

OFFICIAL AND UNOFFICIAL DATA

DREW WESTEN

Department of Psychology, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

Abstract — Unofficial data are empirical findings that guide our research but are generally not reported. This article delineates four forms of unofficial data: casual observation of ourselves and others, unsystematic naturalistic observation, uncodable forms of clinical and phenomenological data, and accidental and nonquantifiable incidents and findings arising during pilot testing and data analysis. The article argues for a broadened conception of empiricism that recognizes unofficial data as data, explores the different contexts of the scientific process in which official and unofficial data are useful, and suggests the implications of the existence and utility of unofficial data for research and publication practices.

Psychological studies are generally reported as stemming from, and building upon, previous studies. In some cases this is obviously accurate, though in many cases — and nearly always in innovative research — the official account of the history of the theory and data in the introduction section, replete with citation of the relevant studies in the relevant literatures, reflects an Orwellian revision of history. This revision is necessary to preserve the psychological community's pre-Kuhnian collective myth that science is a cumulative process in which one study builds on the next, with empirical brick placed upon empirical brick until a solid edifice of Truth has been erected.

This psycho-architectural fantasy, unfortunately, fails to account for two critical materials required for construction of good psychological research. The first is the theoretical mortar that holds the empirical bricks in place and makes them less susceptible to toppling with the slightest wind. The second is a foundation made of what I will call *unofficial data*. Unofficial data are the experiences that lead a researcher to pose a problem, formulate a concept, and pilot-test a method. These unofficial data are typically not cited, or if reported, they are described as cute anecdotes that suggested the site where the real scientific endeavor was to be built.

For example, I was recently piloting some research on affective processes and self-representations, and was trying to compare the effects of activation of affect-laden "ideal self" schemas with activation of relatively neutral self-schemas. I hoped to demonstrate that a discrepancy between observation of one's own behavior and affect-laden ideal self schemas will produce affects, and consequent responses, which differ from those evoked by a similar discrepancy between self-observation and self-schemas without comparable affective valuation. I proceeded to ask subjects to describe a situation in which they acted badly, and a situation in which they did something unusual, unexpected, or out of character. What I found, to my surprise, was that subjects had a great deal of

difficulty recalling incidents in which they surprised themselves but did not evaluate this negatively. With further prodding, some subjects produced instances of anomalous behavior which made them happy, but they were uniformly unable to produce examples to which they were affectively neutral.

Upon completing pilot testing and putting together a codable questionnaire, I noticed that something curious had happened: subjects' difficulty providing memories of cognitively anomalous behavior that were not unpleasant had become a nuisance factor which I handled methodologically by tacking "but not bad" onto the instruction to subjects to describe an event in which they had done something "unusual, unexpected, or out of character." For purposes of publication, the interesting finding about affect and self-schemas was destined to become a footnote.

I have described this event in detail not because it is unusual, but because it is so typical. It is not only common that in pursuing one phenomenon one observes another, but it is equally commonplace that the unexpected phenomenon becomes a nuisance that one tries to eliminate, and that it is either reported as an aside or is never reported at all.

Suppose I were interested enough in this nuisance to follow it up with a study. One possibility would be to record reaction time of subjects trying to retrieve neutral, positive, and negative memories of this sort or ask them to rate relative difficulty of retrieval. The likely outcome is that I would either find the expected relationship (and be one step closer to tenure) or find pilot results discouraging and abandon the project. If the results were positive, I would likely begin an article describing the study by citing the appropriate social-cognitive literature, and would probably not mention what had tipped me off to the phenomenon because that information would be neither quantitative nor replicable.

The initial unanticipated discovery (during piloting of an unrelated study) that subjects have extreme difficulty retrieving neutral self-relevant information from episodic memory is an example of a class of empirical findings I will call *unofficial data*.

An important question is what incremental knowledge is gained by performing a study once a solid foundation has been laid of unofficial data. If the results of the study I perform are positive, while I will have some nice corroborating evidence, I will not have learned anything new: I already believed the hypothesis, I already found corroborating results piloting the study, and I would not have wasted precious tenure-seeking time on the study if I did not have reason to believe — estimating the probability that I am wrong at, say, 0.05? — that results would be positive. If, on the other hand, the results are unpromising, I will indeed have learned something (either that my belief was wrong, or that I am a sloppy methodologist — my data will not tell me which), but the scientific community will never be any the wiser.

Unofficial data are data that either guide research surreptitiously, or do not guide research but should. By contrast, official data are those empirical findings reportable in scientific publications, susceptible to statistical analysis and believed to be replicable. The aim of this paper is to argue for a broadened conception of empiricism that acknowledges that both official and unofficial empirical findings

do, indeed, constitute data, and to suggest the uses and limitations of each. The article will begin by enumerating various types of unofficial data. It will then describe the different contexts of the scientific process in which official and unofficial data are most useful. It will conclude by exploring the implications of their complementarity for psychological practice.

TWO KINDS OF DATA

I will not, in a publication for professional psychologists, outline the characteristics of the official data with which we are all familiar. Instead I will delineate several forms of unofficial data, with which we are equally familiar, but of which we are sometimes officially unaware. The first is casual (i.e., unsystematic) observation of ourselves and others. This is probably the most important form of unofficial data, largely because it provides the empirical basis for the implicit theories of personality that guide most psychological research. When Darley (Darley & Latane, 1968) began his fascinating studies of bystander intervention, he did so because phenomena like the public murder of Kitty Genovese seemed so striking and anomalous. Surprise and anomaly, however, imply discrepancy between the observed and the anticipated, and the anticipated in this case stemmed from an intuitive sense that people tend to be at least somewhat concerned, empathic, altruistic, or at least civil. Darley's research is so interesting to us because we share his intuitive conceptions of human nature, although it would probably be less interesting to a fierce Nuer from the Sudan, whose cultural constructs about personality lead him to see such behavior as less surprising.

Examples of this sort could easily be multiplied. Most social psychology is derived from implicit personality theories, and cognitive research would probably benefit from an excursion into the everyday (see Rogoff & Lave, 1984), to wonder, for example, how people can suspend disbelief in theater. Similarly, it is unlikely that social cognition researchers would ever have come to see self schemas as important in mediating behavior were it not for introspection. Perhaps we should take more seriously the view of human being as intuitive psychologist (Ross, 1977; with a few corrections for affect and motivation, see Westen, 1985) as applied to professional psychologists, who do not, and should not, somehow check all their accumulated unofficial knowledge about human beings at the door when they enter their laboratories. I am not, of course, suggesting that we abandon science for intuitive science, rather, that we can best study human beings if we are aware of our own culturally and idiosyncratically constructed blinders and schemas.

A second form of unofficial data comes from naturalistic observation. If Harlow (see Harlow & Harlow, 1965) had never been to a zoo or had never read (or heard of) accounts of primate behavior in the wild, or if John Bowlby had never visited his lab and made some observations, Harlow might never have known that he was rearing abnormal monkeys in the laboratory. If Lorenz (1952) had never noticed the way goslings follow their mothers, he would never have begun exploring the phenomenon of imprinting.

A third form includes phenomenological data; subtle, less codable aspects of

raw interview data; and clinical data. At times the unofficial database of clinical work, for example, allows the posing of questions that open avenues of more empirically verifiable research which could not have been foreseen without the initial clinical observation. Clinicians working with patients with borderline personality disorders (symptoms of which include unstable work history, difficulty maintaining intimate relationships, labile affects, and impulsive behavior) consistently observed that such patients have difficulty maintaining benignly ambivalent images of significant others, and that when stressed, they begin to see the world as a malevolent place, populated with people who will use, abuse, or abandon them. This leads to an obvious causal hypothesis, that these difficulties regulating and integrating emotion play a central role in producing difficulties in close relationships. Empirical research using projective tests has subsequently been addressed to the quality of object relations (or, if one prefers, interpersonal schemata) in such people (see Kwawer, Lerner, Lerner & Sugarman, 1980), and the clinical observation of these patients may begin to suggest some questions about the organization of cognitive–affective schemas more generally which could be accessible to experimental study. These questions and hypotheses could never have been posed without the unofficial data of clinical practice. What experimentally-based theory looks for the causes of certain interpersonal difficulties in strategies of affect regulation, or suggests that a form of psychopathology could involve the storage or retrieval of multiple experiences with the same stimulus (an intimate other) under different affective valences?

A fourth type of unofficial data includes nonquantifiable observations during pilot testing, accidental findings, and responses of outliers. Our usual response to outliers while pilot testing is to try to refine our instruments so that they will be less likely to dog us in our official studies. Outliers in our official studies make us angry and lead us to search for ways to disqualify them for purposes of statistical analysis. Understanding individual differences in response to standardized procedures is critical in determining the universality of the processes we are studying, yet in statistical accounts they are generally treated as error variance. Why did some of Milgram's subjects refuse the experimenter's orders in every condition? One gets some insight into this by watching filmed interviews with subjects, but not from Milgram's data analysis. I do not mean to point a finger at Milgram or others in this respect, since the asterisk key on my own keyboard has certainly not gone untouched.

TWO CONTEXTS OF SCIENCE

It should now be sufficiently clear that unofficial data not only abound — and guide our research without official recognition — but that were we to rely only on the data we report, our science would be impoverished for it. I would now like to explore the uses and limitations of official and unofficial data, and to argue that they are complementary forms of empirical data which both deserve official sanction.

A useful distinction proposed earlier in this century by philosophers of science is between the context of scientific discovery, in which new theories and hypotheses are created, and the context of justification, in which they are put to

empirical test.* I will argue that unofficial data are, by and large, more useful in the context of discovery, and official data are more useful in the context of justification.

If one looks at the history of psychology, one uncovers the rather startling fact that of the three psychologists who have probably most influenced the questions we ask — Freud, Skinner, and Piaget — none dealt primarily with what are now considered acceptable, official data. Freud's observations and theories continue to shape the way most psychotherapists practice. Skinner's work, largely using small Ns or Ns of one to try to discover new phenomena, rather than large Ns to try to confirm already-formed hypotheses, inspired decades of research and still forms a backbone of contemporary psychological thinking. While Skinner was clearly one of the leading mainstream psychologists of his day, it is questionable whether his brand of small-N research could find a publication outlet today (and his speculations on personality would surely be deleted by any respectable editor, with the comment, "not demonstrated by your data"). Interestingly Skinner, psychology's most strident empiricist, has provided one of the liveliest and most fascinating accounts of the unofficial processes that led to his most significant public conclusions (1959). Similarly, Piaget was far more interested in observing and talking with children, formulating hypotheses, and refining them on small Ns than in testing his conclusions with large samples. One need only glance at the recent edition of the *Handbook of Child Psychology*, especially Vol. 3 (Flavell & Markman, 1983), to see how Piaget's thought has shaped the questions asked by developmentalists. This is not to say that Freud, Skinner, and Piaget were right in all or most of what they argued. A necessary consequence of being someone who proposes bold theories that inspire research and theoretical refinement is that one becomes one of the most significant purveyors of false hypotheses in the business. Yet the history of psychology suggests that it is precisely those psychologists who immerse themselves in unofficial data that thrive in the context of discovery and set the agenda for psychologists who prefer to wade in the cleaner waters of the context of justification.

Indeed, it could not be otherwise. I have already suggested that we frequently learn little that is new from official data because we are necessarily testing pre-formed hypotheses that, if we have any concern for our careers, we already have good reason to suspect are true. (Psychology is currently shaped by an unfortunate process of natural selection which selects those who can produce quickly, which means that one is more likely to test banal, relatively simple hypotheses that one is certain are correct; for empirical evidence of this selection of the unfit, see the studies cited by Mahoney, 1985.) It is true that psychologists are frequently forced to formulate new ideas when empirical researchers produce conflicting findings, and this is the main way that official data are useful in the context of discovery. Yet if a psychologist from another planet, who had

*Karl Popper (1965), in particular, has made much of this distinction, though he has tried to define the context of discovery as outside of empirical science, whereas, as someone influenced by Kuhn (1970), I am arguing that it is an integral part of science and a crucial source of empirical data.

no intuitive notions about human beings and no powers of introspection, came to earth and tried to formulate a theory of personality based on all existing official data in psychology, the theory would probably state that human beings are creatures with reaction times, who as babies frequently find themselves on visual cliffs, who as adults (i.e., age eighteen) seem determined to fill out questionnaires, and who are motivated primarily by the desire to please experimenters and to maintain cognitive consistency. The only reason we, as earthly psychologists, tend to believe differently is that we have access to unofficial data.

In the context of justification, the canons of orthodox empiricism fare much better, and official data are more valuable than unofficial data. By using large samples one can separate the necessary from the contingent, and by using experimental techniques one can more easily test to see if one's causal hypotheses are really reducible to errors of inference or sampling. Even here, however, one must be cautious in emphasizing the relative superiority of official data. First, the use of a theory by a community of practitioners (as in clinical work) is itself a selection process, however sluggish, idiosyncratic, and prone to false confirmation, that weeds out useless ideas and preserves helpful ones.

Second, as has been pointed out numerous times, the bias toward positive results in publication raises serious questions about the validity of statistical probability measures that do not take into account the number of revisions in pilot testing or unpublishable negative results. Perhaps even more significantly, philosophically sophisticated empiricists, notably Sir Karl Popper (1965), have emphasized that positive findings only corroborate, and never prove hypotheses, and that only disconfirmations — which can always be explained away methodologically, or worse still, cannot get published — are actually epistemologically decisive. From a philosophical point of view, Hume's problem of induction does not support present publication practice: after seeing a hundred white swans, one can never predict the color of the next swan with certainty, whereas if one sees a black swan, one *can* with certainty claim that the hypothesis that all swans are white is false. From a more practical and historical point of view, if one studies any hypothesis enough, one can disconfirm aspects of it, and the result of continual refinement, counter-studies, and re-refinements in psychology tends to be loss of interest in the hypothesis rather than increased positive knowledge (Meehl, 1978).

A third caveat is that corroboration and disconfirmation in the context of justification always involve inference, and one can never ascertain through official data whether to abandon the hypothesis or the methodology that failed to support it (see Feyerabend, 1975; Kuhn, 1970). As researchers, we tend to assume that our ideas are right and that if we get negative findings, we need to adjust an imperfect methodology. As readers or reviewers of papers, we typically assume that null findings (if they ever make it to print) reflect inadequate theories. The major determinant of our choice to reject either a theory or a method is probably the degree of our emotional investment in one or the other.

A final caveat is that replication, a key advantage of official data, is largely a chimera. The percentage of studies actually replicated is surely less than one

hundredth of one percent, and the percentage *successfully* replicated reduces that number even further — making it far smaller, I suspect, than the number of times a patient will return to psychotherapy with a new clinician and receive an independent and converging diagnosis of symptoms and etiology.

IMPLICATIONS FOR PSYCHOLOGY AS A SCIENCE

Two practical implications seem to follow from the argument for the complementarity of official and unofficial data, one relating to research practices, and the other, to publication practices. First, the field could greatly benefit from an official sanctioning of interpretive and quasi-deductive tools in our scientific armamentarium. In the other social sciences, investigators like anthropologist Clifford Geertz (1973, 1983) have put hermeneutic methods to good use and have consequently enriched their disciplines. Geertz's essay on common sense as a cultural system (1983) is a good example of the use of symbolic analysis in social science, and cognitive psychology could make considerable use of Geertz's analysis in forming testable hypotheses about everyday cognition, the influence on cognitive processes of culturally constituted "common sense," and intuitive science. The philosopher of law H. L. A. Hart (1961) poses an instructive scenario: if one wished to make an empirical investigation of why people stop at red lights and move forward in their automobiles when the light turns green, one would do well to move beyond the causal statement, amenable to statistical analysis, that red causes stopping and green causes going, by looking at people's *reasons* for stopping or going. This would require an interpretive analysis of how people follow social rules (i.e., looking at the *meaning* of the color based on social conventions), rather than a quantitative analysis of how their behavior was caused. It is important to note that interpretive analysis of this sort involves no less use of empirical data than the quantitative, causal analysis officially sanctioned by the field. It does require more inference, but since as a field we are currently so fascinated with inference processes in our subjects (e.g., Nisbett & Ross, 1980), perhaps we could show more interest in the inference processes of our colleagues.

Similarly, a social cognition researcher interested in self schemas could profitably complement reaction-time and questionnaire studies with a broad reading of anthropological research in ethnoperpersonality, a subdiscipline that explores cultural conceptions of self (see Westen, 1985; Shweder and Bourne, 1982; Smith, 1978). A psychologist interested in moral development could do a careful exegesis, perhaps even using comparative data, of systems of moral belief embedded in myths, in order to explore the alleged universality of certain patterns and the influence of ecology and socialization on moral reasoning.

The implications for publication follow, in part, from the implications for research. A comparative analysis of representations of self in Homer, Lucretius, Flaubert, and Camus should clearly be of interest to psychologists interested in the interplay of self schemas and historical change, and should potentially be publishable in journals concerned with social cognition, such as the *Journal of Personality and Social Psychology*, *Cognition*, or *Social Cognition*. The various forms of unofficial data enumerated here should be publicly acknowledged for what

they are — irreplaceable sources of empirical data in the context of discovery — and should be treated accordingly. Observations in the context of discovery should be reported and examined in discussion sections of experimental studies to enrich quantitative findings and to suggest avenues for future research. Finally, studies that offer, *alongside* quantitative normative data and statistical treatments, qualitative analyses of responses of subjects who do not conform to statistically corroborated results should be given priority for publication over similar studies that provide no insight into individual differences.

I am essentially arguing that we make conscious our scientific unconscious, which guides our research and our theories whether we are aware of it or not. By making our unofficial data official, we could more clearly and self-consciously explore the cultural and personal assumptions underlying our official research, and would therefore be in a better position to consider useful alternatives and revisions that would otherwise remain inaccessible.

In trying to distinguish itself from philosophy and gain a respectability comparable to the natural sciences, psychology has swung too far in its canons of method and has officially sanctioned a narrow form of empiricism that in reality has not, cannot, and should not be institutionalized. Since the rise of behaviorism in the early part of the century, psychologists have treated their own practices in the context of discovery as a black box, reporting and acknowledging the existence only of their final products. It is indeed ironic that a discipline currently so preoccupied with cognition could show so little interest in the inferences, deductions, and analogies that shape our research, and the unofficial data that underlie these complex and remarkable cognitive events.

REFERENCES

- Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, *8*, 377–383.
- Feyerabend, P. (1975). *Against method*. London: Verso.
- Flavell, J. H., & Markman, E. M. (Vol. Eds.) (1983). P. H. Mussen, (1983) (Ed.), *Handbook of child psychology*, Vol. 3. *Cognitive development*. N.Y.: Wiley.
- Geertz, C. (1973). *The interpretation of cultures*. N.Y.: Basic Books.
- Geertz, C. (1983). *Local knowledge: Further essays in interpretive anthropology*. N.Y.: Basic Books.
- Harlow, H., & Harlow, M. (1965). The affectional systems. In A. Schrier, H. Harlow, & F. Stullnitz (Eds.), *Behavior of nonhuman primates*, Vol. 2. NY: Academic Press.
- Hart, H. L. A. (1961). *The concept of law*. Oxford: Clarendon Press.
- Kuhn, T. (1970). *The structure of scientific revolutions* (2nd edition). Chicago: University of Chicago Press.
- Kwawer, J. S., Lerner, H. D., Lerner, P. M., & Sugarman, A. (1980). *Borderline phenomena and the Rorschach test*. NY: International Universities Press.
- Lorenz, K. (1952). *King Solomon's ring*. London: Methuen.
- Mahoney, M. (1985). Open exchange and epistemic progress. *American Psychologist*, *40*, 29–39.
- Meehl, P. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*, 806–834.
- Nisbett, R. E., & Ross, L. C. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, New Jersey: Prentice-Hall.

- Popper, K. (1965). *Conjectures and refutations* (2nd edition). London: Routledge and Kegan Paul.
- Rogoff, B., & Lave, J. (Eds.). (1984). *Everyday cognition: Its development in social context*. Cambridge: Harvard University Press.
- Ross, L. (1977). The intuitive psychologist and his shortcomings. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, Vol. 10. N.Y.: Academic Press.
- Shweder, R. A. & Bourne, E. J. (1982). Does the concept of the person vary cross-culturally? In A. J. Marsella & G. M. White (Eds.), *Conceptions of mental health and therapy*. Boston: D. Reidel.
- Skinner, B. F. (1959). A case history in scientific method. In S. Koch (Ed.), *Psychology: A study of a science*, Vol. 2. NY: McGraw-Hill.
- Smith, M. B. (1978). Perspectives on selfhood. *American Psychologist*, **33**, 1053–1063.
- Westen, D. (1985). *Self and society: Narcissism, collectivism, and the development of morals*. Cambridge: Cambridge University Press.