Preface

When research in an area is in its earliest phases, attention is typically focused on establishing evidence of a relationship between two variables, X and Y, and ascertaining whether the association is causal or merely an artifact of design, measurement, or otherwise unaccounted for influences. But as a research area develops and matures, focus eventually shifts away from demonstrating the existence of an effect toward understanding the mechanism or mechanisms by which the effect operates, as well as establishing its boundary conditions or contingencies. Answering such questions of *how* and *when* result in a deeper understanding of the phenomenon or process under investigation, and gives insights into how that understanding can be applied.

Analytically, questions of how are typically approached using *process* or *mediation analysis*, whereas questions of when are most often answered through *moderation analysis*. The goal of mediation analysis is to establish the extent to which some putative causal variable, X, influences some outcome, Y, through one or more *mediator* variables. For example, there is evidence that violent video game play can enhance the likelihood of aggression outside of the gaming context. Perhaps violent video game players come to believe through their interaction with violent game content that others are likely to aggress, that doing so is normative, that it is an effective solution to problems, or it desensitizes them to the pain others feel, thereby leading them to choose aggression as a course of action when the opportunity presents itself. In contrast, an investigator conducting a moderation analysis seeks to determine whether the size or sign of the effect of X on Y depends in one way or another on (i.e., "interacts with") a moderator variable or variables. In the realm of video game effects, one might ask whether the effect of violent video game play on later aggression depends on the player's sex, age, ethnicity, or personality factors such as trait aggressiveness, or whether the game is played competitively or cooperatively.

Both substantive researchers and methodologists have recently come to appreciate that an analysis focused on answering only *how* or *when* ques-

tions is going to be incomplete. A more fine-grained understanding of a phenomenon comes from uncovering and describing the contingencies of mechanisms—the "when of the how." The analytical integration of moderation and mediation analysis was highlighted in some of the earliest work on mediation analysis, but it is only in the last 10 years or so that methodologists have begun to talk more extensively about how to do so. Described using easily confused terms such as *moderated mediation* and *mediated moderation*, the goal is to empirically quantify and test hypotheses about the contingent nature of the mechanisms by which X exerts its influence on Y. For example, such an analysis could be used to establish the extent to which the influence of violent video game play on aggressive behavior through the mechanism of expectations about the aggressive behavior of others depends on age, sex, the kind of game (e.g., first-person shooter games relative to other forms of violent games), or the player's ability to manage anger. This can be accomplished by piecing together parameter estimates from a mediation analysis with parameter estimates from a moderation analysis and combining these estimates in ways that quantify the conditionality of various paths of influence from X to Y.

Mediation and moderation analysis are two of the more widely used statistical methods in the social, behavioral, and health sciences, as well as business, medicine, and other areas. Some of the most highly cited papers in social science methodology this century are about mediation or moderation analysis. Indeed, it is nearly imperative these days that readers and producers of research understand the distinction between these concepts and know how to implement moderation and mediation analysis in their own work. The book you are now holding is one of the few book-length treatments covering the statistical analysis of both mechanisms and contingencies. The contents of this book, classroom-tested in university courses and workshops I have conducted throughout the world over the last few years, covers the fundamentals of mediation and moderation analysis as well as their integration in the form of *conditional process analysis*, a term I introduced in the first edition of this book. Once you turn the final page, you will be well prepared to conduct analyses of the sort you see here and describe those analyses in your own research.

This is an introductory book, in that I cover only basic principles here, primarily using data from simple experimental or cross-sectional studies of the sort covered in most elementary statistics and research design courses. I do not provide much coverage of longitudinal research, multilevel analysis, latent variables, repeated measures, or the analysis of categorical outcomes, for instance, though I touch on these topics in the final chapter. I presume no special background in statistics or knowledge of matrix algebra or ad-

vanced statistical methods such as structural equation modeling. All the methods described are based entirely on principles of ordinary least squares regression (Chapter 2 introduces and reviews these principles). Most students in the social and behavioral sciences who have taken a first course in statistical methods and research design will be able to understand and apply the methods described here, as will students of public health, business, and various other disciplines.

The examples I use throughout these pages are based on data from published studies which are publicly available on the book's web page at *www.afhayes.com*, so that you can replicate and extend the analyses reported. The archive containing the data files also contains syntax files with most of the SPSS, SAS, and R code found in this book, so you won't have to type it in yourself. To facilitate the implementation of the methods introduced and discussed, in Chapter 3, I introduce a computational aide in the form of a freely available macro for SPSS, SAS, and R called PROCESS. PROCESS combines many of the functions of computational tools about which I have written and published over the years (tools that go by such names as INDIRECT, SOBEL, MODPROBE, and MODMED) into a single integrated command. PROCESS takes the computational burden off the shoulders of the researcher by estimating the models, calculating various effects of interest, and implementing modern and computer-intensive methods of inference, such as bootstrap confidence intervals for indirect effects and the Johnson–Neyman technique in moderation analysis. Example PROCESS commands are provided throughout the book, and SPSS users not interested in using the syntax version of PROCESS can install a dialog box into SPSS that makes the use of PROCESS literally as simple as pointing and clicking. This can greatly facilitate the teaching of the methods described here to students who are just getting started in the use of computers for data analysis.

This book is suitable as either a primary text for a specialized course on moderation or mediation analysis or a supplementary text for courses in regression analysis. It can be used by educators, researchers, and graduate students in any disciplines that use social science methodologies, including psychology, sociology, political science, business, and public health, among others. It will benefit the reader as a handy reference to modern approaches to mediation and moderation analysis, and Appendix A is critical to users of PROCESS, as it is the only official source of documentation for this versatile computational aid. This book will be useful to anyone interested in identifying the contingencies of effects and associations, understanding and testing hypotheses about the mechanisms behind causal effects, and

describing and exploring the conditional nature of the mechanisms by which causal effects operate.

You will find 14 chapters between the front and back covers defining five broad parts of the book. The first part, containing Chapters 1 and 2, introduces the concepts in moderation and mediation analysis and provides an example of their integration in the form of a conditional process model. I also cover a bit about my philosophy on the link between statistics and causality and describe how we should not let the limitations of our data dictate the mathematical tools we bring to the task of trying to understand what our data may be telling us. In Chapter 2, I overview ordinary least squares regression analysis. I assume that most readers of this book have been exposed to least squares regression analysis in some form already, but for those who have not or for whom much time has passed since their last regression analysis, this review will be useful while also introducing the reader to my way of thinking and talking about linear modeling.

The second part focuses exclusively on mediation analysis. In Chapter 3, I describe how linear regression can be used to conduct a simple path analysis of a three-variable $X \rightarrow M \rightarrow Y$ causal system. The estimation and interpretation of direct and indirect effects is the first focus of this chapter, first with a dichotomous causal agent X and then with a continuous X. After an introduction to PROCESS, I then cover inference about direct and indirect effects, with an emphasis on newer statistical methods such as bootstrap confidence intervals that have become the new standard in the 21st century for testing hypotheses about mechanisms in a mediation analysis. Chapter 4 covers dealing with confounds, estimation and interpretation of models with multiple X or Y variables, and quantifying effect size. In this chapter I also provide the rationale for why the historically significant *causal steps* procedure is no longer recommended by people who think about mediation analysis for a living. Chapter 5 then builds on the fundamental of mediation analysis by discussing models with multiple mediators, including the parallel and serial multiple mediator model. Chapter 6 is dedicated exclusively to mediation analysis when X is a multicategorical variable, such as in an experiment with three or more groups constructed through a random assignment procedure.

The third part temporarily puts aside mediation analysis and shifts the discussion to moderation analysis. In Chapter 7, I show how a multiple regression model can be made more flexible by allowing one variable's effect to depend linearly on another variable in the model. The resulting *moderated multiple regression model* allows an investigator to ascertain the extent to which X's influence on outcome variable Y is contingent on or *interacts with* a *moderator* variable W. Interpretation of a moderated multiple

regression model is facilitated by visualizing and probing the moderation, and techniques for doing so are introduced, along with how PROCESS can be used to make the task a lot easier than it has been in the past. Whereas Chapter 7 focuses exclusively on the case where *X* is a dichotomous variable and *W* is a continuum, Chapter 8 continues this line of analysis to models where X is quantitative rather than dichotomous. It also discusses the equivalence between the 2×2 factorial analysis of variance and moderated multiple regression, as well as why it is not necessary to enter variables into a model hierarchically to test a moderation hypothesis. Chapter 9 covers myths and truths about the need to mean-center or standardize variables in a moderation analysis, models with more than one moderator, and comparing conditional effects in complex multiple moderator models. Chapter 10, the last chapter in Part III, is dedicated to testing a moderation hypothesis using regression analysis when *X* or *W* is a multicategorical variable.

The fourth part of the book, Chapters 11 through 13, integrates the concepts and lessons described in the prior two parts by introducing *conditional process analysis*. A model which includes both a mediation and a moderation component is a conditional process model—a model in which either the direct and/or indirect effect of *X* on *Y* through *M* is moderated by or conditioned on one or more variables. Chapter 11 offers an overview of the history of this form of modeling—sometimes referred to as *moderated mediation analysis*—and provides examples in the literature of such conditional processes hypothesized or tested. An introduction to the concepts of conditional direct and indirect effects is provided, along with their mathematical bases, and an example conditional process analysis is provided, including estimation and inference using regression analysis or, more conveniently, using PROCESS. Chapter 12 provides a further example of a conditional process model with moderation of both the direct and indirect effects simultaneously, and shows the equivalence between this one specific model form and something known as *mediated moderation*. But I take a stand in this chapter and argue that unlike moderated mediation, mediated moderation is not a particularly interesting concept or phenomenon and probably not worth hypothesizing or testing. Chapter 13 addresses an example of conditional process analysis when *X* is a multicategorical variable.

The last part of the book contains only one chapter. Chapter 14 addresses various questions that I am frequently asked by readers of prior editions of this book, people who have taken workshops from me, or others who have contacted me by e-mail over the years. The largest section in Chapter 14 is dedicated to writing about mediation, moderation, and conditional process analysis. The rest of the chapter touches on various

miscellaneous issues and questions and a (typically) brief response to each, from my perspective at least.

The appendices are very important, as they are the best source of information about how to use PROCESS. Appendix A is essentially a user's manual for PROCESS that discusses how to set up the macro, construct a PROCESS command, and various options available in PROCESS that vary depending on the analysis being conducted. Appendix B focuses on an important feature in PROCESS that allows you to construct your own model rather than having to rely on one of the many preprogrammed models built into PROCESS.

I have taken care to maintain a light and conversational tone throughout the book while discussing the concepts and analyses, without getting heavily into the mathematics behind them. I believe that maintaining a reader's interest is one of the more important facets of scientific writing, for if one's audience becomes bored or disinterested and attention begins to wander, the power and influence of the message is reduced. Indeed, it is this philosophy about writing that guides the advice I give in Chapter 14, where I talk about how to report a mediation, moderation, or conditional process analysis. Most importantly, the advice I offer in this part of the book is intended to empower you as the one best positioned to determine how you tell the story your data are telling you.

New to the Third Edition

You are holding the third edition of *Introduction to Mediation, Moderation, and Conditional Process Analysis*. This new edition is a bit longer than the second edition but only slightly heavier. The sequence of chapters in the third edition is consistent with the second edition, and the vast majority of material in the second edition has been retained. Yes, most pages are a bit different, most significantly the result of adding syntax support for R users and the release of PROCESS for R. Other changes include the addition of new sections dispersed throughout the book. Below is a nonexhaustive list of some of the changes in this edition relative to the second edition:

- New code for R users accompanying every example, including PROCESS for R released after the second edition of the book was printed.
- A substantially rewritten Appendix A to reflect new features added to PROCESS since the second edition, including a discussion of similarities and differences in the syntax structure in PROCESS for R compared to PROCESS for SPSS and SAS.

- A more detailed discussion of effect scaling and the difference between unstandardized, completely standardized, and partially standardized effects in Chapters 3 and 4 and the implementation of standardized regression coefficients in PROCESS.
- A rebuttal in Chapter 4 to a claim made by Yzerbyt, Muller, Batailler, and Judd (2018) that we should return to the test of joint significance rather than rely on bootstrap confidence intervals for inference about indirect effects.
- A new discussion in Chapter 5 about a method for comparing the strength of two specific indirect effects that are different in sign.
- An expanded discussion and illustration in Chapter 5 of the partial correlation between mediators in a multiple mediator model and the generation of this correlation using a new PROCESS option.
- The introduction in Chapters 11 and 12 of a bootstrap-based Johnson–Neyman-like approach for probing moderation of mediation in a conditional process model, with R code for generating a visual depiction of regions of significance.
- A discussion in Chapter 14 about testing for interaction between a causal antecedent variable X and a mediator M in a mediation analysis and how to test this assumption in a new feature available in PROCESS.
- A section on power analysis and sample size originally in Chapter 4 has been expanded and relocated to Chapter 14.
- Appendix C from the second edition describing MCMED, a macro for the construction of Monte Carlo confidence intervals, has been cut and moved to supplementary materials online at *www.afhayes.com*.

Acknowledgments

I began writing this book well before the first word of it was typed. Several years ago I started receiving invitations from former strangers, many of whom are now colleagues and friends, to come speak on the topic of papers I have published or to conduct workshops on material related to this book. These invitations allowed me to interact with people I otherwise would not likely have ever had opportunity to get to know. Speaking to audiences diverse in background and interests has provided a means to fine-tune and hone my message as my own ideas and philosophy about the contents of

this book evolved. Without those invitations, the hospitality of my hosts, and the time sacrifices they made orchestrating and coordinating my visits, this book would not be anything like what it is, nor would it have evolved from the first to this third edition in the way that it has. So I offer my thanks to Paul Allison, Billy Bai, Julian Barling, H. Onur Bodur, Leslie Booren, Adrian Brügger, Chen Chao, Jonathan Cohen, Roberta Crouch, Grete Dyb, Truls Erikson, Joseph Fletcher, David Gligor, Ryan Googenmas, Shira Dvir Gvirsman, Tilo Hartmann, Andreas Herrmann, Carol Esmark Jones, Kristina Klein, Hans-Joachim Knopf, Catherine Lamberton, Todd Little, Kelley Main, Jörg Matthes, Osvaldo Morera, Peter Neijens, Hristina Nikolova, Harmen Oppewal, Deirdre O'Shea, Jochen Peter, Torsten Pieper, Carolin Plewa, Nicolas Pontes, Stacey Robinson, Michelle Salyers, Jennifer Skeem, Graham Stead, Toon Taris, Annika Tovote, Danielle van Jaarsveld, Jens Vogelgesang, Claes de Vreese, Katherine White, Etty Wielenga-Meijer, Anna Woodcock, Gülden Ülkümen, and everyone else who has spent time with me during my travels talking about their research interests and their lives.

As much as I enjoy speaking and teaching abroad, I do most of my teaching a short walk from my house. Over the years I have had the pleasure of teaching graduate students, both beginning and advanced, working on degrees at The Ohio State University and, more recently, the University of Calgary. Their questions have helped me sharpen my language when it comes to describing abstract concepts in terms that are concrete without being too imprecise or oversimplified. I appreciate their tolerance for numerous typos on PowerPoint slides, patience with my occasional need to repeat myself when I botch an explanation, and now and again waiting attentively as I retype SPSS, SAS, or R code that generates a string of errors when using macros and other tools I invented but can't always remember how to use.

Heartfelt thanks also go to numerous people who have been generously willing to donate their data for use in my classes, workshops, journal articles, and this book. These include Daniel Ames, George Bonanno, Nyla Branscombe, Heike Bruch, Jonathan Cohen, Michael S. Cole, Carsten K. W. De Dreu, Ryan Duffy, Naomi Ellemers, Chip Eveland, Francis Flynn, Donna Garcia, Friedrike X. R. Gerstenberg, Al Gunther, Sri Kalyanaraman, Brian Lickel, Anthony Mancini, Jörg Matthes, Erik Nisbet, Kirsi Peltonen, Jeffrey Pollack, Raija-Leena Punamäki, Michael Schmitt, Michael Slater, Nurit Tal-Or, S. Shyam Sundar, Yariv Tsfati, Eric Van Epps, and Frank Walter. Writing and reading about methodology is much more interesting when examples are based on real data from existing and published studies rather than hypothetical studies made up for the purpose of illustration.

For the record, I should point out that all analyses conducted in this book and claims I make based on others' data are my own and are not necessarily endorsed by those who collected the data in the first place.

C. Deborah Laughton and Seymour Weingarten at The Guilford Press have been very supportive and enthusiastic about this project and others I have worked on for them, and I appreciate their contributions before the writing began, during the production process, and as the first edition evolved into the second. No doubt they will continue to influence its course well into the future. Matthew Fritz offered a review of the manuscript of the first edition prior to production, and I appreciate his insights and recommendations.

The support of my wife, Carole, and kids, Conor and Julia, has been critical. As anyone who has been a part of a research team knows, a study is much more than just the journal article that describes it. There is much that happens behind the scenes of a study that is invisible to outsiders but without which the study just doesn't get done. My family is similar to members of a research lab in that sense. Fortunately, they understand the time commitment that a project like this entails. This is the third time in 15 years they have had to put up with the divided attention that comes with writing a book, especially as the due date approaches, and I appreciate their tolerance.

Finally, I would also like to tell the world in writing about the gratitude I feel toward my father for buying me my first computer in high school, and my mother for allowing me to lock myself away in my room as I taught myself BASIC. I imagine my father didn't think long or deeply about his decision to spend $300 on a Commodore VIC-20 back in the early 1980s, but it is the machine I learned to program on, and it turned out this decision had a big influence on where I ended up in my professional life. Without this early introduction to computer science, I probably wouldn't have chosen this career, I probably wouldn't have written PROCESS, and, as a consequence, this book simply would not exist.