Does mentalization and impulsivity affect suicidal behavior in patients with borderline personality disorder
A structural equation modeling approach

Johanna Andersson & Rasmus Berggren

Handledare: Viktor Kaldo, Institutionen för klinisk neurovetenskap, Göran Rydén, Norra Stockholms Psykiatri, S:t Görans sjukhus
Examinator: Professor Petter Gustavsson, Institutionen för klinisk neurovetenskap
Does mentalizat
tion and impulsivity affect suicidal behavior in
patients with borderline personality disorder
A structural equation modeling approach

Sammanfattning

Nyckelord: impulsivitet, mentaliseringsförmåga, självmord, strukturekvationsmodellering

Abstract
A diagnosis of Borderline personality disorder (BPD) is a risk factor for suicidal behavior. Previous research has indicated a connection between BPD and deficits in mentalization. The present study’s hypothesis is that deficient mentalizing ability and impulsivity increases the risk for suicide among patients with BPD. This hypothesis is tested using a structural modeling equation approach. The sample consist of 40 patients diagnosed with BPD and admitted to S:t Görans hospital in Stockholm during 2003-2008. The model displays an acceptable fit and that mentalization and impulsivity can account for 47% of the variance in suicidality. However, due to factors such as a small sample size and poor operationalization of the constructs, the results cannot be considered to conclusively verify or falsify the hypothesis. There is a more general discussion at the end about the interpretation of latent variables, structural equation modeling and causality and the relation between intra- vs. inter-individual processes.

Key words: impulsivity, mentalization, structural equation modeling, suicide
Does mentalization and impulsivity affect suicidal behavior in patients with borderline personality disorder
A structural equation modeling approach

Johanna Andersson & Rasmus Berggren

Introduction

Suicide

In Sweden in 2008, 1467 people died in suicide. Between 1980 and 2008 the suicide rate decreased; nevertheless it is still a major problem that causes much suffering for families and society at large. The decrease in suicide rate is seen in older ages, aged 25 and older; however, for young people, aged 15-24, not much has changed in the incidence of suicide and in 2008 it was the leading cause of death for men in this age group. For men aged 15-44 it was the second most common cause of death and for women of this age group it was the first (Karolinska Institutets folkhälsoakademi, 2010). These statistics are for completed suicides and the number of suicide attempts is little known. A suicide attempt is not only a risk factor for suicide (Niméus, Alsén, & Träskman-Bendz, 2000); it is also a sign of suffering.

A diagnosis of Borderline personality disorder (BPD) is a distinct risk factor for suicidal behavior (Brodsky, Malone, Ellis, Dulit, & Mann, 1997; Soloff, Lis, Kelly, Cornelius & Ulrich, 1994). Within this group of patients several factors are associated with suicidal behavior, such as comorbid major depressive disorder and substance abuse as well as impulsivity and feelings of hopelessness (Black, Blum, Pfohl, & Hale, 2004).

Borderline Personality Disorder

Borderline personality disorder (BPD) is a common mental disorder that is characterized by high rates of suicide, severe functional impairments and high rates of comorbidity with other mental disorders (Leichsenring, Leibing, Kruse, New, & Leweke, 2011). Depending on the investigator, method and length of follow up, the incidence of suicide in BPD patients ranges up to 10 % (Black et al., 2004). There are two clinical meanings to the term borderline, either as an ‘official’ diagnosis or as a broad category of patients that has been described in psychoanalytical literature (Kernberg & Michels, 2009). As a category, the term describes the group of patients that are in the middle, or on the border, between psychosis and neurosis (Butcher, Mineka, & Hooley, 2010; Kernberg & Michels, 2009). This group is defined by underlying psychological structures, whereas the diagnosis of BPD in the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association, 2002) is defined by surface phenomenology (Kernberg & Michels, 2009). Accordingly, the DSM-IV describes the essential features of BPD as a pervasive pattern of instability of interpersonal relationships, self-image and affect together with notable impulsivity that is present in a range of contexts which is indicated by at least five on a list of nine criteria. According to Leichsenring et al. (2011), suicidal tendency and self-injury (criterion 5) are the most useful indicators of a correct diagnosis. BPD is associated with high levels of
affective instability, visible through intense emotional responses to environmental triggers such as real or perceived abandonment. Another important feature is an extremely unstable sense of self or self-image. Together, these two aspects often lead to highly unstable interpersonal relationships. The marked impulsivity in the disorder combined with an affective instability often leads to self-destructive behaviors such as gambling (Butcher et al., 2010), eating disorders and substance abuse (Leichsenring, et al., 2011) as well as suicide attempts and self-mutilation (Butcher et al., 2010).

BPD continues to be a rather controversial diagnosis (Fonagy, Luyten, & Strathearn, 2011). Due to the minimum of five out of the nine DSM-IV criteria to reach a diagnosis there are 151 theoretical possibilities in diagnosing the disorder (Leichsenring et al., 2011). This lack of homogeneity within the diagnosis, as well as a large comorbidity within and between axes, arbitrary cutoffs and poor test-retest reliability, makes this group difficult to define (Fonagy & Luyten, 2009; Fonagy et al., 2011). Nonetheless, there is growing consensus that there are three core features of BPD. These are dysregulation, impulsivity, and unstable relationships (Fonagy & Luyten, 2009; Sharp et al., 2011). Fonagy and Luyten (2009) theorize that these core features reflect impairments in the ability to mentalize.

Mentalizing

The term mentalization has been used in psychoanalytic literature since the 1960s. It has not had a clear definition and several other terms have been used interchangeably with it (Choi-Kain & Gunderson, 2008). Mentalization is a sort of social cognition. It is an imaginative mental activity through which one is able to perceive and interpret human behavior in terms of mental states (Fonagy & Luyten, 2009). Thus, mentalization is the ability to represent the behavior of one self and that of others in terms of underlying personal desires, needs, feelings, beliefs or reasons (Fischer-Kern et al., 2010; Fonagy & Bateman, 2005). Ultimately, mentalization is a way of understanding one self and others.

In the 1980s and 1990s, the term ‘mentalization’ was used synonymously with the concept of ‘theory of mind’ (Sharp et al., 2011). Mentalization is now considered to be a conceptual derivative of theory of mind that focuses more on internal and interpersonal aspects than theory of mind does (Choi-Kain & Gunderson, 2008; Fischer-Kern et al., 2010). Theory of mind refers to the ability of an individual to realize that there are mental states in others and also to be able to accurately identify what they are. While theory of mind is more oriented towards the other, mentalization adds an internal and more affectively rich dimension to the concept (Choi-Kain & Gunderson, 2008) and an interpersonal understanding of the self, as well as the other (Fischer-Kern et al., 2010).

Fonagy and Bateman include three dimensions in mentalization: 1) two modes of functioning, i.e. implicit and explicit; 2) two objects, i.e. the self and the other; and 3) two aspects of both the content and the process i.e. cognitive and affective (Choi-Kain & Gunderson, 2008). Mentalization can be both implicit, unconscious and automatic, and explicit, conscious and deliberate; and individuals can alternate between the two modes or use them both simultaneously. The two objects, the self and the other, each have his or her own set of mental states. One hypothesis is that the two objects mentalize interactively and that the process of determining one’s own mental state perhaps determines whether one is capable of determining the other’s. The third dimension of the concept deals with the content of the mental state. These mental states can be cognitively or affectively laden to varying extent (Choi-Kain & Gunderson, 2008). The cognitive and affective aspects can be applied to the process
of mentalizing; i.e. optimally mentalization relies on cognition as well as “insight with emotion” (Choi-Kain & Gunderson, 2008, p. 1128). Therefore one can see a connection between the ability to mentalize and intelligence as well as an emotional awareness. When having reached the developmental achievement of mentalization, the individual can presumably effectively navigate in the social world while having a stable sense of self (Fischer-Kern et al., 2010).

This essay assumes a mentalization-based theory of the development of BPD. Mentalization is a developmental accomplishment (Ficher-Kern et al., 2010) and there seems to be several factors that may lead to poor mentalization skills. Fonagy et al. (2011) suggest that both genetic and early environmental factors can negatively affect the development of mentalization. These authors assume that there is an increased risk for BPD when an individual has deficits in mentalization. These deficits in mentalization can result from growing up in a family with low levels of mentalization and too little attention to internal mental states; or as a result of abuse or neglect which has lead the child to inhibit mentalization. Being subjected to abuse or neglect would result in the child’s inhibition of mentalization skills, as it would be too painful to infer that the adult has malicious intent in relation to the child (Fonagy et al., 2011).

Impulsivity

Impulsivity is one of the nine criteria in diagnosing BPD according to DSM-IV. Brodsky et al. (1997) found in their study that the criterion of impulsivity, rather than severity of the BPD pathology as a whole, was associated with suicidal behavior. They also found that impulsivity was associated with more previous suicide attempts, even after controlling for lifetime prevalence of major depression and substance abuse (Brodsky et al., 1997).

Another concept that is closely related to impulsivity is impulsive aggression, a dispositional trait consisting of impulsiveness, irritability as well as a proneness to act aggressively (Fossati et al., 2009). Fossati and colleagues (2009) showed in a study that difficulty in identifying feelings, an element of mentalization, was a central feature of impulsive aggression. Consequently, poor mentalization ability can be associated with a tendency to act in an impulsive aggressive manner.

Purpose

As BPD is associated with an increased risk of suicide, the purpose of this study is to assess the connection between mentalization ($M$), impulsivity ($I$) and suicidality ($S$) among patients diagnosed with BPD. The first hypothesis is that there is a significant direct effect of $M$ on $S$ (see Figure 1).

![Figure 1. Hypothesized model of the association between mentalization and suicidality (Model 1).](image)

Our second hypothesis is that the link between $M$ and $S$ is mediated through another variable $I$. To demonstrate the mediation, an association must first exist between $M$ and $S$. Further, the effect of $M$ on $S$ must be reduced in model 2 compared to model 1. Consequently, the path from $M$ to $S$ is hypothesized to be non-significant...
after having controlled for $I$ (see Figure 2). Thus the model assumes that the effect of mentalization on suicidality is fully mediated through impulsivity.

Figure 2. Hypothesized mediational model of the association between mentalization, impulsivity and suicidality (Model 2).

Method

Participants
A database containing information about 124 patients, on a total of 660 different measures, was used. The patients were all diagnosed with borderline personality disorder and admitted to Karolinska Sjukhuset in Huddinge between 2003 and 2008. The database contains scarce information about age at time of admission; there is only complete information about year of birth. In order to assess various sample characteristics we assume that they were all studied in 2008. Due to the amount of missing data, case-wise deletion was used on any case that did not contain full information on all of our 9 measured variables. The final sample consisted of 36 women and 4 men ($M_{\text{Age}} = 28.95, SD = 6.64$).

Mentalization in the model
In the database the only indicator that could be used for mentalization was the Twenty-item Alexitymia Scale (TAS-20, Bagby, Taylor, & Parker, 1994). Alexitymia is an impairment in describing feelings (Subic-Wrana, 2011) and TAS-20 is a self-report measurement scale of the Alexithymia construct (Bagby et al., 1994). It correlates negatively with psychological mindedness and the scale shows alexithymia to be related to deficiencies in emotional awareness and imagination (Bagby et al., 1994). Alexithymia can be seen as low emotional awareness (Subic-Wrana, 2011), which in turn can be interpreted as poor mentalization. Furthermore, at least two of the three domains of TAS-20 correspond closely to aspects of mentalized affectivity (Fossati et al., 2009). As indicators of mentalizing ability ($M$) we used Recognize Feelings ($X1$), Describe Feelings ($X2$) and Externalize ($X3$). A higher score on TAS-20 indicates poorer mentalizing ability.

Impulsivity in the model
As a measurement of impulsivity in the model the Zanarini Rating Scale for Borderline Personality Disorder was used (ZAN-BPD). It is a clinician-administrated assessment scale to measure the borderline psychopathology. The scale has four sector scores that are meant to reflect the core areas of the borderline psychopathology (Zanarini, 2003): Affect ($X6$ in the model), Cognition ($X7$ in the
model), *Impulsivity* (*X*8 in the model) and *Relation* (*X*9 in the model). The sector scores were used as indicators of a latent construct *Impulsivity* (*I* in the model).

**Suicide in the model**

There are many different meanings connected with suicide in the literature. A generally accepted nomenclature for suicide and suicide-related behaviors does not exist (O’Carroll et al., 1996; Silverman, 2006). Silverman (2006) has merged two lists of frequently used definitions of suicide and found no less than 15 commonly referenced definitions of suicide. Since a single universal definition seems hard to establish for the overall category, it is little wonder that defining suicidal thoughts, feelings and behaviors is so difficult. Niméus et al. (2000) claim that the Suicide Assessment Scale (SUAS) has good predictive ability in respect to future suicide. Thus a high score on the SUAS (> 39) indicates an elevated risk of suicide within the next 12 months (Niméus et al., 2000). As indicators of the latent construct *Suicidality* (*S* in the model) we used the *Total Sum at time of admission* (*X*4) as well as a subscale tapping specifically *Suicide Ideation at time of admission* (*X*5).

**Statistical analyses**

Structural equation modeling (SEM) was used to test the hypothesized models. The analyses were carried out using IBM SPSS Amos 20.0. As fit indices $\chi^2$, CFI, GFI and RMSEA were used. As a general rule of thumb, the $\chi^2$ should be nonsignificant, both CFI and GFI should be higher than .90 and RMSEA should be lower than .10 (Kline, 2011).

**Results**

Testing our first hypothesis seems to indicate that the effect of $M$ on $S$ is significant and moderate ($\beta_1 = .45, p < .05$); the overall model shows a poor fit to data ($\chi^2(4) = 12,772, p = .012; \text{CFI} = .88, \text{GFI} = .89, \text{RMSEA} = .24$), see Figure 3.

**Model 1**

![Observed model of the association between mentalization and suicidality.](image)
Model 2, the mediational model, was tested and the results are shown in figure 4. The effect of M on I is significant and moderate ($\beta_2 = .66, p < .05$); the effect of I on S is moderate, albeit non-significant ($\beta_3 = .66, p = .13$); and the effect of M on S is greatly reduced ($\beta_1 = .03, p = .93$). Model 2 shows a better fit to data ($\chi^2(24) = 32.594, p = .11; \text{CFI} = .91; \text{GFI} = .85; \text{RMSEA} = .10$). Together, M and I explains 47% of the variance in S. The indirect effect of M on S via I is $\beta_2 \times \beta_3 = .66 \times .66$) and 93% ($(\beta_1 \text{in Model 1} - \beta_1 \text{in Model 2})/\beta_1 \text{in Model 1} = .45 - .03)/.45$) of the effect that M has on S is mediated through I.

Discussion

The purpose of this study was to assess the connection between mentalization, impulsivity and suicidality among patients diagnosed with BPD. The first hypothesis is that poor mentalization increases the risk of suicide (Model 1). Testing this hypothesis indicates that there is such an effect, i.e. our model (Model 1) shows that poor mentalization increases suicide risk. However, the $\chi^2$ is significant which indicates a bad model fit. Since $\chi^2$ is sensitive to sample size, it is remarkable to find a significant $\chi^2$ value with such a small sample size ($N = 40$). Thus, a significant $\chi^2$
value, coupled with a small sample size, suggests an exceptionally bad fit (Garson, 2011; Norman & Streiner, 2003).

Our second hypothesis is that the link between mentalization and suicidality is mediated through impulsivity (Model 2). If impulsivity has a mediating effect, the direct effect mentalization has on suicidality, which was seen in Model 1, should become greatly reduced in Model 2. Testing this hypothesis confirms this to some extent. This model shows an acceptable fit; a non-significant $\chi^2$ value as well as other fit indices within or close to their appropriate range (CFI = .91; GFI = .85, RMSEA = .10). The effect of mentalization on suicidality ($\beta_1$) is greatly diminished and non-significant in model 2 ($\beta_1 = .03, p = .93$). There is a significant direct effect of mentalization on impulsivity ($\beta_2$), as well as a non-significant effect of impulsivity on suicidality ($\beta_3$). Both effects are strong ($\beta_2 = .66, p < .05; \beta_3 = .66, p = .13$) and most likely both would be significant had a larger sample been used.

Our study shows that the effect of mentalization on suicidality is almost completely mediated through deficits in impulsivity. Therefore, one might suggest that mentalization-based treatments for patients with BPD that aim to reduce or prevent suicide should focus on increasing the patients’ impulse control directly, rather than focus on increasing their mentalizing ability in itself. However, suicidality is not the only symptom of BPD and this study gives no indication at all of how mentalization affects the other features of BPD.

**Conclusion**

Model 2 fits the data better than Model 1, i.e., the effect of mentalization on suicidality is better explained as mediated through impulsivity rather than mentalization exerting a direct effect on suicidality. Due to several limitations, these results should be interpreted with caution.

**Limitations**

The present study only concerns patients with BPD and thus it does not give the possibility of evaluating the relationship between mentalization, impulsivity and suicidality in the general population. It would be interesting the compare a group of patients with BPD and a group of non-BPD patient controls to see if deficient mentalization ability mediates the effect of BPD on suicidality.

It is possible to claim that poor mentalization skills even without a diagnosis of BPD, through its connection with impulsive aggression, is an indicator for suicide risk. Other factors to take into account that were not included in the model are comorbid affective disorder and substance abuse. These are factors that are known to increase suicide risk (Black et al., 2004) but were omitted as their inclusion would have made the model too complicated for the amount of data available.

The model does not have a clear, theoretically well-defined indicator for mentalization. An alexithymia scale was used as a stand-in since there was no better measurement for mentalization in the database. Theoretically, there seems to be at least some overlap between mentalization and alexithymia (Fossati et al., 2009), so this methodological problem should not be unacceptable. However, it does affect the analysis negatively.

The measurement for suicide is not ideal either. The association in the model between mentalization and suicide is based on the SUAS-measurement and its association with suicide. Since a high result on the SUAS-scale is associated with an increased risk for suicide (Nimëus et al., 2000) it is likely that the association in the model between SUAS and mentalization indicates an increased risk of suicide. Other
ways of predicting suicide is intent, lethality of attempts and number of previous suicide attempts and perhaps any of these measurements could have been used had they been in the database.

Ultimately, the relationship between mentalization, impulsivity and suicidality can be studied only in so far as there are well-defined, valid measures of these constructs. Simply labeling one variable “Mentalization Ability” does not by itself make this a measure of mentalizing ability, and considering the factor structure it is doubtful whether we actually have achieved valid measures thereof. What the variables signify should be interpreted with caution, so as not to commit the naming or reification fallacies discussed in Kline (2011).

Another caveat is that our sample is far too small to support a SEM analysis. Jackson (2003) recommends the ratio between number of subjects and number of model parameters to be 20:1; anything below 10:1 reduces the trustworthiness of the results. Following Jackson’s recommendations, we could use 2 parameter estimates.

All of this combined suggests that the present study does not allow one to draw any far-reaching conclusions. However, the results should not be taken as being completely trivial. Even though it might be questionable to what extent the latent variable M actually corresponds to mentalizing ability, it certainly corresponds to something – if nothing else, than at least to a weighted sub-score on a questionnaire. Even if one does not accept these constructs as corresponding to mentalization, impulsivity and suicidality, we have at least tested a model of the interrelation between scores on these questionnaires. Thus, despite the validity of the constructs, one could potentially use questionnaires as screening tools and leave the interpretation aside.

The discussion in Black et al. (2004) on whether suicide can be predicted in individual patients is important to mention here as well. Most likely it is not possible to predict which patient, on the level of the individual, will ultimately commit suicide. This is especially important to consider when dealing with rare phenomena such as suicide, due to low base rate affecting the accuracy of predictions. However, the intent with the model was to establish whether deficits in mentalization ability would increase the risk for suicide. If so, it could then be possible to use this knowledge when working with this group of patients.

Apart from several limitations pertaining to the current study in particular, there are some more general methodological considerations that further limits the conclusions one can draw from this, and other, studies. Below, there is a discussion about the interpretation of latent variables, structural equation modeling related to causality, correlation and causality, and finally, the relation between inter- and intra-individual processes. The discussion that follows is largely based on Borsboom (2005) and Molenaar and Campbell (2009).

**Interpretation of latent variables**

Psychology has a long-standing empiricist heritage. A distinctive mark of empiricism is a reluctance to go beyond observation or incorporate unobservable entities to account for observable phenomena. In light of this, the widespread use of unobservable entities in psychology, i.e., the liberal use of psychological constructs is rather surprising.

There are two different ways to view the relation between a latent variable and its indicators. The difference between these two positions is mainly a matter of one’s intuition and how much metaphysics one is willing to incorporate, as well as in how one conceptualizes the relationship between the latent variable and its indicators. It
also has some bearing on the causal role of these constructs. Borrowing the terminology from Borsboom (2005) it is possible to distinguish between a formative measurement view and a reflective measurement view.

From the formative measurement view (FMV), also called the operationalist interpretation, a latent variable is nothing over and above its indicators. In a formative model, the latent variable is regressed on its indicators; the arrows go from the indicators to the latent variable. The relationship between the latent variable and its indicators is purely formal, or logical – speaking of latent variables is a convenient way of summarizing already present information, but does not signify anything apart from the measured variables. A prime example of a formative latent variable is that of socioeconomic status (SES) in sociology. Indicators of SES are usually things like years of education, annual income, living arrangement and neighborhood. The intuitive interpretation is that an increase in one of the indicators precedes an increase in the latent variable – an increase in income causes higher SES, but not the other way around. FMV is much more common in sociology and econometrics than it is in psychology. Most psychologists do not conceive of psychological constructs in the formative fashion.

The most common interpretation of latent variables in psychology is the reflective measurement view (RMV), which can also be called the realist interpretation of latent variables. In the RMV, the indicators are regressed on the latent variable; the arrows go from the latent variable to its indicators. Prime examples of latent variables being interpreted in the reflective manner are most well-known psychological constructs, including general intelligence (g) and the Big Five Personality Traits (Conscientiousness, Agreeableness, Neuroticism, Openness and Extraversion). In RMV, a change in the latent variable precedes a change in the manifest variable – in our case, an increase in mentalizing ability precedes (and plausibly cause) an increase in one (or more) of its indicators. In the philosophy of science, FMV corresponds to operationalism whereas the RMV is closer to realism.

In light of psychology’s empiricist tradition, it is noteworthy that psychologists unanimously conceive of latent variables in the reflective fashion rather than the formative fashion. The latter is far more parsimonious because it does not introduce latent variables as being distinct from their indicators, but rather exhaustively defined in terms of them. Occam’s razor taught us that we should choose the least complex interpretation, i.e., one that does not posit unobservable entities beyond necessity. Is there anything that warrants the metaphysically more ambitious realist interpretation of latent variables over the more parsimonious operationalist interpretation?

The most obvious reason to favor RMV over FMV is by appealing to causality. In RMV, the relation between the latent variable and its indicators is a causal relation, not a logical relation. Given that the introduction of latent variables (conceptualized in the realist sense) carries some metaphysical weight, such an invocation would be defensible if it also served some other scientific purpose – such as facilitating scientific explanation. Many psychological traits are supposed to exert a causal influence on behavior at the level of the individual. In the case of general intelligence, we assume that a person’s intelligence is causally related to, and to some extent also explains, his or her responses on an IQ test. This requires a realist interpretation, and we would be justified in assuming the less parsimonious interpretation if the latent variables indeed could be used to explain behavior at the level of the individual. However, as we shall see in a few sections, it is not immediately clear that the latent variables invoked serve our intended purposes and in many cases the realist interpretation may not be warranted.
In this section, the discussion is closely tied to SEM and psychometrics, but can be extended to other areas of psychology without any major modification. To illustrate with an example from a clinical setting: are psychiatric diagnoses merely shorthand ways of describing a collection of symptoms (i.e., operationalism) or are the diagnoses themselves causally related to the symptoms (i.e., realism)? Does patient X have poor appetite, sleep disturbances and a low mood because he or she is depressed (depression here being logically distinct from, and causally related to the symptoms)? Or is saying “patient X is suffering from depression” merely an easier way of saying “patient X suffers from poor appetite, sleep disturbance and low mood”? We may think a diagnosis can explain why the patient suffers from certain symptoms, but in order for this explanation to be viable the explanans (the thing doing the explaining, e.g. depression) has to be distinct from the explanandum (the thing to be explained, e.g. the symptoms). This, in turn, requires a realist interpretation of the latent construct similar to the outlined reflective approach above. However, the way we come to know about the cause is still through its effects so even though they have to be logically distinct, they must also necessarily be epistemologically linked (McDonald, 1999).

Structural equation modeling and causality

In the early days of SEM the technique was commonly referred to as causal modeling. Nowadays, this description is sometimes ridiculed and a common warning to students involved in SEM is that SEM is not a magical tool used to prove causality. The claim is essentially that no statistical model, no matter how mathematically sophisticated, can be used to prove causal relations because causality is not a statistical but an experimental concept. Allegedly, the proper way to establish causal relationships is to conduct randomized controlled trials, and one can never compensate for poor experimental design by using advanced statistical methods to prove causality. The term causal modeling originates from Wright’s (1918) early development of path analysis in order to estimate the magnitudes of effects where the causal relationships were already known. This in an important difference compared to how SEM and path analysis is used in behavioral research today, where the causal mechanisms are rarely known beforehand but instead hypothesized and tested. The proper used of SEM in behavioral research is as a disconfirmatory technique used to reject false models, rather than a technique for proving true models (Kline, 2011).

The claim that “SEM cannot be used to prove causality” should not be treated as a remark about SEM, or causality, but about the term “prove”. Popper (1959) highlighted the difficulty in establishing scientific proofs and the logic behind scientific discoveries, which we’ll try to explain below. In essence, Popper’s point is that no amount of confirming evidence can ever be used to prove a scientific theory, but one single piece of disconfirming evidence is enough to disprove a theory. Analogously in SEM, we have a hypothesized model A₁, the data structure X implied by the model, and our observed data structure Y. Further, we assume that A₁ implies X. What we’re testing is how well Y corresponds to X, i.e., how well the observed data structure corresponds to the data implied by the model. Hopefully, we’ll see that Y and X are quite similar, and so we take this as confirming our hypothesized model A₁. Given that Y is sufficiently similar to X, are we then warranted in concluding that A₁ is the correct model? Unfortunately no, because there are an infinite amount of different models A₂, A₃... that may also generate data X. So even if we find that A₁ is consistent with X this is not sufficient evidence to conclude that A₁ is the correct model.
However, if $A_1$ would be inconsistent with $X$, this is sufficient evidence to conclude that $A_1$ is an incorrect model. Bollen (1989) puts it this way: “If a model is consistent with reality, then the data should be consistent with the model. But, if the data are consistent with the model, this does not imply that the model corresponds to reality.” Therefore, all SEM analyses concerning causal relations should indeed be interpreted with caution (and not with causation!) but not because of any inherent flaw with SEM or causality, but because of the fundamental logic underlying scientific inquiry.

**Within- vs. between-subjects variables**

One of the main arguments for interpreting latent variables in the realist sense is that that would allow us to use them in explaining individual behaviors; specifically, causal explanations. For example, we do think that person A’s intelligence is causally related to his item responses on an IQ-test. This requires that we interpret latent variables in the realist sense.

All conceptions of causality require covariation of some kind (Mill, 1843; Pearl, 2000; Rubin, 1974). Variation can come in one of two ways – either by pooling across subjects, or by repeated sampling over a single individual. The former method is by far the most frequently used method in psychology, and it is also how our data were collected. We may also distinguish between *between-subject* variables and *within-subject* variables. Between-subject variables are measures that vary *between* individuals, like personality traits, sex and general intelligence and it’s what we obtain by sampling across individuals. Within-subject variables are measures that vary within individuals over time, like mood, sleepiness and motivation.

However, if we are to use the latent variables to do causal work at the level of the individual we run into the fact that many psychological variables of interest are only variables in a between-subject sense – not in the necessary within-subject sense. On the within-subject level, traits like general intelligence ($g$) and personality traits aren’t really variables at all, but constants. And a constant cannot cause anything, neither behaviors nor anything else, because a constant, by definition, does not exhibit variation. The main reason for interpreting latent variables in the realist sense, namely that they would facilitate causal explanations at the level of the individual, now seems untenable because the latent variables we invoked aren’t really the ones we require. In order to use them in causal explanations, they have to be right kind of variables – namely within-subject variables.

**References**


