Generalizability of Research Results

9

9.1 Empirical Research and Generalizability

Section 2.5 outlines ways of gaining scientific knowledge. There are different approaches used to develop new theories, among them the attempt to generalize from observations (→ induction, see Sect. 2.5). This is followed by tests and, if necessary, negation or modification of the theory.

No matter how many observations are the same, one cannot draw *definite* conclusions as to corresponding lawlike generalizations. At some point, an unknown and deviant case can occur. On the other hand, one of the central aspects of scientific realism, as noted in Sect. 3.1, is reminiscent:

If a theory and the statements contained in it prove themselves long-term and often in appropriate tests and in practical applications, then there is obviously much to suggest that these statements are relatively likely to be approximately true, although, of course, one cannot achieve any certainty.

This chapter deals with approaches in which different findings about the same research question are summarized (→ meta-analyses) or new studies are carried out to check previous results (→ replications). Such approaches focus on the **generalizability** of research results.

The generalizability of test results relates to the question of how well one can extrapolate from a particular result to other *subjects* (e.g., sample → population), *research objects* (e.g., success of companies during the introduction of a product → success of companies in general), *contexts* (e.g., USA → Europe, present → future) and when using other *methods* (e.g., laboratory experiment → field study).

> Kerlinger and Lee (2000, p. 474) characterize the importance of generalizability as follows:
>
> "Can we generalize the results of a study to other participants, other groups, and other conditions? Perhaps the question is better put: How much can we generalize the results of the study? This is probably the most complex and difficult question that can be asked of research data because it touches not only on technical matters (like sampling and research design), but also on larger problems of basic and applied research."

Of course, the ability to generalize results is critical to many application-oriented investigations. The results of a customer survey are usually only relevant if one can apply the results to the behavior of customers in general. Typically, the focus in application-oriented research is primarily about generalizing from a sample to a population of interest.

We previously presented one generalizability approach regarding the quality criteria of measurement in Sect. 6.3. Here, one aims at the generalizability of a measurement by reducing measurement errors. This is typically a question of reliability, whereby classic reliability tests take into account only one source of error (for example, time of study). A consideration of different measurement errors that can occur simultaneously, as well as their interaction, is performed in the context of **generalizability theory** (Cronbach et al. 1972, see also Sect. 6.3.4). This chapter is not about the generalizability of measurements, but, in line with the previous chapters about theory and hypotheses testing, about the **generalizability of research results**, which usually refers to the relationships of variables according to the proposed hypotheses. Of course, the validity, reliability, and generalizability of measurement instruments is an essential prerequisite for the generalizability of test results, since the results can only be meaningfully interpreted if the measurement error is as small and controllable as possible. In scientific research, replication studies and meta-analyses are commonly carried out in order to arrive at generalizable statements regarding study results.

9.2   Replication Studies

Replication studies are repetitions of empirical studies aimed at demonstrating the reproducibility of the results. Replication studies do not differ from the original studies in terms of the study object or research question, but often in some aspects of the procedure. Through (successful) replication studies, one can achieve a certain independence for the study's results from sampling errors, the specifics of the research methods, and, at a minimum, from the time the original study was conducted. Here, we again consider the requirements for inductive reasoning mentioned in Sect. 2.5—that observations under different conditions should lead to the same result.

In general, the possibility of replication or the reproducibility of results of empirical studies is an *essential criterion for the scientific value* of studies. Therefore, it is common in the natural sciences that results must be replicable. This achieves the independence of the results of a particular study context and study method, as well as some protection against results that may be biased by the research process and the researchers (see Sect. 10.2). In the context of the discussion of paradigms and relativism (Chap. 3), it was suggested that empirical results could also be systematically influenced by the view of the researchers (theory-ladenness) and by the applied methods. By replicating the use of different methods performed by different researchers, independence from such influences is more likely to be ensured. McCullough and Vinod (2003, p. 888) describe the replicability of studies as a basic requirement of science: "Replication is the cornerstone of science. Research that cannot be replicated is not science, and cannot be trusted either as part of the profession's accumulated body of knowledge or as a basis for policy."

Hunter (2001) distinguishes the following *types of replications*:

- **Statistical replications** refer to exact repetitions of previous studies with the aim to increase the accuracy of statistical results by reducing the sampling error.
- **Scientific replications** refer to studies that use equivalent but not identical methods when repeating previous studies.
- **Conceptual replications** are replication studies with deliberate changes made to the original study. The change occurs, for instance, by including additional variables for the purpose of examining further potential influencing factors or by so-called moderator variables, which either limit or generalize the scope of the previous findings.

> Kerlinger and Lee (2000, p. 365), on the nature and significance of replication studies, state:
>
> "Whenever possible, replicate research studies...The word replication is used rather than repetition because in a replication, although the original relation is studied again, it might be studied with different kinds of participants, under somewhat different conditions, and even with fewer, more, or even different variables."

Although replicability of studies is obviously an important prerequisite for the scientific acceptance of research results, replication attempts often fail. For example, in a large-scale replication project, hundreds of studies published in leading psychology journals were replicated (Open Science Collaboration 2015). Only 36% of the studies showed an effect consistent with the effect of the original study as the effects in the replication studies tended to be weaker than the effects in the original studies. This indicates the existence of a publication bias (see Sect. 9.3).

Also, in marketing research, *the success rates of replication studies* are similarly low (e.g., Hubbard and Vetter 1996). To conclude from the failed replication

attempts that one cannot trust the results of science would be premature. There are many reasons why the findings of a replication study differ from those of the original study (see Eisend et al. 2016; Lynch et al. 2015):

- Empirical studies usually rely on random sampling, and their results are therefore subject to a *sampling error*. That a replication study based on a particular sample produces a non-significant result may be due to chance. The significant result of the original study can then still be the result that would usually be obtained with several repeated replication attempts. However, based on conflicting results (i.e., based on a significant result from the original study and a non-significant result from the replication study), how can we know whether the original study or the replication study provides the "true" result? For this purpose, the two results can be summarized and integrated by means of a meta-analysis (Lynch et al. 2015; for meta-analysis see Sect. 9.3). If the integrated result is significant, then it confirms the significant result of the original study.
- Some studies cannot be replicated exactly because the *documentation of the methodological details of the original study* is often insufficient to repeat a study in the same detail. One of the reasons for insufficient documentation is that the presentation of studies in many scientific journals is very condensed. Even small deviations from the original study design; for example, the time of day when an experimental study is conducted that measures the emotions or performance of the participants may influence the results of a study.
- In the case of conceptual replications in particular, one attempts to extend the scope of study results beyond the context of the original study by adapting or expanding the original study accordingly. This could be, for instance, a study in a different cultural context, using a different demographic group of people or stimuli other than in the original study. If the results of the replication study differ from the original study, then this may be due to the *contingency of the results.* This means that the results of the original study are valid only in the context of the original study (e.g., in the USA) but not in the context of the (conceptual) replication study (e.g., in Asia).
- Ultimately, of course, there is the possibility that the original study's results were collected and/or analyzed sloppily, or that the researchers even manipulated or falsified the results. In this case, attempts to replicate results may be a way to identify potentially fake results. *Mistakes made by researchers or fabrications of results* tend to limit confidence in science. They also represent a significant ethical problem for science, which Chap. 10 discusses in more detail.

Despite the importance of replication studies to the scientific process, *relatively few replication studies are published*. Evanschitzky et al. (2007) reported a replication rate of 1.2% from 1990 to 2004 in the leading marketing journals (*Journal of Marketing*, *Journal of Marketing Research*, and *Journal of Consumer Research*). That means that only 1.2% of all studies published in these three journals during this period were replication studies. In comparison, replication studies conducted in the

period from 1974 to 1989 were at 2.3%, meaning that the replication rate was cut by nearly 50% over time.

Why is it that so few replications are published, even though their importance to science is central? Hunter (2001) highlights two possible reasons for the low interest in replication studies by researchers and journals and mentions the corresponding counterarguments:

- *Low creativity in replication studies*; Counter argument: Sound research requires a solid knowledge base; creativity is not the only criterion for the quality of research.
- *Little increase in knowledge*; Counter argument: A single study with the systematic problems and contingencies of their results is too weak as a knowledge base. Recall the inductive-realistic model of theory testing in Sect. 5.3.

If the chance of getting replication studies published is low, most scientists will have no incentive to conduct the studies. In addition, a replication study that questions a well-published and widely accepted outcome in another study may appear as offensive or even as a personal attack on the authors of the original study. This may also explain why the willingness of researchers to help their colleagues replicate one of their studies is rather low (Reid et al. 1982; Wicherts et al. 2006).

In recent years, marketing research has shown an increased awareness of the need for replication studies. For example, some marketing journals have set up a "Replication Corner" (e.g., *International Journal of Research in Marketing* and *Journal of Marketing Behavior*) or published special issues on replication studies (e.g., *Journal of Advertising*). In special issues of journals or journal sections, replication studies are not in direct competition with original studies, and replication studies thus have a better chance of successfully passing through the peer review process and being published despite the lower creativity of the results. Nevertheless, outside of these outlets for replication studies, the number of replication studies in leading marketing journals has further decreased over the years, while the prevalence of **intra-study replications**, that is, replications of empirical studies within the same project that are actually not considered true replicative research, has increased over years (Kwon et al. 2017).

9.3   Meta-Analysis

A particularly comprehensive and methodologically advanced approach of empirical generalizations is **meta-analysis**. Glass (1976, p. 3) defines a meta-analysis as the "analysis of analyses...the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings." The procedure is to summarize as many as possible (ideally all) relevant empirical results for a particular research question or hypothesis and, to a certain extent, calculate a "common" result, taking into account the different sample sizes. To that end, the

presentation of the results from all studies must be comparable and uniform. This occurs by means of so-called **effect sizes**. As explained in Sect. 7.2, effect sizes provide a quantitative assessment of the magnitude of some phenomenon that is used to address a particular research question (Kelley and Preacher 2012). Common effect size measures are correlation coefficients, standardized mean differences, or odds ratios.

> Lehmann et al. (1998, p. 746), on the relevance of meta-analyses for empirical marketing research, state:
>
> "One of the most fruitful avenues for analysis is exploring what can be learned from past studies. For example, an advertising agency that has studied the impact of increasing advertising 237 times can learn more from synthesizing the information in the 237 studies than from running the 238th. The process of combining information from past studies is known as empirical generalization and/or meta-analysis (that is the analysis of past analyses). The basic premise is that we can learn from other (past) situations."

The meta-analysis not only integrates results, but also examines their diversity and variability (**heterogeneity**). If the results based on different studies are quite consistent (*homogeneous*), then the overall result that was integrated in the meta-analysis can be regarded as a generalizable finding and reused as such in further research and practice. If the individual results are very different (*heterogeneous*), then this difference can be investigated and (partially) explained in the context of a meta-analysis. This is achieved by applying so-called **moderator** variables (see Sect. 8.2) that are used to explain the variability of effect size values (that is, the realization of a particular effect size measure). Figure 9.1 illustrates this relationship. If there is high variability (heterogeneity) in the effect size values that measure the strength of the relationship between two variables (e.g., attitude toward a product as the independent variable and purchase behavior as the dependent variable), a moderator variable (e.g., product type) might be able to reduce the heterogeneity
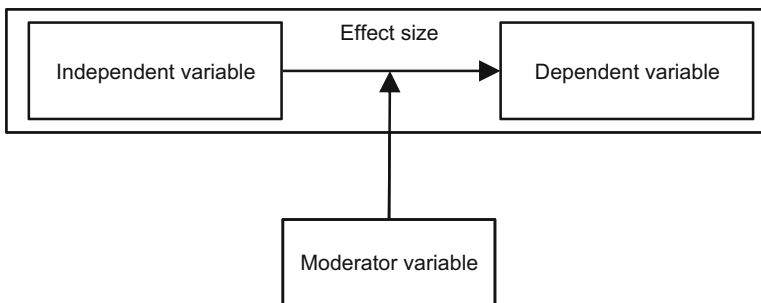

Fig. 9.1 Explaining heterogenous findings in a meta-analysis by means of moderator variables

by explaining under which conditions the effect size values becomes stronger or weaker.

A common moderator variable is the study design, which distinguishes whether the study results were determined in a controlled laboratory experiment or in a field study. If we divide all empirical results into two groups (results from laboratory experiments and results from field studies), we can compare the results from the two groups. If there is a statistically significant difference, we can assume that the overall results cannot be generalized across the different research designs, but they must be differentiated. If there is no statistically significant difference, the study results can be generalized over different study designs. Therefore, the moderator variables might be able to explain the heterogeneity of empirical results. The difference in studies that are integrated in a meta-analysis is therefore not a disadvantage, but rather an advantage because the meta-analysis can show whether the overall result is independent of the specifics of individual studies or if the influence of the differences in the studies is relevant. Moderator variables can refer to different dimensions of generalization. Depending on whether the moderator analysis *reveals homogeneity or heterogeneity of findings* within a particular dimension, the findings can either be *generalized* across that particular dimension (e.g., the findings do not depend on the research method and can be generalized across research methods) or have to be *distinguished* (e.g., the findings depend on the research method and cannot be generalized across research methods). Figure 9.2 illustrates how the assessment of heterogeneity and homogeneity in a meta-analysis can contribute to the generalization of findings across different dimensions.
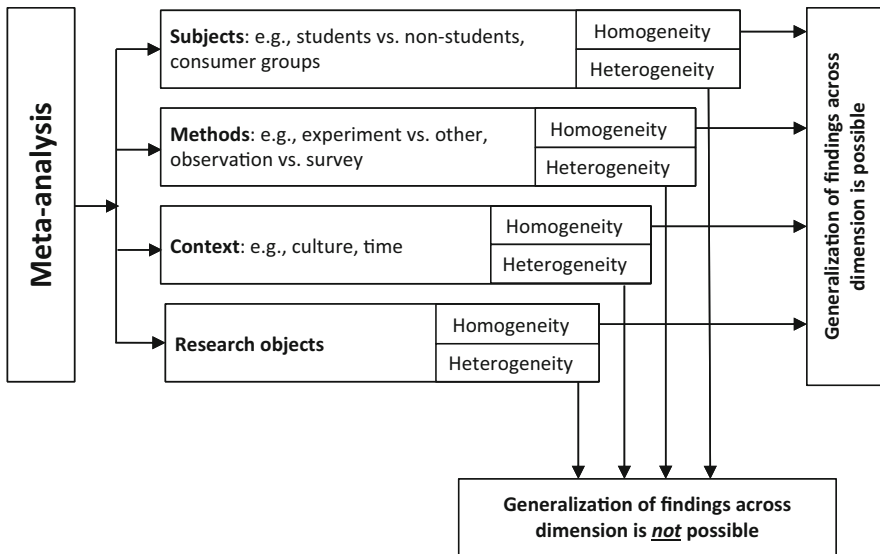

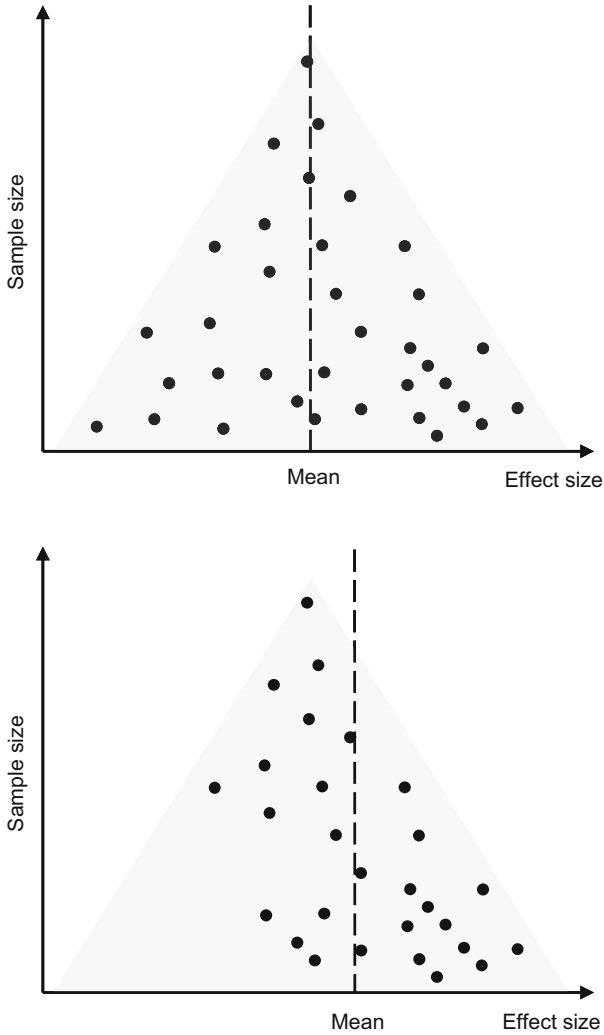**Fig. 9.2** Meta-analysis and generalization across various dimensions

Fig. 9.3 Publication bias and funnel graph

Meta-analyses can help analyze the so-called **publication bias**. This is the frequently empirically confirmed phenomenon that non-significant results are reported less frequently in studies than significant results (for studies on publication bias, see Ferguson and Brannick 2012; Kepes et al. 2012; Renkewitz et al. 2011). Researchers tend to skip reporting insignificant results (see the ethical issue in Sect. 10.2), because these non-significant findings will be less likely to successfully pass the peer review process. As a consequence of publication bias, the results in published studies are upward biased; that is, they are usually "too strong" because

the "weak" results are not published at all. In this way, empirical generalizations, which are about the size of an effect, become questionable.

The data in a meta-analysis can be analyzed with regard to the existence of a publication bias and this bias can be corrected. Figure 9.3 illustrates this procedure. In the **funnel graph**, the size of the effect and the sample size of the study from which the effect originates are compared against each other. Very small samples have very large sampling errors, so the variance of the effect size values (e.g., correlation coefficient estimates) around the mean is quite large. The upper funnel graph shows the result of a meta-analysis in which effect size values are scattered, as expected, around the mean according to their sample size. The distribution looks like a triangle or a funnel (hence, the term funnel graph). The lower funnel graph lacks some effect size values, namely, small effect size values based on small samples. These effect size values are more likely to be non-significant compared with large effect size values or effect size values based on large samples. The lower funnel graph thus shows a publication bias: the empirical distribution of the effect size values deviates from the expected distribution in the upper funnel graph. The deviation is systematic, since non-significant findings are missing. The plotted average also illustrates that, in the presence of a publication bias, the integrated (i.e., average) effect found in a meta-analysis is upward biased.

From this funnel graph, it is possible to determine whether there is a publication bias. There is a comprehensive set of methods (documented in detail by Rothstein et al. 2005), with which, among other things, a theoretical distribution can be analytically restored and, thus, an upward biased mean (as shown in the lower graph in Fig. 9.3) can be corrected.

9.4   Generalizability and Theory Development

In many cases, the process of theory development does not take place in such a way that, at the beginning, a "final" theory is formulated and then empirically tested, ultimately leading to rejection or acceptance. Rather, the literature (e.g., Weick 1995) speaks of a **theory continuum** that, in addition to the building of theories and their (empirically-based) acceptance or rejection, also includes phases of modification and refinement of the theories. Essential steps in the theory continuum and the role of generalizations require a brief characterization:

- **Draft**: Theory building is the subject of the fourth chapter and explicitly deals with (in Sect. 4.3.4) the relevance of empirical generalizations. In the *context of discovery* (see Sect. 1.1), generalizations can be used as an inductive approach to theory building. Usually, for the publication of a new theory draft, some preliminary empirical confirmations are needed.
- **Acceptance or rejection** of a theory: These steps are already known from the inductive-realistic model presented in Sect. 5.3. In the case of multiple empirical evidences and confirmations ($\rightarrow$ generalization), we decide on an (provisional) acceptance of the theory; if "empirical failures" dominate, we usually reject the

theory. In the *context of justification* (see Sect. 1.1), generalizations help to reduce the risk of errors and mistakes compared with testing a hypothesis in a single study. The point here is that in the case of statements based on a large number of results, it is possible to conclude with greater certainty the "truth" of a theory.

If a theory is broadly accepted after a number of successful tests, and its relevance has been demonstrated, it may be included in relevant *textbooks*. Such textbooks reflect the current status of scientific knowledge concerning a certain field, are the basis for general information and are used in the education of students and young scientists.

Two leading philosophers of science with very different positions characterized the relevance of textbooks for scientific information with the following statements.

Richard Boyd (2002, p. 1) noted:

"For example, if you obtain a good contemporary chemistry textbook you will have good reason to believe (because the scientists whose work the book reports had good scientific evidence for) the (approximate) truth of the claims it contains about the existence and properties of atoms, molecules, sub-atomic particles, energy levels, reaction mechanisms, etc. Moreover, you have good reason to think that such phenomena have the properties attributed to them in the textbook independently of our theoretical conceptions in chemistry."

Thomas Kuhn (1970, p. 43) states:

"Close historical investigation of a given specialty at a given time discloses a set of recurrent and quasi-standard illustrations of various theories in their conceptual, observational, and instrumental applications. These are the community's paradigms, revealed in its textbooks, lectures and laboratory exercises. By studying them and by practicing with them, the members of the corresponding community learn their trade."

- **Refinement**: Section 2.1 characterizes theories by making statements about relationships of concepts with respect to a particular study object (e.g., building customer relationships). Many details of these relationships (e.g., linear or non-linear relationships, magnitude of effects; see Sect. 7.2) are not well-known when the first draft of a theory is designed. This requires numerous detailed studies, the results of which can be summarized and generalized, which then allows statements about typical value ranges of correlations, regression coefficients, etc. In view of the given theoretical framework and with regard to the procedure, such research resembles the "normal science" as characterized and discussed by Thomas Kuhn (1970).
- **Modification**: Modification can be the change of a theory through the addition or elimination of concepts and relationships; this includes moderators and mediators (see Sect. 8.2). For example, if a large number of studies show that a theoretically presumed relationship rarely occurs, this is a reason to rethink the corresponding

variables or the assumed relationships. Moderators influence the strength of the relationship between/among variables; mediators represent the connection between the corresponding variables that are indirectly linked.

In Hunt's inductive-realistic model (2012; see also Sect. 5.3), "acceptance" and "rejection" directly reflect the corresponding "status of the theory"; "refinement" and "modification" are more likely to be assigned to the status of "working acceptance." The latter can also mean that more and clearer results are required for a decision to be made about the acceptance or rejection of the theory.

The process of theory development can be illustrated by the phenomenon of "loss aversion."

Draft: Kahneman and Tversky (1979) introduced the concept of "losses loom larger than gains" and values as assigned by people based on gains and losses, as a critique of the classical economic utility model. They started with simple experiments that showed the enhanced loss sensitivity of consumers.

Refinement: Over the years, many studies have been conducted that attempted to specify, generalize, or modify the concept. For instance, the concept was initially linked to decisions with risks and later extended to riskless choices (Tversky and Kahneman 1991). In marketing, the concept has been used to investigate brand choice of consumers by introducing gains and losses for price variables (Briesch et al. 1997), as well as other quality attributes (Kivetz et al. 2004). Further studies have applied the idea of reference-dependent choices to different areas such as transportation or health care.

Modification: Several variables have been investigated as potential moderators of loss aversion. For instance, to evaluate a product's value, consumers can use either internal reference points (i.e., past information in their memory) or external reference points (i.e., current information provided at the point of purchase). When consumers use external reference points, they show greater price-loss aversion than those using internal reference points (Mazumdar and Papatla 2000).

Acceptance: In 2014, Neumann and Böckenholt presented a meta-analysis and summarized prior research on loss aversion in product choice. They showed how generalizable the concept is and that the degree of loss aversion depends—among other factors—on product and consumer characteristics. In the meantime, the concept of loss aversion is an accepted and important theory that became an integral part of most consumer behavior textbooks (e.g., Hoyer et al. 2018).

How can we describe the role of *generalizations in the research process*? Chapters 4 and 5 focus on the process from theory building to empirical testing; in this chapter, generalizations are more concerned with a larger number of empirical
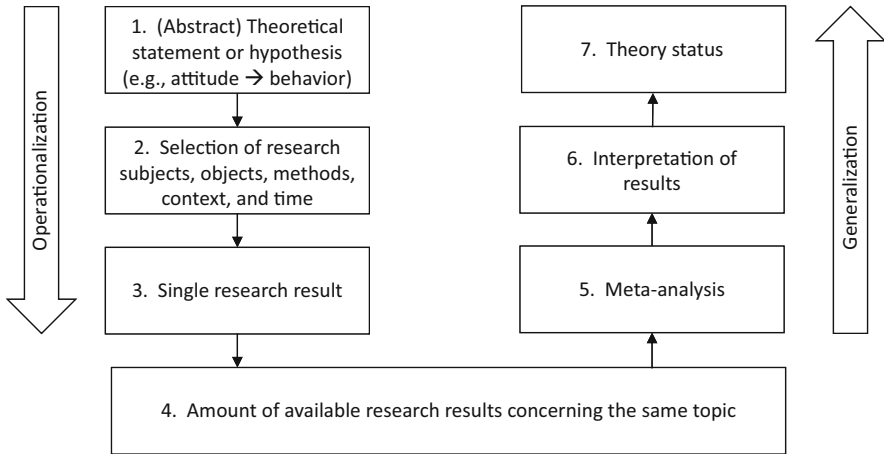
Fig. 9.4 Relationship between operationalization and generalization

findings that help to make more generalized statements (beyond the results of a single study). The following considerations intend to show differences and relationships between the two types of research processes (operationalization and generalization). Figure 9.4 illustrates the relationship.

Operationalization plays a central role in the empirical test of theoretical statements. This is because, in order to verify theories by empirical data, it is necessary to assign measurements to the (abstract) theoretical concepts by using appropriate methods and to analyze the results of these measurements with regard to the suggested hypotheses (see Sect. 6.1). The process of operationalization is, at the same time, a process of concretization and thus of narrowing the research object. For example, in this way, a general question about the relationship between attitude and behavior (Fig. 9.4, step 1) may become a concrete question of the correlation between the attitude toward a particular brand and brand choice. In addition, the corresponding study is performed at a specific time, in a specific context, using certain methods, etc. (Fig. 9.4, step 2), thus leading to a single research result of a specific study dealing with a concrete rather than a general research question (Fig. 9.4, step 3). The question arises as to what meaningfulness such a specific study has for the more general initial question of the generalizability of the research results.

The link to **generalization** is that the result of a single study is no longer considered in isolation, but in relation to other results of studies on the same topic that have already been conducted (meta-analysis) or are generated through replication studies (see Sect. 9.2). This is described by steps 4 and 5 in Fig. 9.4: having a database of the already available study results is the basis for performing a meta-analysis (see previous section). A meta-analysis provides results about the extent to which the theoretically expected relationships have been confirmed (effect sizes) and to what extent changes in variables lead to changes in other variables (magnitude of

effects). The interpretation (Fig. 9.4, step 6; see also Sect. 5.2) of such results leads to assessments of the theory status (Fig. 9.4, step 7), that is, the question of whether a theory is accepted, rejected, or in the status of "working acceptance" and therefore should be modified, refined, and tested further. We can see that the process starts with a large number of single results and leads to more general (and more abstract) statements by summarization and integration (using meta-analysis).

Some of the central ideas and their relationships are summarized in Fig. 9.5. At the beginning of the process of theory development is the step of "theory building" (see Chap. 4). Its result is a "theory draft" that is ready for an empirical test (see Chap. 5). A single study (from "hypotheses" to "results") might lead to insights (→ "interpretation"). In many cases, several empirical studies are performed by different researchers and/or in different contexts (e.g. psychology, consumer behavior, behavioral economics). The results of these multiple studies can be collected and analyzed by means of "meta-analysis" (see Sect. 9.3) and interpreted in the next step. Interpretation in this context means to evaluate the existing theory ("theory draft") in the light of the empirical results. The strengthened impact on interpretation provided by *number of results* compared to a single result is symbolized in Fig. 9.5 by a dark bold arrow.

Results of a single study or a meta-analysis reveal empirical successes or failures (see Sect. 5.3; Hunt 2012). If a theory meets a general corroboration due to the dominance of *empirical successes*, some empirical results may be used to modify and/or to refine the theory. Of course, a dominance of *empirical failures* would increase the doubts about the truth of a theory and weakens the theory. Based on this a researcher or the scientific community has to decide whether to reject the theory or not.
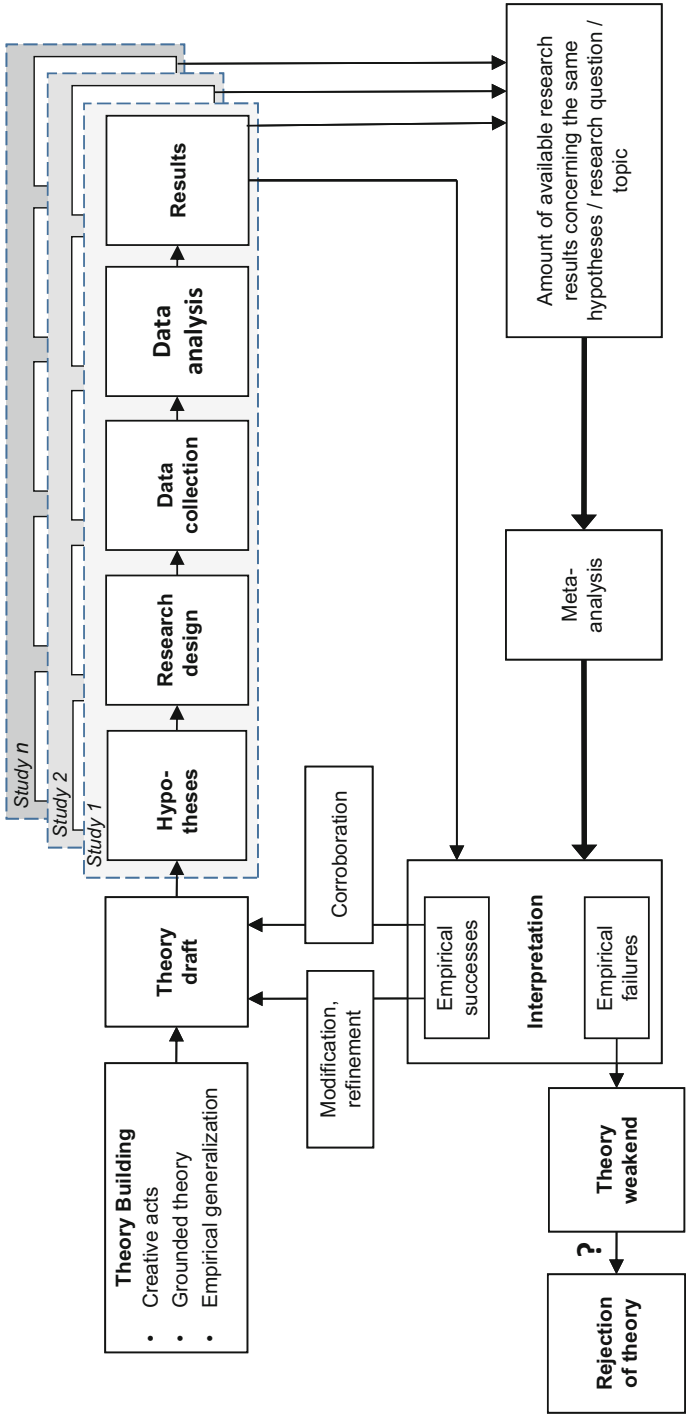
Fig. 9.5 Theory development and empirical research

References

Boyd, R. (2002). Scientific realism. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. http://plato.stanford.edu.

Briesch, R. A., Krishnamurthi, L., Mazumdar, T., & Raj, S. P. (1997). A comparative analysis of reference price models. *Journal of Consumer Research, 24*(2), 202–214.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of Generalizability for scores and profiles*. New York: Wiley.

Eisend, M., Franke, G. R., & Leigh, J. H. (2016). Reinquiries in advertising research. *Journal of Advertising, 45*(1), 1–3.

Evanschitzky, H., Baumgarth, C., Hubbard, R., & Armstrong, J. S. (2007). Replication research's disturbing trend. *Journal of Business Research, 60*(4), 411–415.

Ferguson, C. J., & Brannick, M. T. (2012). Publication Bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analysis. *Psychological Methods, 17*(1), 120–128.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3–8.

Hoyer, W. D., MacInnis, D. J., & Pieters, R. (2018). *Consumer behavior* (7th ed.). Boston, MA: Cengage.

Hubbard, R., & Vetter, D. E. (1996). An empirical comparison of published replication research in accounting, finance, management and marketing. *Journal of Business Research, 35*(2), 153–164.

Hunt, S. D. (2012). Explaining empirically successful marketing theories: The inductive realist model, approximate truth, and market orientation. *Academy of Marketing Science Review, 2*(1), 5–18.

Hunter, J. E. (2001). The desperate need for replications. *Journal of Consumer Research, 28*(1), 149–158.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–291.

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods, 17*(2), 137–152.

Kepes, S., Banks, G. C., McDaniel, M., & Whetzel, D. L. (2012). Publication Bias in the organizational sciences. *Organizational Research Methods, 15*(4), 624–662.

Kerlinger, F., & Lee, H. (2000). *Foundations of behavioral research* (4th ed.). Melbourne: Wadsworth.

Kivetz, R., Netzer, O., & Srinivasan, V. S. (2004). Alternative models for capturing the compromise effect. *Journal of Marketing Research, 41*(3), 237–257.

Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.

Kwon, E. S., Shan, Y., Lee, J. S., & Reid, L. N. (2017). Inter-study and intra-study replications in leading marketing journals: A longitudinal analysis. *European Journal of Marketing, 51*(1), 257–278.

Lehmann, D. R., Gupta, S., & Steckel, J. (1998). *Marketing research*. Reading, MA: Addison-Wesley.

Lynch, J. G., Jr., Bradlow, E. T., Huber, J. C., & Lehmann, D. R. (2015). Reflections on the replication corner: In praise of conceptual replications. *International Journal of Research in Marketing, 32*(4), 333–342.

Mazumdar, T., & Papatla, P. (2000). An investigation of reference price segments. *Journal of Marketing Research, 37*(2), 246–258.

McCullough, B. D., & Vinod, H. (2003). Verifying the solution from a nonlinear solver: A case study. *American Economic Review, 94*(1), 873–892.

Neumann, N., & Böckenholt, U. (2014). A meta-analysis of loss aversion in product choice. *Journal of Retailing, 90*(2), 182–197.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), 943.

Reid, L. N., Rotfeld, H. J., & Wimmer, R. D. (1982). How researchers respond to replications requests. *Journal of Consumer Research, 9*(2), 216–218.

Renkewitz, F., Fuchs, H. M., & Fiedler, S. (2011). Is there evidence of publication Bias in JDM research? *Judgment and Decision making, 6*(8), 870–881.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication Bias in meta-analysis. Prevention, assessment and adjustments*. Chichester: Wiley.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics, 106*(4), 1039–1061.

Weick, K. (1995). What theory is not, theorizing is. *Administrative Science Quarterly, 40*(3), 385–390.

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist, 61*(7), 726–728.

Further Reading

Hunter, J. E. (2001). The desperate need for replications. *Journal of Consumer Research, 28*(1), 149–158.

Lipsey, M. W., & Wilson, D. T. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), 943.