# 1 Introduction



Research that establishes the mechanism or mechanisms by which effects operate or the conditions that facilitate and inhibit such effects deepens our understanding of the phenomena scientists study. Mediation analysis and moderation analysis are used to establish evidence or test hypotheses about such mechanisms and boundary conditions. Conditional process modeling is used when one's research goal is to describe the boundary conditions of the mechanism or mechanisms by which a variable transmits its effect on another. Using a regression-based path-analytic framework, this book introduces the principles of mediation analysis, moderation analysis, and their unification as conditional process analysis. In this initial chapter, I provide a conceptual overview of moderation and mediation and describe an example of a conditional process analysis that combines elements of both mediation and moderation analysis. After articulating my perspective on the use of statistical methods when testing causal processes, I end with a synopsis of the chapters that follow.

## 1.1 A Scientist in Training

As an undergraduate student studying psychology at San Jose State University back in the 1980s, one of the first empirical research projects I undertook was a study on the relationship between students' attitudes about college and their selection of seat in the classroom. I developed an instrument that purportedly (although in hindsight, not really) measured whether a person felt getting a college education was generally a good and important thing to do or not. After the participants in the study completed the instrument, I presented each of them with a diagram of a generic college classroom, with seats arranged in a 6 (row) by 5 (column) matrix, and I asked them to mark which seat they would choose to sit in if they could choose any seat in the classroom. Based on which row he or she selected, I scored how close to the front of the room that participant preferred (6 = front row, 5 = second row, 4 = third row, and so forth).

With these two measurements collected from over 200 students attending San Jose State, I could test my prediction that students with a more positive attitude about college (i.e., who scored higher on my attitude scale) would prefer sitting closer to the front of the classroom. Indeed, when I calculated Pearson's coefficient of correlation between the two measurements, I found the relationship was positive as expected, $r = 0.27$. Furthermore, a hypothesis test revealed that the probability of obtaining a correlation this extreme or more extreme from zero (positive or negative, as I tested the hypothesis two-tailed even though my prediction was directional) was too small ($p < .001$) to consider it just a fluke or "chance." Naturally, I was excited, not realizing as I do now that *any* result is exciting whether consistent with a prediction or not. Unfortunately, three anonymous reviewers did not share my enthusiasm, and the then-editor of the *Journal of Nonverbal Behavior* let me know in no uncertain terms that this finding was neither of sufficient interest nor derived with sufficient rigor to warrant publication. Rather than rewriting the paper and resubmitting elsewhere, I filed the paper away and moved to upstate New York to pursue a PhD in social psychology.

After more than 20 years, I still have this paper, and now and then I take it out of my file drawer when reflecting on where I have been in my professional life and where I am going. Looking at it now, it is clear to me that the reviewers were correct and the editor's decision sound and justified. Even if the study had been conducted with the kind of rigor I now ask of myself and my own students, in the paper I offered nothing but speculation as to why this association existed. Furthermore, I could not establish the direction of cause, if any. Although I argued that variations in attitudes caused variation in seat choice, it is just as plausible that where one sits influences one's attitude about college. For example, perhaps students who sit closer to the front receive more attention and feedback from the instructor, can hear and see better and therefore learn more, and this in turn leads them to feel better about the college experience in general. Even if I was able to ascertain why the association exists or the direction of cause, I was in no position to be able to describe its boundary conditions, such as the type of people in whom this relationship would be expected to be larger or smaller. For instance, no doubt there are many bright students who love the college experience but for one reason or another choose to sit in the back, just as there are students who sit in the front even though they would much rather be somewhere else—anywhere else—than in that classroom.

I have learned many lessons about research over the years—lessons that began with that first early and unsuccessful attempt at academic publishing.

I have learned that research is tough, that it takes patience, and that our egos often get too involved when we interpret feedback from others. Although this particular study never was published, I have learned that resilence to rejection combined with persistence following failure often does lead to success. But I think one of the more important lessons I've learned being both a producer and a consumer of research is how much more impressive a study is when it can speak to more than just whether an effect exists, whether a relationship is different from zero, or whether two groups differ from each other. Instead, some of the best research I have done and the best research I have read goes further by answering not only "whether" or "if," but also "how" and "when." Approaches to analyzing one's data with the goal of answering these latter two questions is the topic of this book.



## 1.2 Questions of Whether, If, How, and When

Questions of "whether" or "if" focus primarily on whether two variables are related, causally or otherwise, or if something is more or less likely to happen in one set of circumstances or conditions than another. Such questions are often the first ones a scientist-in-training asks, sometimes merely by observing the social world around him or her and wondering about it. For example, I occasionally teach a large undergraduate course on research methods to students in the social sciences. In this course, I require the students to conceive a study, collect some data, and write up the results. By far the most popular research topic proposed by students in this class is the effects of exposure to the thin ideal on self-esteem and body dissatisfaction. Term after term several groups of students want to design a study to see if women who are exposed to images of women depicted in beauty magazines, the Internet, popular television, and music videos—as thin and beautiful—suffer in some way from this exposure. I believe this is such a popular topic because it is nearly impossible to avoid the daily bombardment by the media of depictions of what the ideal woman should look like and, by extension, what society seems to value. Naturally, many wonder whether this is bad for women and society—if women's sense of worth, image of their bodies, and likelihood of disordered eating are affected by this exposure.

Questions of the whether or if variety also serve as a starting point in our quest to understand the effects of something that has happened in society, when a new technology is developed, when a new problem confronts the people of a community or nation, and so forth. After the twin towers of the World Trade Center in New York City were brought down by terrorists on September 11, 2001, researchers started asking whether and what kind of physical and psychological health effects it had on those who experienced it (e.g., Cukor et al., 2011; DiGrande et al., 2008), those who only observed it from a distance (e.g., Mijanovich & Weitzman, 2010), or how people's behavior changed after the event (e.g., Richman, Shannon, Rospenda, Flaherty, & Fendrich, 2009). And a relatively new genre of television known as *political entertainment* has spawned much research about its viewers and whether shows like *The Daily Show* or *The Colbert Report* serve to politically educate, mobilize, or demotivate those who view them (e.g., Baumgartner & Morris, 2006; Xenos & Becker, 2009).

The empirical literature in most every scientific discipline is replete with research that provides answers to questions of whether or if, and for good reason. Many theoretical and applied questions in the sciences focus on whether there is evidence of association between some presumed causal antecedent $X$ and some putative consequent or outcome $Y$. Is a particular therapeutic method effective at reducing depression (e.g., Hofmann & Smits, 2008)? Does combining drugs with psychotherapy work better than therapy alone (e.g., Cuijpers, van Straten, Warmeredam, & Andersson, 2009)? Does playing violent video games or watching violent television make people aggressive (e.g., Anderson & Bushman, 2001; Anderson et al., 2010)? Does exposure to negative political advertisements turn people off from participating in the political process (e.g., Lau, Silegman, Heldman, & Babbit, 1999)? Are the children of divorced parents more prone to behavioral or psychological problems than children of married parents (e.g., Amato, 2001; Amato & Keith, 1991)? Does rewarding performance at work increase employee satisfaction and reduce turnover (e.g., Judge, Piccolo, Podsakoff, Shaw, & Rich, 2010)? What sets science apart from armchair speculation is that we can answer such questions by collecting data. Being able to establish that two variables are associated—that an effect or relationship of some kind exists—is in part what science is about, and research that does so is worth undertaking. Indeed, the drive to answer questions of this sort is one of the things that motivates scientists to get up in the morning.

But establishing association does not translate into deep understanding even when a causal association can be established. We know that we better understand some phenomenon when we can answer not only whether $X$ affects $Y$, but also *how* $X$ exerts its effect on $Y$, and *when* $X$ affects $Y$ and when it does not. The "how" question relates to the underlying psychological, cognitive, or biological process that causally links $X$ to $Y$, whereas the "when" question pertains to the boundary conditions of the causal association—under what circumstances, or for which types of people, does

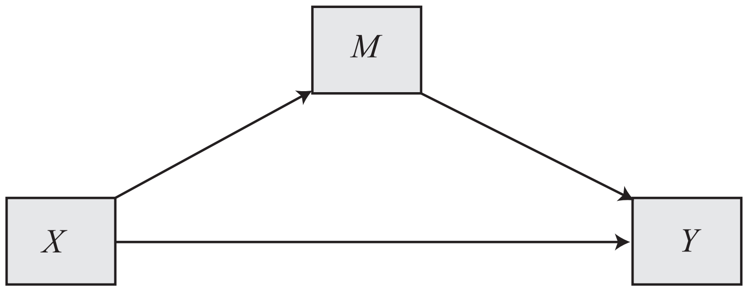


**FIGURE 1.1.** A simple mediation model with a single mediator variable $M$ causally located between $X$ and $Y$.



$X$ exert an effect on $Y$ and under what circumstances, or for which type of people, does $X$ not exert an effect?

## Mediation

A researcher whose goal is to establish or test how $X$ exerts its effect on $Y$ frequently postulates a model in which one or more intervening variables $M$ is located causally between $X$ and $Y$. One of the simplest forms of such a model is depicted in Figure 1.1. These intervening variables, often called *mediators*, are conceptualized as the mechanism through which $X$ influences $Y$. That is, variation in $X$ causes variation in one or more mediators $M$, which in turn causes variation in $Y$. For example, there is clear evidence that exposure to the thin ideal through the mass media is a risk factor if not an actual cause of body dissatisfaction in women (e.g., Grabe, Ward, & Hyde, 2008; Levine & Murnen, 2009). But how does this occur? Research suggests that internalization of the norm functions as a mediator of this relationship (Lopez-Guimera, Levine, Sanchez-Cerracedo, & Fauquet, 2010). Women who report greater exposure (or who are given greater exposure experimentally) to the thin-as-ideal image of women are more likely to internalize this image and seek thinness as a personal goal than those with less exposure. Such internalization, in turn, leads to greater body dissatisfaction (Cafri, Yamamiya, Brannick, & Thompson, 2005). So internalization of the standard portrayed by the media is one mechanism that links such exposure to body dissatisfaction. Of course, other mechanisms may be at work too, and Lopez-Guimera et al. (2010) discuss some of the other potential mediators of the effect of such exposure on women's beliefs, attitudes, and behavior.

Investigators interested in examining questions about mechanism resort to *process modeling* to empirically estimate and test hypotheses about the two pathways of influence through which $X$ carries its effect on $Y$ depicted in Figure 1.1, one *direct* from $X$ to $Y$ and the other *indirect* through $M$. More popularly known as *mediation analysis*, this type of analysis is extremely common in virtually all disciplines. Some of the most highly cited journal articles in methodology both historically (e.g., Baron & Kenny, 1986) and more recently (e.g., MacKinnon, Lockwood, Hoffman, & West, 2002; Preacher & Hayes, 2004, 2008a) discuss mediation analysis and various statistical approaches to quantifying and testing hypotheses about direct and indirect effects of $X$ on $Y$. I describe the fundamentals of mediation analysis in Chapters 4 through 6.

## Moderation

When the goal is to uncover the boundary conditions for an association between two variables, moderation analysis is used. An association between two variables $X$ and $Y$ is said to be moderated when its size or sign depends on a third variable or set of variables $M$. Conceptually, moderation is depicted as in Figure 1.2, which depicts moderator variable $M$ influencing the magnitude of the causal effect of $X$ on $Y$. Moderation is also known as *interaction*. For example, experimental studies of exposure to the thin-as-ideal standard reveal that such exposure tends to have a larger effect on body dissatisfaction and affect among women who have already internalized the thin-as-ideal standard (see, e.g., Groetz, Levine, & Murnen, 2002). In other words, relative to women who strive for thinness as a personal goal, women who buy in less to the social norm that thinner is better are less likely to show evidence of body dissatisfaction after exposure to thin models through media images. So internalization of the norm ($M$) functions as moderator of the effect of exposure to images reflecting the thin-as-ideal norm ($X$) on body dissatisfaction ($Y$).

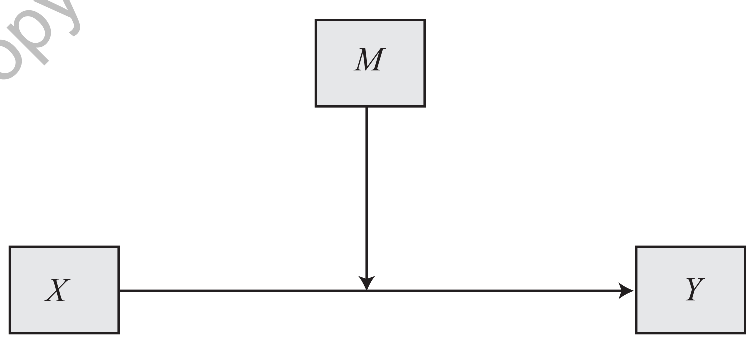


**FIGURE 1.2.** A simple moderation model with a single moderator variable $M$ influencing the size of $X$'s effect on $Y$.

Statistically, moderation analysis is typically conducted by testing for *interaction* between $M$ and $X$ in a model of $Y$. With evidence that $X$'s effect on $Y$ is moderated by $M$, the investigator typically will then quantify and describe the contingent nature of the association or effect by estimating $X$'s effect on $Y$ at various values of the moderator, an exercise known as *probing an interaction*. The basic principles of moderation analysis are introduced in Chapters 7 to 9.

This example illustrates that the answers to how and when questions can be intertwined. A variable could function as either a mediator or a moderator, depending on how the phenomenon under investigation is being conceptualized and tested. And in principle, the same variable could serve both roles simultaneously for certain processes that evolve and operate over long periods of time. For instance, early exposure to media images that portray the thin-as-ideal norm can persuade adolescents that thin is indeed better, which results in body dissatisfaction given that few women can live up to this unrealistic and even unhealthy standard. Of course, not all young women will buy into this message. Among those who do, once this norm has been internalized and adopted as a personal goal, it is more likely to influence how such women perceive themselves following later exposure to this norm relative to those who don't believe thinner is better.



## 1.3 Conditional Process Analysis

It is not difficult to find examples of mediation and moderation analysis in the empirical literature, and there have been numerous papers and book chapters emphasizing the value of moderation and mediation analysis to further understanding processes of interest to researchers in specific disciplines, many of which also provide methodological tutorials (e.g., Baron & Kenny, 1986; Breitborde, Srihari, Pollard, Addington, & Woods, 2010; Bryan, Schmiege, & Broaddus, 2007; Dearing & Hamilton, 2006; Eveland, 1997; Fairchild & McQuillin, 2010; Frazier, Tix, & Barron, 2004; Gogineni, Alsup, & Gillespie, 1995; Holbert & Stephenson, 2003; James & Brett, 1984; Kraemer, Wilson, Fairburn, & Agras, 2002; Krause, Serlin, Ward, & Rony, 2010; Lockhart, MacKinnon, & Ohlrich, 2011; MacKinnon, Fairchild, & Fritz, 2007; Magill, 2011; Maric, Wiers, & Prins, 2012; Preacher & Hayes, 2008b; Ro, 2012; Whisman & McClelland, 2005). However, rather infrequently is the combination of the two discussed in the same article. Researchers are advised to estimate indirect effects and look for interactions, but rarely both in an integrated analytical model. This lack of attention to the integration of moderation and mediation analysis may be due in part to the fact that analytical procedures that combine moderation and mediation were introduced in a systematic fashion to the research community only in the last 10 years or so. For instance, Muller, Judd, and Yzerbyt (2005) write about the mediation of a moderated effect and the moderation of a mediated effect, Edwards and Lambert (2007) provide a framework for testing hypotheses that combine moderation and mediation using path analysis, and Preacher, Rucker, and Hayes (2007) introduce the concept of the "conditional indirect effect" as a quantification of the contingent nature of a process or mechanism and provide techniques for estimation and inference (additional articles include Morgan-Lopez & MacKinnon, 2006; Fairchild & MacKinnon, 2009).

In part as a result of these articles, researchers are now throwing around terms such as "mediated moderation," "moderated mediation," and "conditional indirect effects" relatively frequently, but often are only somewhat awkwardly implementing the corresponding analytical methods because of a lack of clear guidance from methodologists for how to properly do so and write about it. To be sure, the few methodology articles that do exist attempt to speak to the user, and some provide statistical software code or tools to ease the implementation of the methods discussed, but only so much can be accomplished in a single journal article. Furthermore, the advice that does exist is fragmented and spread across multiple articles in different journals. Part IV of this book is dedicated to the analytical integration of mediation and moderation using a data-analytical strategy I have termed *conditional process modeling* or *conditional process analysis*.

Conditional process analysis is used when one's research goal is to describe the conditional nature of the mechanism or mechanisms by which a variable transmits its effect on another and testing hypotheses about such contingent effects. As discussed earlier, mediation analysis is used to quantify and examine the direct and indirect pathways through which an antecedent variable $X$ transmits its effect on a consequent variable $Y$ through one or more intermediary or mediator variables.[1] Moderation analysis is used to examine how the effect of antecedent variable $X$ on a consequent $Y$ depends on a third variable or set of variables. Conditional process analysis is both of these in combination and focuses on the estimation and interpretation of the conditional nature (the moderation component) of the indirect and/or direct effects (the mediation component) of $X$ on $Y$ in a causal system. Although not known by this name, the methodology articles mentioned earlier have prompted an increasingly widespread adoption of this analytical method. It is not difficult to find examples of conditional process modeling in the empirical literature of many disciplines, including social psychology (Popan, Kenworthy, Frame, Lyons, & Snuggs,

[1] *Antecedent* and *consequent* variables will be formally defined in Chapter 3.

2010; van Dijke & De Cremer, 2010), health psychology (Luszczynska et al., 2010), biological psychology (Oei, Tollenaar, Elzinga, & Spinhoven, 2010), developmental psychology (Parade, Leerkes, & Blankson, 2010), clinical psychology and psychiatry (Goodin, McGuire, Stapleton, et al., 2009; Rees & Freeman, 2009), cognitive psychology (Naumann, Richter, Christmann, & Groeben, 2008), public health (Blashill & Wal, 2010), sociology (Li, Patel, Balliet, Tov, & Scollon, 2011), women's studies (Sibley & Perry, 2010), neuroscience (Goodin, McGuire, Allshouse, et al., 2009), business and management (Cole, Bedeian, & Bruch, 2011; Cole, Walter, & Bruch, 2008), and communication (Antheunis, Valkenburg, & Peter, 2010; Jensen, 2008, 2011; Palomares, 2008), among others.

A concrete example will help to clarify just what conditional process analysis is all about. Just prior to writing this chapter, the U.S. Congress held the American and world economies hostage over largely politically motivated disagreements and fighting over the conditions under which the amount of money the government is allowed to borrow can be raised—the so-called *debt ceiling*. In part as a result of this political bickering and a failure of Congress to adequately address spending and revenue problems, Standard & Poor's lowered the credit rating of the U.S. government for the first time in history, from AAA to AA+. U.S. unemployment is currently at a recent high at over 9%, housing prices are falling, and so too is the value of people's retirement portfolios. The Greek economy was recently bailed out by the International Monetary Fund, the European Union is facing economic instability, and a few months ago a major earthquake followed by a tsunami and near-nuclear meltdown at a power plant in Japan roiled the Japanese people and its economy. Not to downplay the significance of a bad economy for the public at large, but imagine owning a business in this kind of environment, where your economic livelihood and your ability to pay your workforce and your creditors depends on a public that is reluctant to let go of its money. Personally, I'd seriously think about finding another profession. Perhaps that is why I chose the relatively recession-proof profession of university professor.

It is in this context that Pollack, VanEpps, and Hayes (2012) conducted a study examining the affective and cognitive effects of economic stress on entrepreneurs. Of primary interest was whether economic stress prompts business owners to contemplate pursuing other careers, giving up their entrepreneurial roles, and just doing something else instead. But they went further than asking just whether economic stress is related to such "withdrawal intentions." They proposed that such economic stress leads to depressed affect, which in turn enhances their intention to leave entrepreneurship and pursue another vocation. This is a question about not whether but *how*. On top of this, they proposed that entrepreneurs who are more socially connected to others in their field would be less susceptible to the deleterious effects of economic stress. Having the support of other entrepreneurs in your business community could help to buffer the effects of that stress on depression and, in turn, the desire to leave the business. This proposed explanation addresses a question of *when*. Under what circumstances, or for which type of people, is the effect of stress on depression and business withdrawal intentions large versus small or even zero?

To conduct this study, Pollack et al. (2012) sent a survey to members of Business Networking International, a social networking group for small business owners. The 262 respondents were asked a series of questions used to score the economic stress they felt related to their business (higher score = more stress), whether and how much they thought about withdrawing from entrepreneurship (higher score = greater intentions to leave), the extent to which they felt various emotions (e.g., discouraged, hopeless, inadequate) related to their business over the last year (higher score = more depressed affect), and how many people they spoke to, e-mailed, or met with face-to-face about their business on a daily basis from this networking group (higher score = more social ties).

Somewhat surprisingly perhaps, there was no evidence of an association between economic stress and withdrawal intentions. Entrepreneurs who reported feeling more economic stress were no more or less likely to report greater intentions to withdraw from their business than those who felt less stress ($r = 0.06, p > .05$). But that is not the whole story, for this finding belies what is a more interesting, nuanced, and, ultimately, conditional process. A moderation analysis revealed that those who reported relatively higher stress did report relatively higher withdrawal intentions compared to those with lower stress (i.e., the relationship was positive), but this was true only among those with relatively few social ties with network members. Among those who reported relatively more social ties, there was little or even a *negative* association between economic stress and withdrawal intentions. So social ties seemed to buffer the effects of stress on desire to withdraw from their business enterprise. This is *moderation*; social ties moderates the effect of economic stress on withdrawal intentions.

Pollack et al. (2012) proposed that the effect of economic stress on entrepreneurial withdrawal intentions operated through negative affect. That is, economic uncertainty and the resulting stress it produces bums business owners out, makes them feel inadequate and helpless, and leads them to choose to pursue other careers. This is *mediation*. In fact, participants who reported more economic stress did report more depressed affect ($r = 0.34, p < .01$), and those who reported more depressed affect

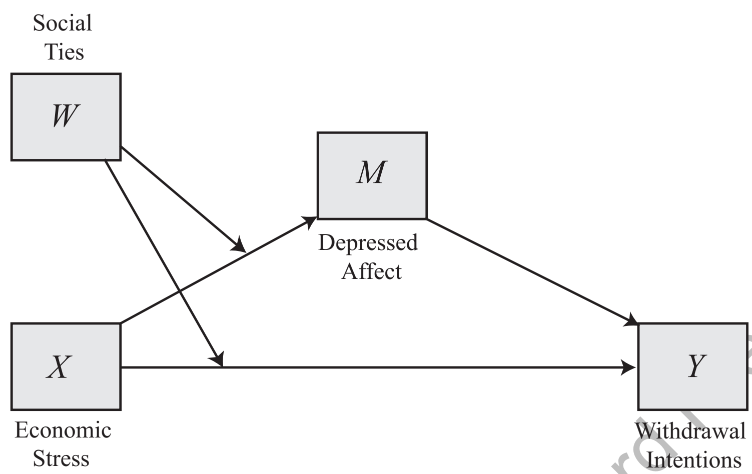


**FIGURE 1.3.** A conceptual diagram of a conditional process model corresponding to the Pollack et al. (2012) study.



reported greater intentions to withdraw ($r = 0.42, p < .01$). But this process, according to Pollack et al. (2012), can be "interrupted" by strong social ties. Having people you can lean on, talk to, or bounce ideas off to manage the business-related stress can reduce the effects of such stress on how you feel and therefore how you think about your future as a business owner. The evidence was consistent with the interpretation that economic stress affects how business owners feel, depending on their social ties. Entrepreneurs under relatively more economic stress who also had relatively few social ties reported relatively more business-related depressed affect. But among those with relatively more social ties, economic stress was unrelated or even negatively related to negative affect. So social ties moderated the effect of stress on negative affect as well as on withdrawal intentions.

A conceptual diagram of a conditional process model corresponding to this example can be found in Figure 1.3. This diagram depicts what some have called *moderated mediation* and others have called *mediated moderation*. In fact, it depicts both. It has been given other labels as well, such as a *direct effect and first stage moderation model* (Edwards & Lambert, 2007) or simply "model 2" (Preacher et al., 2007). Regardless, observe that this diagram depicts two moderated relationships, one from economic stress to depressed affect ($X \rightarrow M$), and the other from economic stress to withdrawal intentions ($X \rightarrow Y$), both of which are diagrammed as moderated by social ties ($W$). In addition, there is an indirect effect of economic stress on withdrawal intentions through depressed affect depicted ($X \rightarrow M \rightarrow Y$), but because this indirect effect includes a component that is proposed as moderated (the $X \rightarrow M$ association), the indirect effect is also moderated

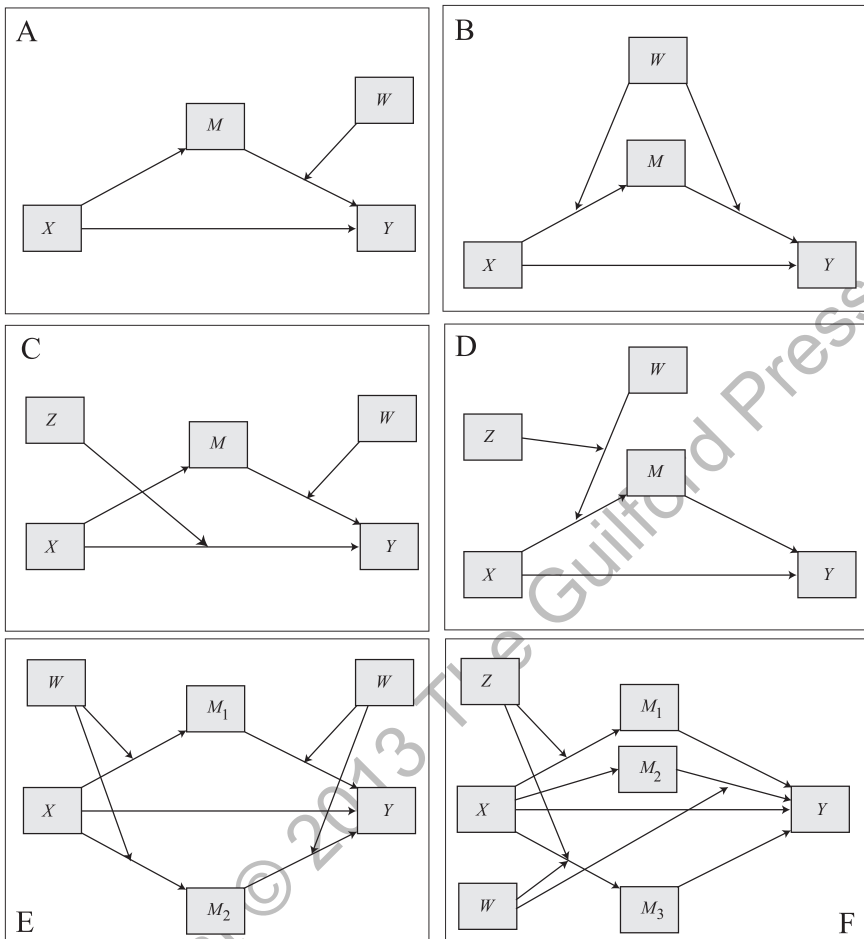


**FIGURE 1.4.** Some variants of a conditional process model, from quite simple (A) to fairly complex (F).



or *conditional*. The direct effect of economic stress on withdrawal intentions ($X \rightarrow Y$) is also depicted as moderated. According to this diagram, it too is conditional, for it depends on social ties. Thus, the process linking economic stress to withdrawal intentions through depressed affect is moderated or conditional, hence the term *conditional process model*. Throughout this book I describe how to piece the components of this model together and estimate and interpret direct and indirect effects, moderated as well as unmoderated.[2]

[2]It turns out in this case that there was no evidence that the direct effect of economic stress on withdrawal intentions was moderated by social ties, even though the so-called *total effect* was moderated. The moderation of the total effect of economic stress on withdrawal

The example depicted in Figure 1.3 is only one of the forms that a conditional process model can take. A few additional possibilities can be found in Figure 1.4, but these still represent only some of the many, many ways that moderation and mediation can be combined into a single integrated model. Panel A depicts a model in which the $M \rightarrow Y$ effect is moderated by $W$, called a *second stage moderation model* in terms introduced by Edwards and Lambert (2007). For examples of this model in published research, see Cole et al. (2008) and Antheunis et al. (2010). The model in panel B adds moderation of the $X \rightarrow M$ effect to the model in panel A, yielding a *first and second stage moderation model* (Edwards & Lambert, 2007). Parade et al. (2010) provide an example of this model. Panel C is like the model in panel A but adds moderation of the direct effect of $X$ ($X \rightarrow Y$) by $Z$. Panel D depicts moderation of the $X \rightarrow M$ effect by $W$, which itself is moderated by $Z$. See Chang (2010) for an example. Panels E and F show models with two mediators. The model in panel E is similar to panel B but includes moderation by $W$ of all effects to and from $M_1$ and $M_2$ (see, e.g., Takeuchi, Yun, & Wong, 2011). Panel F depicts a complex model (see Andreeva et al., 2010) with three mediators and two moderators. In this model, the $X \rightarrow M_3$ effect is moderated by both $W$ and $Z$, the $X \rightarrow M_1$ effect is moderated by $Z$, and the $M_2 \rightarrow Y$ effect is moderated by $W$.

## 1.4 Correlation, Causality, and Statistical Modeling

The study of economic stress in entrepreneurs just described illustrates what conditional process modeling is all about, but it also illustrates what some construe as a weakness of mediation analysis in general, as well as how liberally people often attribute causality as the mechanism producing the associations observed in any kind of study. These findings come from a cross-sectional survey. This study is what is often called called *observational* rather than experimental. All measurements of these entrepreneurs were taken at the same time, there is no experimental manipulation or other forms of experimental control, and there is no way of establishing the causal ordering of the relationships observed. For example, people who are feeling down about their business might be more likely to contemplate withdrawing, and as a result they work less, network less often with other business leaders, and feel more stress from the economic pressures that build up as a result. The nature of the data collection makes it impossible to establish what is causing what. In terms of the three criteria often described as necessary conditions for establishing causation (covariation,

intentions is not depicted in Figure 1.3. The distinction between a total effect and a direct effect will be introduced in Chapter 4.

temporal ordering, and the elimination of competing explanations), this study establishes, at best, only covariation between variables in the causal system.

Experimentation and, to a lesser extent, longitudinal research offer some advantages over cross-sectional research when establishing causal association. For example, suppose economic stress was experimentally manipulated in some way, but otherwise the same results were found. In that case, we would be in a much better position to argue direction of cause, at least in part. Random assignment to levels of economic stress would ensure that neither social ties, depressed affect, nor withdrawal intentions could be affecting the stress the study participants felt. It also guarantees that economic stress and depressed affect are not spuriously associated, meaning they share a common cause. But random assignment would not help establish the correct temporal ordering of depressed affect and withdrawal intentions. Although it could be that economic stress influences depressed affect which, in turn, influences withdrawal intentions ($X \rightarrow M \rightarrow Y$), it remains possible that economic stress influences withdrawal intentions, which then influences depressed affect ($X \rightarrow Y \rightarrow M$).

To deal with this limitation of one-shot experimental studies, a sequence of experimental studies can help to some extent (see Stone-Romero & Raposa, 2010). First, one attempts to establish that $X$ causes $M$ and $Y$ in one experimental study. Success at doing so can then be followed with a second experimental study to establish that $M$ causes $Y$ rather than $Y$ causing $M$. The estimates from such analyses (perhaps including a moderation component as well) could then be pieced together to establish the nature (conditional or not) of the indirect effects of $X$ on $Y$ through $M$. But as Spencer, Zanna, and Fong (2005) note, it is not always easy or even possible to establish convincingly that the $M$ measured in the first study is the same as the $M$ that is manipulated in the second study. Absent such equivalence, the ability of a sequence of experiments to establish a causal chain of events is compromised.

Collecting data on the same variables over time is an alternative approach to studying causal processes, and doing so offers some advantages. For instance, rather than measuring entrepreneurs only once, it would be informative to measure their experience of economic stress on multiple occasions, as well as their depressed affect and intentions to withdraw from entrepreneurial activity. If economic stress influences withdrawal intentions through its effect on depressed affect, then you'd expect that people who are under more stress *than they were before* would express stronger intentions to withdraw *than they expressed earlier* as a result of feeling more depressed affect *than they were feeling earlier*. But covariation over time does not imply cause, just as covariation at a single time fails to establish a causal association. There are statistical procedures that attempt to disentangle contemporaneous from time-lagged association (e.g., Finkel, 1995), and there is a growing literature on moderation and mediation analysis, as well as their combination, in longitudinal studies (e.g., Bauer, Preacher, & Gil, 2006; Cole & Maxwell, 2003; Cheong, MacKinnon, & Khoo, 2003; Selig & Preacher, 2009). However, I do not address this literature or corresponding methods in this book.

One could advance the argument that scientists really should not attempt to model purportedly causal processes with data that do not afford causal interpretation. However, I could not make that argument convincingly because I don't believe this. We don't use statistical methods to make causal inferences. Establishing cause and effect is more a problem in research design than it is in data analysis. Statistical methods are just mathematical tools that allow us to discern order in apparent chaos, or signals of processes that may be at work amid random background noise or other processes we haven't incorporated into our models. The inferences that we make about cause are not products of the mathematics underneath the modeling process. Rather, the inferences we make are products of our minds—how we interpret the associations we have observed, the signal we believe we have extracted from the noise. To be sure, we can and should hold ourselves to a high standard. We should strive to design rigorous studies that allow us to make causal inferences with clarity when possible. But we won't always be able to do so given constraints on resources, time, the availability of data, the generosity of research participants, and research ethics. We should not let the limitations of our data collection efforts constrain the tools we bring to the task of trying to understand what our data might be telling us about the processes we are studying. But we absolutely should recognize the limitations of our data and couch our interpretations with the appropriate caveats and cautions.

Causality is the cinnamon bun of social science. It is a sticky concept, and establishing that a sequence of events is a causal one can be a messy undertaking. As you pick the concept apart, it unravels in what seems like an endless philosophical spiral of reductionism. Even if we can meet the criteria of causality when testing a simple $X \rightarrow M \rightarrow Y$ model, what is the mechanism that links $X$ and $M$, and $M$ to $Y$? Certainly, those causal processes must themselves come into being through some kind of mechanism. What are the mediators of the individual components of the causal chain? And what mediates the components of those components? And if those mediators can be established as such, what mediates those effects?

In other words, we have never really explained an association entirely, no matter how many intervening variables we propose and account for linking $X$ and $Y$. This does not mean that it is not worth thinking deeply about what cause means or discussing and debating what kinds of standards we must hold ourselves to as scientists in order to accept causal interpretations. But that isn't going to happen in this book. There are other books and journal articles on the topic of causality if you want to explore the concept on your own (e.g., Davis, 1985; Holland, 1986; Morgan & Winship, 2007; Pearl, 2009), and there is a growing chorus of quantitative social scientists who reject the regression-based orientation I outline here on the grounds that linear modeling and statistical adjustment simply don't do the job many people claim it does. That said, this book is about statistically modeling relationships—relationships that may but may not be causal in the end—and I think you will find the techniques and tools described here useful in your quest to understand your data and test some of your theoretical propositions and hypotheses. Just how large an inferential chasm between data and claim you attempt to leap is your decision to make, as is how you go about justifying your inference to potential critics. I will not, nor should I or anyone else, forbid you to use the methods described here just because your data are *only* correlational in nature.

## 1.5 Statistical Software

I believe that the widespread adoption of modern methods of analysis is greatly facilitated when these methods are described using software with which people are already familiar. Most likely, you already have access to the statistical software I will emphasize in this book, primarily SAS and IBM SPSS Statistics (the latter of which I refer to henceforth simply as SPSS). Although other software could be used (such as Mplus, LISREL, AMOS, or other structural equation modeling programs), most of these don't implement at least some of the procedures I emphasize in this book. And by eliminating the need to learn a new software language, I believe you more quickly develop an understanding and appreciation of the methods described herein.

Throughout the pages that follow I will emphasize estimation of model parameters using ordinary least squares (OLS) regression. Although any program that can conduct OLS regression analysis can estimate the parameters of most of the models I describe, such programs can only get you so far when taken off the shelf. For instance, no program I am aware of implements the Johnson–Neyman technique for probing interactions, and neither SPSS nor SAS can generate bootstrap confidence intervals for products of parameters, a method I advocate for inference in mediation analysis and conditional process analysis. Over the last several years, I have been publishing on moderation and mediation analysis and providing various tools for SPSS and SAS in the form of "macros" that simplify the analyses I describe in this book. These go by such names as INDIRECT (Preacher & Hayes, 2008a), MODMED (Preacher et al., 2007), SOBEL (Preacher & Hayes, 2004), MODPROBE (Hayes & Matthes, 2009), and MED3/C (Hayes, Preacher, & Myers, 2011). But each of these tools was designed for a specific task and not others, and keeping track of which tool should be used for which analysis can be difficult. So rather than confuse you by describing the ins-and-outs of each of these tools, I have designed a new macro for this book called PROCESS that integrates most of the functions of my earlier macros into one handy command or dialog box, and with additional features not available in my other macros. My prediction is that you will come to love PROCESS and will find yourself turning to it again and again in your professional life. The PROCESS procedure is freely available and can be downloaded from my home page at *www.afhayes.com*, and documentation describing its use and features can be found in Appendix A.

The advent the the graphic user interface (GUI) in the 1980s made data analysis a point-and-click enterprise for some and turned what is a distasteful task for many into something that is actually quite fun. Yet I still believe there is value to understanding how to instruct your preferred software package to perform using syntax or "code." In addition to providing a set of instructions that you can easily save for use later or give to collaborators and colleagues, syntax is easier to describe in books of this sort than is a set of instructions about what to click, drag, point, click, and so forth, and in what sequence. Users of SAS have no choice but to write in code, and although SPSS is highly popular in part because of its easy-to-navigate user interface, and I do provide a GUI-based version of PROCESS, I nevertheless will describe all SPSS instructions using syntax. In this book, all code for whatever program I am using or describing at that moment will be denoted with `courier` typeface in a shaded box, as below.



```
process vars=attitude exposure social intent/y=intent/x=exposure/m=attitude
    /w=social/model=8/wmodval=1.25/boot=5000/save=1.
```

Some commands will not fit in a single line in this book and must be carried below to the next line. When this occurs, it will be denoted by indentation of the continuing text, as above. A command has ended when you see a *command terminator*. In SPSS, the command terminator is a period ("."), whereas in SAS it is the semicolon (";"). A failure to include a command terminator at the end of your command is likely to confuse your software, and a string of errors is inevitable.

## 1.6 Overview of This Book

This book is divided into four broad parts. The first part, which you are reading now, consists of the introductory material in this chapter as well as an overview of the basic principles of linear models using OLS regression in Chapters 2 and 3. These chapters should be considered important prerequisite reading. If you are not familiar with the fundamentals of linear modeling, almost nothing else in this book will make any sense to you. So although the temptation to skip the material in this section may be strong, do so at your own risk.

Chapters 4 through 6 define the second part, which is devoted to mediation analysis. Chapter 4 illustrates the basic principles of elementary path analysis, with a focus on the partitioning of the total effect of antecedent variable $X$ on consequent variable $Y$ into direct and indirect effects, as well as means of making statistical inference about direct and indirect effects. Chapter 5 extends the principles and methods introduced in Chapter 4 into the realm of multiple mediator models—models of causal influence that are transmitted by two or more intervening variables operating in parallel or in sequence. Chapter 6 discusses miscellaneous issues in mediation analysis such as measures of effect size, confounding and causal order, and models with multiple causal antecedent or consequent variables.

The third part is Chapters 7 through 9, and the topic is moderation analysis. In Chapter 7 I define the concept of a conditional effect and show how to set up a linear model that allows the effect of one variable on another to depend linearly on a third variable. I illustrate how a hypothesis of moderation is tested and the parameter estimates of the corresponding model interpreted. I also introduce a few methods of dissecting the conditional nature of association and show how to construct a visual representation of moderation. Chapter 8 illustrates the generality of the procedure introduced in Chapter 7, including interaction between quantitative variables or between dichotomous moderators and focal predictors. Chapter 9 ends the section on moderation with discussions of miscellaneous issues in the estimation of models that allow one variable's effect to depend on another, such as models with multiple interactions, and a debunking of myths and misunderstandings about centering and standardization in moderation analysis.

Chapters 10 through 12 end the book with an introduction to conditional process analysis, the fourth and final part. Chapter 10 provides numerous examples of conditional process models proposed and estimated in the literature, introduces the important concepts of conditional and unconditional direct and indirect effects, describes how they are defined mathematically, and shows how they are estimated. Chapter 11 provides a slightly more complex analytical example of conditional process analysis while also illustrating the distinction between moderated mediation and mediated moderation. Chapter 12 addresses various miscellaneous issues and questions about the analysis of the contingencies of mechanisms.

This is an introductory book, and so there are many important, interesting, and some could say critical points and controversies that I gloss over or completely ignore. For example, the majority of the analyses I illustrate will be done using OLS regression-based path analysis, which assumes fixed effects, continuous outcomes, and the absence of random measurement error. Of course, we generally don't measure without error, and it is well known that a failure to account for random measurement error in the variables in a linear model can produce bias and misleading results. And often our outcomes of interest are not continuous. Rather, they may take one of two values or perhaps are measured on a course ordinal scale. In such cases OLS regression is not appropriate. I also neglect multilevel models, modeling change over time, or even the most basic of repeated measures designs. These are interesting and important topics, to be sure, and there is a developing literature in the application of mediation and moderation analysis, as well as their combination, to such problems. But assuming you don't plan on abandoning OLS regression any time soon as a result of some of its weaknesses and limitations, I believe you will be no worse for the wear and, I predict, even a bit better off once you turn the last page and have developed an understanding of how to use OLS regression to model complicated, contingent processes.



## 1.7 Chapter Summary

The outcome of an empirical study is more impressive, more influential, and more helpful to our understanding of an area of scientific inquiry if it establishes not only *whether* or *if* $X$ affects $Y$ but also *how* and *when* that relationship holds or is strong versus weak. If all effects exist through some kind of mechanism, and all effects have some kind of boundary conditions, then the most complete analysis answers both the how and when question simultaneously. In this chapter I have introduced the concepts of mediation (how $X$ influences $Y$) and moderation (when $X$ influences $Y$) and their combination in the form of a conditional process model. Although data analysis cannot be used to demonstrate or prove causal claims, it can be used to determine whether the data are consistent with a proposed causal process. Thus, the methods described in this book are useful for testing causal processes even absent data that lend themselves to unequivocal causal interpretation. My emphasis throughout this book is on the use of regression-based path analysis as a means of estimating various effects of interest (direct and indirect, conditional and unconditional). In order to grasp the material throughout this book, the basic principles of linear modeling using regression analysis must be well understood. Thus, the next two chapters provide on overview of the fundamentals of OLS regression.