

8

DATA TO DIE FOR

Archival Research

Brett W. Pelham

Baseball great Yogi Berra once noted that “you can observe a lot just by watching.” Berra’s point is consistent with a great deal of archival research. Archival research is merely passive observational research that relies on data collected or created by others. Birth, marriage, or death records, for example, can become archival data. Thus, good archival research can be even easier than Yogi Berra’s casual research. In the case of archival research, others have usually done the watching for you.

Berra was also saying that good observations don’t always have to be complicated. The simplicity of archival methods stands in sharp contrast to the sophistication of many other modern research techniques. Consider computer programs that can (a) parse human speech into linguistic categories (Tausczik & Pennebaker, 2010), (b) disentangle participant- and group-level patterns in hierarchically organized data (Raudenbush & Bryk, 2002), or (c) analyze a video of a dance and calculate the degree of synchrony between dancer and music (Brick & Boker, 2011). Archival research usually stands in contrast to such high-tech observations. Nonetheless, archival studies can yield crucial information that cannot be obtained any other way. Archival studies can tell us how well a theory travels to the real world. Further, the best archival studies do not merely tell us *that* something can happen in the real world. They tell us *when* and even *why* (Cook & Campbell, 1979; McGuire, 1989).

Strengths and Weaknesses of Archival Studies

Archival studies have other strengths. A big problem with experiments is the fact that they can make people behave unnaturally. This may happen, for example, because people know they are being watched. In fact, Bateson and colleagues found that people behave better than usual when something merely *reminds them*

of being watched. When posters portray watchful human eyes rather than flowers, people more often (a) pay for coffee on the honor system and (b) clean up after themselves after eating (Bateson, Nettle, & Roberts, 2006; Ernest-Jones, Nettle, & Bateson, 2011). Because human beings are sensitive to being watched, experimenters must usually work very hard to get people to behave naturally (Aronson & Carlsmith, 1968).

Archival research bypasses the dilemma of unnaturalness by studying natural behavior (Pelham & Blanton, 2013). It also bypasses experimenter bias by eliminating the experimenters who unwittingly create this problem. For example, I know that U.S. consumers purchase regular octane gasoline more often than higher octane gasoline. This is because I have noticed that the buttons for regular gasoline typically show excessive signs of wear. This is a very casual archival observation, but it still bypasses concerns that consumers select regular gasoline because they are trying to impress an experimenter. Likewise, archival studies of criminal behavior solve a major ethical problem when they examine public records of murders—rather than trying to induce murder in the laboratory.

This chapter examines strengths and weaknesses of archival research, with an emphasis on some of the poorly appreciated strengths. For example, I argue that archival research has gotten a bit of a bad rap when it comes to establishing internal validity. Next, I suggest two simple recipes for being a solid archival researcher. The first recipe (the OOPS heuristic) is a checklist for critically analyzing and maximizing the external validity of archival research. The second recipe (the GAGES heuristic) tackles internal rather than external validity. It suggests that if researchers address five pesky confounds, they will often have gone a long way toward maximizing internal validity. After sketching out these two heuristics, I discuss a series of archival research studies that focus on topics as diverse as social judgment, the self-concept, and longevity. My goal in so doing is not merely to extoll the virtues of archival research. It is also to show that even novice researchers can conduct highly rigorous research. OOPS and GAGES are recipes for success.

John Stuart Mill and Internal Validity

As the British philosopher John Stuart Mill (1863) would be quick to note, archival researchers often face serious challenges establishing causality (Pelham & Blanton, 2013). If we update Mill's insights to incorporate modern statistical language, Mill argued that there are three requirements for establishing causality.

Covariation

Mill's first requirement is covariation. For Y to cause Z, changes in Y must covary with changes in Z. At a minimum Y and Z must be correlated. Distress appears to cause divorce. It certainly covaries with it. On the other hand, divorce can cause

at least some people to become distressed (Booth & Amato, 1991). Covariation is consistent with either temporal sequence.

Temporal Sequence

Mill also argued that if Y causes Z, changes in Y must *precede* changes in Z. In many passive observational studies (including archival studies) researchers have little or no information about temporal sequence. Consider the Neanderthal diet. Microbiological and isotopic analyses of the teeth, bones, and even the dried-up feces of Neanderthals reveal that the typical Neanderthal diet was about 80% meat. Our extinct cousins ate much more meat than *Homo sapiens* did (Sistiaga, Mallol, Galván, & Summons, 2014; Wißing et al., 2015). At one level it's obvious that specific Neanderthals were Neanderthals *before* they ate so much meat. They were born that way. On the other hand, thinking about temporal sequence on an evolutionary scale, some have argued that Neanderthals took the particular evolutionary route they did because their ancestors migrated to a part of the earth in which (a) a very cold climate and (b) the preponderance of megaherbivores set them on a specific evolutionary path. Neanderthals may have *become* the short, muscular, cold-adapted hominids they were because they were surrounded by so much steak—and so little salad.

But things aren't *always* this tricky. In fact, some archival studies do include information about temporal sequence. Consider archival studies of sports. In a baseball game, innings clearly indicate temporal sequence. Further, even if we ignore innings in baseball, we can often eliminate worries about reverse causality. Archival studies of aggression in baseball show that pitchers are more likely to hit batters with “bean balls” on hotter days than on cooler days (Reifman, Larrick, & Fein, 1991). We do not need to worry that a pitcher's decision to throw a ball at a batter changes the ambient temperature. We also do not need to worry that pitchers simply become less accurate as the temperature rises. On hotter days pitchers actually walk *fewer* batters than usual, and throw slightly fewer wild pitches. An archival study of more than 57,000 baseball games also showed that a pitcher is more likely to hit a batter when one of his own players was just hit by the pitcher from the opposing team (Larrick, Timmerman, Carton, & Abrevaya, 2014). That's payback. Archival research also shows that athletes repay good deeds as well bad ones (Willer, Sharkey, & Frey, 2012). Information about temporal sequence is not embedded in all archival records. But many archival data sets fare surprisingly well when it comes to temporal sequence.

Eliminating Confounds

When it comes Mill's third rule of causality—eliminating all possible confounds—archival studies do *not* usually fare as well as thoughtfully designed surveys. Survey designers usually choose exactly which questions people answer. Thus,

well-informed survey researchers can measure and statistically control for any confounds about which they are worried. In contrast, when researchers rely on data collected by others, they often see that these others were rarely obsessed with eliminating confounds.

Despite this gloomy observation, there are reasons for optimism. First, some sources of archival data do exist that include measures of important confounds. Some archival data sets were created by scientists rather than sports enthusiasts. The standard cross-cultural sample (Murdock & White, 2006) and the World Values Surveys (www.worldvaluessurvey.org) are two invaluable examples. In these rich archival data sets, one researcher's survey question or ethnographic observation is another researcher's confound. Second, some archival studies are natural experiments, which solve the problem of confounds by giving people a manipulation at random (or nearly so). Third, after-the-fact coding in an archival study can yield important insights into why a specific outcome occurred. For example, archival data provide rankings of the market share of many brands of a specific product (e.g., fast food). It is possible to have consumers rate these stimuli (e.g., McDonald's, Wendy's, Arby's) on multiple dimensions (e.g., price, customer service) to see which dimensions do the best job in a statistical footrace that predicts the brand rankings.

External Validity and the OOPS Heuristic

John Stuart Mill would surely love lab experiments. But some features that make lab experiments high in internal validity make them low in external validity. External validity is about generalizability to the real world, and archival studies usually examine real world events. I've argued elsewhere (Pelham, 2017) that concerns about external validity almost always fall into one of four categories. Each letter of the OOPS heuristic represents one of these categories. Archival research usually does a good job of addressing the OOPS concerns. What are these concerns?

Operationalizations

Because testability is a cherished scientific canon, psychologists acknowledge that we can only study things scientifically if we specify operational definitions (Pelham & Blanton, 2013). But there are many ways to operationalize most hypothetical constructs. Consider sexual arousal. In men, a useful operationalization is change in penis size. Thus, researchers in human sexuality developed the plethysmograph, which measures penis volume (Adams, Wright, & Lohr, 1996). But if we wish to study sexual arousal in the other 52% of the earth's population, plethysmographs are useless. For this reason, experts developed a measure of sexual arousal that works for women as well as men. Thermographic stress analysis (TSA) involves assessing changes in genital temperature. A thermography camera can

detect changes of about one fourteenth of a degree Celsius in a very brief period. Thermography works just as well for women as it does for men (Kukkonen, Binik, Amsel, & Carrier, 2010).

We can place greater confidence in any research finding when it holds up well across multiple operationalizations. *Altruism*, for example, could be defined as (a) giving food to another organism or (b) risking your own safety to protect another organism (Dawkins, 1976). Nursing your infant daughter fits the first definition. Saving a drowning stranger fits the second. Along these lines, kin selection (making sacrifices for organisms with whom you share genes) is now widely accepted because there is good evidence for kin selection regardless of which definition of altruism one adopts (Hamilton, 1964a, 1964b; Smith, Kish, & Crawford, 1987). One reason why Pinker's (2010) argument that human violence has declined over the centuries is so convincing is that Pinker's archival studies use *many* operational definitions of violence, from killing or enslaving people to spanking children, burning witches, or hurting animals in films. Across numerous operational definitions, violence has declined, especially in this century.

Occasions

"To everything there is a season." Human behavior varies greatly across the day, across seasons, and across millennia. College students report having sex much more often between 11 p.m. and 1 a.m. than at any other time of day (Refinetti, 2005). People's hormone levels also vary naturally over time. This variation often has important consequences. Welling and colleagues (2008) showed that men rated highly feminine female faces to be more attractive than usual on days when the men's testosterone levels were higher than usual.

Looking at timing over a broader window, both births and deaths vary with the seasons. Consider the archival data in Figure 8.1. They show that Americans more often die in winter than in summer (despite the fact that most deaths by accident are more common in the summer; Rozar, 2012). There is debate about exactly why this seasonal pattern occurs, but the pattern is clearly *seasonal* rather than calendrical. The pattern disappears at the equator and is reversed in the Southern hemisphere. Marriage rates, too, vary over the course of the year. As you already knew, June is the most popular month for American weddings. As you probably did *not* know, people are also more likely to get married during the month of their own birthdays than in other months (Pelham & Carvallo, 2015).

Time also matters century by century. Two thousand years ago, Romans died more often in the summer than in the winter (because diseases like malaria were much more common in summer; Scheidel, 2009). At that time, the entire population of the earth was smaller than the current population of Indonesia. Turn the clock back to 10,000 years ago, when agriculture barely existed, and the earth's human population was smaller than the current population of Chicago.

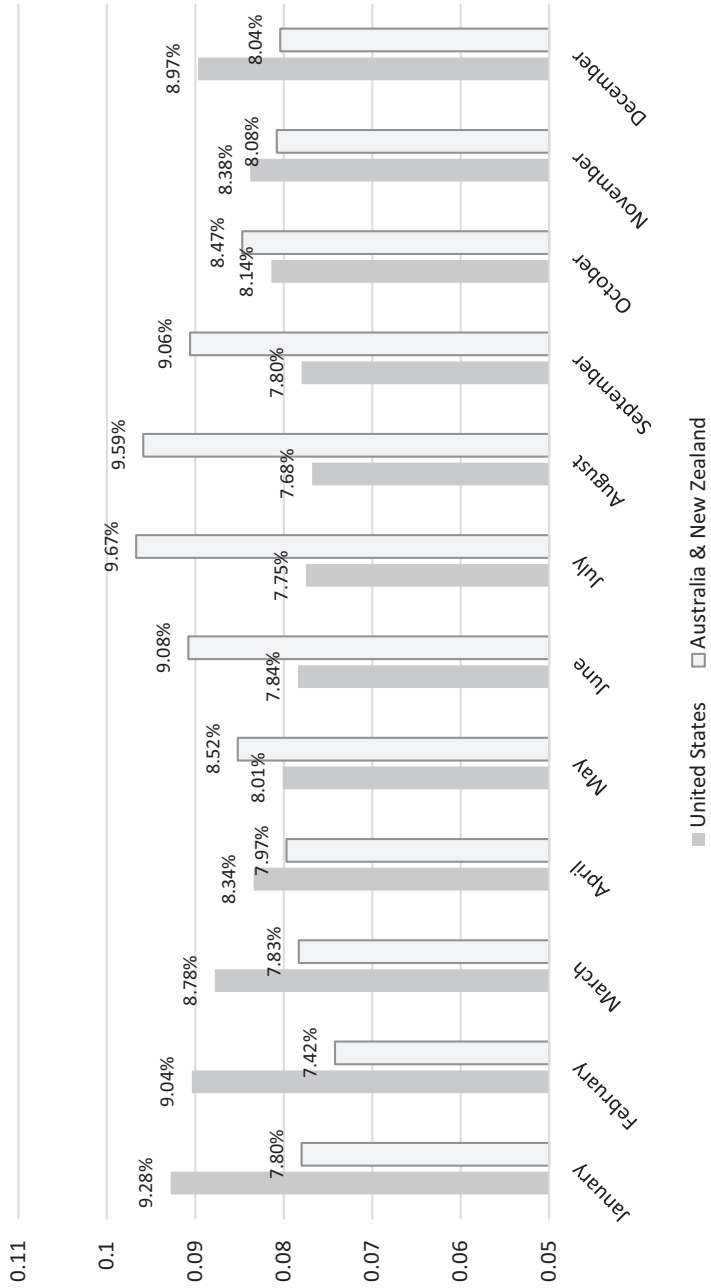


FIGURE 8.1 Likelihood of Dying by Month in the United States versus Australia and New Zealand, 1990–2010 (Adjusted for the Number of Days in a Month)

Note: U.S. Data were harvested from the Social Security Death Index (SSDI). Australia and New Zealand data were harvested from the Australia and New Zealand “Find a Grave Index.” Both archival data sources were accessed at ancestry.com.

So time matters. Thus, when evaluating any research finding, we must ask if that finding would hold true at other times. Showing that something interesting is true is impressive. Showing that it was also true 1,000 years ago is more impressive. Time is particularly important in developmental and evolutionary psychology. What cognitive skills do most toddlers possess that most infants do not? How have hominids changed over the past million years? Does interest in sex vary temporally? Gangestad and colleagues (2010) found that women's degree of sexual opportunism varies across their ovulatory cycle. When women are at the most fertile point in their cycles, they're more likely than usual to endorse attitudes such as "I believe in taking my sexual pleasures where I find them."

Populations

Almost no research finding applies to every imaginable population. People see most colors better than dogs do, and dogs smell things people can barely imagine. European magpies recognize themselves in mirrors but most other birds do not (Prior, Schwartz, & Gunturkun, 2008). Chimps are more sexually promiscuous than people or gorillas. If chimps could talk, they might tell us they believe in taking their sexual pleasures where they find them. Even if we limit ourselves to people, the validity of a specific research finding can also vary dramatically with populations. Presumably very few nuns believe in "taking their sexual pleasures where they find them," even if they are currently ovulating. Many research findings vary by population. Zajonc (1965) documented social facilitation in ants, cockroaches, parakeets, puppies, and monkeys as well as in people. Archival data sets rarely include multiple species, but they do very often include highly diverse human populations. This is a big strength.

Situations

A final aspect of external validity focuses on generalization across situations. All research takes place in a specific context, and that context can influence what researchers observe. The way people think and reason seems to vary based on the way in which an experimenter dresses. When experimenters dress casually people seem to *think* casually (i.e., intuitively; Simon et al., 1997). When people dress more formally, others are more likely to obey them (Bickman, 1974). It is a central tenet of social psychology that the specific situations in which people find themselves (e.g., a synagogue vs. a singles bar) can have a huge impact on how people think, feel, and behave. To know how robust a research finding is, you need to know how well it holds up in a wide variety of situations.

To summarize, to the degree that a specific archival study shows that an effect holds up well using different operational definitions, in multiple temporal windows, for different populations, and in different situations, there can be no doubt that the study has capitalized well on one of the strengths of archival research. But to return

to an earlier point, plenty of external validity in the absence of internal validity is not very informative. Are there any rules of thumb for assessing *internal* validity?

GAGES: Five Common Confounds That Can Undermine Archival Research

There is a reason why census takers, epidemiologists, and marketers have long focused on a handful of regional and demographic variables when doing their jobs. Five of the cardinal ways in which human beings vary include geography, age, gender, ethnicity, and socioeconomic standing (education and/or income). I refer to these five key variables using the acronym GAGES (geography, age, gender, ethnicity, and socioeconomics).

Geography

Geographically speaking, knowing where a person lives can be telling. From red states vs. blue states to latitude vs. altitude, location matters. According to the 2010 Census, the average Maryland resident had almost twice the family income of the average West Virginia resident. New Jersey is about 1,000 times more densely populated than Alaska. Personal beliefs and values also vary widely across U.S. states. Residents of Vermont are more than five times as likely as residents of Mississippi to report that they are not religious (Newport, 2014). In fact, research on cultural evolution suggests that properties of the physical environments in which people live predict variables as different as what kind of language people speak, whether people cook with spices, whether women are allowed to have multiple husbands, and whether people are xenophobic (Billing & Sherman, 1998; Everett, Blasib, & Roberts, 2015; Fincher, Thornhill, Murray, & Schaller, 2008).

Age

Demography can matter just as much as geography. Beginning with age, older Americans worry less than their younger counterparts (Newport & Pelham, 2009). They also eat healthier diets, exercise less frequently, and care more deeply than young people do about nurturing close, established relationships (Carstensen, Isaacowitz, & Charles, 1999; Dugan, 2013). Older Americans are also substantially more likely than their younger counterparts to be religious, and to be wholly unfamiliar with Lil Wayne. Differences such as these are why there is a field called developmental psychology.

Gender

Moving on to gender, across the globe, men are more likely than women to assault or kill others, to commit suicide, to work in dangerous jobs, and to abuse drugs.

Conversely, women are more likely than men to suffer from depression and to serve as caretakers, both at home and at work. The list of ways in which gender matters is so long that there is entire branch of research known as gender studies.

Ethnicity

Ethnicity matters, too. Both Blacks and Latinos are more likely than Whites to suffer from clinical depression (Dunlop, Song, Lyons, Manheim, & Chang, 2003). Relative to Whites, Blacks are also much more likely to lack confidence in the police (Jones, 2015), to vote Democratic, and to be familiar with Lil Wayne. More than 50 years after the passage of the U.S. Civil Rights Act, there are still large ethnic differences in income, unemployment, and education.

SES

Above and beyond ethnicity, one of the best predictors of longevity and well-being is socioeconomic standing (SES; Bosworth, Burtless, & Zhang, 2015). Socioeconomic standing also predicts important attitudes and values (Pelham, 2018; Pelham & Hetts, 1999) as well as serious problems such as suicide risk and automobile accident rates (Sehat, Naieni, Asadi-Lari, Ferooshani, & Malek-Afzali, 2012). There is a reason many sociologists study SES. It matters.

Given the importance of the GAGES, researchers who conduct archival research will ideally be able to show that an archival research finding goes above and beyond any effects of the GAGES. Of course, the list of possible confounds about which researchers should worry does not end with GAGES. Specific confounds vary with the specific research question at hand. But the five major worries summarized by the GAGES are a great place to start. Because census takers, public health officials, marketers, and people seeking dates often care about GAGES, there is often good information about GAGES in archival data sets.

Is GAGES WEIRD? Henrich, Heine, and Norenzayan (2010) argued that a great deal of psychological research fails to consider the tremendous cultural diversity of the planet. Specifically they noted that the great majority of past research in psychology focused on WEIRD people, those who come from “Western, Educated, Industrialized, Rich, and Democratic” societies. GAGES is distinct from WEIRD. First, WEIRD expresses a concern about external validity (e.g., would shopkeepers in India behave like students in Indiana?). By contrast, GAGES is all about internal validity. That being said, GAGES does have some overlap with WEIRD. Cultures, after all, have geographies. Cultures also vary in age, ethnicity, SES, and even gender ratios. Furthermore, one could treat GAGES variables as cultural moderators rather than confounds. Conceptually, however, WEIRD overlaps more with OOPS than with GAGES. One key difference here is that WEIRD focuses on cultures whereas OOPS usually focuses on individual people. Further, WEIRD includes populations and situations but is largely silent

regarding operationalizations and occasions. In a sense, then, OOPS means that the WEIRD critique may not go quite far enough. A complete understanding of the strengths and weaknesses of archival research requires more than WEIRD insights.

Moderation and Theory

Researchers who rely on archival data rarely have the luxury of putting their preferred predictor in a footrace with all possible confounds. However, archival researchers do sometimes have access to theoretically derived moderator variables (including WEIRD ones) that ought to predict when an effect grows stronger versus weaker. For example, laboratory experiments on modeling (i.e., social learning) show that people are more likely than usual to imitate targets who resemble them (e.g., see Ariely, 2012; Bandura, 1977). Phillips's archival research on copycat violence capitalizes on exactly this logic. First, people do copy highly publicized suicides and homicides. Second, this copycat effect is stronger than usual when these tragedies get more media attention (Phillips & Carstensen, 1986). Third, consistent with principles discovered in lab experiments, people are more likely than usual to copy the suicidal behavior of others when they belong to the same gender, age, or ethnic group as a target.

The points made thus far suggest that it is possible to conduct archival research that strikes an impressive balance between internal and external validity. In the remainder of this chapter, I briefly summarize archival research on five different topics, ranging from social cognition to health and mortality. The thread that unites these archival studies is the fact that the authors all found creative ways to maximize both internal and external validity. Although I focus mainly on research topics I happen to study, I hope readers will realize that an appreciation of OOPS and GAGES could be applied effectively to almost any topic.

Archival Studies of Social Cognition

False Consensus

One of the first researchers to use archival data to study social cognition was Mullen (1983), who studied the false consensus effect (Ross, Greene, & House, 1977). This is the tendency for people to overestimate the percentage of others who share their beliefs or behaviors. Mullen believed this bias would still appear when avoiding it could help people win thousands of dollars. Mullen also suspected (correctly) that the false consensus effect is larger for people whose attitudes or behavior place them in the statistical minority rather than the majority. To study the false consensus effect, Mullen capitalized on data from a TV game show ("Play the Percentages"). The key data points provided by game show participants were their estimates of the percentage of studio audience members who would be

able to answer specific trivia questions (e.g., “What state did Hubert Humphrey represent in Congress?”) Back when people still remembered Humphrey, 72% of audience members were able to answer this question correctly.

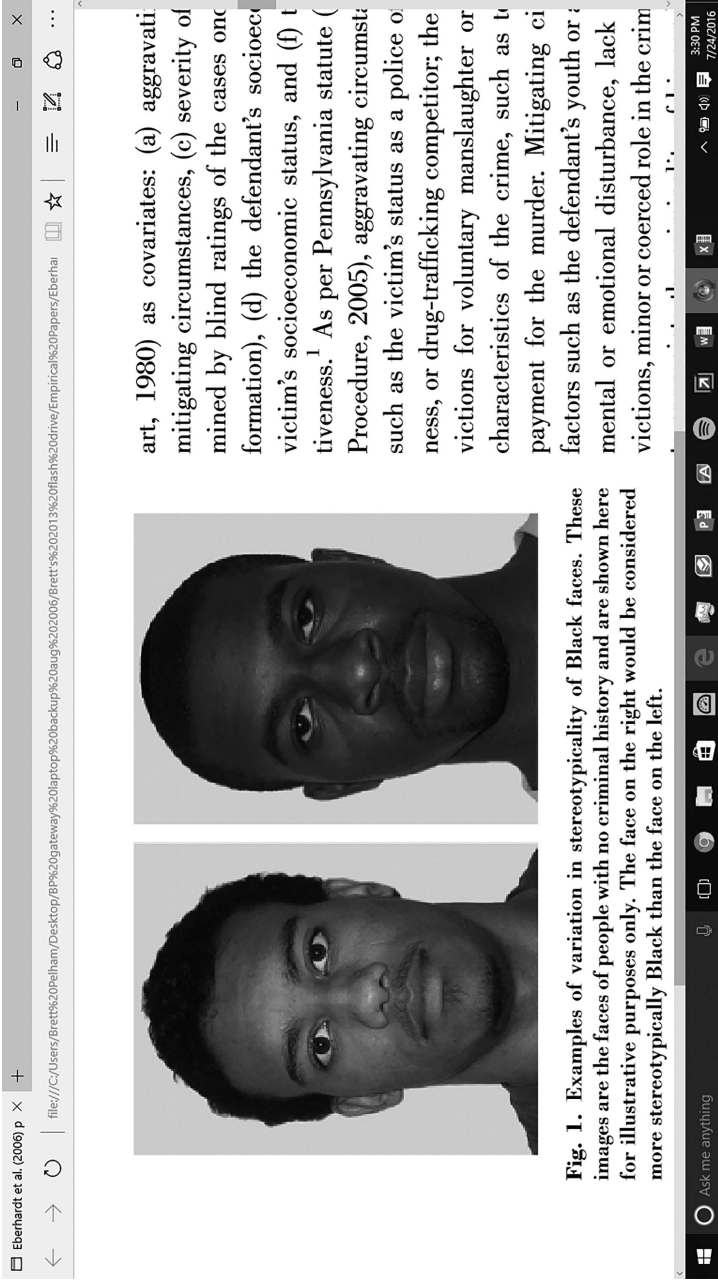
Mullen observed clear evidence of the false consensus effect. Participants overestimated the percentage of others who knew the answers to questions when they themselves had known the answers to the questions. Second, false consensus effects were larger than usual when people’s own answers placed them in the statistical minority. The rare people who knew the answer to a difficult question were especially likely to assume that other people shared their esoteric knowledge.

Mullen’s documented a false consensus effect with a slightly different operational definition than the one usually used in the lab, with a novel population, and in a very different situation than the lab, satisfying three of the four OOPS criteria. Further, because this study included equal numbers of men and women, and because men and women both showed the effect, we cannot attribute Mullen’s effects to a gender confound. Finally, it is hard to imagine that any other GAGES confounds could apply to Mullen’s archival study without also applying to laboratory studies. More highly educated participants *may* have known more of the answers to the trivia questions, but there is no reason to believe that being educated *in and of itself* would make people offer higher *consensus* estimates—or that this would happen in games show but not in laboratories.

Ethnic Stereotyping

A more sobering example of archival research in social cognition is Eberhardt, Davies, Purdie-Vaughns, and Johnson (2006) research on stereotypes and capital punishment. Eberhardt and colleagues identified criminal records from more than 600 men who had been convicted of murder in greater Philadelphia between 1979 and 1999. They then identified all of the cases ($n = 44$) in which a Black defendant had been convicted of killing a White victim. Based on previous work by Blair and colleagues, they suspected that Black men convicted of killing White victims would be more likely to receive a death sentence when they had a more stereotypically Black appearance than when they did not. The researchers showed photographs of all of the selected Black defendants to students who knew nothing of the men’s criminal status. These judges assessed “the stereotypicality of each Black defendant’s appearance and were told they could use any number of features (e.g., lips, nose, hair texture, skin tone) to arrive at their judgments” (Eberhardt et al., 2006). Figure 8.2 shows two Black male volunteers who vary in stereotypicality.

One of the most methodologically impressive aspects of these archival findings is that Eberhardt and colleagues controlled for six potential confounds all known to be important predictors of sentencing decisions. These included “(a) aggravating circumstances, (b) mitigating circumstances, (c) severity of the murder (as determined by blind ratings of the cases once purged of racial information),



art, 1980) as covariates: (a) aggravating circumstances, (c) severity of mitigating circumstances, (d) severity of formation), (d) the defendant's socioeconomic status, and (f) the victim's socioeconomic status (Procedure, 2005), aggravating circumstances such as the victim's status as a police officer, or drug-trafficking competitor; the victims for voluntary manslaughter or characteristics of the crime, such as the payment for the murder. Mitigating factors such as the defendant's youth or mental or emotional disturbance, lack of victims, minor or coerced role in the crime

FIGURE 8.2 Two Black men with no criminal records who vary in the stereotypicality of their appearance.

Note: Eberhardt's findings suggest that if both men were to commit a crime, the man on the right would be judged more harshly.

(d) the defendant's socioeconomic status, (e) the victim's socioeconomic status, and (f) the defendant's attractiveness" (Eberhardt et al., 2006, p. 384). Further, their operational definitions of constructs a–e were based on established Pennsylvania statutes. Because the archival records did not include information on defendant physical attractiveness, the research team got blind raters to judge this. Even after controlling statistically for all six of these confounds, Eberhardt et al. (2006) found that Black men with a more stereotypically Black appearance were more likely than Black men with a less stereotypic appearance to be given a death sentence. A follow-up study showed that when defendants had been convicted of killing *Black* rather than *White* victims, the stereotypicality of the men's appearance no longer made a difference for their sentences.

A consideration of GAGES reveals that this study controlled for many geographic confounds by staying near Philadelphia. Of course, OOPS dictates that it would have been even better to study more than one region of the United States. But no single study can do everything. The authors also appear to have controlled for the defendants' ages because age is often considered a mitigating factor. They controlled for gender by studying only men. The authors not only controlled for ethnicity but also deconstructed it (ethnic stereotypicality was an independent variable). They also controlled for the SES of both the defendants and the victims. The additional factors for which Eberhardt and colleagues controlled reveals that GAGES is not an exhaustive list. But the fact that these authors left no GAGES stone unturned attests to the importance of these variables, as well as to the sophistication of this study.

Counterfactual Thinking and Emotions

Not all research in social cognition focuses on tragedies. Some of it focuses on triumphs. Medvec, Mavey, and Gilovich (1995) studied athletic triumphs, including triumphs that don't always make people feel very good. Laboratory research on counterfactual thinking shows that when something good or bad happens, people often consider counterfactual (alternative) realities. Counterfactual thoughts sometimes create counterintuitive emotions. For example, missing a flight by two hours usually produce regret. But missing a flight by two minutes usually produces a lot more of it (Roese, 1997). When Medvec and colleagues conducted their archival studies of counterfactual thinking and emotions, most previous studies had been conducted in the lab. Further, many of these studies were based hypothetical scenarios ("How would you feel if . . .?") rather than real outcomes. Medvec and colleagues put the factual into the study of counterfactuals.

They did so by considering the emotional implications of earning gold, silver, and bronze medals in major athletic competitions. Most Olympic gold medalists must surely be on top of the world after their victories. At a bare minimum they end up on top of the medal stand, and their gold medals often bring them fame

and fortune. By contrast, many silver medalists may feel the pain of knowing how close they came to winning. For bronze medalists, however, *two* things would have had to have gone differently for them to have won gold (e.g., Usain *and* Justin). The most salient counterfactual for bronze medalists is probably that they could have easily finished in fourth place, earning no medal at all. This logic suggests that athletes might typically be happier with an inferior outcome (a bronze medal) than with a superior one (a silver medal).

To test this prediction, Medvec et al. recorded NBC's televised coverage of the 1992 Olympics. They then extracted every scene that showed a bronze or silver medalist (in any sport NBC chose to cover) the moment the athletes learned they had finished second or third. They did the same thing for the period when athletes stood on the medal stand. Finally, they showed all of the video clips to a group of raters who were kept blind not only to Medvec et al.'s predictions but also to athletes' order of finish. They also turned the volume to zero for all of the ratings so that raters would not be biased by the comments of any of the NBC sports analysts, especially Bob Costas. The raters simply judged each athlete's expressed happiness on a 10-point scale.

Medvec and colleagues found that, despite finishing third rather than second, Olympic bronze medalists looked happier than their slightly faster, stronger, and more coordinated peers. This was true both immediately after their performances and on the medal stand. Of course, these results alone do not say whether *counterfactual thinking* was responsible for the observed emotions. To address this, Medvec et al. performed a second set of archival analyses from the same Olympic TV coverage. This time they selected all of the available *interviews* with bronze and silver medalists and asked blind raters to judge the "extent to which the athletes seemed preoccupied with thoughts of how they did perform versus how they almost performed." This follow-up study suggested that bronze medalists were more focused on what they "at least" did whereas silver medalists were focused on what they "almost did." A replication study focusing on a state-level athletic competition confirmed this result.

These results seem safe from any obvious GAGES confounds. It is highly unlikely, for example, that men (a) more often finish third than women and (b) are chronically happier than women. One not-so-obvious confound, however, is that in some Olympic events (e.g., wrestling, basketball), bronze medalists have just *won* a competition whereas silver medalists have just *lost* a competition. That's a real confound. In a supplemental analysis, Medvec et al. focused solely on events (e.g., track and field) in which there was no such confound. The bronze medalists still looked happier than the silver medalists. This archival research is also a standout when it comes to OOPS. It used novel operationalizations, it examined behavior in athletic events that took place on two different occasions, the participants came from all over the globe, and the situation in which people were studied was radically different than the lab. In my view the authors of this study struck methodological gold.

Self-Concept and Identity

Implicit Egotism: Early Studies of Career Choice

Archival research can also be a rich source of information about the self-concept. About 15 years ago, my colleagues and I became interested in the idea that people resemble the legendary Narcissus. Laboratory studies had already shown that people have an unconscious preference for the letters in their own names (Nuttin, 1985). Kitayama and Karasawa (1997) extended this to show that people prefer the numbers in their own birthdays (see also Beggan, 1992). So if something is part of the self, it must be good. Inspired by such findings we began to study implicit egotism, an unconscious preference for people, places, and things that resemble the self. We began by using archival data to study careers and street addresses.

In the early days of this work, we did not fully appreciate the risks inherent in archival research. When I began this work, for example, I knew that Carlos was a Latino first name. I also knew that there is a good chance Jeff Goldstein is Jewish. But I did not know, for example, that a person whose last name is Jefferson is about 180 times as likely to be Black as a person whose last name is Carlson. I eventually learned just how strongly people's first and last names can be confounded with GAGES.

This being said, my colleagues and I were not completely oblivious. In one of our first studies of implicit egotism we focused on career choice (Pelham, Mirenberg, & Jones, 2002). Our analyses of professional membership records for dentists and lawyers showed that people with names like Dennis, Denise, Lawrence, and Laura gravitated toward the jobs that resembled their names. We controlled to some degree for geography by limiting our searches (for both occupations) to the eight most populous U.S. states. We controlled for gender by separating male and female names. We further controlled for the frequency of our first names. After completing this initial study, however, we were disappointed to see how difficult it was to locate nationwide records that reliably identified people by name and occupation (but see Abel, 2010).

Implicit Egotism and Choice of a Residence

In contrast to the paucity of national data on names and professions, there are plenty of archival data that identify millions of people by name and place of residence. We have thus conducted numerous studies of implicit egotism and choice of a residence. Relying on 66 million Social Security Death Index records, Pelham et al. (2002) showed that just as people whose first or last name is Thomas are overrepresented in cities named St. Thomas, people whose first or last name is Peter are overrepresented in cities named St. Peter. Although we argued that our focus on cities with "Saint" in their names should have reduced ethnic confounds,

this study was still susceptible to geographic confounds. Although we conducted follow-up studies to address this concern, eliminating all of the GAGES concerns entirely is no easy matter. Some studies of implicit egotism are also open to critiques based on reversed causality (e.g., are there a lot of women in Georgia named Georgia because they moved there or because parents in Georgia prefer this first name?)

One way to address problems such as these is to blend archival methods with more traditional methods. Pelham and Carvalho (2015) did exactly this. First, using archival death records, we showed that men named Cal and Tex had disproportionately lived in the large U.S. states that resembled their names. In a replication study with living participants, we identified more than 800 men named Cal or Tex who lived in either California or Texas. A survey showed that these men, too, disproportionately lived in states closely resembling their names. Further, this was strongly true even when we focused exclusively on men who reported *moving* to the states in which they lived.

In addition to the usual problems of confounds and reverse causality research on implicit egotism is also susceptible to sampling problems. Although we always specified in advance how we had matched names with states, cities, or street addresses, we often faced arbitrary decisions. This problem largely evaporated on April 2, 2013. This is when the 130 million records that made up the entire 1940 U.S. Census became public (Pelham & Carvalho, 2015). At about this same time, the genealogical website ancestry.com also made available a tremendous number of American and international birth and marriage records.

Back to Career Choice

The 1940 U.S. Census data allowed us to see if people preferred occupations that doubled as their exact surnames (Pelham & Carvalho, 2015). To test this idea, we examined the 2,000 most common U.S. surnames. From this list, we identified the 11 surnames that constituted common, traditionally male occupations (e.g., Baker, Carpenter, Farmer, Mason). We then calculated the expected number of men with these exact surnames who *would* have reported working in each of these 11 occupations if there were no association between surname and career choice. We compared each expected frequency with the observed frequency that corresponded to a surname-occupation match. As shown in Figure 8.3, to at least a small degree, all 11 of the surname-occupation pairs yielded support for implicit egotism (as indicated by ratios greater than 1.0).

The score of 1.34 for Porter means that men name Porter were 34% more likely to have worked as porters compared with the entire set of men who had the other 10 surnames. But might these results reflect an ethnic confound? If Black men disproportionately worked as porters, and if Black men were disproportionately *named* Porter, this could create artifactual support for implicit egotism. A similar argument could be made for other GAGES.

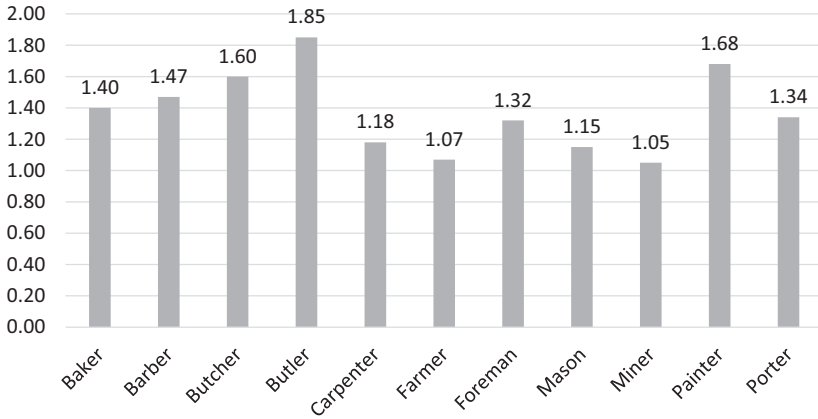


FIGURE 8.3 Ratio of Observed to Expected Surname-Occupation Matches for Men With Common Surnames That Also Serve as Male Occupation Names

Source: 1940 Census.

Because of the sheer size of these census data, we were able to evaluate these results separately for Blacks and Whites, and for men who were equated perfectly for education. Even after we separated men by ethnicity and focused separately on men with exactly 6, 8, 10, or 12 years of education, there was always robust support for implicit egotism. For example, White men named Farmer who had an eighth grade education gravitated toward farming relative to all other White men in the 1940 Census who were not named Farmer but who also had an eighth grade education.

We did not publish this finding, but we also found that implicit egotism in career choice held up for men of different ages. This leaves only geography. If lots of people named Farmer happen to live in Nebraska as compared with Massachusetts, this could lead to artifactual support for implicit egotism. Although we did not consider this geographic confound in our original report, I recently conducted state-by-state analyses to see if the association between surname and occupational choice would hold up at this level. It did. Further evidence that this effect is not likely due to a geographic or ethnic confound is based on the fact that the effect replicated strongly—using exactly the same 11 career-surname pairs—in both the 1880 U.S. Census and the 1911 England Census. Taken together, these three studies—which span a 60-year window—also do a decent job of addressing the OOPS issue of occasions.

Implicit Egotism and Marriage

Perhaps the strongest archival evidence for implicit egotism comes from research that focuses on birthdays rather than names. Most people prefer their birthday

numbers. Unlike a person's first or last name, however, which can easily be confounded with GAGES, a person's birth month or birthday number is much more arbitrary. Birthday numbers also range from 1–31, which greatly simplifies decisions about how to sample specific numbers. If possible, sample them all. If people gravitated toward things that matched their birthday numbers, this would constitute rigorous support for implicit egotism.

People do, and the thing toward which they gravitate is other people. Pelham and Carvallo (2015) found that people were disproportionately likely to marry other people who happened to share either their birthday number or their birth month. As shown in the top panel of Figure 8.4, brides in Summit County, Ohio, were 6.5% more likely than they should have been by chance to marry a groom who shared their exact birthday number. As shown in the bottom panel of Figure 8.4, this bias increased to 37.6% for the subset of brides who *married* on their birthday numbers. Both of these effects, as well as conceptually identical effects for birth *month*, replicated well in a very large set of more recent (1958–2001) Minnesota marriage records.

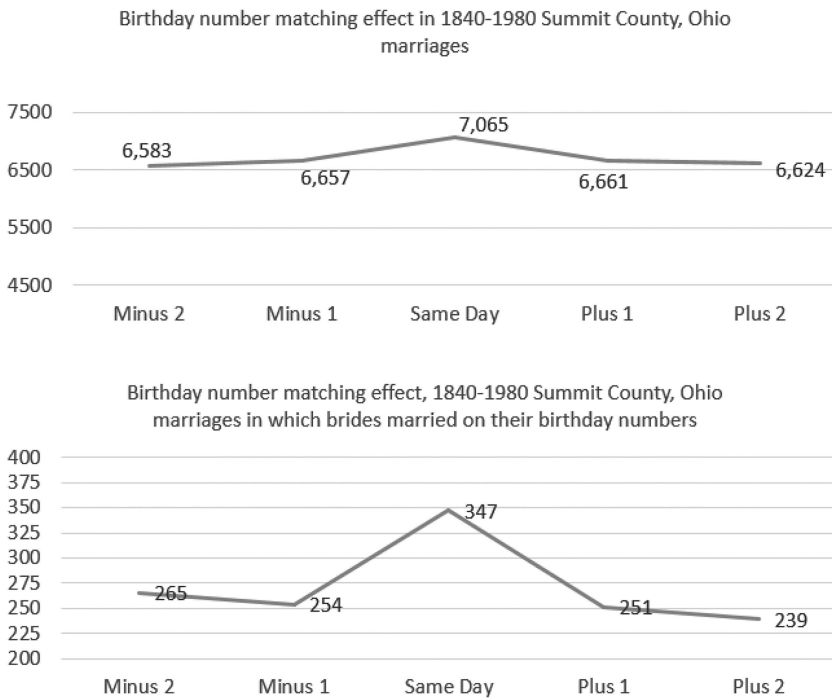


FIGURE 8.4 Implicit Egotism in Marriage

Note: People with the same birthday number were more likely than usual to marry. This effect was much stronger than usual for brides who married on their birthday number. Adapted from Pelham and Carvallo (2015).

I hasten to add that implicit egotism is not the only aspect of the self that has been studied using archival data. Research by Twenge and colleagues—most of it archival—has shown that self-esteem (“I am capable.”) and narcissism (“I will build a great, great wall.”) have been increasing in the United States over the past few decades. Twenge typically examines archival data *collected by psychologists*, which has allowed her to document changes in constructs that would be hard to operationalize using birth or marriage records. That being said, Dewall, Pond, Campbell, and Twenge (2011) did examine changes in egotism over the past few decades without asking anyone to fill out any surveys. They did so by tabulating changes in the relative use of the words “I,” “me,” and “mine” in popular song lyrics between 1980 and 2007. Across this window there was a linear increase in how often popular singers sang about themselves (via these pronouns). If Twenge and colleagues are correct that U.S. culture is becoming more narcissistic over time, this bodes well for the future of research on implicit egotism.

Longevity and Death

In addition to archival research on social cognition and the self-concept, there is also archival research on longevity and dying. This research focuses on how psychological variables influence the age at which people join all the others who have gone before them—to that hallowed place I call archival death records. This archival research reveals that psychological variables have more to do with longevity and dying than one might think. Levity seems to promote longevity.

The Nuns Study

To see if viewing the world favorably helps people live longer, Danner, Snowdon, and Friesen (2001) studied 180 nuns. Around 1930, when these women became nuns, they had to write a brief autobiography—which the church dutifully retained. In the late 1990s, Danner and colleagues got permission to analyze these records. The research team blindly coded each young nun’s life story for how many positive emotions it included. It’s pretty easy to spot happiness in these essays. Sister 1 wrote things like, “With God’s grace, I intend to do my best . . .” Sister 2 wrote things like, “I look forward with eager joy to . . . a life of union with Love Divine.”

Based on a simple count of the number of positive emotion words the nuns used, those in the top quartile of the positive emotion distribution lived 9.4 years longer than those in the bottom quartile! Because the nuns lived in one of only two cities (and because the researchers took city differences into account) we do not have to worry much about a geographical confound. The research team also controlled for the age at which women wrote their essays and their eventual education. It also seems safe to assume that the lifestyle most nuns lead went a

long way toward holding many other important variables constant. I'm guessing that few, if any, of the nuns died because of a lack of food or basic medical care. Because this was a sample of nuns, gender was held constant. The research provided no information on ethnicity, but I am pretty confident that the large majority of them were White (and Catholic, too). In short, this study seems to have controlled well for the GAGES confounds. This is particularly important in a study of longevity because each of the GAGES variables has a well-documented association with longevity.

The Baseball Players Study

Any nun with a decent sense of humor will tell you that she tries to fulfill her calling in life by making sacrifices and saving souls—all while having to wear a funny outfit. Baseball players also care deeply about saving and sacrificing, and they, too, wear funny outfits. Like nuns, baseball players also appear to live longer when they express more positive emotions. Baseball players don't have to write autobiographies to play in the majors, but they do have to get their pictures taken. Abel and Kruger (2010) took advantage of this fact by conducting an archival study of professional baseball players whose photos appeared in the 1952 *Baseball Register*. They blindly rated all 230 of the official player photos for how happy the players looked. Photos in which the men were not smiling were given the lowest score, photos in which the men were smiling a polite but unnatural smile got a middle score, and those in which the men expressed a truly happy ("Duchenne") smile got the highest score.

Abel and Kruger reasoned that these photos would reflect a player's characteristic emotional state. Apparently, they did. By the time, Abel and Kruger wrote their report in 2009, the very large majority of these men had died, and Abel and Kruger extracted these dates of death from archival sources. The men who showed a true smile lived an average of seven years longer (mean age 79.9) than the men who did not smile at all (age 72.9). The men who smiled politely (mean age 75.0) also lived a bit longer than the non-smilers. Impressively, these differences in longevity held up even after controlling for a pretty hefty list of competing predictors of longevity (e.g., body mass index, education, marital status, length of playing career).

The fact that Abel and Kruger used player photos from 1952 means that they came pretty close to controlling for ethnicity. In 1952, 94.4% of professional baseball players were White—with the remaining 5.6% being split pretty evenly between Black and Latino players. Blacks do not live as long as Whites, and Latinos actually live a bit *longer* than Whites. For ethnicity to be a serious confound, then, almost all the Blacks would have to be looking solemn and almost all of the Latinos would have to be smiling very happily. It is unlikely—though not impossible—that an ethnic confound could be responsible for their findings.

Further, if a critic were still worried about this confound, it would be easy to code the 230 photos for ethnicity and add this to a covariance analysis. Finally, just as the nun study controlled for gender by studying only women, this study controlled for gender by studying only men.

In light of the GAGES criteria, Abel and Kruger seem to have knocked this one out of the park. They did not address geography, but any critic who was worried about it could add geographical codes to the data and control for them. If there is any real weakness to this study, it is the small number of (deceased) players with usable data ($n = 23$) who posed with Duchenne smiles (samples sizes were greater than $n = 60$ for the other groups). In contrast, the nun study relied on a continuous rather than a categorical coding scheme for positive emotionality. Taken together, these two archival studies are worth smiling about.

A Distressing Natural Experiment

As strong as these two archival studies are, one could still argue that some unknown confound played a role in the results. Perhaps healthier baseball players smiled more. Perhaps nuns from wealthier family backgrounds used more positive emotion words. To know with greater certainty that psychological variables can influence well-being or longevity, one would have to use random assignment to decide that one group of people experienced a dramatic event that others just like them did not. Ideally this ethically callous experiment would have a huge sample size. Further, the design might be particularly powerful if the experimental manipulation involved some kind of horrific negative experience. Being forced to fight in what many considered an unjust war might fit the methodological bill. During the Vietnam War, the U.S. government created exactly this situation as a natural experiment. As Hearst, Newman, and Hulley (1986) noted, the Vietnam War draft meant randomly assigning a subset of young American men to fight in Vietnam. This draft thus created a huge natural experiment.

Hearst and colleagues identified 14,000 men born in California or Pennsylvania between 1950 and 1952 whose draft numbers did versus did not come up. They then checked to see what happened to the men in the 10-year window after the war (1974–1983). They found, among other things, that suicide rates for the unlucky men whose draft numbers had come were 13% higher than they were for the lucky men whose draft numbers had *not* come up. This sounds like a modest effect of this natural manipulation. In fact, the effect of actually serving in combat was certainly much larger. This is because fully 74% of the drafted men found a way to *avoid* going to war. Further, a patriotic 9% of those who had *not* been drafted enlisted as volunteers. After correcting statistically for these facts, the estimated impact of serving as a soldier in Vietnam was an 86% increase in the risk of suicide. Some archival studies come extremely close to the gold standard of true experiments. In natural experiments, the GAGES, like all other confounding variables, are held constant.

Holidays and Mortality

Archival research also suggests that psychological variables can determine the exact day on which people die. Phillips and colleagues conducted several archival studies that suggested that people can sometimes postpone their own deaths to experience important events. Phillips and Feldman (1973) focused on religious holidays. They found that in cities with many Jewish residents (i.e., New York City and Budapest, Hungary), people were more likely to die shortly after than shortly before Yom Kippur (the Jewish Day of Atonement). I think it's safe to say that Phillips' work was met with great skepticism.

To address the skeptics, Shimizu and Pelham (2008) harvested data from the same Social Security Death Index discussed elsewhere in this chapter. We reasoned that the two major holidays on which most Americans would least want to die would be Thanksgiving and Christmas. On these two major holidays many Americans go to great lengths to be with those they love. We thus reasoned that people might be less likely to die on the exact day of either of these holidays than on any of the immediately surrounding days. If people who are gravely ill try to hold on to make it to a major holiday, we also reasoned that death rates should be a little higher for the two days immediately *after* these two holidays than for the two days immediately before them. For both holidays we focused on the 16-year window between 1987 and 2002. Later data were not yet available, and earlier data did not usually include an exact date of death. Our results were very similar for Thanksgiving and Christmas, with the effect for Christmas being predictably larger.

As shown in Figure 8.5 people did defer their deaths until after Christmas Day. In this 16-year window about 4,000 more Americans died on December 27 than on December 25. Because these data provided no demographic information we

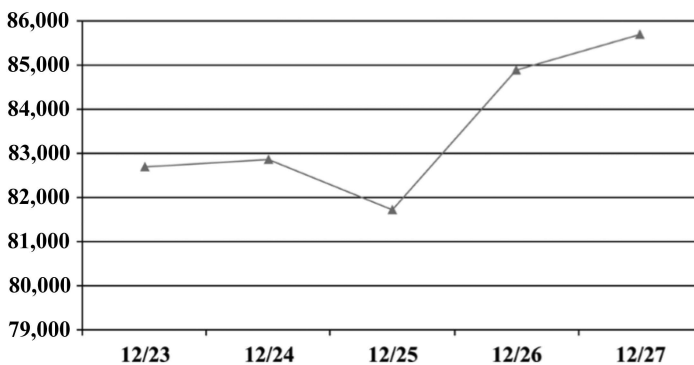


FIGURE 8.5 Number of Americans Dying on Christmas Day versus on any of the Four Neighboring Days

Source: SSDI, 1987–2002. Figure is adapted by permission from Shimizu and Pelham (2008).

were not able to control for GAGES. However, it is worth noting that the exact day on which Christmas occurs varies arbitrarily. Further, poor weather is no more likely to occur on Christmas Day than on any of the days surrounding it. Christmas also comes to women as well as to men. One worrisome confound that applies to both Thanksgiving and Christmas, however, is that people are probably much less likely to take long trips on the exact days of these holidays than on the surrounding days. In a series of supplemental analyses using California mortality data we were able to address this confound. The effects all remained when we removed all of the deaths due to accident.

We were also able to identify two potential moderators of the Christmas Day death-deferral effect. First, using Phillips and King's (1988) list of 106 actuarially Jewish surnames (e.g., Goldberg, Silverstein), we focused on Americans who were very likely to be Jewish. The pattern shown in Figure 8.5 almost completely disappeared in the Jewish surname sample. In contrast to Jews, most American (non-Jewish) children are absolutely enamored of Christmas. If the Christmas death deferral effect is grounded in a desire to experience Christmas, children should show a larger than usual Christmas death deferral effect. Although the SSDI records do not include age, it was possible to derive ages from dates of birth and death. The magnitude of the death deferral effect for children (aged 5–20) was 10 times as large as the effect observed for adults. At least some people appear to exercise at least some control over the exact timing of their own deaths.

Conclusions

Archival data are a rich source of information about human behavior. From birth to death, people leave behind a great deal of evidence of their daily behavior. As the technical barriers that restrict access to archival data become smaller, the potential for new archival discoveries keep growing. In the past year alone, I have explored archival data from the World Values Survey, DHS Surveys, the CIA Factbook, the UNDP data site, Google Correlate, and dozens of birth, census, marriage, divorce, and death records. The possibilities of using archival data extend well beyond social cognition and health psychology. Although I have focused here on three topics that interest me, archival research methods have also shed light on many other topics, from climate change to ethnic biases in birth outcomes.

I hope that as researchers seek new opportunities to conduct archival research they will focus on the crucial role of theory in scientific discovery. The best theory testing often involves testing for statistical moderation, and archival data sets are very well-suited to this—because they often involve huge and diverse samples. Archival data sources also provide opportunities for replication that are almost impossible to match with other data sets. Finally, for researchers who have difficulty with delay of gratification, archival studies can often provide rapid answers to pressing questions. Some archival studies can be done in hours if not minutes. Regardless of how long archival research takes, it can tell us not just whether a

theory holds up outside the lab but also for whom, in what way, when, where, and sometimes even why.

References

- Abel, E. L. (2010). Influence of names on career choices in medicine. *Names: A Journal of Onomastics*, 58, 65–74. doi:10.1179/002777310X12682237914945
- Abel, E. L., & Kruger, M. L. (2010). Smile intensity in photographs predicts longevity. *Psychological Science*, 21, 542–544.
- Adams, H. E., Wright, L. W., & Lohr, B. A. (1996). Is homophobia associated with homosexual arousal? *Journal of Abnormal Psychology*, 105, 440–445.
- Ariely, D. (2012). *The (honest) truth about dishonesty: How we lie to everyone—especially ourselves*. New York, NY: HarperCollins.
- Aronson, E., & Carlsmith, J. M. (1968). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (pp. 1–78). Reading, MA: Addison-Wesley.
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, 2, 412–414.
- Beggan, J. K. (1992). On the social nature of nonsocial perception: The mere ownership effect. *Journal of Personality and Social Psychology*, 62, 229–237. doi:10.1037/0022-3514.62.2.229
- Bickman, I. (1974). The social power of a uniform. *Journal of Applied Social Psychology*, 4, 47–61.
- Billing, J., & Sherman, P. W. (1998). Antimicrobial functions of spices: Why some like it hot. *The Quarterly Review of Biology*, 73, 3–49.
- Booth, A., & Amato, P. (1991). Divorce and psychological stress. *Journal of Health and Social Behavior*, 32(4), 396–407. Retrieved from <http://www.jstor.org/stable/2137106>
- Bosworth, B., Burtless, G., & Zhang, K. (2015). *Later retirement, inequality in old age, and the growing gap in longevity between rich and poor*. White paper, full text. Retrieved from www.brookings.edu
- Brick, T. R., & Boker, S. M. (2011). Correlational methods for analysis of dance movements. *Dance Research, Special Electronic Issue: Dance and Neuroscience: New Partnerships*, 29, 283–304.
- Carstensen, L. L., Isaacowitz, D. M., & Charles, S. T. (1999). Taking time seriously: A theory of socioemotional selectivity. *American Psychologist*, 54, 165–181.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Danner, D. D., Snowdon, D. A., & Friesen, W. V. (2001). Positive emotions in early life and longevity: Findings from the nun study. *Journal of Personality and Social Psychology*, 80, 804–813.
- Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.
- DeWall, C. N., Pond, R. S., Campbell, W. K., & Twenge, J. M. (2011). Tuning in to psychological change: Linguistic markers of self-focus, loneliness, anger, anti-social behavior, and misery increase over time in popular U.S. song lyrics. *Psychology of Aesthetics, Art, and Creativity*, 5, 200–207.
- Dugan, A. (2013). *Fast food still major part of U.S. diet*. Retrieved from www.gallup.com/poll/163868/fast-food-major-part-diet.aspx
- Dunlop, D. D., Song, J., Lyons, J. S., Manheim, L. M., & Chang, R. C. (2003). Racial/ethnic differences in rates of depression among preretirement adults. *American Journal of Public Health*, 93, 1945–1952.

- Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. J. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*, 383–386.
- Ernest-Jones, M., Nettle, D., & Bateson, M. (2011). Effects of eye images on everyday cooperative behavior: A field experiment. *Evolution and Human Behavior*, *32*, 172–178.
- Everett, C., Blasib, D. E., & Roberts, S. G. (2015). Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. *PNAS*, *112*, 1322–1327.
- Fincher, C. L., Thornhill, R., Murray, D. R., & Schaller, M. (2008). Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism. *Proceedings of the Royal Society*, *275*, 1279–1285.
- Gangestad, S. W., Thornhill, R., & Garver-Appara, C. E. (2010). Fertility in the cycle predicts women's interest in sexual opportunism. *Evolution and Human Behavior*, *31*, 400–411. doi:10.1016/j.evolhumbehav.2010.05.003
- Hamilton, W. D. (1964a). The genetical evolution of social behavior I. *Journal of Theoretical Biology*, *7*, 1–16.
- Hamilton, W. D. (1964b). The genetical evolution of social behavior II. *Journal of Theoretical Biology*, *7*, 17–52.
- Hearst, N., Newman, T. B., & Hulley, S. B. (1986). Delayed-effects of the military draft on mortality—A randomized natural experiment. *New England Journal of Medicine*, *314*, 620–624.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–135.
- Jones, J. M. (2015). In U.S., confidence in police lowest in 22 years. Retrieved from www.gallup.com/poll/183704/confidence-police-lowest-years.aspx
- Kitayama, S., & Karasawa, M. (1997). Implicit self-esteem in Japan: Name letters and birthday numbers. *Personality and Social Psychology Bulletin*, *23*, 736–742. doi:10.1177/0146167297237006
- Kukkonen, T., Binik, Y., Amsel, R., & Carrier, S. (2010). An evaluation of the validity of thermography as a physiological measure of sexual arousal in a nonuniversity adult sample. *Archives of Sexual Behavior*, *39*, 861–873.
- Larrick, R. P., Timmerman, T. A., Carton, A. M., & Abrevaya, J. (2014). Temper, temperature, and temptation: Heat-related retaliation in baseball. *Psychological Science*, *22*, 423–428.
- McGuire, W. J. (1989). A perspectivist approach to the strategic planning of programmatic scientific research. In B. Gholson, W. R. Shadish, Jr., R. A. Neimeyer, & A. C. Houts (Eds.), *Psychology of science: Contributions to metascience* (pp. 214–245). Cambridge: Cambridge University Press.
- Medvec, V. H., Madey, S. F., & Gilovich, T. (1995). When less is more: Counterfactual thinking and satisfaction among Olympic medalists. *Journal of Personality and Social Psychology*, *69*, 603–610.
- Mill, J. S. (2002/1863). *A system of logic*. Honolulu, HI: University Press of the Pacific.
- Mullen, B. (1983). Egocentric bias in estimates of consensus. *The Journal of Social Psychology*, *121*, 31–38.
- Murdock, G. P., & White, D. R. (2006). *Standard cross-cultural sample: On-line edition*. Working Papers Series, Permalink. Retrieved from <http://escholarship.org/uc/item/62c5c02n>
- Newport, F. (2014). *Mississippi most religious state, Vermont least religious*. Retrieved from www.gallup.com/poll/167267/mississippi-religious-vermont-least-religious-state.aspx

- Newport, F., & Pelham, B. W. (2009). *Don't worry, Be 80: Worry and stress decline with age*. Retrieved from www.gallup.com/poll/124655/dont-worry-be-80-worry-stress-decline-age.aspx
- Nuttin, J. M., Jr. (1985). Narcissism beyond Gestalt and awareness: The name letter effect. *European Journal of Social Psychology, 15*, 353–361. doi:10.1002/ejsp.2420150309
- Pelham, B. W. (2018). *Evolutionary psychology: Genes, environments, and time*. London, UK: Palgrave.
- Pelham, B. W., & Blanton, H. (2013). *Conducting research in psychology: Measuring the weight of smoke* (4th ed.). Pacific Grove, CA: Cengage Publishing.
- Pelham, B. W., & Carvallo, M. R. (2015). When Tex and Tess Carpenter build houses in Texas: Moderators of implicit egotism. *Self and Identity, 14*, 692–723.
- Pelham, B. W., Mirenberg, M. C., & Jones, J. T. (2002). Why Susie sells seashells by the seashore: Implicit egotism and major life decisions. *Journal of Personality and Social Psychology, 82*, 469–487. doi:10.1037/0022-3514.82.4.469
- Phillips, D. P., & Carstensen, L. L. (1986). Clustering of teenage suicides after television news stories about suicide. *New England Journal of Medicine, 315*, 685–689.
- Phillips, D. P., & Feldman, K. A. (1973). A dip in deaths before ceremonial occasions: Some new relationships between social integration and mortality. *American Sociological Review, 38*, 678–696.
- Phillips, D. P., & King, E. W. (1988). Death takes a holiday: Mortality surrounding major social occasions. *Lancet, 2*, 728–732.
- Pinker, S. (2010). *Better angels of our nature: Why violence has declined*. New York: Penguin Books.
- Prior, H., Schwarz, A., & Gunturkun, O. (2008). Mirror-induced behavior in the magpie (*Pica pica*): Evidence of self-recognition. *PLoS Biology, 6*, e202.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Refinetti, R. (2005). Time for sex: Nycthemeral distribution of human sexual behavior. *Journal of Circadian Rhythms, 3*. Retrieved from <http://doi.org/10.1186/1740-3391-3-4>
- Reifman, A. S., Larrick, R. P., & Fein, S. (1991). Temper and temperature on the diamond: The heat-aggression relationship in Major League Baseball. *Personality and Social Psychology Bulletin, 17*, 580–585.
- Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin, 121*, 133–148.
- Ross, L., Greene, D., & House, P. (1977). The false consensus effect: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13*, 279–301.
- Rozar, T. L. (2012, May 23). *Impact of seasonality on mortality*. Society of Actuaries Conference. Retrieved from www.rgare.com/knowledgecenter/Pages/MortalitySeasonalityPresentation.aspx
- Scheidel, W. (2009). *Disease and death in the ancient city of Rome Version 2.0*. Working Paper. Retrieved from www.princeton.edu/~pswpc/pdfs/scheidel/040901.pdf
- Sehat, M., Naieni, K. H., Asadi-Lari, M., Foroushani, A. R., & Malek-Afzali, H. (2012). Socioeconomic status and incidence of traffic accidents in metropolitan Tehran: A population-based study. *International Journal of Preventive Medicine, 3*, 181–190.
- Shimizu, M., & Pelham, B. W. (2008). Postponing a date with the grim reaper: Ceremonial events, the will to live, and mortality. *Basic and Applied Social Psychology, 30*, 36–45.

- Simon, L., Greenberg, J., Harmon-Jones, E., Solomon, S., Pyszczynski, T., Arndt, J., & Abend, T. (1997). Terror management and cognitive experiential self-theory: Evidence that terror management occurs in the experiential system. *Journal of Personality & Social Psychology*, *72*, 1132–1146.
- Sistiaga, A., Mallol, C., Galván, B., & Summons, R. E. (2014). The Neanderthal meal: A new perspective using faecal biomarkers. *PLoS ONE*, *9*, e101045.
- Smith, M. S., Kish, B. L., & Crawford, C. B. (1987). Inheritance of wealth as human kin investment. *Ethology and Sociobiology*, *8*, 171–182.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*, 24–54.
- Welling, L. L. M., Jones, B. C., DeBruine, L. M., Smith, F. G., Feinberg, D. R., Little, A. C., & Al-Dujaili, E. A. S. (2008). Men report stronger attraction to femininity in women's faces when their testosterone levels are high. *Hormones and Behavior*, *54*, 703–708.
- Willer, R., Sharkey, A., & Frey, S. (2012). Reciprocity on the hardwood: Passing patterns among professional basketball players. *PLoS ONE*, *7*(12), e49807.
- Wißing, C., Rougier, H., Crevecoeur, I., Germonpré, M., Naito, Y. I., Semal, P., & Boche-rens, H. (2015). Isotopic evidence for dietary ecology of late Neandertals in North-Western Europe. *Quaternary International*. doi:10.1016/j.quaint.2015.09.091
- Zajonc, R. B. (1965). Social facilitation. *Science*, *149*, 269–274.