grammar instruction book are by William Shakespeare or Jane Austen or J. K. Rowling; the latter might be preferred (at the current time) on grounds of accessibility, but the content of the passages would have little bearing on the material being taught.

This paradigm does not, however, fit our own teaching—where we find it helpful if not essential to engage with our examples in detail—nor our research. A key reason the first author of this paper relies on Bayesian inference (and on particular model classes within that framework) is that this approach seems to have been effective in his and others' experiences in applied examples. The hundreds of problems on which an applied statistician works throughout his or her career serve as a continuously-developing nonrandom sample of test cases. Different statisticians work on different sets of problems, and the many useful contributions of a good statistical consultant or collaborator will often be attributed to the statistician's methods or philosophy rather than to the artful efforts of the statistician himself or herself (Gelman, 2010). Similarly, any general principles of writing are ultimately derived from important special cases. The second author of this paper is a professional writing consultant and draws much of his worldview from his personal experiences as reader and writer.

The present article got written though a serendipitous chain of events. As a researcher in organization theory and management, Basbøll had become aware of instances of plagiarism by Karl Weick, a well-known scholar in the field who in several articles (beginning with Swieringa and Weick, 1982) presented, as fact, a story of doubtful veracity that had been published by someone else several years earlier. Over the years, Weick has used the anecdote as a basis for social-science theorizing and for practical recommendations to businessmen. (See Basbøll, 2012a, 2012b, for further background.)

In reflecting on this chain of events, we felt that Weick's obscuring of the anecdote's source served to give the copied story a higher standing than it otherwise deserved. Through various informal discussions (see, for example, Felin, 2006, and Gelman, 2012) and analogies to formal statistical inference (Gelman and Basbøll, 2013), we became interested in the meta-argument of whether the plagiarism mattered to the larger questions being studied by Weick. In the words of Lizardo (2006), why care about an "obscure example from an equally obscure poem"?

We felt that the hiding of the story's source did matter for Weick's arguments, not because it was copied from a poem (obscure or otherwise) and not merely on the grounds of morality or trust (the reason why, for example, we might be less likely to believe a dean's report on a university budget issue if we were to learn that he had cheated on his taxes), but because the improper sourcing stripped the story of much of its value. In particular, the unsourced story was *no longer anomalous* (something can only be an anomaly with respect to its context, something that was removed when the source was left out) and it was *no longer immutable*: once the sourcing was removed, the story was open to arbitrary reinterpretation. As discussed in section 1 of the present article, a good story represents a problem with existing models and an opportunity, in its stubborn reality, to rip holes in new theories. In contrast, Weick's unsourced story was all too able to be adapted to whatever message he wanted to send. This is not to deny the importance of the story in contributing to Weick's thought processes, but the obscuring of the source material limited the story's use in impersonal (scientific) inquiry.

To us, the series of events was anomalous: At first we wondered why anyone would defend or deny an obvious breach of scholarly ethics; then we asked ourselves why *we* should be so sure the inadequate sourcing tainted Weick's conclusions. And the story was immutable: The more we looked into it and engaged with its details, the more we were forced to develop our own ideas.

An irritated (we presume) Karl Weick might feel that we are milking the story of his small mistake for all we can get, but we actually think that pushing this case hard, examining our own thoughts as well as the actions of others, can give us useful insight into the process of science. Indeed, Weick himself has been entirely unrepentant: "Other than to insert a footnote saying 'source unknown', I would not have done anything different were I in the same position today" (Weick, 2006).

V. Stories as evidence

During the First World War, a medical student named Albert Szent-Gyorgyi served with the Hungarian army in the Alps. He quickly decided that an extended military career would be the death of him and, certain that he had a larger contribution to make to humanity, shot himself in the arm in order to be discharged from active duty. His sense of his potential proved accurate, and by 1937 he had won the Nobel Prize for his discovery of vitamin C (Moss, 1988). But his military experience stayed with him and he was known to entertain his fellow scientists with a story that circulated among the troops when he was stationed in the mountains (Basbøll 2012b). It is such a good story that the Czech immunologist and poet Miroslav Holub turned it into a poem that was published in the Times Literary Supplement in 1977:

"Brief Thoughts on Maps"

Albert Szent-Gyorgyi, who knew a lot about maps
according to which life is on its way somewhere or other,
told us this story from the war
due to which history is on its way somewhere or other:

The young lieutenant of a small Hungarian detachment in the Alps
sent a reconnaissance unit out into the icy wasteland.
It began to snow
immediately, snowed for two days and the unit
did not return. The lieutenant suffered: he had dispatched
his own people to death.

But the third day the unit came back.
Where had they been? How had they made their way?
Yes, they said, we considered ourselves
lost and waited for the end. And then one of us
found a map in his pocket. That calmed us down.
We pitched camp, lasted out the snowstorm and then with the map
we discovered our bearings.
And here we are.

The lieutenant borrowed this remarkable map
and had a good look at it. It was not a map of the Alps
but of the Pyrenees.

Goodbye now.

Here is how Weick told the story in an influential 1987 paper:

Definitions notwithstanding, I can best show what I think strategy is by describing an incident that happened during military maneuvers in Switzerland. The young lieutenant of a small Hungarian detachment in the Alps sent a reconnaissance unit out into the icy wilderness. It began to snow immediately, snowed for two days, and the unit did not return. The lieutenant suffered, fearing that he had dispatched his own people to death. But the third day the unit came back. Where had they been? How had they made their way? Yes, they said, we considered ourselves lost and waited for the end. And then one of us found a map in his pocket. That calmed us down. We pitched camp, lasted out the snowstorm, and then with the map we discovered our bearings. And here we are. The lieutenant borrowed this remarkable map and had a good look at it. He discovered to his astonishment that it was not a map of the Alps but of the Pyrenees. (Weick 1987: 223)

Crucially, he provides no reference to Holub's poem and uses no quotation marks, even though it is, of course, essentially a verbatim transcription of the poem with the line breaks removed. That is, Weick obscures the relationship between the story he is telling and his source. Perhaps more importantly, however, he subtly distorts the story, both in its content, by situating it in Switzerland (which is unlikely to be true), and in its historicity. What Holub describes as a "story from the war" becomes "an incident that happened while on military maneuvers," which suggests that the events have been well-documented by historians.

For his part, Szent-Gyorgyi told his story to remind his peers that in science even errors can lead to progress. He did not offer it as a "true story," and certainly not as

advice to mountaineers or managers. Another, less famous, story should suffice to tell us why.

It begins like Szent-Gyorgyi's story.⁶ The young lieutenant of a small Hungarian detachment in the Alps sent a reconnaissance unit out into the icy wasteland. It began to snow immediately, snowed for two days and the unit did not return. Here, too, the lieutenant suffered, and this time with good reason; he had, in fact, dispatched his own people to death. A week later, the unit was found at the bottom of a ravine, having fallen off the edge a precipice. In the log book, the final entry read: "We considered ourselves lost and waited for the end. And then one of us found a map in his pocket. That calmed us down. We pitched camp, lasted out the snowstorm and then with the map we tried to get our bearings. And here we are." When the search team returned, the lieutenant asked to have a look at the map they had recovered from the dead soldiers. It was not a map of the Alps but of the Pyrenees.

A third story goes as follows. The young lieutenant of a small Hungarian detachment in the Alps sent a reconnaissance unit out into the icy wasteland. It began to snow immediately, snowed for two days and the unit did not return. Here, too, the lieutenant suffered and here, too, he had good reason to do so; he had also dispatched his own people to death. They had considered themselves lost and were waiting for the end, when one of them found a map in his pocket. That calmed them down. They pitched camp, lasted out the snowstorm and then tried to get their bearings using the map. Unfortunately, it was not a map of the Alps but of the Pyrenees, and the unit was never seen again.

These three stories, taken together, tell us something about the validity of anecdotal evidence. If, like some organization theorists and management consultants, you take the moral of the first story to be that "any old map will do" (and, by extension, any old plan), then you are reading it as representative of the experience of following a map in mountainous terrain, which means that you believe that the second and third stories relate exceptional experiences, deviations from the norm. But this reading is often not made explicit, even by the theorists or consultants to themselves. They simply find the story "compelling."

You would never hear the third story: all the witnesses are dead. And, while the second story will have witnesses, it is less likely to have audiences. It tells the mundane story of following a false map to the wrong place. This makes it much more likely that the first story, despite its counterintuitive moral, will be told and believed, than a story that urges caution in mountainous terrain and care in the use of maps and compasses. In order to determine which story we should use to represent reality, then, we should take this selection bias into account. The mere fact that soldiers of the first world war such as Szent-Gyorgyi were telling this story, does not make it an accurate account of what they believed about using maps. In fact, the unlikeliness of the story is part of its appeal.

⁶ The two stories that follow are, of course, entirely fictional. The veracity of Szent-Gyorgyi's story has never been satisfactorily tested, but there is good evidence to suggest that it did, in fact, circulate among scientists in the mid-1970s. Our variations on the theme, however, are entirely made up. We are grateful to David Huelsbec, who suggested a similar variation on Gelman's blog as an illustration of the point about selection bias that we are trying to make here.

Perhaps this can be better seen in a puzzling story that Weick also tells management theorists, this time in an interview in the *Harvard Business Review*.

One organization that has struggled with reliability is Union Pacific. Back in the 1990s, the company suffered repeatedly from managerial paralysis—even the employees began to call it the Utterly Pathetic railroad. At that time, the following story started circulating among employees and customers: A locomotive engineer got so fed up with the railroad's incompetence that he decided to commit suicide. So he went outside, lay down on the railroad tracks—and starved to death. That kind of urban myth was a perfect way to express just how frustrated people had become with the railroad not doing anything during a period of intense upheaval. (Weick as quoted by Coutu, 2003)

Weick describes this story as an “urban myth” but it seems clearly to be a joke. It is not at all surprising that a company that is in trouble will have jokes circulating among the employees about its management. Indeed, it is common for companies, no matter how well they are doing, to have jokes circulating about the bosses. This hardly tells us “just how frustrated people had become,” whereas a story that might conceivably count as a rumor (i.e., a story that people tell in order to be believed) would be much more representative. If he here slides from what is really a joke to something he presents as a “myth,” in the Alps case he slides from what is clearly a myth (a “story from the war”) to a historical event (“an incident that happened while on military maneuvers”).

One could construct a rough hierarchy of stories along the following lines:

- a. Joke, or fiction such as Harry Potter,
- b. Urban legend,
- c. Dubious story (the lost soldiers in the Alps),
- d. Plausible but undocumented story (e.g., Iraq’s weapons of mass destruction),
- e. Documented anecdote.

The further you go along this scale, the more it would seem reasonable to take the story as evidence for a theory of social behavior. But even the culmination—stories that are universally recognized to be true—do not directly provide evidence unless we understand the process by which the story is selected. As we say, this relates to the well-known statistical problem of selection bias: a statistical model is complete only when it describes the data-selection process as well as the data themselves.

VI. Cooperation in trench warfare

In 1984, political scientist Robert Axelrod published *The Evolution of Cooperation*, an extremely well-received book⁷ that combined game theory, computer experiments,

⁷ The back cover of the paperback features rave reviews from decision theorist R. Duncan Luce, computer scientist Douglas Hofstadter, medical essayist Lewis Thomas, and others, in publications including the *New York Times* book review, *Times Literary Supplement*, *Wall Street Journal*, and *Journal of Policy Analysis and Management*.

and historical reasoning to argue that cooperation can and will develop spontaneously even in environments that would seem highly unpropitious.

Here is some brief background (taken from Gelman, 2008). From a psychological or economic point of view, why do people cooperate with each other (instead of acting purely selfishly, which would give them short-term benefits, at the very least)? A historian might study cooperation as it has existed in past societies, and a sociologist might consider the settings in which individuals follow cooperative and non-cooperative roles. From a political-science perspective, the natural question is how to *promote* cooperation—this is behavior that is essential to the functioning of any political system, to avoid a Hobbesian war of all against all. From a game-theoretic standpoint, motivating cooperation has always been viewed as a challenge, and this puzzle has been given various names, such as the free-rider problem, the tragedy of the commons, and the prisoner's dilemma (a term due to Albert Tucker; see Maurer, 2003). In these settings, cooperation is advantageous in the long term, but, in the short term, individuals unambiguously benefit from not cooperating.

In situations where cooperation is important (for example, in a business) or even a matter of life and death (for example, in the military), it is considered crucial to set up a team spirit. However, in other settings, most notably in the economic sphere (recall the overfishing example), the incentives to not cooperate are so strong that psychological motivation does not seem enough. Cooperation can then be enforced through governmental action or private binding agreements (which themselves typically require governmental presence to be enforceable).

Axelrod's interest was slightly different, however. Rather than study settings where cooperation is automatic, or where cooperation needed outside enforcement, he was interested in intermediate scenarios in which cooperative behavior was risky and unstable but developed anyway. This seems to describe much of human interactions—when the rules break down, people can act brutally, but stable societies are greased by a layer of trust.

Axelrod's most striking example comes from the First World War, this time on the Western rather than the Italian front. British and German soldiers were facing each other in a fight to the death, yet in many situations the two sides avoided fighting, with various informal truces (most notably at Christmas, 1914) and more elaborate arrangements with soldiers shooting to miss on purpose (and at the same time demonstrating their ability to do harm by aiming at precise targets). Throughout the war, the commanders had to develop different strategies—for example, rotating troops more quickly in and out of the front lines and later, sending soldiers on raids of the opposing trenches—to stop them from getting friendly with the enemy.

The first author of the present paper encountered this example at the suggestion of his undergraduate thesis advisor, Hayward Alker, who suggested looking into the trench-warfare story a bit more closely to see how it matched with Axelrod's framing of it. We began by reading Ashworth (1980), the source for the information about cooperation in the trenches. Axelrod's argument built upon details of the First World War experience, and to study his argument we first immersed ourselves in his source.

Our first characteristic of a good story is that it is *anomalous*, and, indeed, Ashworth's story of cooperation in trench warfare is unexpected and does not seem well explained by existing theories. Theories of selfish behavior would seem to predict noncooperation—why shoot to miss if you can kill one of your enemies?—but simple theories of cooperation would, conversely, not explain the violence that was the norm in the war. (The preceding argument is necessarily sketchy. We are not attempting to go through all of Axelrod's reasoning; we are just showing how it fits into our theory of stories.)

Our second characteristic of a good story is that it is *immutable*. And, indeed, after a careful re-examination of the history of the "live and let live system" of the trenches, we came to the conclusion that there was more to the story than Axelrod's theory allowed. To continue from Gelman (2008)—itself an elaboration of an undergraduate thesis from twenty-two years earlier—the pattern of soldiers who do not want to fight and commanders who force them to do so—has been reported throughout history. In the Second World War, a famous study by Colonel S. L. A. Marshall estimated that only one-quarter of the U.S. soldiers in a position to fire their rifles actually did so.⁸ This behavior has been attributed to fear and a sense of isolation, as well as simple self-preservation, since firing your gun can make you a target.

Now we return to Axelrod's argument, which is an attempt to explain theoretically the cooperation described by Ashworth. Given the immediate risks of cooperation, how did the soldiers so many times develop cooperative social structures without the possibility of binding agreements? Axelrod's model was that the soldiers on each side (British or German) would have an immediate gain from shooting: if you shoot, you can take out a potential enemy, and that holds whether or not he shoots. From this perspective, cooperation is the anomaly that Axelrod explains by analogy to the repeated-play prisoner's dilemma. In the context of his book, the trench warfare story both motivates the study of this game-theoretic problem and points to a potential larger solution to a general social problem, reducing conflict by embedding settings of conflict in repeated-play games. Much of the rest of Axelrod's book is devoted to a computer experiment demonstrating the robust persistence of cooperative strategies in repeated prisoner's dilemma games.

This is all fine, but what if it was *not* actually in the immediate interest of soldiers to shoot? Based on our reading of Ashworth, we think the "Western Front game" was much simpler for individual soldiers. A stable equilibrium on both sides was to duck down in the trench, not stick your head out, and not shoot. It was this equilibrium that had to be altered by the commanders on both sides to enforce action.

To the extent our explanation (which is similar to that of Gowa, 1986) is convincing, the immutability of the trench-warfare story implies a weakness of Axelrod's argument, in that he was (in our opinion) giving a complicated rationale for cooperation when a simpler explanation would do. This does not invalidate his book, but it suggests that, when looking at obstacles to cooperation, one might look not just at local incentives that create prisoner's dilemmas but also at outside actors (in this

⁸ Ironically (given the subject of the present article), Marshall's finding has itself been disputed as a possible fabrication; see Spiller (1988).

case, the leaders of the armies on both sides) who benefit from conflict and have the ability to alter the rules of the game.

If you disagree with our analysis of incentives in trench warfare, that is fine too. Our point still holds that the story takes its value from being anomalous, and that its immutability makes it a powerful tool for interrogating theories of cooperation. Axelrod's full sourcing of the story (by citing the then-recent book by Ashworth) allowed this all to happen. Had Axelrod instead, for example, simply referred to "a well-known story from the Western Front," it would have been more difficult for him and later political scientists to think fully about the hidden assumptions involved in applying his game-theoretic model to real-world conflict.

VII. The role of stories in statistical reasoning

One might imagine a statistician criticizing storytellers for selection bias, for choosing the amusing, unexpected, and atypical rather than the run-of-the-mill boring reality that should form the basis for most of our social science. But then how can we also say the opposite, that stories benefit from being anomalous? We reconcile this apparent contradiction by placing stories in a different class of evidence from anecdotal data as usually conceived. The purpose of a story is not to pile on evidence in support of one theory or another but rather to shine a spotlight on an anomaly—a problem with an existing model—and to stand as an immutable object that conveys the complexity of reality.

We have argued that statisticians and quantitative social scientists learn from stories. But this cannot be happening in the usual manner of statistical inference from sampling and experimentation. The stories we tell are not random samples. Rather, stories are chosen for content and message: we sometimes tell stories of unusual and surprising happenings and sometimes tell stories that satisfyingly confirm our expectations. People say "the plural of anecdote is not data" to emphasize that selected stories cannot and should not be taken as representative of the larger world.

Stories do not qualify as statistical data in the usual sense (given that authors have nearly complete freedom in choosing the stories to include and how to present them, the possibilities for bias are essentially unlimited), so what is their proper role in developing social-science understanding?

Stories do not in general allow for statistical inference, but they can be useful for model checking. Return again to the idea that good stories are anomalous. An anomaly can only be defined with reference to a model: the idea is that an observed story can be surprising, compared to expectations. In Bayesian statistics, we use the term "posterior predictive checking" to refer to the procedure of comparing observed data to simulated replications from a fitted probability model. The idea is that any given statistical model will have many different implications, and detailed-enough data will reveal the model's inevitable oversimplifications (in the way that an extreme enlargement of a photograph will look grainy).

For a story to serve in this way as an (informal) posterior predictive check, it does not need to be representative; rather, interesting stories often represent the surprising

cases the represent the limits of our understanding. But it is important for the story to be immutable: if the content of a story can change with each telling, it can all too easily fit any existing theory and thus be useless for model checking.

The other connection of storytelling to posterior predictive checking is in the ultimate goal, which is not to reject an existing model of the world but rather to gain insights that lead to model improvements. Consider the story of cooperation in trench warfare, which was used by Robert Axelrod not just to reject a simple model of zero-sum adversaries in combat but to advance a new theory based on cooperation in the prisoner's dilemma. The stories from the book of Tony Ashworth provided evidence contradicting the implicit zero-sum model and motivating the development of Axelrod's framework. This process is hardly automatic: as we have noted earlier, Joanne Gowa and others have placed the cooperation stories in a more complex multi-player model of conflict within the war. In any case, rejection of an existing model is intended as a step toward a new model and deeper understanding.

VIII. Stories in management science

At an address to the Academy of Management in 1979, Karl Weick, who was then editor of the prestigious *Administrative Science Quarterly*, argued for what he called "affirmation as inquiry" (Weick 1982). He wanted to get social scientists to abandon their tendency to always criticize what practitioners and their fellow social scientists say, and instead to simply believe them. Criticism, he said, is a "wet blanket," and our persistent efforts to "poke holes" in the images we construct of our situations might cause us to lose sight of something fundamental, and perhaps even universal, in them. Siding with Dean Acheson against Richard Neustadt in a discussion of whether presidents should be given "confidence" or "warning" when making critical decisions, he proposed a "climate of affirmation" in which momentous actions are not abandoned out of a defensiveness born of criticism. He cited John Steinbruner's description of leaders as "confidence givers . . . without [whom] surely no government could manage." The shades of the recent Iraq war become still more disconcerting when we consider Weick's 1998 address to the Academy of Management, in which he said that "any old story will do" in our attempts to solve "the puzzles of the human condition" (quoted by Czarniawska 2005: 274).

The examples in this paper illustrate how a story can perhaps too easily be used as an excuse to propound general theories in political and management sciences, and how a knowledge of the details and context of a story can facilitate an active critical examination. We were able to use the well-documented trench warfare story, with its richly detailed background, to interrogate Axelrod's model of cooperation. In contrast, Weick's unsourced Alpine story is too slippery to pin down and thus has supported all too many variations of the "any map will do" theory with no easy entry point for critical engagement. In fact, fellow organization theorist Barbara Czarniawska (2005: 274) argues that such criticism is beside the point, that story-based reasoning should merely be appreciated, not criticized: she says we should "suspend disbelief" and simply "trust" Weick.

Weick's view is not a marginal one in administrative science today. And administrative science, in turn, is an increasingly important shaper of how

practitioners and theorists understand social life. Business management advice is typically supported by plausible theory and illustrated with examples demonstrating the successful implementation of recommended strategies. Weick (2003: 381) and Czarniawska (2005: 277) agree that the scholar's task is to give managers the merely "plausible" stories they want. More generally, books by popularizers of social research such as Malcolm Gladwell are full of stories that are used to both to illustrate theories and suggest their validity. Sometimes such works find their way back into the scientific literature as support for the theories they popularize. For example, in a recent book, organizational psychologist James March cites Gladwell on "the tipping point," and even goes on to argue that there is little difference between fiction, like Anton Chekhov's short stories, and social research, like Karl Weick's analyses of crises (March, 2010: 22, 69).

We have no objection to the use of stories—real or fictional—to dramatize points that have been demonstrated via solid research. Our concern here is with stories that are employed to make the demonstration themselves. We take neither the extreme hard-line position that stories are merely communicational and decorative, nor the anything-goes view that all sources of inspiration are equal. Rather, we recognize that we (and others) use stories all the time to evaluate our ideas, and we feel it is important to take such stories seriously, as it is through their anomalous aspects and their immutability that we learn from them. It does stories no favors to accept them uncritically.

Similar concerns have been raised in historical analysis, where stories have a special role in that they are often the only evidence about certain events. In a discussion of the value of stories as evidence in studies of classical history, Saller (1980) writes, "In the early part of this century authors of German textbooks on historical method warned students against using anecdotes on the grounds that their form was not fixed and their contents fluctuated since the narrators exercised their imaginations to improve stories with each telling." Saller goes on to discuss the risks of trying to learn from stories but also the potential gains that can be achieved if the sources of an anecdote can be tracked down, if common elements appear in multiple tellings of a story, and if an anecdote sheds light on "ideology or beliefs about reality." To be useful in historical research, a story need not be anomalous, but it should be immutable—or else its mutations should themselves be documented and understood.

IX. Discussion

As discussed, the value we see in stories does not rely on them being random or representative samples of the world. That said, the issue of selection bias cannot be ignored. If a story is too weird and too anomalous, we cannot expect it to provide general insight. For example, even a well-documented anecdote of some soldier in the Alps who happened to find his way back to camp with no map but only a blank sheet of paper, would not provide useful evidence for a claim such as, "when you are lost, any old sheet of paper will do."

This last example brings us to a possible criticism of our article, that we are being too literal and missing the larger point. Surely Karl Weick does not *really* believe that any old map will do. When traveling in central London, for example, we can only

assume he would prefer the classic Tube map rather than, say, a plan of the Paris Metro. The real point of his advice was allegorical, that business managers and other decision makers would do better to act, even on a flawed plan, than to sit still. And of course any systematic evaluation of such a claim would be best based on statistical data analysis rather than anecdotes. Conditional on the claim being accepted, why not back it up with a story?

Our response is twofold. First, a careful examination of Holub's telling of Szent-Gyorgyi's anecdote suggests that the soldiers survived not by the strategy of acting under a wrong plan but rather by the despised tactic of waiting until the storm went away. The map (according to the story) did help, but only by giving the soldiers the calmness required to *not* act hastily. The analysis we have just performed illustrates the benefits of immutability.

Our second response to the possible criticism that we are too picky is to point out that Weick referred to the anecdote as an actual event that "happened" (without giving the sourcing that would easily cause the reader to doubt this claim). Had Weick instead simply made up a story to fit his theory, or had he referred to a clearly fictional story (perhaps an episode from a John Le Carré novel or a Bullwinkle cartoon in which the hero succeeds by acting resolutely under what turned out to be false premises), we would have nothing to criticize. We believe that Weick was using his story not merely to *illustrate* his model of reality but to *support* it. Again, we see such support arising from the specificity of the anecdote, its possible veracity and its rough edges that get smoothed away upon unsourced retellings. Unmoored to its original source, the story gets altered by the tellers so that it can be used to make any point that people want to make from it.

We conclude with some comments on political ideology. Storytelling has been championed by a wide range of scholars who would like to escape the confines of rigor. On the academic left, storytelling is sometimes viewed as a humane alternative to the impersonal number-crunching of economists, while the academic right uses stories to connect to worldly business executives who have neither the time nor patience for dry scholasticism. Karl Weick seems to us to express an unstable mix of these attitudes, championing the creative humanism of story-based social reasoning while offering his theories as useful truths for the business world. And indeed he may be correct in both these views: perhaps organization-science data are weak enough that narrative-based intuition is a better guide to practice, and perhaps these intuitions are indeed most valuable for leaders of hierarchical organizations such as big business and the military. Our concern in the present paper is with the way that stories are presented or misrepresented in the social-science process, not so much with any particular conclusions which of course can be valid even if developed in ways that we do not consider methodologically sound.

As students of logical expression and statistical methods, we are interested in how storytelling can move us forward in our social-science understanding. As already noted, stories are here to stay, and we think we and others can make the best use of anecdotes by considering them in all their complexity, going backward to their sources, forward to their political implications, and sideways to interrogate our models. Our own attitudes toward what works and what doesn't in statistics, writing, and teaching, are so strongly formed by our experiences that it would be foolish of us

to deny the centrality of stories in our understanding. But to learn the most from stories, we should take them seriously, just as professionals such as doctors, lawyers, and police officers develop a mixture of practical expertise and general understanding by immersing themselves and caring about the details of each new case.

References

- Abbott, A. (1983). Sequences of social events: Concepts and methods for the analysis of order in social processes. *Historical Methods* 16, 129-147.
- Abbott, A. (1995). Sequence analysis: New methods for old ideas. *Annual Review of Sociology* 21, 93-113.
- Abell, P. (2004). Narrative explanation: An alternative to variable-centered explanation? *Annual Review of Sociology* 30, 287-310.
- Ashworth, T. (1980). *Trench Warfare, 1914-1918: The Live and Let Live System*. New York: Holmes and Meier.
- Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.
- Basbøll, T. (2012a). Legitimate peripheral irritations. *Journal of Organizational Change Management* 25, 220-235.
- Basbøll, T. (2012b). Any old map won't do; Improving the credibility of storytelling in sensemaking scholarship. Copenhagen Business School, WMO Working Paper No. 4. <http://www.cbs.dk/forskning/institutter-centre/institut-organisation/what-makes-organization/working-papers>
- Basbøll, T., and Graham, H. (2006). Substitutes for strategy research. *ephemera* 6, 194-204.
- Bearman, P. S., and Stovel, K. (2000). Becoming a Nazi: A model for narrative networks. *Poetics* 27, 69-90.
- Bruner, J. (1986). *Actual Minds, Possible Worlds*. Cambridge, Mass.: Harvard University Press.
- Bruner, J. (2002). *Making Stories*. New York: Farrar, Straus, and Giroux.
- Coutu, D. L. (2003). Sense and reliability: A conversation with celebrated psychologist Karl E. Weick. *Harvard Business Review*.
- Czarniawska, B. (2005). Karl Weick: Concepts, style and reflection. *Sociological Review* 53, 267-278.
- Felin, T. (2006). Charges of plagiarism in org theory. Orgtheory blog, 22 Jul. <http://orgtheory.wordpress.com/2006/07/22/charges-of-plagiarism-in-org-theory/>

Ganz, M. (2008). Why stories matter: The art and craft of social change. Friends of Justice website, 18 Feb. <http://friendsofjustice.wordpress.com/2009/02/18/marshall-ganz-why-stories-matter/>

Gelman, A. (2004). Exploratory data analysis for complex models (with discussion). *Journal of Computational and Graphical Statistics* 13, 755-787.

Gelman, A. (2008). Methodology as ideology: Some comments on Robert Axelrod's "The Evolution of Cooperation." *QA-Rivista dell'Associazione Rossi-Doria* 2, 167-176.

Gelman, A. (2010). Bayesian statistics then and now. *Statistical Science* 25, 162-165.

Gelman, A. (2011). Going beyond the book: Toward critical reading in statistics teaching. *Teaching Statistics* 34 (3), 82-86.

Gelman, A. (2012). Another day, another plagiarist. Statistical Modeling, Causal Inference, and Social Science blog, 16 Apr. <http://andrewgelman.com/2012/04/another-day-another-plagiarist/>

Gelman, A., and Basbøll, T. (2013). To plagiarize is to throw away data. *American Scientist* 101, 168-171.

Gelman, A., Meng, X. L., and Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica* 6, 733-807.

Gowa, J. (1986). Anarchy, egoism, and third images: The evolution of cooperation and international relations. *International Organization* 40, 167-186.

Heath, C., and Heath, D. (2007). *Made to Stick: Why Some Ideas Survive and Others Die*. New York: Random House.

Kearney, R. (2002). *On Stories*. New York: Routledge.

Legare, C. H., Gelman, S. A., and Wellman, H. M. (2010). Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child Development* 81, 929-944.

Lizardo, O. (2006). Coment on Felin (2006). Orgtheory blog, 22 Jul. <http://orgtheory.wordpress.com/2006/07/22/charges-of-plagiarism-in-org-theory/#comment-1417>

March, J. G. (2010). *The Ambiguities of Experience*. Ithaca, N.Y.: Cornell University Press.

Marshall, S. L. A. (1947). *Men Against Fire*. New York: William Morrow and Co.

- Maurer, S. B., and Tucker, A. W. (1983). An interview with Albert W. Tucker. *Two-Year College Mathematics Journal* 14, 210-224.
- Moss, R. W. (1988). *Free Radical: Albert Szent-Gyorgyi and the Battle Over Vitamin C*. New York: Paragon Books.
- Saller, R. (1980). Anecdotes as historical evidence for the principate. *Greece & Rome*, second series 27, 69-83.
- Schank, R. P., and Abelson, R. P. (1977). *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Sloman, S. (2005). *Causal Models: How People Think About the World and Its Alternatives*. Oxford University Press.
- Spiller, R. J. (1988). S. L. A. Marshall and the ratio of fire. *RUSI Journal* 133, 63-71.
- Swieringa, R., and Weick, K. E. (1982). An assessment of laboratory experiments in accounting. *Journal of Accounting Research* 20 (supplement), 56-101.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.
- Weick, K.E. (1982). Affirmation as Inquiry. *Small Group Research* 13 (4): 441-450.
- Weick, K. E. (1987). Substitutes for strategy. In *The Competitive Challenge: Strategies for Industrial Innovation and Renewal*. D. J. Teece, ed. Cambridge, Mass.: Ballinger. pp. 222-233.
- Weick, K. E. (1995). *Sensemaking in Organizations*. Thousand Oaks, Calif.: Sage.
- Weick, K.E. (2003). Commentary on Czarniawska. *Management Learning* 34, 379-382.
- Weick, K. E. (2006). Dear editor: A reply to Basbøll and Graham. *Ephemera* 6, 193.