original article

What is wrong with mediators and moderators?*

Robert A. Roe
Maastricht University

Mediator and moderator analyses are enjoying great and growing popularity among psychological researchers. However, their use as tools for causal analysis is alarming since this is exactly what these analyses are unsuited for. In this article, I posit that mediator and moderator models are based on a temporal illusion and that alternative arrangements of variables produce models that may explain relations between variables equally well. In fact, for each 3-variate model (whether mediator or moderator) five alternative models can be devised. This is empirically demonstrated with data from a large-scale study on employee attitudes and behaviors. The conclusion is that mediator and moderator analyses lead to inferences that are at best unfounded and at worst wrong, and that the only way to examine sequence and causal order is by means of temporal research. A possible remedy is two-dimensional modeling: it serves as a prophylactic against temporal illusions and as a tool that helps choosing proper methods of analysis.

The analysis of mediator and moderator effects has become one of the most popular analytical methods in psychology. A count based on journals covered by PsychInfo[1] shows that during the past two decades the numbers of articles analyzing mediation and/or moderation have increased exponentially. While the numbers of articles using mediation and moderation analysis were 537 and 209 in 1991, these numbers had grown to 3472 and 1979 by 2010. These figures are based on articles explicitly mentioning mediator and moderator variables and do not include studies using structural equation modeling that may also involve mediation or moderation.

The notion of "*mediation*" grew out of research on "intervening mechanisms" that began in the 1920s, when researchers became interested explaining relationships between independent and dependent variables from hidden, non-observable mechanisms of the human mind (e.g. "drive"; Hull, 1943). The terms "mediation" and "mediator variable" only emerged during the 1950s (see for instance: Cofer, 1958; Hilgard, 1958; Rozeboom, 1956). The abbreviation "mediator" was adopted in later years (e.g. Birnbaum & Mellers, 1979). It is worth noting that "intervening variables" were originally seen as theoretical and non-observable, whereas "mediator variables" were typically conceived as observable and measurable. The notion of "*moderation*" comes from the world of statistics, where it was used in connection with regression analysis. The term "moderator variable" was introduced by Saunders (1955, 1956). It referred to a third variable that modifies or moderates the regression of one variable on another one. The moderator variable

*Abbreviated text based on: R.A. Roe (2011). What is wrong with mediators and moderators? *15th European Congress of Work & Organizational Psychology*. Maastricht, Netherlands, May 24-28, 2011.

[1] Counts conducted by October 19, 2011.

was originally seen as defining "groups" of subjects for which different regressions would hold. Gender, age, race and socio-economic level are examples of moderators used in early research. Expressions like "moderating effect" and "moderated regression" became widely adopted in the 1960s and 1970s. The original meaning of a variable "moderating" the relationship between two other variables was gradually replaced by the more generic notion of interaction effect, in which two or more independent variables can be seen as moderating each others effect on the dependent variable.

The fact that mediation and moderation are nowadays seen as related seems a matter of historical coincidence. Three developments are worth mentioning in this context. First, researcher's growing focus on variables rather than constructs, which diminished the conceptual distinction between mediators and moderators. Second, innovations in multivariate regression analysis, permitting layers of dependent variables and inclusion of interaction terms, which allowed bringing mediation and moderation together in one statistical framework. Third, the advent of causal modeling, based on the idea that partial regression coefficients allow making causal inferences from data obtained at one moment in time (Blalock, 1960). Mediators and moderators first appeared together in an article by James and Brett (1984). They also feature together in the often-cited article by Baron & Kenny (1986), which appeared two years later[2].

In spite of their different roots, mediation and moderation methods are currently used for a similar purpose, i.e. establishing causal relationships between three or more variables. Researchers typically postulate models in which one or more antecedent variables are hypothesized to "exert an influence" on one or more consequent variables, with mediator and/or moderator variables determining how this influence is exerted. Researchers test these models by examining the covariation between variables across subjects. The aim of this article is to remind researchers of the weaknesses of this approach, and the logical impossibility of inferring causal relations from between-subject differences, regardless whether the variables are measured at one point in time or at multiple points in time. I will argue that mediator and moderator analysis are based on a temporal illusion that thwarts the possibility to make causal inferences in the proper way, i.e. by means of temporal research (Roe, 2008).

Below, I briefly discuss how temporal illusions affect research practice and particularly the way of drawing mediator and moderator models. Next, I point out that alternative models can be drawn, which may theoretically be equally acceptable. I give examples of reciprocal mediation and moderation effects, using selected results from a study on work motivation and quality of working life, conducted in a sample of 2660 workers from three European countries, published in 2000 (Roe, Zinovieva, Dienes, & Ten Horn, 2000). Acknowledging that not all models may fit the data equally well, I next discuss the case of non-reciprocity and the issue of model fit. I demonstrate that information extracted from such cases provides no evidence of causal order. Finally, I propose two-dimensional modeling as a means to prevent confusion of between-subjects and within-subjects analysis and indicate how it can help doing causal analysis in a proper way.

Temporal illusions

The term temporal illusion is used here to denote the belief among researchers that the

[2] The improved test of mediation proposed by these authors made this article into the most cited in the field, counting 16,362 citations by October 19, 2011.

flow of time is present when it is not. Examples of temporal illusions are numerous. It is, for instance, customary among researchers to theorize about events and processes, which by definition unfold over time, to gather and analyze cross-sectional data from which time is lacking, and interpret results in terms of events and processes (e.g. Wang & Takeuchi, 2007). This practice is typically accompanied by the ritualistic statement "that the cross-sectional findings should be confirmed by longitudinal research" (cf. Schaubroeck, Lam, & Cha, 2007). Likewise, researchers often interpret between-subject correlations between variables X and Y as showing "an influence" of X on Y. They do this either for variables measured at one moment or at different time moments, which is equally unjustified. In the same spirit, and focal in this article, researchers wrongly infer mediator and moderator effects from between-subject correlations of variables measured at the same point in time.

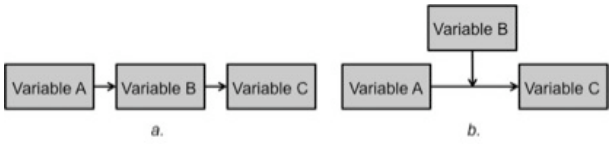

Figure 1. Basic mediator model (a) and moderator model (b)

Temporal illusions are unwanted and dangerous. They foster the misconception that differential (between-subject) and temporal (within-subject) research are two ways of testing the same theory that will—in the long run—produce the same results. Thereby they obstruct temporal research and hinder advance in psychological theory development. Moreover, they lead to inferences that are at best unfounded and at worst wrong, and they solicit interventions that are at best ineffective and at worst damaging.

Modeling mediation and moderation

Mediator and moderator analysis is based on models such as shown in Figure 1. In accordance with the direction of writing in Western cultures, these models are drawn from left to right. Mediators are positioned in the middle to suggest that they "*transmit the influence*" of the antecedent variable on the consequent variable. Moderators are inserted at some intermediate location and supposed to "*influence*" the relationship between adjacent variables.

It is important to note that, at least in cross-sectional research, this way of drawing mediator models is arbitrary and at the same time misleading. The models should rather be drawn like in Figure 2, which implies that there are multiple ways of defining the relationships, which might all account for given empirical evidence.

From the generic mediator model in Figure 2a six three-variable mediator models can be obtained. Figure 3 gives these six models and illustrates which parameters and fit would be obtained for each model for data on Responsibility, Meaningfulness and Performance taken from Roe, et al. (2000).

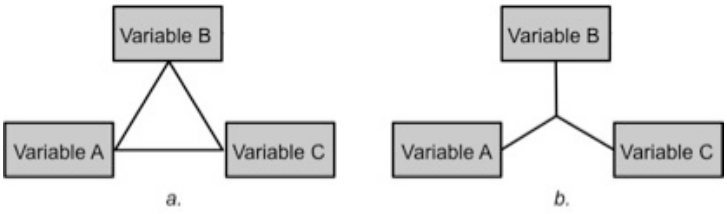

Figure 2. Alternative generic mediator model (a) and moderator model (b)

All six models show evidence of (partial) mediation effects of a magnitude similar to what is typically reported in the literature. This example demonstrates that variables can have *reciprocal* mediation effects: each variable mediates the relations between the other variables. All six models make sense from a

theoretical (and practical) point of view. Considering these models will make the reader realize that—given the cross-sectional nature of the data—*sequence is in the eye of the beholder*.

Figure 4 shows similar results for moderators. From the generic model in Figure 2b six moderator models are derived. Using data on Self-efficacy, Meaningfulness and Satisfaction from Roe et al. (2000) we, again, find that alternative models support equally acceptable interpretations, and that moderation effects can be reciprocal.

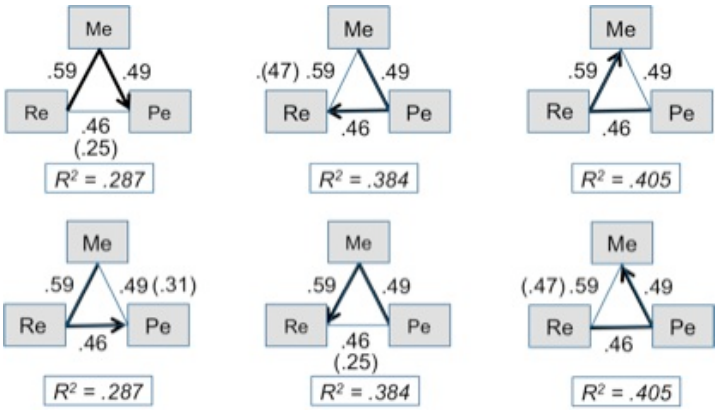

Figure 3. Six mediator models with correlations and partial correlations after partialing out the mediator (between brackets). Re = Responsibility, Pe = Performance, Me = Meaningfulness, R^2 = multiple correlation

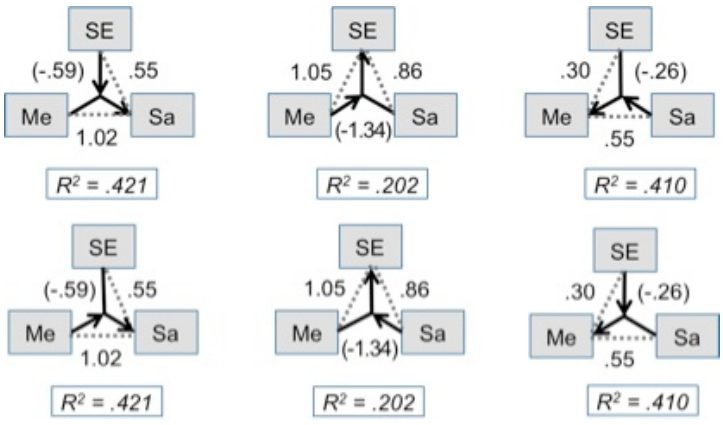

Figure 4. Six moderator models with regression coefficients for the predictor, moderator, and interaction term (between brackets). Me = Meaningfulness, Sa = Satisfaction, SE = Self-efficacy, R^2 = multiple correlation

Non-reciprocity: a best model?

Reciprocal mediation and moderation will not always occur. Conventional reasoning, based on the temporal illusion, might give rise to the idea that the model with the highest fit (% variance explained) shows the real sequence. Logically, there is no ground for the conclusion that the "influence of A on C" is "transmitted through B" or is "moderated by D" on the basis of model fit. The argument is clearly wrong, since all models merely reflect a single pattern of statistical associations between simultaneous measures. That is, the fit merely shows the proportion of *between-subject* variance explained and has no bearing whatsoever on within-subjects relationships. Any psychological interpretation in terms of processes, states or actions with a particular causal order is illusory.

This can perhaps best be illustrated with a practical example, namely that of the spatial dimensions of suitcase. Using a particular set of suitcases we might find that (due to different degrees of variation within this set) Depth and Width give a better prediction of Height than Width and Height predict Depth. The difference in % variance explained would obviously have *no temporal meaning whatsoever*.

The above examples are confined to the case of variables that are measured simultaneously, at a single moment in time. Some readers might believe that these problems vanish when a longitudinal rather than a cross-sectional design is used. However, this is not the case. Correlational analysis, even if it involves variables measured at different moments in time, is merely capturing between-subject differences. It fails to provide information on what happens within subjects, unless the processes referred to by the variables are "ergodic", that is, stationary (time-invariant) and homogeneous (identical for each subject). Thus, the problems of inference remain, unless researchers would be able to show

that the extraordinary conditions of ergodicity apply (Molenaar & Campbell, 2009).

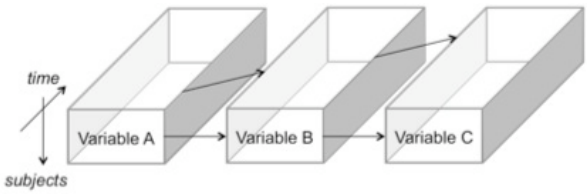

Figure 5: Two-dimensional model with arrows indicating covariation across subjects and across time (with lags).

A remedy: Two-dimensional modeling

A simple remedy to avoid the above problems is to change the way of drawing models by adding a time dimension. Figure 5 gives an example for the case of three variables.

This way of modeling makes a clear distinction between within-subjects and between-subjects covariation of variables. It suggests that, in order to investigate *mediation* effects, researchers should look for within-subject covariation between the antecedent variable, the mediator variable, and the consequent variable. It also makes researchers aware of the need to address sequence and time lags—something lacking from classical mediation analysis. If causal links are to be established, *there is a logical necessity for time lags to be larger than zero*, since temporal sequence is one of the necessary conditions for causality.

The way in which *moderation* has to be established is less obvious. If we go back to the earlier conception of moderator variable as a variable that defines subgroups of people showing differences in the regression between two other variables, a moderator variable would be one showing between-subject differences that are associated with between-subject differences in the within-subject covariation of an antecedent and consequent variable. In practical terms: the parameters of the regression over time of each subject would covary with their scores on the moderator variable.

Recommended analyses

Two-dimensional modeling helps in finding better methods of analysis than conventional mediator and moderator analysis. *Mediator analysis* requires an assessment of three variables[3] over time. The three time-series obtained would need to be regressed on each other in the proper sequence (mediator on antecedent, consequent on mediator), with certain time-lags. The resulting set of regression parameters (within-subject, one for each subject) could then be subject to a clustering procedure (between-subject) to identify groups of subjects with similar mediation. If mediation effects are assumed to be similar for all subjects, one could, alternatively, estimate a single set of regression parameters for the whole sample of subjects using multilevel growth modeling techniques (e.g. Singer & Willett, 2003).

Moderator analysis would be slightly different. With a moderator conceived as a stable individual-difference variable, such as ability or personality, one would do the same as above with two rather than three variables within each subject. Next, subjects would be grouped by their within-subject regression parameters, and groups would be compared on the moderator variable. In case of a moderator conceived as varying over time, one would again have to establish three time series for each individual, but now one would define an interaction-term of the antecedent and the moderator and use that as a single (time-lagged) predictor of the consequent variable. Again, multilevel growth modeling could provide the needed techniques.

[3] I confine myself to the simplest case with three variables.

Next to correlational designs, researchers may also use experimental designs with repeated measurements in which antecedents precede mediators and mediators precede consequents. This is compatible with the suggestions of Kraemer et al. (Kraemer, Kiernan, Essex, & Kupfer, 2008) for the study of mediation and moderation in clinical research. Referring to experimental designs with two or three measurement moments, they point out that antecedent or "target" variables (to be understood as discrete treatments) must precede mediators and that moderators must precede antecedent variables[4].

A brief note seems in place about cross-lagged panel analysis (Campbell & Stanley, 1963), which researchers also use for causal inferences. Although its logic seems compelling since it implies measurements at two or more moments in time (and thereby sequence), it must be noted that cross-lagged panel designs suffer from confusion of between-subject and within-subject variation as well. The magnitude of the correlation between an antecedent variable at time 1 and a consequent variable at time 2 carries no information on within-subject changes (except for the case of ergodicity). Logic does not permit making inferences about causal order along the within-subject dimension from the proportion of variance between subjects.

Conclusions

Mediator and moderator analysis as we know and use it today is based on a temporal illusion, and not suited to make causal inferences. Establishing mediator and moderator effects requires models and designs that separate lagged covariation over time within subjects from differential covariation between subjects. Such models and designs allow researchers to avoid temporal illusions and engage in research that gives a valid image of how behavior unfolds and which factors govern it.

References

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

Birnbaum, M. H., & Mellers, B. A. (1979). One-mediator model of exposure effects is still viable. *Journal of Personality and Social Psychology, 37*(6), 1090-1096. doi: 10.1037/0022-3514.37.6.1090

Blalock, H. M., Jr. (1960). Correlational analysis and causal inferences. *American Anthropologist, 62*, 624-631. doi: 10.1525/aa.1960.62.4.02a00060

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching*.

Cofer, C. N. (1958). The Mediation Hypothesis in the Analysis and Description of Behavior. In R. A. Patton (Ed.), *Current trends in the description and analysis of behavior.* (pp. 120-141). Pittsburgh, PA US: University of Pittsburgh Press.

Hilgard, E. R. (1958). Intervening variables, hypothetical constructs, parameters, and constants. *The American Journal of Psychology, 71*, 238-246. doi: 10.2307/1419211

Hull, C. L. (1943). *Principles of behavior: an introduction to behavior theory*. Oxford England: Appleton-Century.

James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology, 69*(2), 307-321. doi: 10.1037/0021-9010.69.2.307

Kraemer, H. C., Kiernan, M., Essex, M., & Kupfer, D. J. (2008). How and why criteria defining moderators and mediators differ between the Baron & Kenny and MacArthur approaches. *Health Psychology, 27*(2, Suppl), S101-S108. doi: 10.1037/0278-6133.27.2(Suppl.).S101

Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science, 18*(2), 112-117. doi: 10.1111/j.1467-8721.2009.01619.x

[4] This condition is easily fulfilled with stable moderators, like e.g. gender. It is equivocal when moderators can change over time.

Roe, R. A. (2008). Time in applied psychology: Studying what happens rather than what is. *The European Psychologist, 13*, 37-52.

Roe, R. A., Zinovieva, I. L., Dienes, E., & Ten Horn, L. A. (2000). A comparison of work motivation in Bulgaria, Hungary, and the Netherlands: Test of a model. *Applied Psychology: An International Review, 49*(4), 658-687.

Rozeboom, W. W. (1956). Mediation variables in scientific theory. *Psychological Review, 63*(4), 249-264. doi: 10.1037/h0043718

Saunders, D. R. (1955). The 'moderator variable' as a useful tool in prediction. *Proceedings of the Conference on Testing Problems, Educational Testing Service*, 54-58.

Saunders, D. R. (1956). Moderator variables in prediction. *Educational and Psychological Measurement, 16*, 209-222. doi: 10.1177/001316445601600205

Schaubroeck, J., Lam, S. S. K., & Cha, S. E. (2007). Embracing transformational leadership: Team values and the impact of leader behavior on team performance. *Journal of Applied Psychology, 92*(4), 1020-1030. doi: 10.1037/0021-9010.92.4.1020

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis. Modeling change and event occurence.* Oxford: Oxford University Press.

Wang, M., & Takeuchi, R. (2007). The role of goal orientation during expatriation: A cross-sectional and longitudinal investigation. *Journal of Applied Psychology, 92*(5), 1437-1445. doi: 10.1037/0021-9010.92.5.1437

Robert A. Roe

is emeritus-professor of Organisational Theory and Organisational Behaviour at Maastricht University, the Netherlands. He studied psychology and obtained his doctorate at the University of Amsterdam. He was professor of Work & Organisational Psychology in Delft, Tilburg and Nijmegen, director of the Work & Organization Research Center in Tilburg, director of the Netherlands Aeromedical Institute, and founding president of the European Association of Work & Organisational Psychology. He is currently president of the European Federation of Psychologists' Associations. His publications cover a broad range of topics in work & organizational psychology and in research methodology.

R.Roe@maastrichtuniversity.nl