A process model of replication studies: on the relation between different types of replication

IJzendoorn, M.H. van

CHAPTER 3

A Process Model Of Replication Studies: On The Relation Between Different Types Of Replication*

Marinus H. van IJzendoorn
Center for Child and Family Studies
Leiden University
The Netherlands

INTRODUCTION

In this chapter we argue that the "received view" of replication as the exact or algorithmic repetition of an original study has become obsolete. Exact repetition only appears to be a boundary case of a series of more or less "varied" replications, that is, replications that systematically vary one or more parameters of the original study to see whether its outcome remains stable or changes in a predictable way. Only in case of failure to produce any predicted outcome through a series of varied replications does the suspicion of an irreproducible effect arise, and therefore the need for a more or less exact replication. From this "constructivistic" perspective it is to be expected that exact replications are rather rare, although methodologically replicability is considered to be one of the most important cornerstones of science. The production of so-called "objective knowledge" (Popper, 1980) would be dependent on the intersubjectivity of observations in accordance with the hypothesis. Intersubjectivity of observations means that observational reports would only be taken seriously if they could in principle be reproduced by any competent researcher. A theory would only be considered falsified in case of a replicable observation contradicting one of the implications of that

*This chapter was partly written during a stay at the National Institute of Child Health and Human Development, Bethesda, MD, as a senior Fulbright fellow. Preparation of this chapter was supported in part by the Netherlands Organization of Scientific Research (NWO) through a Pioneer grant. The author would like to thank Michael Lamb for his constructive criticism on an earlier version of this chapter.

theory (Popper, 1980). We argue here that the central idea of replicability should not be interpreted literally, that is, as the necessity of exact replications. Only if we take a more constructivistic perspective does it become possible to notice replication studies being carried out as part and parcel of everyday research. To emphasize its central role in methodology as well as in research practice, we present a process model of replication in which different kinds of replication efforts are systematically interrelated. Against the background of this process model of replication it becomes clear that research programs allow for continuous replications of their data and results, although isolated studies may remain unreplicated.

LACK OF EXACT REPLICATIONS AND THEIR IMPACT

It seems as if a tremendous lack of replication studies exists in the natural sciences as well as in the behavioral and social sciences. A few examples may illustrate this point for the behavioral and social sciences. The famous Harlow (1958) experiments with primates raised by an "iron" mother or by a "cloth" mother never have been exactly replicated, although their impact on the theory of primary attachments in animals and in humans have been substantial (Paul & Blumenthal, 1989). According to L.L. Ainsworth (1984) we even should doubt the possibility of replicating the experiments because their design has been described with too little detail. Maybe more influential still is J.B. Watson's conditioning experiment with Little Albert. This experiment showing the force of conditioning in human learning has never been replicated, neither by Watson himself nor by his numerous students and followers. Nevertheless there is some reason to believe that the experiment's exact replicability should be doubted. Its design and procedures have been described in a superficial way, and with sometimes contradicting details. Furthermore, the experiment with Albert was preceded by many failures to show that operant conditioning played a major role in human learning (Samelson, 1980).

Sometimes very influential studies are replicated after a long period of time. Examples are: Burt's twin studies on the heredity of intelligence, Mead's Samoa study focusing on the cultural context of adolescence (see Chapter 2), and Efron's media-sociological study on the political biases in television news about the candidates in an American presidential election campaign. On the basis of a detailed and scrutinous (secondary) analysis of Burt's data set, Kamin (1974) concluded that Burt had at least partially faked his data and had invented imaginary subjects and experimenters. Freeman (1983) showed that Mead probably had not taken into account all relevant and available information, and had sketched a somewhat unidimensional and optimistic picture of an idyllic tropical island. Stevenson and his colleagues (Stevenson, Eisinger, Feinberg, & Kotok, 1973) reanalyzed part of Efron's raw data after coding these data with a more detailed and reliable coding system. They concluded that Efron's theory could be falsified by her own data.

These replications seemed to disconfirm at least partially the original studies, but they could not restrict their influence. On the contrary, the replication studies provoked sometimes heated discussions about the merits of replication (Holmes, 1983; Kloos, 1988), about the meaning and possibility of "real" replications, and about their impact on scientific discourse (Collins, 1985).

It would seem, then, that replication studies do not seriously influence discourse about the value of certain theories and their underlying data, and that the request for more replications remained unanswered. Caplow (1982) found that about 1% of the sociological studies is replicated. Shaver and Norton (1980) observed that replications are nearly absent in educational research: They found that only 14% of 151 research projects reported on in the *American Educational Research Journal* could be considered replications. Bahr and his colleagues (Bahr, Caplow, & Chadwick, 1983) traced 300 replications (half of which are from psychology) in the Social Science Citation Index over a period of almost 10 years (1973–1981), looking for the concept of replication or its equivalents in the title of the papers. Studies that refer explicitly to the concept of replication may be often considered exact replications. It seems, therefore, that only a small amount of behavioral and social research is being exactly replicated, and that even very influential studies remain unreplicated in the strict sense. Furthermore, if a study has been replicated, and results diverge from the original outcome, more discussion about the merits of the replication and its author(s) seems to be provoked than about the original study and the theory by which it was originated. On the one hand, replications are being considered the cornerstones of modern science, but on the other hand they seem to be carried out seldomly and they seem to have little impact. How can this paradox be explained? We think that a naive definition of replication as exact repetition of a single original study is the reason replications remain unnoticed.

REPLICATIONS IN THEORY: ALGORITHMIC REPLICATIONS

Many researchers think that studies can be replicated in an exact way (Collins, 1985). These so-called exact or algorithmic replications would be carried out strictly according to the technical rules of the original author(s); an identical outcome should consequently be attained (Collins, 1975). From a methodological perspective, however, such replications cannot exist, simply because technical guidelines will always leave room for minor variations, and because spatiotemporal parameters will have changed anyway (Musgrave, 1975). Whether the same outcome will be produced will, therefore, depend on the robustness of the original results in the presence of minor variations in the research design, and of spatiotemporal variations. A replication will be called an exact replication if it is essentially similar to the original study. Every comparison, however, is being made from a certain perspective (Popper, 1980, p. 421). As we have said, replications necessarily change the design of the original study, but on the basis of available background knowledge it will be assumed that the changing parameters will not have any influence on the outcome. For example, in a replication of a study on the conditioning of emotions, spatial parameters are being considered irrelevant; that is, the conditioning study may be carried out in different parts of the world and is nevertheless supposed to yield the same kind of results.

Our background knowledge, however, is not infallible, and can be considered "true" only for the time being (Popper, 1980). If the original study and its replication show divergent results, differing parameters constitute possible alternative hypotheses to interpret the divergencies. With respect to cultural-anthropological research, for example, Kloos (Chapter 2) composed a list of important parameters: Replications usually are being carried out in another social group, historically in another society, from a different cultural and personal background of the researcher, and often also from a different theoretical perspective. Differences between Mead's (1961) and Freeman's (1983) study, for example, may be explained by the differences in social group in which both researchers carried out their observations and interviews, differences in historical period (before and after the second World War), and differences in theoretical perspective (cultural determinism versus sociobiology). Last but not least, it certainly made a difference whether a woman or a man carried out interviews on sexual issues (Chapter 2).

If little knowledge about a certain domain exists, and if therefore our background knowledge is still minimal, alternative hypotheses about the causes or reasons for divergent replication results will be mushrooming. In uncharted territory small differences in age of subjects, in apparatus, in physical environment, in psychological atmosphere, and in the experimenter's personality may imply cumulative differences in outcome (Smith, 1970). For example, research on attachment relationships between a caregiver and an infant is usually based upon the Strange Situation procedure (Ainsworth, Blehar, Waters, & Wall, 1978). This procedure consists of a series of infant's separations from and reunions with the caregiver in a strange environment, as well as a confrontation with an unknown person who tries to interact with the infant. This procedure is often used in research on the effects of infant day care and other practically relevant issues, but it is not clear to what extent variations in playroom, play material, personality, and sex of experimenter and "stranger," and the order in which the procedure is carried out during a measurement session influence coding results. In the influential Bielefeld study (Grossmann & Grossmann, 1990), the large playroom with the life-sized play material may be hypothesized to be one of the causes of the infants' seemingly avoidant behavior to the parent when they enter the room after being away for 3 minutes (Sagi, Van IJzendoorn, & Koren, 1989). It also has been suggested that the overrepresentation of resistantly attached infants in Israeli kibbutzim may partly be explained by the stressful sociability test that preceded the Strange Situation (Sagi et al., 1989). Because a strong background theory about essential ingredients of the Strange Situation is lacking, many alternative hypotheses are possible in case of diverging attachment classification distributions.

In sum, strictly seen it is impossible to copy an original study in the strict sense of the word; necessarily some potentially relevant parameters will inevitably change, and our background knowledge determines whether these changes are essential or not. The stronger our background knowledge is, that is, the more frequently confirmed or at least not falsified this knowledge is, the more exact our replications may be considered to be. If background knowledge implies that spatiotemporal parameters are irrelevant in producing a certain effect or outcome, the experiment can more easily be replicated: In this case, we do not have to control for evasive spatiotemporal variables, as these variables cannot influence our evaluation of the identity of the original and the replication study.

THE PRACTICE OF REPLICATION: A CONSTRUCTIVISTIC VIEW

The constructivistic view on replication studies is based on sociological research in scientific laboratories. Through participant observation in eminent laboratories, and through in-depth interviews with leading scientists, a description of the practice of replication has been developed that contradicts "received views" of the replication process. Research on "gravitational radiation" (Collins, 1975), on the construction of the "ETA laser" (Collins, 1985), and on conditioning of "planarian worms" (Travis, 1981), for example, has been the object of sociological studies. The results of these studies can be summarized as follows: First, interviews show that exact replications do occur only very infrequently in the natural sciences, and they constitute a low-status activity (Mulkay, 1986). Researchers are much more inclined to change some parameters of the original experiment systematically, so as to be able to discover new applications, and they are much less inclined to make a "carbon copy" of the study. The outcome of the original study is accepted—until the contrary has been proven. A falsification appears to be taken into consideration only if an outcome repeatedly cannot be established, although the variations on the original theme seem to be minor. Exact or algorithmic replications should be considered a boundary case of "varied" replication. Starting with rather large variations of the design of an important study, researchers first try to add something new to the original outcome, for example by showing that the same effect can be produced in some other species or with another chemical substance. If such larger variations do not yield the expected results, more refined variations are being practiced, until some doubt about the validity of the original study arises. In the end, "exact" replication is applied to (dis-)confirm the doubts, and to check the assumptions of the varied replications. In sum, these social studies imply that many scientists feel that exact replications may be carried out, but usually are irrelevant for scientific progress.

Participant observations in laboratories also show that exact replications are seldom carried out, and that every replication is subject to a negotiation process on the relevance of changed parameters for diverging results (Collins, 1985). In other words, replications are not accepted without discussion about the relevance of changed parameters for the production of (un-)expected results. The so-called "enculturation model" is assumed to be a more adequate description of replication than the algorithmic model. The enculturation model emphasizes the process of socialization of the researchers who implicitly learn to differentiate between relevant and irrelevant parameters. Complicated measurement procedures are not considered to be replicable by just reading a pertinent paper. Because of lack of space, in published papers essential details about the procedures are being left out; sometimes the researcher intentionally gives a global description to defend him- or herself against premature and incompetent replication, or to prevent colleagues from producing new facts before he or she finds time to "discover" them. Researchers who want to apply the procedure are urged to contact their colleague for further information, and maybe even for training through which unreported presuppositions and details are being transferred. Until such enculturation has taken place, a researcher who cannot replicate a certain result may expect to be accused of being not competent enough to carry out the experiment (Travis, 1981).

In the case of attachment theory, the same enculturation process appears to exist for procedures like the Strange Situation (Ainsworth et al., 1978), and the Adult Attachment Interview (Main & Goldwyn, in press). A training of several weeks in one of the American research centers is considered necessary for the reliable and valid coding of the observations or interviews, and therefore for a plausible and persuasive contribution to the international discourse on attachment. When central theses of attachment theory are in danger of being falsified by "untrained" researchers, this "incompetence" and lack of enculturation will be explicitly used against the "dissident" (see, for an example, Waters, 1983). On the one hand, such arguments appear to belong to the category of *ad hominem* arguments (Fogelin, 1987), that should be forbidden in rational scientific discourse; on the other hand, however, some measurement procedures can be so complicated that an intensive training indeed is necessary to get insight into essential elements, to be able not only to code the material reliably, but also in agreement with the investigators who constructed the procedure. Sometimes it is hard to know which procedural variations indeed are relevant and which are irrelevant, unless the "novice" is being instructed in personal training sessions with the constructors. Background knowledge cannot always be made explicit as Popper (1980) assumed. We should leave room for "tacit knowledge" (Polyani, 1973) that can be crucial for the adequate application of complicated measures, not only in the domain of the natural sciences but also in that of the behavioral sciences.

Accepting that background knowledge always remains fallible, and partly implicit, and that, therefore, the application of measurement procedures is subject to an enculturation process the effectiveness of which can be doubted, the possibility of a so-called "experimenter regression" has to be considered. This paradox means that replication logically can not be the ultimate test of a statement's truth. Replication is a matter of trained application of measurement procedures that cannot be described algorithmically. Therefore, the application of the procedures always can be criticized, and, in fact, a further test would be necessary to see whether the quality of the replication meets the accepted standards or not. But this test would also be based on a complicated measurement procedure, and would therefore again be subject to criticism. The "experimenter" regression could be stopped if the final decision about the adequateness of a measure's application would be taken by the investigator who constructed the measure. Such a short circuit, however, would imply methodological solipsism: Every researcher would in the end have the right to decide about the truth value of his or her own propositions. Constructivists, however, locate the decision on a social level: The scientific forum—or at least its most powerful part—is supposed to finally decide about the validity of the replication. Collins (1985) and Travis (1981) describe this process of consensus construction as a rhetorical discourse resulting in a paradigm shift or continuation (Kuhn, 1962). Replications are social constructions; they are the final product of a discourse on what is considered adequate and relevant in a certain domain of research.

THE PARADOX OF REPLICATION AS NEGOTIATION

The constructivistic argumentation for replications as the result of negotiations is elegant. According to constructivists, participant observation studies in different domains of natural science have shown that replications can seldom be carried out algorithmically, and that replication results are considered "true" only after intensive debate about relevant parameters. This view on replication appears to demystify the idea of exact replications in the natural sciences, and in this respect seems to close the gap between the social and the natural sciences. If we apply the constructivistic sociology of science on its own methods and results, however, a paradox seems to be implied (Mulkay, 1984). Through several replications it was established that replication studies constitute contingent social constructions, and can never be considered as unproblematic validations of knowledge claims. The same holds true for constructivistic replication studies: In these cases, too, rhetorics are being used to make clear that several different replications did only differ in irrelevant aspects, and, therefore, should be seen as real replications. Metastudies should not be exempted from the constructivistic interpretation of replication. And if every researcher—or better still, every forum of researchers—would construe his or its own interpretation of reality and, therefore, his or its own idea of replication, why should we take the constructivistic view more seriously than the view of the researchers in the field?

Although constructivism criticized the concept of exact replication, and showed the relativity of replication outcome, this theory of science did not succeed in rejecting replication as a regulatory criterion for science. Replications can always be heavily discussed because of changing parameters explaining divergent results, but replication studies remain a very strong motive for the spiral of ever more detailed discussions about knowledge claims, and the studies that constitute their foundation. For example, the constructivist Collins (1985) still concludes that "Replicability . . . is the Supreme Court of the scientific system." The simplistic view on replication as the ultimate criterion for the validity of the original study, however, should be replaced by the idea that all results of scientific research, and, therefore, also the replication results, can only get their impact on scientific developments through the medium of discourse. Replications may lead to refinements of original results, or to their rejection, but only as a result of debates in the scientific forum. A study that fails to replicate the original results, or that succeeds in replicating them, cannot determine the outcome of the scientific discourse. The possibility of replication, however, appears to be a condition sine qua non if a study is to be taken seriously in the discourse. That is, every research project should in principle be replicable; if this is not the case, the study should not be taken into account by the forum (Popper, 1980). Whether a study will in fact be replicated is a different question. Only if a series of varied replications does not succeed in producing new "facts," researchers feel urged to replicate the original as exactly as possible. And in that case, it would still be difficult to know whether an exact replication is being carried out: Replications never completely succeed in controlling all potentially relevant parameters (for example, spatiotemporal parameters). Therefore, the exactness of the replication is the result of competent researchers discussing the replication against the background of available knowledge in the field.

A PROCESS MODEL FOR REPLICATION

Replications can be classified in several different ways. Lykken (1968) for example introduced the concepts of "literal replication," "operational replication," and "constructive replication," and they constitute a continuum along which ever more parameters are being varied. La Sorte (1972) developed a somewhat more extensive classification that also includes research on the validity of measures and longitudinal research (repeated measurements). Bahr et al. (1983) present the most systematic and complete classification. They differentiate four important parameters that may vary in replication studies—time, place, method, and sample—and they show that combinations of these parameters result in 16 different kinds of replication. These types of replication can be placed on a continuum from constant parameters (the exact replication) to completely different parameters. They found that about 27% of 300 replications could be classified in the category: constant sample and method, but different time and place. About 21% could be classified as keeping the sample characteristics constant, and changing all the other parameters. These two types of replication, therefore, include about half of the replication studies.

Such classifications do not take into account two important types of replication—that is, secondary analysis and meta-analysis. Secondary analysis is a kind of replication in which all parameters except the researcher and the method of data analysis are kept constant. Secondary analysis starts with the data as collected in a certain way in a certain sample. The replication of Burt's studies by Kamin (1974) in fact can be considered a secondary analysis, in which only Burt's statistics were tested against probability. If such a secondary analysis is carried out with minimal methodical and theoretical variations, and nevertheless leads to divergent results, it can be considered one of the most powerful falsifications through replication. Secondary analysis also is one of the most inexpensive and efficient types of replication, because it is based on existing data sets. One of the main barriers to secondary replication is, however, the accessibility of the original data sets. Wolins (1962) reported that 24% of 37 authors were willing to make their data sets available. Twenty-one researchers claimed that their data sets were lost. Craig and Reesen (1973) found that 38% of 53 researchers were unwilling to make their raw data available for secondary analysis (see also Miedema, 1986).

The results of the secondary analyses are sometimes rather disappointing.

Results do not always appear to be replicable, because the first researcher used other methods that now sometimes seem outdated (Bryant & Wortman, 1978). We would like to state, therefore, that secondary analysis should have the status of an intermediate type of research situated between the original study and its actual replication. If secondary analysis of the existing data set already leads to falsification of the results, further replication including data collection should be considered an intolerable waste of time and money (see Figure 3.1).

Secondary analysis can be subclassified in two categories. First, secondary analysis may include recoding of the original raw data. Stevenson et al. (1973), for example, did recode part of the Efron's material on media reports of a presidential campaign. We propose to call this type of replication the "complete secondary analysis," because two phases of processing the raw data are involved: the coding and analyzing of the data. In many studies in the behavioral sciences data are recorded on video- or audiotape; such raw data always can be used for a complete secondary analysis, at least during the time that the recordings still have a reasonable quality. Second, secondary analysis may be restricted to coded data; in this case, we propose to use the concept of "restricted secondary analysis." In this type of secondary analysis the coding system is not changed but only the methods of analyzing the data, to see whether the original results survive statistical criticism or the application of refined methods of statistical analysis (see, for example, Cronbach & Webb's [1975] unsuccessful reanalysis of a classical Aptitude-Treatment Interaction study).

If the restricted or complete type of secondary analysis shows convergent results, replications should be carried out in which new data under different conditions are being collected (see Figure 3.1). From the start, the original study will be "trusted" so much that rather significant variations in the design will be applied. Larger variations may lead to more interesting discoveries in addition to the original study, but they will be followed by smaller variations if more global replications fail to produce new "facts." This process of varied replications seems to be inherent to modern science, and to be especially characteristic for the phase of "normal" science (Kuhn, 1962). This type of replication is not algorithmically carried out, but it is tried, on the contrary, to determine the robustness of the original results for spatiotemporal and other variations, and possibly to refine the original results with additional hypotheses. If even modest variations fail to reproduce the results, a more or less exact replication is needed (see Figure 3.1).

We consider meta-analyses to be replications because these analyses test the replicability of the original study in a series of varied replications. Until recently, the integration of the results of a great number of empirical studies on a certain domain seemed exclusively to be a matter of qualitative analysis in which the reviewer structures his material intuitively, and reaches a global conclusion. In the last decade a trend toward formalizing this qualitative process by means of meta-analyses can be observed (Glass, McGaw, & Smith, 1981; Chapter 6). Through meta-analysis a combined effect size for different studies may be computed that gives a more precise and replicable description of the outcome of a series of related studies (Rosenthal, 1991). Furthermore, a meta-analysis can trace the effects of systematic variations in design on the outcome of a series of studies, thereby locating the spatiotemporal boundaries of replicability.

In attachment theory, for example, several instructive examples of such meta-analyses can be found. Goldsmith and Alansky (1987) showed through meta-analysis that the average effect size for the relation between maternal responsiveness and infants' attachment classification appeared to be much smaller in a series of replications (with more or less strong variations in methods and samples) than Ainsworth and her colleagues (1978) had found in their original study. Although this meta-analysis was based on studies that were very divergent in quality of design and instrument, its outcome should be taken into account as a warning not to foreclose the discussion about the central theme of attachment and responsiveness. Van IJzendoorn and Kroonenberg (1988) were able to show through meta-analysis that cross-cultural differences of attachment classification distributions were smaller than the intracultural differences. Before this meta-analysis it was assumed that distributions were culture-specific, just on the basis of a few samples from Germany, Japan, and Israel. McCartney and Phillips (1988) showed that the average effect size for the relation between type of care (day care or "homecare") and quality of attachment, specifically avoidance to the caregiver, was very small, although a few isolated studies seemed to have proven a negative effect of day care on attachment. The authors also fruitfully used the variations in replication studies to analyze their influence on the relation between type of care and attachment. They showed that neither sex of experimenter nor sex and age of infant were related to the effect size, but that it did matter whether the coders were blind to the infants' membership of type of care. If the coders were not blind to group membership, effect sizes were relatively higher and pointed toward a negative influence of day care on attachment quality (McCartney & Phillips, 1988, p. 167). In this way, varied replications can be combined to get insight into the overall effect size as well as into the influences of the variations on the relations studied.

As is shown by Lytton (Chapter 6), however, meta-analyses require many decisions to be taken in several different parts of the study. These decisions may or may not be made explicit, and other meta-analysts may or may not reach the same conclusions based upon the same "raw data" but departing from different presuppositions and decision rules. For example, in a meta-analysis on parental punishment strategies the exact definition of this construct should be given, and the researcher should decide what measures are considered to be valid operationalizations of this construct. Different meta-analysts may propose different definitions and measures, and therefore select different (sub-)samples of pertinent studies, yielding different effect sizes (Chapter 6). Therefore, meta-analyses should also be replicated, and through variations in decision rules, it should be tested how robust meta-analytic outcomes in fact are. Only after intensive discourse on implicit and explicit decision rules—against the background of available knowledge (temporarily) considered true—the scientific forum may decide on one or the other outcome, and consider this result as (provisional) part of true background knowledge (see Figure 3.1).

CONCLUSIONS

Although many researchers and philosophers of science think that replications represent an important cornerstone of science, much confusion exists as to what we should consider replication studies to be, and what kinds of replication efforts can be distinguished. In this chapter, we tried to show that replications should not be narrowed down to the rather small category of exact or algorithmic replications. Usually replications consist of variations upon the theme of the original study, and the scientific discourse focuses upon the question in what sense it may theoretically be supposed that varied

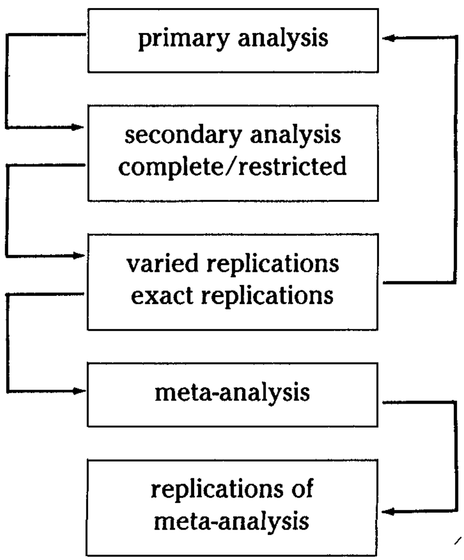

Figure 3.1. A process model of different types of replications.

replications produce results converging with those of the original study. In a broader interpretation of replication, we should not only consider the partial or complete repetition of the original study through secondary analysis or through collecting new data as "real" replication, but also meta-analysis of a series of related studies. Through meta-analysis it can be tested what variations in relevant parameters influence the results, and whether the outcome of the original study indeed can be considered to be replicated—not in an isolated replication study, but in a series of varied replications. In our process model of replication, the relations between the different kinds of replication studies have been shown, and the discursive context of replications has been emphasized. Replicability is a necessary condition for every study that is intended to have an impact on the scientific discourse, but an exact replication is only a boundary case indicating that varied replications did not yield fruitful discussions about the theory to which the original study contributed.

REFERENCES

Ainsworth, L. L. (1984). Contact comfort: A reconsideration of the original work. *Psychological Reports*, *55*, 943–949.

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment*. Hillsdale, NJ: Erlbaum.

Bahr, H. M., Caplow, T., & Chadwick, B. A. (1983). Middletown III: Problems of replication, longitudinal measurement, and triangulation. *Annual Review of Sociology*, *9*, 243–264.

Bryant, F. B., & Wortman, P. M. (1978). Secondary analysis. The case for data archives. *American Psychologist*, *33*, 381–387.

Caplow, T. (1982). La répétition des enquetes: Une methode de recherche sociologique. *L'Année Sociologique*, *32*, 9–22.

Collins, H. M. (1975). The seven sexes: A study in the sociology of a phenomenon, or the replication of experiments in physics. *Sociology*, *9*, 205–224.

Collins, M. H. (1981). Son of seven sexes: The social destruction of a physical phenomenon. *Social Studies of Science*, *11*, 33–62.

Collins, H. M. (1985). *Changing order. Replication and induction in scientific practice*. London: Sage.

Craig, J. R., & Reese, S. C. (1973). Retention of raw data: A problem revisited. *American Psychologist*, *28*, 723–730.

Cronbach, L. J., & Webb, N. (1975). Between-class and within-class effects in a reported Aptitude x Treatment interaction: Reanalysis of a study by G.L. Anderson. *Journal of Educational Psychology*, *67*, 717–772.

Fogelin, R. J. (1987). *Understanding arguments. An introduction to informal logic*. New York: Harcourt.

Freeman, D. (1983). *Margaret Mead and Samoa. The making and unmaking of an anthropological myth*. Cambridge, MA: Harvard University Press.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills: Sage.

Goldsmith, H. H., & Alansky, J. A. (1987). Maternal and infant temperamental predictors of attachment: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, *55*, 805–816.

Grossmann, K. E., & Grossmann, K. (1990). The wider concept of attachment in cross-cultural research. *Human Development, 33*, 31–48.
Harlow, H. F. (1958). The nature of love. *American Psychologist, 13*, 673–685.
Holmes, L. (1983). A tale of two studies. *American Anthropologist, 85*, 929–936.
Hunt, K. (1975). Do we really need more replications? *Psychological Reports, 36*, 587–593.
Kamin, L. J. (1974). *The Science and Politics of I. Q.* New York: Wiley.
Kloos, P. (1988). *Door het oog van de antropoloog. Botsende visies bij heronderzoek.* Muiderberg: Coutinho.
Kuhn, T. S. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press.
La Sorte, M. A. (1972). Replication as a verification technique in survey research: A paradigm. *The Sociological Quarterly, 13*, 218–227.
Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70*, 151–159.
Main, M., & Goldwyn, R. (in press). Interview-based adult attachment classifications: Related to infant-mother and infant-father attachment. *Developmental Psychology.*
McCartney, K., & Phillips, D. (1988). Motherhood and child care. In B. Birns & D. F. Hay (Eds.), *The different faces of motherhood* (pp. 157–183). New York: Plenum.
Mead, M. (1961). *Coming of age in Samoa.* New York: The New American Library.
Miedema, s. (1986). *Kennen en handelen. Bijdragen aan het theorie-praktijk-debat in de opvoedingswetenschap* [Knowledge of action]. Leuven: Acco.
Mulkay, M. (1984). The scientist talks back; A one-act play, with a moral, about replication in science and reflexivity in sociology. *Social Studies of Science, 14*, 265- 282.
Mulkay, M. (1986). Relication and more replication. *Philosophy of Social Science, 16*, 21–37.
Musgrave, A. (1975). Popper and "diminishing returns from repeated tests." *Australian Journal of Philosophy, 53*, 248–253.
Paul, D. B., & Blumenthal, A.L. (1989). On the trail of little Albert. *The Psychological Record, 39*, 547–553.
Polyani, M. (1973). *Personal knowledge. Toward a post-critical Philosophy.* London: Routledge & Kegan Paul.
Popper, K. R. (1980). *The logic of scientific discovery.* London: Hutchinson.
Rosenthal, R. (1991). *Meta-analytic procedures for social research* (rev. ed.). Newbury Park, CA: Sage.
Sagi, A., Van IJzendoorn, M. H., & Koren, K. (1991). *Primary appraisal of the Strange Situation: A cross-cultural analysis of preseparation episodes. Developmental Psychology, 27*, 587–596.
Samelson, F. (1980). J.B. Watson's little Albert, Cyril Burt's twins, and the need for a critical science. *American Psychologist, 35*, 619–625.
Shaver, J. P., & Norton, R. S. (1980). Randomness and replication in ten years of the American Educational Research Journal. *Educational Researcher, 9*, 9–15.
Smith, N. C. (1970). Replication studies: A neglected aspect of psychological research. *American Psychologist, 25*, 970–975.
Stevenson, R. L., Eisinger, R. A., Feinberg, B. M., & Kotok, A. B. (1973). Untwisting the NewsTwisters: A replication of Efron's study. *Journalism Quarterly, 50*, 211–219.
Travis, G. D. L. (1981). Replicating replication? Aspects of the social construction of learning in planarian worms. *Social Studies of Science, 11*, 11–32.
van IJzendoorn, M. H., & Kroonenberg, P. M. (1988). Cross-cultural patterns of attachment. A meta-analysis of the Strange Situation. *Child Development, 59*, 147- 156.
Waters, E. (1983). The stability of individual differences in infant attachment. Comments to the Thompson, Lamb, and Estes contribution. *Child Development, 54*, 516–520.
Wolins, L. (1962). Responsibility for raw data. *American Psychologist, 17*, 657–658.