



Published in final edited form as:

J Appl Res Mem Cogn. 2023 June ; 12(2): 189–194. doi:10.1037/mac0000121.

Rigorous exploration in a model-centric science via epistemic iteration

Berna Devezer¹, Erkan O. Buzbas²

¹Department of Business, University of Idaho

²Department of Mathematics and Statistical Science, University of Idaho

Abstract

Urgent attention is needed to address generalizability problems in psychology. However, the current dominant paradigm centered on dichotomous results and rapid discoveries cannot provide the solution because of its theoretical inadequacies. We propose a paradigm shift towards a model-centric science, which provides the sophistication to understanding the sources of generalizability and promote systematic exploration. In a model-centric paradigm, scientific activity involves iteratively building and refining theoretical, empirical, and statistical models that communicate with each other. This approach is transparent, and efficient in addressing generalizability issues. We illustrate the nature of scientific activity in a model-centric system and its potential for advancing the field of psychology.

Keywords

generalizability; modeling; model-centric; result-centric; epistemic iteration; exploration

Establishing empirical regularities that apply to multiple populations, in a multitude of conditions, and recognizing the limits of such generalizations seem critical to progress in some sciences. Generalizability problems in psychology as outlined by Bauer (2023) are hard to deny and need to be addressed with urgency. However, we are not convinced by Bauer (2023)'s framing of the problem, which emphasizes a predicament that needs addressing before meaningful discussions on generalizability can take place.

Bauer (2023) frames the problem as separate from issues of research rigor and implies that generalizability can be achieved independently from statistical inference validity. Scientists' subjective constraints on generality (COG) assessments are seen as essential to drive scientific progress. This framing overlooks the interconnectedness of scientific theory, experiments, and statistical inference and treats statistics as a post-hoc tool solely for data analysis, prioritizing interpretation of results over evaluation of theory, models, and methods. It undermines active engagement, critical readership of research, and methodical and proactive scientific approaches.

Correspondence concerning this article should be addressed to Berna Devezer, College of Business and Economics, University of Idaho. bdevezer@uidaho.edu.

The authors have equally contributed to this work.

We believe a shift from result-centric paradigm to model-centric paradigm is necessary in re-framing generalizability problems to devise meaningful solutions. This paradigmatic shift can be outlined as follows:

1. Quantitative psychology largely adheres to a result-centric paradigm.
2. Model-centric paradigm which considers the *model* as the focal unit of progress provides a valuable shift in perspective
3. The model-centric paradigm aligns rigorous theoretical, experimental, and statistical modeling iteratively, assessing each iteration against some fixed criteria.
4. Generalizability is achieved via explicit model specification and connections among different models; COG are determined by assessment of model assumptions.

This paper delves into these points. As an illustration of the model-centric paradigm, we introduce a new example inspired by John von Neumann's elephant problem of finding parameters for a function that draws an elephant shape in two-dimensions (Wikipedia contributors, 2023). Our example goes: We assume a *target system*, an elephant. We want to estimate the *target subsystem*, the shape of an average elephant using noisy observations of elephants. To this end we: i) build a *theoretical model*, M_T , to capture the shape of an elephant in two-dimensions, ii) build an *experimental model*, M_E to sample the elephants, and iii) use a *statistical model*, M_S to analyze the data. We evaluate the models and if not satisfied we reiterate.

Dominant empirical paradigm of quantitative psychology is result-centric

The misconception that science is merely a collection of facts, rather than a collaborative process of generating explanations, is widespread (see Marks, 2009, Ch. 1) and represents the prevailing paradigm in psychology. Even the recent science reform has focused on the irreproducibility of results rather than critical assessment of psychological theories, models, methods. Viewing results as the epistemic target of science and the principal unit of progress underlies problematic patterns in scientific culture: Unrealistic expectation that each study should generate new scientific facts; overreliance on exclusively post-hoc use of ad-hoc statistics and their misuse as sole arbiters of scientific truths; desire to make scientific discoveries without systematic exploration; loss of long-term, iterative assessment of scientific progress; wishful thinking that fixing a set of results reported in the literature will improve science.

When dichotomized fact-like results such as the absence/presence of an effect are placed at the center of scientific process, generalizability becomes a property of empirical activity alone, which promotes pursuing only questions such as: "Are these effects *true* for other populations, contexts, tasks, measures?" However, a science that can only reflect on itself by evaluating the truth value of results but cannot acknowledge epistemic progress beyond accumulating facts is impoverished.

An abused tool of result-centric paradigm and its potential culprit is null hypothesis significance tests (NHSTs). Let us consider our elephant example. Without even invoking sampling error, think of deterministic inference. A candidate M_T has four true non-null complex-valued parameters; their Fourier transform plotted on $(0, 2\pi)$ are shown in figure 1 A. With these parameters and an appropriate mathematical model (Mayer et al., 2010), we can build a good estimate of the shape of an average elephant (model stage 4 of Figure 1 C). Now, consider the null hypothesis: *The four model parameters are null*, and the case in which we correctly reject this null hypothesis jointly for all four parameters. With results given solely by figure 1 A, we have not garnered much regarding the shape of an average elephant. The reason is that our focus was on testing a hypothesis about parameters, or *results* specifically; and not about building *models* involving the relationship among the parameters. A model would have been more useful for our goal of performing inference about the shape of an average elephant. The lesson is that there are severe theoretical limitations to result-centric paradigm even when it returns true results.

For decades, psychological science has witnessed the “replacement of good scientific practice by a statistical ritual that researchers perform not simply on the grounds of opportunism or incentives but because they have internalized the ritual and genuinely believe in it” (Gigerenzer, 2018). This ritual has led to substituting rigorous assessments of research quality with quantitative surrogates. Oftentimes it is grounded in a flawed understanding of NHST and p-values (Greenland et al., 2016; Wasserstein & Lazar, 2016). Accumulating fact-like effects is unlikely to generate theoretical understanding of target phenomena (Cummins, 2000; P. E. Meehl, 1990; P. Meehl, 1986; Newell, 1973; van Rooij & Baggio, 2021). Without moving beyond the assurances afforded by pseudo-truth claims, efforts to enhance generalizability will likely fail to foster cumulative scientific progress.

Back to Bauer (2023): Their framing of the issue of generalizability makes mostly sense within a result-centric paradigm where rigor is limited to valid statistical inference and replicability of results. Only in this weak paradigm, statistical inference is a post-hoc tool to extract scientific facts from data and researchers’ verbal COG statements may be considered progressive. Since result-centric psychology treats statistically significant effects practically as true and a proper basis for new studies, COG statements offer post-hoc relief from overinterpretation of effects. Sampling of experimental units, operationalizations, and contexts are also treated as post-hoc concerns for generalizing the results to *other* populations, rather than a prior consideration for ensuring valid inference to the sampled population.

Paradigm shift to a model-centric science

A paradigm shift seems overdue in how we read, evaluate, design, and perform scientific research. At the core of this shift lies model-thinking as a requisite for a rigorous, credible, generalizable psychological science. Models are abstractions and idealizations² we use to *represent* aspects of real-world target systems (e.g., Godfrey-Smith, 2006; Potochnik, 2015).

²*Abstractions* are simplifications and subtractions we make to suppress aspects of the target system while *idealizations* involve smoothing out certain features to create a more perfect, albeit intentionally false version of the system (Woods & Rosales, 2010).

The notion of model-based science has been around for decades (e.g., Levins, 1966) and successfully pursued in other disciplines (e.g., population biology, Weisberg, 2006).

In this paradigm, a *model* is the epistemic target of scientific activity and the focal unit of progress. Model-centric science does not aim to accumulate knowledge using a collection of scientific results or facts as building blocks but focuses on *explicitly* building, evaluating, refining scientific models (see Figure 2 in Devezer et al., 2019, for an illustration of a model-centric scientific process) because models encapsulate our capacity to understand/explain, predict, or modify nature (Levins, 1966). Modeling calls for making assumptions explicit so they can be scrutinized and assessed independently (Epstein, 2008). A scientist who imagines the world, theorizes about it, generates or tests hypotheses, designs and conducts a study, or performs a statistical analysis is running a model regardless of their assumed scientific paradigm. A major problem with result-centric science is that the models are implicit, informal, and difficult to interpret due to hidden assumptions. Valid inference relies on valid model assumptions. If they are violated, model-centric paradigm focuses on building larger models with fewer assumptions and robust inferential methods. Results serve an instrumental, not a terminal, role in building better models for better explanations, predictions, or interventions. Progress is not measured by the discovery of some truth but by improvements in model performance regarding a specific aim (Gelman et al., 2020). Models of target subsystems are not true or false. They are useful or not. For example, one can build an M_T of our elephant using a regression model, which may estimate the shape of an elephant equally well (Wei, 1975) as the Fourier transform approach.

Rigor and progress via iterative modeling

Principled modeling at each stage of the research process helps us bridge scientific theories and empirical phenomena to answer questions of interest. In a model-centric paradigm Figure 1 B would present a typical research process in quantitative psychology conceptualizing levels of modeling. The *target system* of interest (step 0a) could be a real-world phenomenon (e.g., elephant, human memory) or a theoretical, even fictitious entity (van Rooij, 2022). A research question typically concerns only some features of the *target system*, the *target subsystem* (step 0b). In step 1, the target subsystem is abstracted, idealized, and represented by M_T . M_T is built relative to our research question and scientific aim (e.g., mechanistic explanation, prediction). M_T specifies the observables, unobservables, and the relationships among them as deemed relevant for our aim. M_T explicitly states the assumptions under which these relationships arise. M_T can exist as a mental representation or be expressed verbally, visually, or formally (i.e., mathematical or computational). As a standalone research activity under model-centric science, building a useful M_T may require programmatic effort and epistemic iterations. Model-centric science aims to iteratively increase the precision of M_T by way of formalization, and its accuracy by testing and refining its assumptions.

Figure 1 C shows how four parameters of M_T given in Figure 1 A add iteratively a feature to estimate the shape of an average elephant, and how our elephant is reasonably obtained in model stage 4. This iterative process may involve fitting a model routinely, but the focus of

the model-centric science is not these results. It is the relationship between them (how they contribute to the aim of estimating the shape of an average elephant). This aim is achieved via a systematic process of model-building, model-evaluation, and model-refinement.

Testing the performance of M_T requires data which are generated by designing and performing an experiment⁴ Thus M_E is developed to capture a physical realization of M_T in the lab or field (step 2a), by sampling from many possible instantiations of M_T . M_E becomes the mechanism generating the data. In psychological science, the view of M_E as a data generating mechanism is largely missing. The result-centric paradigm assumes that the data are generated by a natural mechanism and hence can reveal a ground truth. In reality, M_E is a fabricated filter through which we can observe a target subsystem, yet may or may not represent it accurately. The mapping between M_T and M_E might be unknown. Lack of accuracy in M_T and practical constraints in specifying M_T may result in a misspecified model. A given M_E may not account for all variables of interest or capture key parameters, may introduce measurement error, fix certain parameters inadvertently, or introduce new ones, for instance. Building an M_E that effectively samples M_T is necessarily an iterative, exploratory process. As Gelman et al. (2020) state “The hopelessly wrong models and the seriously flawed models are, in practice, unavoidable steps along the way toward fitting the useful models.”

The data generated by M_E contains uncertainty which can only be accounted for by an M_S (step 2b). A formal M_S , buttressed by mathematical statistics, aims to accurately represent M_E while providing guarantees afforded by statistical theory. M_E and M_S need to be constructed in the same step, not sequentially: Ideally, they *are* equivalent models. To see this, imagine an in-silico simulation experiment where data are generated directly from an M_S defining the experimental design accurately. In practice, we may first construct M_S and then design an experiment satisfying M_S 's assumptions (e.g., specified sampling procedure) or first design M_E and then choose an M_S to statistically express it (e.g., a measurement error model). Either way, they are distinct representations with distinct assumptions which may lead to misspecified M_S for many reasons (e.g., nonrandom sampling, confounds). Model-centric science focuses on iteratively refining M_S in a way to accurately capture M_E and sometimes to deliberately account for its shortcomings by explicitly modeling the sources of error introduced by M_E . We need the combined forces of M_E and M_S to perform valid inference generalizable to properties of M_T . In much of result-centric psychology, M_S is a post-hoc consideration that arises after data collection and is not built in a principled way, defaulting to a conventional off-the-shelf model unquestioningly. Due to lack of explicit, careful model building and refinement, we end up with an M_S whose assumptions are not satisfied by the M_E generating the data, leading to invalid statistical and scientific inference.

In step 3, the scientist performs statistical inference about a property of M_T under M_S , using data generated by M_E . In result-centric science, this step constitutes the focus of epistemic activity, without scrutinizing any of the models contributing to its production. The caveat

⁴Here *experiment* refers to a probability experiment, a data generating mechanism with an uncertain outcome. An *experiment* encompasses observational studies and controlled experimental studies.

is that this inference can only be evaluated under a model and is only valid to the extent the model assumptions are satisfied. Statistical inference under default models is fraught (see examples in Gelman et al., 2020; Yarkoni, 2022). In contrast, a model-centric science aims to explicitly build and refine models at every stage to iteratively remove discrepancies and errors, and to develop inferential tools to handle inevitable modeling problems so as to improve the quality of statistical inference. Building a principled workflow based on models is the key to achieving scientific rigor in a model-centric paradigm. Many resources are available for psychologists ready for this paradigm shift (e.g., Gelman et al., 2020; Schad et al., 2021).

Generalizability in a model-centric paradigm

Result-centric science creates confusion in scientists' mind regarding what the key elements of generalizability are. The discussion on generalizability is substantial in Bauer (2023)'s and yet no definition of it is provided. We provide below a definition for sciences that rely on inference under uncertainty:

A set of inferences drawn from a representative sample in *population A* is *generalizable to population B* with respect to some fixed criteria, if it is equivalent with respect to that criteria.

First, the population here does not refer to a real entity that is sampled, but a well-defined set on which a variable takes values. A set of mathematical objects living on an infinitely many dimensional space might constitute a population as valid as the set of human population. Second, note the need for fixed criteria, say G for generalizability, satisfying our standards. In the absence of G , statements about generalizability are nonsensical.

Model-centric view makes a population under study explicit, prevents ill-defined categorizations, thereby providing a way to evaluate multiple generalizability issues under a single framework. Consider the two experiments consisting of (M_T, M_E, M_S) and (M_T, M_E^*, M_S) , where M_E and M_E^* are two non-equivalent experimental models. If these two experimental models work well in a target subsystem conditional on G , then the set of results obtained under them are generalizable to $(M_T, \{M_E \cup M_E^*\}, M_S)$, conditional on G . Now consider (M_T^*, M_E, M_S) or (M_T, M_E, M_S^*) where the starred and unstarred models are not equivalent have the same interpretation of generalizability. For M_T^* and M_S^* , we generalize results to the larger set of models $(\{M_T \cup M_T^*\}, M_E, M_S)$ and $(M_T, M_E, \{M_S \cup M_S^*\})$ respectively. Theoretical generalizations are of the former type to produce deeper commonalities among formal approaches to sciences, whereas statistical generalizations are of the latter type to provide robust procedures to sciences. Generalization is about making the scope of the results applicable to larger models with a keen interest in modeling.

Perhaps we consider a target subsystem involving features of human behavior or cognition (e.g., working memory). Here, typically an M_T is a verbal description of a relationship among variables and is not well-specified. Now generalizability issues arise because M_T is underspecified. The population of observations M_E samples from is statistically not well-defined, and M_E -to- M_T relationship is not well-understood for example due to lack

of probability sampling and/or arbitrary operationalization. Choice of a default M_S then may exacerbate potential model misspecification, leading to invalid statistical inference. Each level of modeling fails to specify and account for the sources of variation in the population of interest. Generalizability in a model-centric paradigm is achieved by explicitly connecting statistical inferences to the target subsystem by formalizing the process of iterating and aligning M_T , M_E , and M_S . Explicit assumptions are integral part of a statistically healthy model. Reiteration involves examining how model assumptions differ from reality and redefining them to make them healthier.

Model-centric paradigm makes the need for explicit modeling of some generalization gremlins identified by Bauer (2023) clear. Non-representativeness of samples is a concern at three modeling steps. A well-specified M_T defines a natural population of interest. M_E samples from M_T modifying this population, and M_S represents the sampling scheme M_E employs and forms the bridge between M_E and M_T by making additional assumptions, modifying the population a second time (e.g., as in post-stratification, Kennedy & Gelman, 2021). Biased operationalizations and threats to construct and ecological validity need to be modeled a priori. The focus on mean differences at the expense of individual variability dissolves almost automatically when we shift from a result-centric to a model-centric perspective, as we are now interested in explaining a target subsystem rather than obtaining an answer regarding the existence of a treatment effect. Building M_T means specifying exactly which aspects of the system we want to understand and how we represent the relationship among these aspects while M_S will explicitly account for uncertainty at different levels in the model and aim to reduce it over iterations.

Model-centric science also affords an appropriate appraisal of COG statements, because it is explicit. Proposing to add COG on all empirical papers, (Simons et al., 2017) appear to suggest that researchers should discuss the boundary conditions of their results and make declarations regarding populations, stimuli, procedures, and contexts they may generalize to without describing *how* these generalizations are defined. The examples they provide include statements about researchers' expectations, beliefs, and speculations, consistent with the result-centric view. In a model-centric science, none of this is informative. Which population the authors believe their results will hold for is irrelevant; what is relevant is the formal sampling scheme. Which procedures the authors speculate to be necessary for the results to replicate is also irrelevant when M_T and M_E are explicitly connected. All COG reduce down to a continuous assessment of model connections and assumptions.

Model-centric science not only guides scientific research, but also shapes how we approach reading and consuming science. Model-centric reader does not view a set of experimental results as absolute truth, nor do they expect every experiment to yield a groundbreaking discovery or conclusive answer to a research question. They do not base their own research solely on prior results, nor do they try to ignore uncertainty. Instead, they evaluate results against models and assumptions, carefully examining the plausibility and implications of those assumptions. They build M_T , M_E and M_S that incrementally improve upon prior work. They understand that exploring the model space in an iterative and systematic manner is crucial for potential discovery, and that most scientific activity will be exploratory.

Model-centric reader is not a passive recipient of scientific facts but a critic and co-creator of scientific models. When operating in a model-centric research culture, they can independently assess the limits of generalizability of any reported inferential claim by evaluating models and assumptions. When faced with a result-centric paradigm lacking transparency regarding model specification and assumptions, the model-centric reader holds judgment rather than taking reported results on faith.

Disclaimer

Figure 1 B is not a normative prescription for sciences. There is more to science beyond this simplistic idealization and not all scientific activity concerns layers of modeling or statistical inference. Examples include formal sciences and qualitative methods. Here, we limited our discussion to sciences that rely on performing statistical inference from data. Some scientists work exclusively on developing M_T , others on M_E or M_S . Some scientists try to bridge the gaps between M_T , M_E , and M_S . Epistemic iteration is not achieved by one scientist but as a community. Iterative modeling takes numerous model refinements between levels. The shift to model-centric science may ensure that scientists communicate with each other with precision and accuracy, engage with each others' models, and refine them in a principled workflow to cumulatively build knowledge. When model-based thinking informs science, we potentially produce much more than a list of scientific facts and generalizability becomes an in-built feature of the workflow rather than a post-hoc concern.

References

- Bauer PJ (2023). Generalizations: The grail and the gremlins. *Journal of Applied Research in Memory and Cognition*.
- Cummins R (2000). "how does it work?" vs. "what are the laws?" two conceptions of psychological explanation. keil and wilson (eds), *explanation and cognition*, mit press, 2000, pp 117–145. robert cummins.
- Devezer B, Nardin LG, Baumgaertner B, & Buzbas EO (2019). Scientific discovery in a model-centric framework: Reproducibility, innovation, and epistemic diversity. *PloS one*, 14(5), e0216125. [PubMed: 31091251]
- Epstein JM (2008). Why model? *Journal of artificial societies and social simulation*, 11(4), 12.
- Gelman A, Vehtari A, Simpson D, Margossian CC, Carpenter B, Yao Y, Kennedy L, Gabry J, Bürkner P-C, & Modrák M (2020). Bayesian workflow. *arXiv preprint arXiv:2011.01808*.
- Gigerenzer G (2018). Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science*, 1(2), 198–218.
- Godfrey-Smith P (2006). The strategy of model-based science. *Biology and philosophy*, 21, 725–740.
- Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, & Altman DG (2016). Statistical tests, p values, confidence intervals, and power: A guide to misinterpretations. *European journal of epidemiology*, 31, 337–350. [PubMed: 27209009]
- Kennedy L, & Gelman A (2021). Know your population and know your model: Using model-based regression and poststratification to generalize findings beyond the observed sample. *Psychological methods*, 26(5), 547. [PubMed: 33793269]
- Levins R (1966). The strategy of model building in population biology. *American scientist*, 54(4), 421–431.
- Marks J (2009). *Why i am not a scientist: Anthropology and modern knowledge*. Univ of California Press.
- Mayer J, Khairy K, & Howard J (2010). Drawing an elephant with four complex parameters. *American Journal of Physics*, 78(6), 648–649.

- Meehl PE (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological reports*, 66(1), 195–244.
- Meehl P (1986). 14 what social scientists don't understand. *Metatheory in social science: Pluralisms and subjectivities*, 315.
- Newell A (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium.
- Potochnik A (2015). The diverse aims of science. *Studies in History and Philosophy of Science Part A*, 53, 71–80.
- Schad DJ, Betancourt M, & Vasishth S (2021). Toward a principled bayesian workflow in cognitive science. *Psychological methods*, 26(1), 103. [PubMed: 32551748]
- Simons DJ, Shoda Y, & Lindsay DS (2017). Constraints on generality (cog): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, 12(6), 1123–1128. [PubMed: 28853993]
- van Rooij I (2022). Psychological models and their distractors. *Nature Reviews Psychology*, 1(3), 127–128.
- van Rooij I, & Baggio G (2021). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science*, 16(4), 682–697. [PubMed: 33404356]
- Wasserstein RL, & Lazar NA (2016). The asa statement on p-values: Context, process, and purpose.
- Wei J (1975). Least square fitting of an elephant. *Chemtech*, 5(2), 128–129.
- Weisberg M (2006). Forty years of 'the strategy': Levins on model building and idealization. *Biology and Philosophy*, 21, 623–645.
- Wikipedia contributors. (2023). Von neumann's elephant — Wikipedia, the free encyclopedia [[Online; accessed 10-April-2023]]. https://en.wikipedia.org/w/index.php?title=Von_Neumann%27s_elephant&oldid=1136353945
- Woods J, & Rosales A (2010). Virtuous distortion: Abstraction and idealization in model-based science. *Model-Based Reasoning in Science and Technology: Abduction, Logic, and Computational Discovery*, 3–30.
- Yarkoni T (2022). The generalizability crisis. *Behavioral and Brain Sciences*, 45, e1.

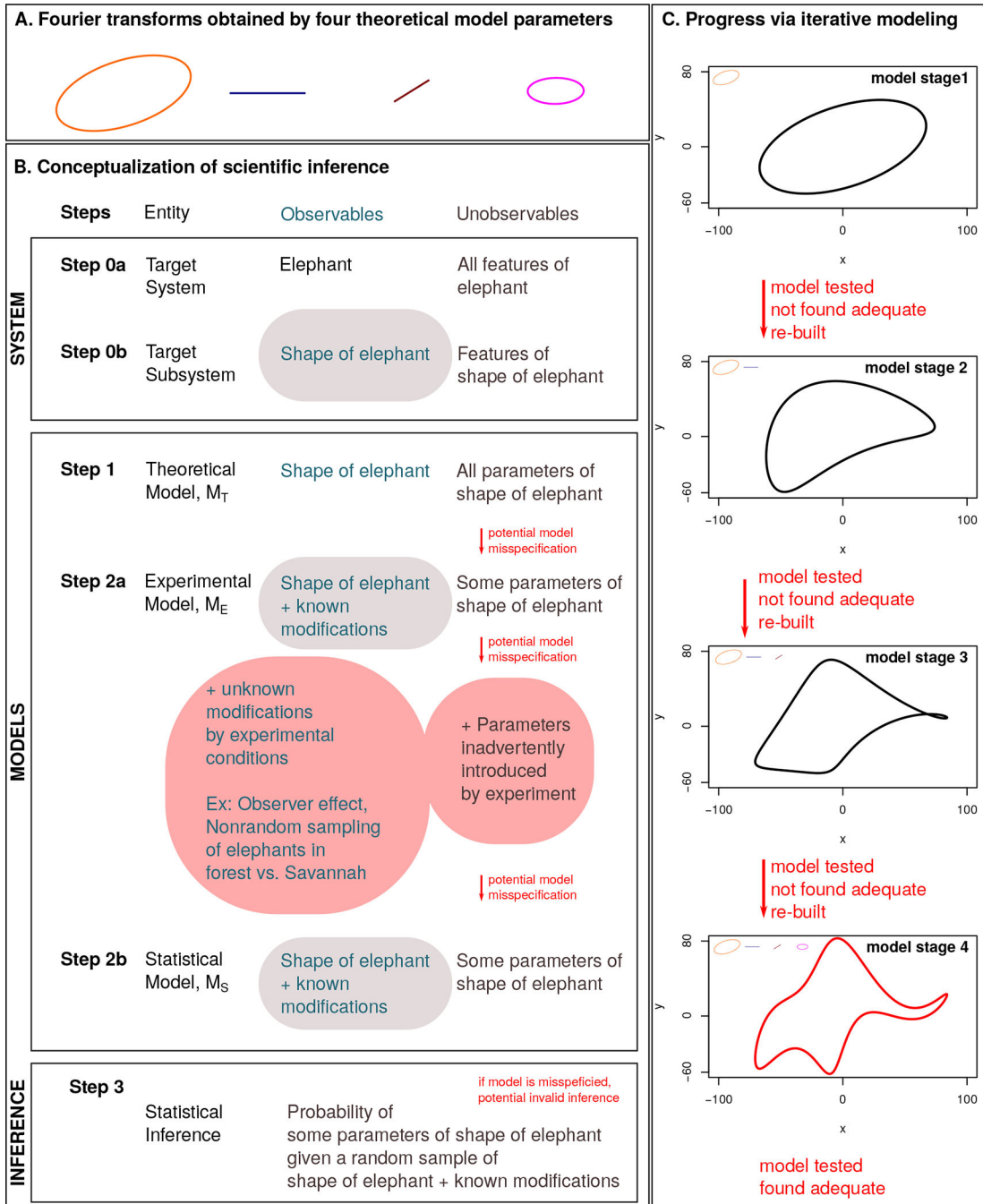


Figure 1. An example showing some of the differences between a result-centric and model-centric approach inspired by von Neumann’s elephant. **A.** Fourier transforms of four parameters (without their useful M_T) to illustrate the limitations of a result-centric approach focusing only on parameters. **B.** Conceptualization of models, observable and unobservable quantities involved in model-centric scientific inference. **C.** Iterative modeling starting with a one-

parameter model and progressing to four-parameters model to obtain the shape of an average elephant reasonably well with four parameters.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript