MULTIVARIATE BEHAVIORAL RESEARCH
2025, VOL. 60, NO. 6, 1116–1150

Equilibrium Causal Models: Connecting Dynamical Systems Modeling and Cross-Sectional Data Analysis

Oisín Ryan[a#] and Fabian Dablander[b#]

[a]Department of Data Science and Biostatistics, University Medical Center Utrecht, Utrecht, The Netherlands; [b]SEVEN, University of Amsterdam, Amsterdam, The Netherlands

ABSTRACT

Many psychological phenomena can be understood as arising from systems of causally connected components that evolve over time within an individual. In current empirical practice, researchers frequently study these systems by fitting statistical models to data collected at a single moment in time, that is, cross-sectional data. This raises a central question: Can cross-sectional data analysis ever yield causal insights into systems that evolve over time—and if so, under what conditions? In this paper, we address this question by introducing Equilibrium Causal Models (ECMs) to the psychological literature. ECMs are causal abstractions of an underlying dynamical system that allow for inferences about the long-term effects of interventions, permit cyclic causal relations, and can in principle be estimated from cross-sectional data, as long as information about the resting state of the system is captured by those measurements. We explain the conditions under which ECM estimation is possible, show that they allow researchers to learn about within-person processes from cross-sectional data, and discuss how tools from both the psychological measurement modeling and the causal discovery literature can inform the ways in which researchers collect and analyze their data.

KEYWORDS

Dynamical systems; causal discovery; cross-sectional data; ergodicity; structural equation modeling

Introduction

Over the last two decades, many areas of psychological science have moved toward characterizing psychological phenomena as arising from systems that evolve over time within an individual, perhaps best typified by *network* and *idiographic process* approaches (Borsboom, 2017; Hamaker, 2012; Molenaar, 2004; van der Maas et al., 2006; Wichers, 2014). To gain insight into these systems, empirical researchers have relied on fitting *statistical models* to a variety of different observational data types, such as cross-sectional data, gathered from many individuals at a single point in time (Borsboom & Cramer, 2013; Epskamp et al., 2018; Robinaugh et al., 2020), and time series data, gathered from one or more individuals at many points in time (Bringmann et al., 2013; Epskamp, 2020; Hamaker et al., 2005; Hamaker & Wichers, 2017; Wichers, 2014).

There are two central challenges that these areas of psychological research face. The first challenge concerns how cross-sectional data can be used to learn about within-person processes. Recent treatments of the *ergodicity* problem have suggested that inferences from cross-sectional to within-person statistical dependencies may rarely be warranted (Hamaker & Wichers, 2017; Molenaar, 2004). While this has led a number of researchers to question the utility of gathering cross-sectional data to learn about within-person processes at all (Bos et al., 2017; Fisher et al., 2018; Hofmann et al., 2020), others have argued that patterns of statistical dependency in cross-sectional data provide an important and complementary source of information (Borsboom & Cramer, 2013; von Klipstein et al., 2021). The second challenge pertains to the question of how observational data can yield insights into the effects of *interventions* (Cronbach, 1957; Diener et al., 2022; Foster, 2010; Grosz et al., 2020; Pearl, 2009; Rohrer, 2018; Ryan & Hamaker, 2022). For example, does reducing avoidance behavior reduce social anxiety? If so, is this change short-lived, with anxiety eventually returning to its usual level, or

long-lived, with typical anxiety levels permanently reduced? Statistical models alone cannot provide an answer to these questions. Instead, we need a causal model of the system (Briganti et al., 2022; Dablander & van Bork, 2021; Pearl et al., 2016; Peters et al., 2017; Ryan et al., 2022; Schölkopf & von Kügelgen, 2022). Recent work in the causal modeling literature has shown that there are typically many different valid causal models for any given dynamical system, and that these models may differ with respect to the type of data they can be learned from, and the types of interventions they are informative about.

In this paper, we introduce *Equilibrium Causal Models* (ECMs) to the psychological literature and investigate how this model class may help address both of the challenges outlined above. An ECM is a causal model of a dynamical system that allows for inferences about the *long-term* effects of interventions, even though it does not describe or allow inferences about the moment-to-moment relations that produce those effects. The ECM can be represented as a causal graph (Bongers et al., 2022). ECMs have been studied before (Bongers et al., 2022; Dash, 2005; Iwasaki & Simon, 1994; Spirtes, 1995; Strotz & Wold, 1960; Weinberger, 2020, 2023)—forming the target of inference in many cyclic causal discovery methods (Bongers et al., 2021; Lacerda et al., 2012; Mooij et al., 2013; 2020; Mooij & Claassen, 2020; Richardson, 1996; Rothenhäusler et al., 2015)—but are virtually unknown in psychology. Crucially, the ECM can in principle be estimated or learned from cross-sectional data, and thereby provides the missing link between dynamical systems modeling and cross-sectional data analysis. Although most contemporary causal modeling work focuses on Directed Acyclic Graphs (DAGs; Pearl, 2009), ECMs may contain patterns of cyclic causal effects which reflect feedback relations in the underlying dynamical system. This makes them an attractive model class in situations where feedback relationships are expected (Borsboom, 2017; Borsboom & Cramer, 2013; Haslbeck et al., 2022; Park et al., 2023; Schmittmann et al., 2013), but makes them more difficult, on a practical level, to estimate and learn from data. Focusing on the most basic linear case, we introduce and develop the intuition behind ECMs and—drawing on both the psychological measurement and causal discovery literature—discuss how equilibrium causal modeling can be made feasible for empirical research.

This paper is structured as follows. In Section 1, we introduce a linear dynamical system model that we will use as a running example, and describe the ways in which different interventions affect the system on different timescales. In Section 2, we use this example model to introduce the general concept of an ECM, showing how they allow us to make inferences about long-term intervention effects, and how they can in principle be estimated from cross-sectional data under idealized conditions. In Section 3, we examine how these measurement conditions can be relaxed, showing how ECMs fit into the ergodicity debate, and how the psychological measurement literature can aid in estimating ECMs. In Section 4, we discuss key challenges of applying ECMs in practice and show how modern causal discovery methods can be used to overcome problems of unobserved confounding and model identifiability. Finally, we discuss the implications of the ECM perspective for empirical practice moving forward. In addition, throughout we provide example code for estimating and discovering ECMs using relevant SEM and causal discovery packages in *R* (R Core Team, 2021).

Dynamical systems and causal effects

Dynamical systems theory provides a unifying framework for studying how systems as disparate as the climate and the behavior of humans change over time. A *dynamica system model* is a set of equations which describe how the (multivariate) process of interest evolves over time (Hamilton, 1994; Strogatz, 2014). Often, dynamical systems models are formulated as *generative* or *computational* models, formalizing substantive assumptions about the exact nature of the mechanisms and moment-to-moment relationships which drive the system under investigation (Haslbeck et al., 2022; Robinaugh et al., 2021; Scholkopf et al., 2021; van Rooij, 2022). This is in contrast to, say, *statistical* models, which only describe patterns of co-occurrence in a particular data type, and *causal* models, which we can think about as a mid-point between statistical and computational models (Haslbeck et al., 2022; Schölkopf & von Kügelgen, 2022). Causal models are more informative than statistical models, but less informative than computational models: They allow for inferences about statistical relationships *and* about the effects that certain interventions might have on the system, without necessarily allowing inferences about the exact mechanistic relations in the underlying system (Peters et al., 2017; Scholkopf et al., 2021; Schölkopf & von Kügelgen, 2022).

In the first part of the paper, we will focus on developing the intuition behind having a causal model of a dynamical system. We introduce a simple multivariate dynamical

systems model, and describe various interventions that are possible in this system. We will use this model throughout the remainder of the paper to build intuition for ECMs and what they can tell us about dynamical systems. We also focus our analysis of how to recover ECMs from data around this specific type of dynamical system (for discussions about ECMs in general, see Mooij et al., 2013; Bongers et al., 2022; Blom et al., 2020; Weinberger, 2020; Dash, 2005; Iwasaki & Simon, 1994).

A basic dynamical systems model

A dynamical systems model consisting of *linear* relationships between the variables of the system can be represented using the auto-regressive equation

$$\mathbf{X}_t = \mathbf{c} + \mathbf{\Phi}\mathbf{X}_{t-1} + \boldsymbol{\epsilon}_t, \tag{1}$$

where $\mathbf{c}$ is a vector of *intercepts*, $\mathbf{\Phi}$ is a matrix of parameters encoding *lagged effects* of $\mathbf{X}_{t-1}$ on $\mathbf{X}_t$, and $\boldsymbol{\epsilon}_t$ is a vector of *error terms*, representing perturbations to the system drawn from a Gaussian distribution $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \mathbf{\Sigma}_\epsilon)$. In psychological modeling terms, this can be considered a model for how a system evolves over time *within* a particular person.

Readers may recognize Equation (1) as the first-order vector auto-regressive or VAR(1) model, which is a popular choice for the statistical analysis of time series data across a number of domains, and in particular is widely used to analyze psychological time series data (Bringmann et al., 2013; Epskamp, 2020; Hamaker, 2012; Hamilton, 1994; Vanhasbroeck et al., 2021). In those settings we typically have many repeated self-report measures of some psychological phenomena, such as *stress* or *anxiety*, and the VAR(1) model is used to model the wave-to-wave auto- and cross-covariances between these different variables. In the present paper, we treat Equation (1) not as a statistical model but as a computational (i.e., generative) model, describing how future values of the process $\mathbf{X}_t$ are produced by a combination of time-constant forces $\mathbf{c}$ acting on the system; past values of the process $\mathbf{X}_{t-1}$ through the matrix $\mathbf{\Phi}$; and random exogenous perturbations to the system $\boldsymbol{\epsilon}_t$. As an example, suppose $\mathbf{X}_t$ represents the multivariate process *stress*. Then we could interpret $\mathbf{c}$ as representing the stable effect of one's environment (e.g., having a stressful job), $\boldsymbol{\epsilon}_t$ as representing time-varying effects of one's environment (e.g., entering or leaving a stressful interaction), and $\mathbf{\Phi}$ as representing the way in which current stress levels produce or regulate future stress levels, as well as the levels of the other psychological processes in our system. The diagonal elements of $\mathbf{\Phi}$ are termed auto-regressive effects and the off-diagonals cross-lagged effects, with ϕ_{jk} representing the extent to which $X_{k,t-1}$ determines $X_{j,t}$. The values of the lagged parameters $\mathbf{\Phi}$ can be represented in the form of a network. This is shown in Figure 1(a), where we see for example a cyclic relationship $X_2 \rightleftarrows X_3$ due to the presence of non-zero values of the cross-lagged effects ϕ_{23} and ϕ_{32}. This implies that X_2 at time point $t-1$ has an effect of size ϕ_{32} on X_3 at time point t, which in turn has an effect on X_2 of size ϕ_{23} at time point $t+1$.[1]

A key concept in dynamical systems modeling is that of an *equilibrium*. We can think of an equilibrium as a resting or steady state of the system—if the system is at equilibrium, then it will stay there unless it experiences some shock or perturbation that pushes it away from equilibrium. In general, a dynamical system can have more than one equilibrium, and these equilibria can be stable (pulling nearby trajectories toward it) or unstable (pushing nearby trajectories away; Strogatz, 2014; Dablander, 2020). The linear dynamical system model we introduced above can be considered simple in the sense that, if the eigenvalues of $\mathbf{\Phi}$ are non-zero, and all smaller than 1 in absolute value (denoted $|\lambda| < 1$), then the system fluctuates around a single stable equilibrium position $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}_t]$ over time, as depicted in Figure 1(b). $\mathbb{E}$ denotes the expectation operator yielding the long-run average of the process. We can write

$$\begin{aligned} \mathbb{E}[\mathbf{X}_t] &= \mathbb{E}[\mathbf{c} + \mathbf{\Phi}\mathbf{X}_{t-1} + \boldsymbol{\epsilon}_t] \\ \boldsymbol{\mu} &= \mathbf{c} + \mathbf{\Phi}\boldsymbol{\mu}, \end{aligned} \tag{2}$$

since the noise term $\boldsymbol{\epsilon}_t$ is independent of $\boldsymbol{X}_t$ and has expectation zero and $\mathbf{c}$ is a constant.[2] Equation (2) is another way of stating that if we are at equilibrium (right-hand side), then we will we stay at equilibrium (left-hand side) in the absence of any external perturbation. If the eigenvalues of $\mathbf{\Phi}$ are all smaller than 1, perturbations $\boldsymbol{\epsilon}_t$ push the system away from equilibrium, but the lagged parameters $\mathbf{\Phi}$ ensure that the system returns to equilibrium over time. In the present paper we will consider only stable systems, that is, systems that return to their equilibrium after a perturbation. The expression for how the equilibrium position in a stable system is determined by the model parameters can be found by re-arranging Equation (2) as

[1] While we use an auto-regressive equation rather than a linear differential equation (a continuous-time model; Driver et al., 2017; Ryan et al., 2018; Ryan & Hamaker, 2022) for simplicity, the results presented here generalize to the continuous-time case if we interpret the auto-regressive equation as defined over an infinitesimal time-step.

[2] Equation (1) is sometimes also written in its mean-centered form, using $(\mathbf{X}_t - \boldsymbol{\mu})$. However, that form can sometimes imply that the mean of the process is determined independently from the rest of the parameters of the model, an assumption we explicitly do not make in the current setting.

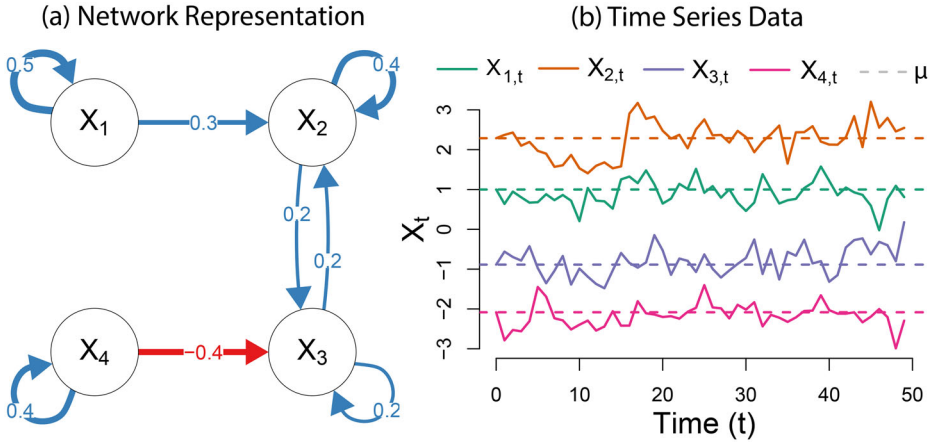

Figure 1. Left: Example of a linear dynamical system depicted as a network. Each arrow $X_k \rightarrow X_j$ represents the value of a lagged parameter ϕ_{jk}. Right: Data generated by the dynamical system with $\mathbf{\Phi}$ as depicted in Panel (a), $\boldsymbol{c} = [0.50, 1.25, -2.00, -1.25]$, and $\mathbf{\Sigma}_\epsilon = \boldsymbol{I}\sigma^2$ with $\sigma = 0.30$. The dashed lines represent the equilibrium positions $\boldsymbol{\mu}$.

$$\boldsymbol{\mu} = (\mathbf{I} - \mathbf{\Phi})^{-1}\mathbf{c}, \tag{3}$$

which shows that the location of the equilibrium position $\boldsymbol{\mu}$ is determined both by the value of the lagged parameters $\mathbf{\Phi}$ and by the time-constant intercept terms $\mathbf{c}$.

Causal effects and interventions

Now that we have a dynamical systems model in place, we can consider how the behavior of this system reacts to different *interventions*. In other words, we can consider how the system can be understood in terms of different *causal relations* and its reactions over time to different interventions. These causal relations will allow us to understand *causal models* of a dynamical system in the following sections. Note that for now we will focus primarily on defining causal effects and causal models, and will return to the problem of estimating or learning that causal model from data, and the strict assumptions necessary for that type of inference, in later parts of the paper. To avoid any potential confusion when reading this section, recall that we treat the model described in Equation (1) as a causal, not a statistical model.

To define a causal relationship we need to specify three components. First, we need to define the specific type of intervention or action which we will apply to the system. Following the modern causal inference literature, we limit ourselves to considering interventions that are *local* and *modular* (Pearl, 2009; Peters et al., 2017). That is, we assume it is possible to intervene on one component of our system at a specific time without altering other parts of the system. In other words, we assume that we can intervene to force $X_{j,t}$ to obtain a particular value (say, $X_{j,t} = 1$), without also changing the value of $X_{k,t}$ (locality), and that the manner in which the effect variable $X_{k,t+1}$ reacts to $X_{j,t} = 1$ is not itself changed by the intervention (modularity). The second component we need to define is the *target* of the intervention, that is, the part of the system on which we intervene, such as the value of X_j at time point t. Third, we need to define the property of the system on which we want to evaluate the *effect* of the intervention (see also, Gische & Voelkle, 2022). An intervention is said to have a causal effect on the system if the intervention leads to changes in the short- and/or long-term properties of the system, such as the value of the system at a future time point, or the equilibrium positions. For our purposes we will focus on three possible types of intervention that we could apply to our dynamical system model, and outline the effects each intervention has on different properties of the system.

Pulse interventions

A *pulse intervention* can be understood as an intervention that changes the value of a single variable at a single point in time (Bender et al., 1984; Hamilton, 1994; Pearl, 2009). Consider the dynamical system visualized in Figure 1(a). Applying a pulse intervention by forcing $X_{1,t} = 1$ results in $X_{2,t+1}$ increasing in value by $\phi_{21} = 0.30$ compared to the situation in which $X_{1,t}$ is forced to obtain a value of zero. In the causal modeling literature, such an intervention would be denoted $do(X_{1,t} = 1)$, and the causal effect on $X_{2,t+1}$ would be defined as the contrast between the expected values of $X_{2,t+1}$ under two different intervention conditions, e.g., $\mathbb{E}[X_{2,t+1}|do(X_{1,t} = 1)] - \mathbb{E}[X_{2,t+1}|do(X_{1,t} = 0)]$, which, keeping all other variables fixed,

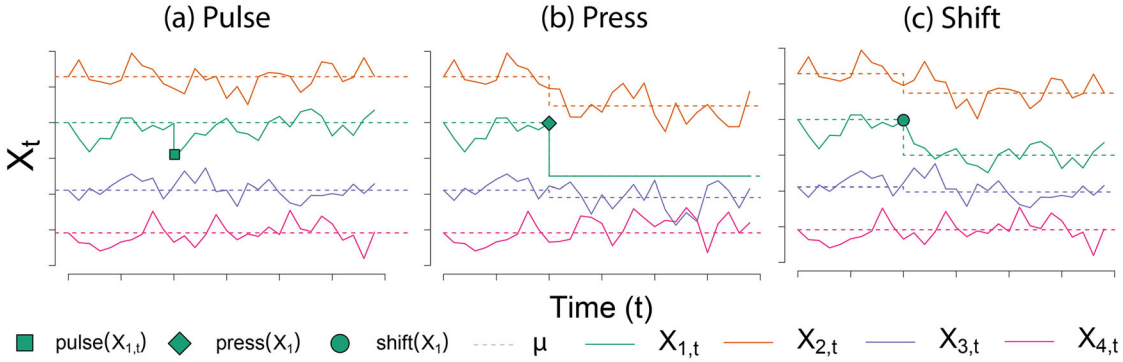

Figure 2. Effect of a pulse (left), press (middle), and shift (right) intervention on our example system.

yields a value of 0.30 (Ryan & Hamaker, 2022). To avoid confusion with other interventions, we will in this paper denote such an intervention $pulse(X_{1,t} = 1)$. Cross-lagged relationships ϕ_{ij} have a natural interpretation in terms of causal effects, as they directly determine the direct effect of a pulse interventions on the value of the system at the next time point.[3] In stable, linear dynamical systems, pulse interventions only have a short-lived effect on the system. They never result in any long-term changes in the equilibrium position or other properties of the system. Mathematically, this can be understood by noting that the effect of a pulse intervention can be found by taking the appropriate power of the lagged effects matrix, $\mathbf{\Phi}^s$, where s represents the time since the intervention took place. Since in stable systems the eigenvalues of $\mathbf{\Phi}$ are smaller than one, $\mathbf{\Phi}^s$ converges to a matrix of zeros as s gets larger and larger—the effect of the pulse intervention is zero at a long enough timescale. An example of this can be seen in Figure 2(a), where the pulse intervention described above is applied at $t = 10$, and the behavior of the system as it reacts to this intervention is visualized. For stable linear systems, the pulse intervention has no long-term effect—the equilibrium position of each variable remains the same.[4] For a mathematical treatment of pulse interventions, see Appendix A.1 and for code to calculate and simulate the effects of all interventions described in this paper, see https://github.com/fdabl/Equilibrium-Causal-Models.

Press interventions

A *press intervention* is an intervention that sets a single variable to a constant value over *an interval of time*, rather than at just a single moment in time (Bender et al., 1984; Hyttinen et al., 2012; Pearl, 2009). While a pulse intervention can be thought of as a temporary shock to the system, the press intervention could be thought of as "clamping down" on the system over a longer period of time. Formally, it can be expressed as $do(X_{j,t} = 1)$ for $t \in \{T, T + 1, ..., T + k\}$ (Pearl, 2009; Ryan et al., 2022). For clarity, we will here denote such an intervention using $press(X)$. The press intervention has both a short and long-term effect on the system, as we can see in Figure 2(b). Here we intervene in the system to force $X_1 = 1$ starting at $t = 10$ for an indefinite period of time. In the short-term, that is, from t to $t + 1$, this press intervention has a similar effect to the pulse intervention, changing the value of $X_{2,t+1}$ by the same amount. However, the press intervention changes the equilibrium positions of the system: During the period where the press intervention is active, the equilibrium positions of both X_2 and X_3 change, in both cases decreasing in value. This happens because forcing X_1 to obtain a constant value over time effectively alters the lagged relationships in the system, setting all of the lagged effects pointing toward X_1 in $\mathbf{\Phi}$ to zero. This in turn yields a new vector of equilibrium positions when plugged into Equation (3). Note that in this case the press intervention results in a system which is still stable, but for some linear systems, applying a press intervention can result in an even more dramatic change in the long-term behavior of the system, changing it from stable to unstable. For simplicity, we will consider only systems that are stable and stable under press interventions, a point we will return to in the discussion. For a mathematical treatment of press interventions, see Appendix A.2.

Shift interventions

Finally, we may consider interventions that target parts of the system other than the value of X_j at one or more points in time. A *shift intervention* represents

[3]This intervention also has an indirect effect on $X_{3,t+2}$, which we can see by reading off the directed path $X_1 \rightarrow X_2 \rightarrow X_3$ from the network in Figure 1(a). The effect of this intervention on $X_{3,t+2}$ can be calculated by taking the product of cross-lagged effects involved. However, this intervention has no effect on $X_{4,t+1}$ since $\phi_{42} = 0$ and there are no indirect paths from $X_{1,t}$ to $X_{4,t+s}$ in the network.

[4]In systems with multiple equilibria, pulse interventions may push the system into a different equilibrium (see e.g., Dablander, 2020).

Table 1. Characterizations of the effects of different interventions.

Intervention	Action	Short-term effect	Long-term effect
Pulse	Force $X_{k,t} = a$ at time t	$\mathbb{E}[X_{j,t+1}] = \phi_{jk} \times a$	No effect
Press	Force $X_{k,t} = a$ for $t \in \{T, T+1, \ldots\}$	$\mathbb{E}[X_{j,t+1}] = \phi_{jk} \times a$	$\boldsymbol{\mu} = (\mathbf{I} - \boldsymbol{P}_k\Phi)^{-1}(\boldsymbol{P}_k\boldsymbol{c} + \boldsymbol{a}_k)$
Shift	Force $c_k = c_k + a$ for $t \in \{T, T+1, \ldots\}$	$\mathbb{E}[X_{j,t+1}] = \phi_{jk} \times (c_k + a + \boldsymbol{\Phi}\mathbf{X}_t)$	$\boldsymbol{\mu} = (\mathbf{I} - \boldsymbol{\Phi})^{-1}(\boldsymbol{c} + \boldsymbol{a}_k)$

The short-term effect describes the effect of $X_{k,t}$ on $X_{j,t+1}$. The long-term effect is defined with respect to the new equilibrium positions produced by the intervention. $\boldsymbol{P}_k$ represents the $p \times p$ identity matrix with the k^{th} diagonal set to zero, and $\boldsymbol{a}_k$ represents a $p \times 1$ vector with the k^{th} element set to a and zeros elsewhere. The derivation of the long-term effects is detailed in Appendix A.

an intervention to change the intercept of a variable c_j in our system, that is, a change to the time-invariant force acting on a particular variable in the model, and which we will denote *shift*(X) (Eberhardt & Scheines, 2007; Peters et al., 2016; Rothenhäusler et al., 2015).[5] Figure 2(c) depicts a shift intervention to set c_1 (the intercept term of X_1) to a value of 1 from time point $t = 10$ onwards. Unlike in the case of a press intervention, we can see that the variable X_1 is free to vary when the shift intervention is applied. However, the equilibrium position of X_1 changes due to the change in the intercept, and this in turn results in a change to the equilibrium positions of X_2 and X_3. The change in the intercept yields a new vector of equilibrium positions $\boldsymbol{\mu}$, which can be obtained by plugging the new intercept vector into Equation (3). As such, shift interventions have a long-term effect on the dynamical system, and this long-term effect is distinct from that of press interventions. Unlike press interventions, shift interventions in linear systems never change the overall stability of the system, since shift interventions do not change the lagged relationships $\boldsymbol{\Phi}$, which determine stability. For a mathematical treatment of shift interventions, see Appendix A.3.

Summary

As we can see, different interventions have different effects on different properties of a system. This is summarized in Table 1, where we show the three different intervention types described above, whether these interventions can in principle have a short or long-term effect on the system, and if so, how to calculate this effect. Of course, not all interventions will have an effect on the target variable, since this depends on the weights matrix $\boldsymbol{\Phi}$ of the system at hand: In our example system in Figure 1, no intervention on X_1 will have an effect on X_4, while all types of interventions on X_1 will have some effect on X_2. Notably, however, in any stable linear system, pulse interventions can have no long-term effects on the equilibrium, while press and shift interventions can affect the equilibrium. To continue our stress example, we could imagine that pulse interventions might produce short-term changes to stress levels for example by reminding one of an upcoming public speech. Press interventions may map onto much stronger interventions, for example the administering of a hypothetical drug which completely suppresses the stress response, while shift interventions change stable parts of one's environment which produce stress, for example by switching to a less stressful profession. Because we are interested in long-term changes to the system in the current paper, we will focus on press and shift rather than on pulse interventions in the remainder. Specifically, in the following we will describe how, under certain conditions, we can use observations of a dynamical system taken at a single point in time to learn about the effects of these interventions, and show how it is possible to do so *without* knowing the moment-to-moment generating parameters of the dynamical system.

Equilibrium causal models

In the previous section, we introduced a basic dynamical system governed by linear moment-to-moment dynamic relationships and showed how we can define and compute the effects of different interventions in that system. A natural implication of this is that, if we wish to learn about the effects of interventions in our system, we could try to do this by collecting suitable repeated measures time series data of each process, and in recent years, a number of psychological researchers have called both for an increasing focus on theories for how psychological processes evolve over time within an individual and a complimentary increase in empirical approaches which collect and analyze psychological time series to gain insight into these processes (Hamaker, 2012; Molenaar, 2004; Robinaugh et al., 2019).

It turns out, however, that this is not the *only* way in which researchers might hope to learn about the effects of interventions in a dynamical system. In this section, we describe the *Equilibrium Causal Model*

[5]This interpretation allows us to consider c_j as a latent time-invariant variable to which we apply an intervention. Alternatively, we could also interpret a shift intervention as a *soft* intervention or mechanism change, which alters part of the system (e.g., the mean of a variable) without forcing the variables to obtain a constant value (Eberhardt & Scheines, 2007).

(ECM), which is a type of model which allows for inferences about the effects of certain interventions in a dynamical system, but which in principle can be learned from single-time-point observations, that is, cross-sectional data. ECMs allow inferences about these interventions *without* capturing the moment-to-moment dynamics of the underlying dynamical system.

A causal model of a dynamical system

To define the ECM, it is first necessary to clarify what we mean by a *causal model* in the first place. A *causal model* of a dynamical system is any model that yields predictions about the effect of *one or more* types of interventions in that system. In the context of the example model introduced above, a causal model would correctly inform us about at least one cell of Table 1: the short and/or long-term effect of pulse, press and/or shift interventions. The equations which govern how the dynamical system evolves over time, as specified in Equation (1), would of course represent the best causal model of the system we can hope for: If we know all of the parameters of these equations, then we know the exact moment-to-moment dynamics governing the system, and so can derive the short and long-term effects of all intervention types. However, a causal model need not be quite as complete a description of the system to still be useful.

An Equilibrium Causal Model is a causal model of the system which allows inferences about the long-term effects of interventions, that is, the effects of interventions that act on and affect the equilibrium positions of the system. While in principle we could think about defining an ECM for many different types of dynamical systems, in the current paper we will only consider ECMs of linear dynamical systems such as the one presented in the previous section. An ECM of our example system would allow us to assess the effect of applying press and/or shift interventions (the bottom-right and middle-right cells of Table 1) on the equilibrium positions of other variables. To see how this works, recall that the equilibrium is the resting state of the system. If the system is at equilibrium, it will—in expectation—take on the same value at the next point in time, as stated in Equation (2). An ECM of this system can thus be understood as any set of equations, that is, any *new* set of parameters, replacing $\mathbf{\Phi}$ and $\mathbf{c}$, which (a) yields the same equilibrium positions as the original system, that is, satisfies Equation (2) and (b) yields the same effects of press and shift interventions as described in Table 1, but which (c) does not describe the exact moment-to-moment dynamics of the system, as defined by the original parameters of Equation 1.

Hyttinen et al. (2012) showed that there exists an ECM representation of the linear dynamical system that we consider here, and described how the parameters of that ECM can be obtained as functions or transformations of the original systems parameter matrices as follows. First, we can obtain a matrix of *equilibrium direct effects* $\tilde{\mathbf{\Phi}}$ by (a) deleting the auto-regressive parameters or self-loops ϕ_{ii} such that $\tilde{\phi}_{ii} = 0$, and then (b) computing the off-diagonal parameters of this matrix as a function of the cross-lagged parameters in the dynamic process

$$\tilde{\phi}_{jk} = \frac{\phi_{jk}}{(1 - \phi_{jj})}, \tag{4}$$

as shown in Appendix B. Intuitively, the equilibrium direct effects can be seen as *re-scaling* or *standardizing* the cross-lagged effects, according to how stable the outcome variable X_j is. The more stable the outcome variable (the closer ϕ_{jj} is to one), the bigger the equilibrium direct effect $\tilde{\phi}_{jk}$. As such, the matrix $\tilde{\mathbf{\Phi}}$ can be interpreted as long-run or *equilibrium direct effects*, rather than the moment-to-moment direct effects of the original system (although derived in a different context, the same interpretation is given to re-scaling effects by $(1 - \phi_{jj})$ by Shamsollahi et al., 2022). Intuitively, we can understand this re-scaling as a path-tracing operation through the auto-regressive effect of X_j over-time (for more details, see Appendix B.1). Second, we obtain the intercepts of the equilibrium model in a similar way, by transforming (re-scaling) the original intercepts through

$$\tilde{c}_j = \frac{c_j}{(1 - \phi_{jj})}, \tag{5}$$

which again means that, the more stable the variable, the larger the re-scaled intercept term becomes. We can again interpret these new intercept terms as representing the cumulative or long-run effect of the time-constant forces $\mathbf{c}$ in the original system. The intuition for this interpretation is given in Appendix B, where we show how these parameters arise from path-tracing through (or equivalently, marginalizing over) the auto-regressive parameters.

Taken together, these new parameter matrices define an ECM of the system, which can be written as

$$\boldsymbol{\mu} = \tilde{\mathbf{c}} + \tilde{\mathbf{\Phi}}\boldsymbol{\mu}. \tag{6}$$

Using the results of Hyttinen et al. (2012), it can be shown that these transformed parameters imply the same equilibrium positions $\boldsymbol{\mu}$ as the original dynamical system in Equation (2) (see Appendix B.2 for

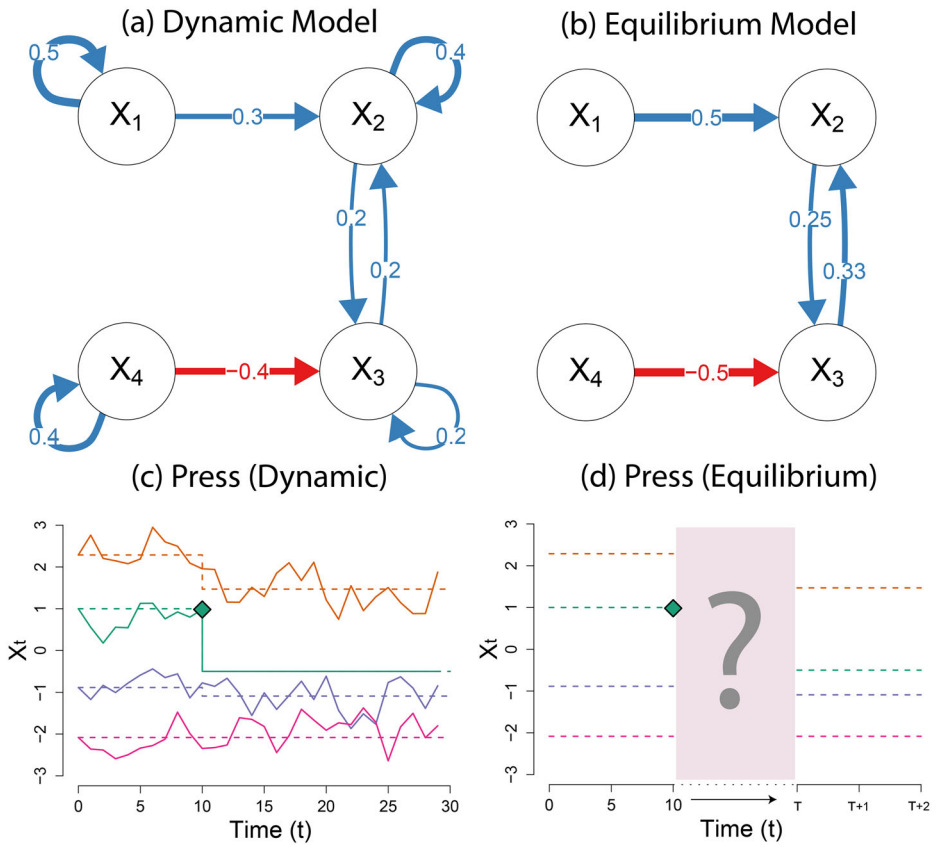

Figure 3. Example of a linear dynamical system (a) and its Equilibrium Causal Model (b). In panel (c) we see the evolution of the dynamical system following a press intervention at time $t = 10$, where dashed lines represent the equilibrium positions of each variable, and solid lines represent simulated trajectories of each variable. In panel (d) we illustrate how the ECM predicts the same long-term effect of the press intervention, that is, the change in equilibrium positions (dashed lines). However, the equilibrium model cannot be used to make inferences about the actual trajectories of each variable over time.

details). We want to reiterate that the parameters $\tilde{\mathbf{c}}$ and $\tilde{\mathbf{\Phi}}$, while arrived at through re-scaling of the parameters $\mathbf{c}$ and $\tilde{\Phi}$ of the original dynamical system, are not parameters of this dynamical system. Instead, they are parameters of its Equilibrium Causal Model. The parameters in Equations (4) and (5) can naturally be obtained from the parameters of the dynamical system. However, we will show in the remainder of the paper that they can also be obtained from cross-sectional data under particular conditions.

While this may seem somewhat abstract at the moment, this definition of the ECM is useful for two reasons. First, it implies that the ECM can be informative about certain intervention effects, and second, it implies that the ECM could in principle be estimated from single-time-point observations, without the need to estimate either the auto-regressive effects or the cross-lagged effects. Before we examine how to estimate the ECM, however, we will first focus on understanding what kinds of inferences the ECM allows, that is, why we might want to estimate it in the first place.

To understand what else the ECM can and cannot tell us about a dynamical system, we first visualize the parameters of the ECM for our example system in Figure 3(b) alongside the original parameters in panel (a). We can see that the ECM contains a direct effect $\tilde{\phi}_{jk} \neq 0$ only when there is a corresponding moment-to-moment direct effect $\phi_{jk} \neq 0$, which means that the ECM parameters *in this situation* are informative about the presence or absence of direct relationships.[6] As a result, while the ECM in Equation (6) can be read as static model, relating stable equilibrium positions to each other, the model itself contains cyclic relationships, since the dynamical system from which it is derived also contains cyclic relationships. Besides this, however, the ECM is not very informative about the moment-to-moment dynamics of the system. Since the intercepts and cross-lagged parameters have been re-scaled, the values of the ECM parameters can be quite different to the moment-to-moment direct

[6]According to Equation (6), if $\phi_{jj} \neq 1$, as we would expect in a stable system, then $\tilde{\phi}_{jk} \neq 0$ if and only if $\phi_{jk} \neq 0$.

effects: (in)equality relations (such as $\phi_{23} = \phi_{32}$ and $|\phi_{21}| < |\phi_{34}|$) are not necessarily preserved in the ECM ($\tilde{\phi}_{23} > \tilde{\phi}_{32}$ and $|\tilde{\phi}_{21}| = |\tilde{\phi}_{34}|$). Since the parameters are all re-scaled, and since $\tilde{\mathbf{\Phi}}$ contains no autoregressive parameters, if we were to plug these model matrices into Equation (1), we would simulate time series which look very different from those produced by the original system.

Despite these limitations, the ECM is a useful and informative representation of the dynamical system because the ECM makes the same predictions about the long-term effects of interventions as the original system. The proof of this statement is adapted from the work of Hyttinen et al. (2012) and is outlined in detail in Appendix B.3 and B.4. For a press intervention, these predictions can be made by simply plugging in the ECM parameters into the expression for the press intervention in Table 1. This is shown for our example in panels (c) and (d) of Figure 3. First, we see that for $t < 10$, both models imply the same equilibrium positions in the no-intervention setting. In panel (c) we show the trajectory of the system following a press intervention to force $X_2 = a$, the effect of which on the long timescale is to alter the equilibrium positions of X_1, X_2, and X_3. We see in panel (d) that the ECM, while not allowing us to predict the trajectories of each variable following the intervention, does allow us to predict the new equilibrium positions following the intervention. The post-intervention equilibrium positions predicted by the ECM are identical to those of the original system.

The importance of these insights are so central to the rest of the developments presented here that they deserve repeating. The ECM is informative about the long-term effects of interventions in our system, while simultaneously being almost entirely uninformative about the short-term effects of those interventions. In other words, although we are studying a system that evolves over time, we can potentially make correct predictions about how to intervene in that system without knowing the moment-to-moment dynamics that govern the system. In the next section, we extend the ECM to multiple individuals and then discuss how to estimate its parameters from cross-sectional data.

Equilibrium causal models and multiple individuals

Now that we have defined what an Equilibrium Causal Model is and shown that it yields important insights into the effects of interventions, we can begin to turn to the question of how ECMs might be useful in empirical research using cross-sectional data. Before doing so, however, we must first extend our dynamical systems model to multiple individuals. The dynamical systems model in Equation (1) describes a within-person model, that is, how process values $\mathbf{X}$ evolve over time t for a single individual. To extend this model to include multiple different individuals we need to impose some structure on the ways in which these individuals differ from and are similar to one another, that is, we need to add between-person component(s) to our dynamical systems model. We consider the following model:

$$\mathbf{X}_{i,t} = \mathbf{c}_i + \mathbf{\Phi}\mathbf{X}_{i,t-1} + \boldsymbol{\epsilon}_{i,t}, \tag{7}$$

where the parameters have the same meaning as in Equation (1) and the subscript i denotes individuals so that parameters or variables with that subscript vary across individuals. The process values $\mathbf{X}_{i,t}$ and perturbations $\boldsymbol{\epsilon}_{i,t} \sim \mathcal{N}(0, \mathbf{\Sigma}_\epsilon)$ both vary across time and differ between individuals. In terms of the parameters of the model, individuals are similar in the sense that they share the same lagged parameter matrix $\mathbf{\Phi}$. However, individuals are allowed to differ in their intercepts, that is, time-constant forces $\mathbf{c}_i \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{c}}, \mathbf{\Sigma}_{\boldsymbol{c}})$. Note that this model is conceptually similar to a VAR(1) model with random intercepts across individuals. The equilibrium positions for each individual satisfy the expression

$$\boldsymbol{\mu}_i = (\boldsymbol{I} - \mathbf{\Phi})^{-1}\boldsymbol{c}_i, \tag{8}$$

which implies that variation between people in their equilibrium positions arises because of the variation in time-constant terms $\mathbf{c}_i$. As in the within-person case, we consider only those systems that are *stable*, which, as outlined above, in the linear case means assuming that $\mathbf{\Phi}$ has eigenvalues $|\lambda| < 1$.

We can define different interventions in the system—and hence different causal effects—in much the same way as we did for the single-individual model in the previous section. Since the time-constant forces differ across individuals, so too does the effect of press and shift interventions on the equilibrium positions of those individuals, and these can be computed using the expressions shown in Table 1. As in the single-individual case, we can write down the ECM of this system as

$$\boldsymbol{\mu}_i = \tilde{\boldsymbol{c}}_i + \tilde{\mathbf{\Phi}}\boldsymbol{\mu}_i, \tag{9}$$

where $\tilde{\mathbf{\Phi}}$ is the matrix of equilibrium direct effects, re-scaled and fixed across individuals. The term $\tilde{\boldsymbol{c}}_i$ represents a vector of person-specific intercepts, which are each re-scaled as described in the previous section.

This implies that these new intercept terms are also normally distributed, $\tilde{\boldsymbol{c}}_i \sim \mathcal{N}(\tilde{\boldsymbol{\mu}}_{\boldsymbol{c}}, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{c}})$, with the mean and covariance matrix of the intercepts across individuals ($\tilde{\boldsymbol{\mu}}_{\boldsymbol{c}}$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{c}}$) also re-scaled in a similar way (see Appendix B for details).

As in the single-individual case, the ECM predicts the same equilibrium positions as the original model and makes equivalent predictions about the long-term effects of press and certain (standardized) shift interventions. This is discussed in detail in Appendix B.3 and B.4. The implication of this is that, if we can learn or estimate this model from data, then we can infer the long-term effects of these interventions.

Equilibrium causal models from cross-sectional data

With these concepts in place, we can now turn our attention to how and under what conditions ECMs can be estimated from different data types. As is well known from the causal modeling literature, inference about causal effects from observational data is only feasible under a number of strict conditions, such as the absence of unobserved confounding variables and selection bias, and the possibility of local or modular interventions, which we have introduced above (see also Pearl, 2009). In addition, causal inference in practice often relies on the estimation of a statistical model, and so assumptions necessary for valid statistical inference, such as linearity and distributional form assumptions, are also required. While these assumptions deserve critical evaluation in any causal modeling context (and we return to them in more detail later in the current paper), there are two additional types of conditions that we must contend with when considering inference about equilibrium causal effects. The first of these concerns the nature of the underlying dynamical system we are studying. As we have stated above, in the current paper we consider only linear dynamical systems which are stable and stable under different interventions, and we consider that we are studying a population of individuals who are relatively homogeneous, differing only in the values of their individual intercepts. The second condition concerns the nature of the information that cross-sectional measurements captures about the underlying dynamical system.

The crucial insight we can take from our treatment above is that, because the ECM implies the same equilibrium positions as the original system, if we observe those equilibrium positions when collecting cross-sectional data then we can potentially estimate the ECM on that data. As such, ECMs provide a potential bridge between dynamical systems modeling and cross-sectional (i.e., multiple individuals, single-time point) data analysis.

Stated more formally, a sufficient (but as we will see later, not necessary) condition is that cross-sectional measurements consist of direct observations of the equilibrium positions of the system for each individual, $\boldsymbol{\mu}_i$. That is, we need single time-point observations across individuals to represent the resting state or long-run average value of each variable for each person. The importance of this condition becomes clear when we rewrite the model in Equation (9) as

$$\boldsymbol{\mu}_i = \tilde{\boldsymbol{\mu}}_{\boldsymbol{c}} + \tilde{\boldsymbol{\Phi}}\boldsymbol{\mu}_i + \boldsymbol{\zeta}_i, \tag{10}$$

where $\boldsymbol{\zeta}_i \sim \mathcal{N}(0, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{c}})$ represents the person-specific deviation from the population average intercept $\tilde{\boldsymbol{\mu}}_{\boldsymbol{c}}$.[7] On the left-hand side we have a vector of equilibrium positions per person, which are regressed on their direct causes on the right-hand side. The parameters of this model are a vector of constants $\tilde{\boldsymbol{\mu}}_{\boldsymbol{c}}$, a square matrix of regression parameters, and a mean-zero residual with variance covariance matrix defined above.

By re-writing the model in this form, it becomes clear that the ECM can be seen as a structural equation model (SEM; Bollen, 1989) defined with respect to *equilibrium data*. From the usual SEM variance-covariance decomposition it follows that the (co)variance of the equilibrium positions is given by

$$\boldsymbol{\Sigma}_{\boldsymbol{\mu}} = (\mathbf{I} - \tilde{\boldsymbol{\Phi}})^{-1}\boldsymbol{\Sigma}_{\boldsymbol{c}}(\mathbf{I} - \tilde{\boldsymbol{\Phi}})^{T}, \tag{11}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{c}}$ is the covariance matrix describing how the time-constant forces $\boldsymbol{c}$ of the individuals relate to each other, and $\boldsymbol{\Sigma}_{\boldsymbol{\mu}}$ is the covariance matrix of the equilibrium data. This equation implies that, if we observe the equilibrium positions for all of the variables in our system, then the problem of estimating the ECM reduces down to the problem of estimating the corresponding SEM. The estimated ECM then allows us to make predictions about the long-term effects of different interventions, without knowing the moment-to-moment dynamics of the system, as we have outlined above.

To illustrate that this indeed works, we simulate equilibrium data from the dynamical system in Figure 1 for sample sizes n between 50 and 1000. For each individual we have a single observation, representing the equilibrium value for that variable for that

[7] This is because $\tilde{\boldsymbol{c}}_i \sim \mathcal{N}(\tilde{\boldsymbol{\mu}}_{\boldsymbol{c}}, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{c}})$ can be re-written as $\tilde{\boldsymbol{c}}_i = \tilde{\mu}_c + \boldsymbol{\zeta}_i \sim \mathcal{N}(0, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{c}})$

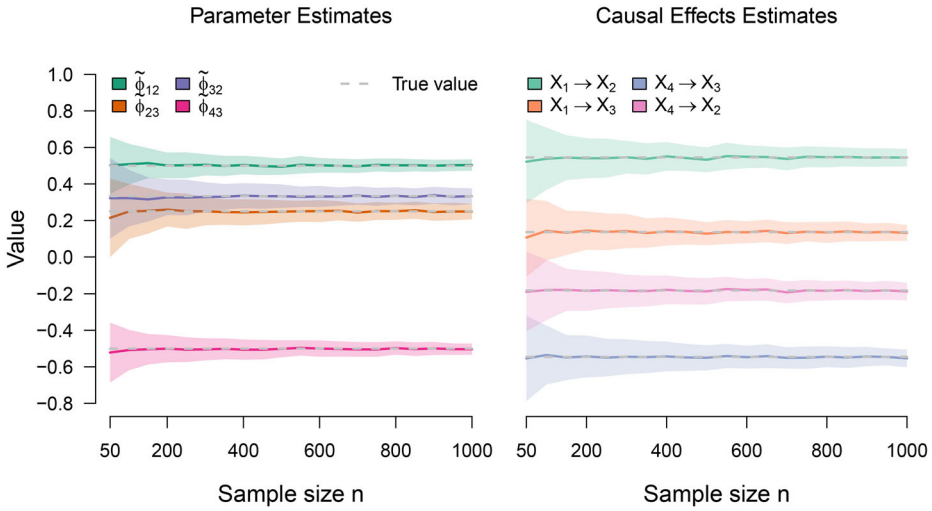

Figure 4. Mean (solid lines) and standard deviation (shaded areas) of parameter estimates (left) and causal effect estimates (right) as a function of sample size *n*. True values are shown as dashed grey lines. Causal effects refer to two different press interventions, $press(X_1 = 1)$ and $press(X_4 = 1)$, with the size of the causal effect in each case evaluated with respect to the change in the resulting equilibrium values of X_2 and X_3. The figure illustrates that both parameter estimates of the equilibrium model and estimates of the effects of press interventions are unbiased when using the appropriate SEM model fit to equilibrium data.

individual.[8] We use *lavaan* (Rosseel, 2012) to fit a cyclic linear SEM model in which X_2 is regressed on (X_1, X_3) and X_3 is regressed on (X_2, X_4). In causal modeling terms, this reflects the situation in which the structure of the causal model is known, and we want to estimate the values of the effects themselves. Because this model is statistically identified, its parameters can be estimated from the equilibrium data just as in any standard SEM application. We repeat this 250 times for each sample size. The left panel in Figure 4 shows that, as expected, the parameters of the model are estimated without bias, with a sampling variance that decreases with sample size. The right panel shows that the effects of the press interventions on X_1 and X_4 on the equilibrium positions of X_2 and X_3 are estimated without bias, too. Naturally, they exhibit a higher variance given that they are computed from parameters that are themselves estimated. The code to reproduce these (and all further) simulation results and figures is available from https://github.com/fdabl/Equilibrium-Causal-Models.

This short exercise illustrates that single time-point measurements can, in principle, yield a model that allows valid and useful inferences about long-term causal effects of an underlying dynamical system. Of course, in showing how this works *in principle* we have made a number of simplifying assumptions, notably regarding the nature of psychological measurements, the statistical identifiability of the model, and the degree of knowledge available on the causal system of interest. In the following sections, we will turn our attention to the potential of using ECMs in psychological research in practice. We will first examine previous research on the nature of psychological measurements and whether the measurement assumption outlined here can be relaxed. We will then discuss issues around statistical identifiability, and the discovery and interpretation of causal models from observational data in general.

Psychological measurement and the ergodicity problem

In showing how equilibrium models can be estimated from data in the previous section we have relied on the assumption that measurements perfectly capture the equilibrium position of the underlying process. In this section, we assess to what degree this is a reasonable assumption considering what is known about the nature of psychological measurement. In particular, We discuss the ergodicity problem, which is commonly interpreted to imply that no useful inferences can be made about dynamical systems from single-time-point cross-sectional observations (Hamaker for more detailed treatments of the ergodicity problem, we refer readers to 2012; Molenaar for more detailed treatments of the ergodicity problem, we refer readers to 2004). Drawing on research from the latent-state-trait modeling literature, we will show that cross-sectional data can in fact be informative about dynamical systems if sufficient knowledge about the variance components underlying our measurements is available.

[8]Intercepts are drawn from a multivariate Gaussian with mean zero and the identity matrix as the covariance matrix.

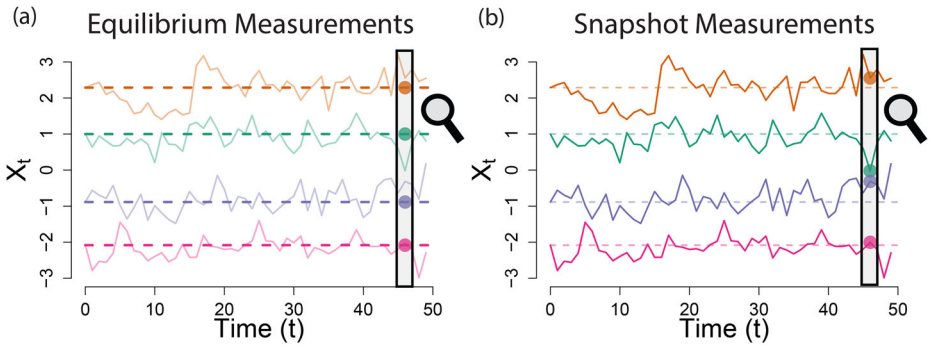

Figure 5. Left: Shows the equilibrium measurement as capturing the expected value or resting state of the process $\boldsymbol{X}$ at a single point in time t. Right: Shows the snapshot measurement as capturing the value or position of a process $\boldsymbol{X}$ at a single point in time t.

Ergodicity as a measurement problem

In the previous section we showed that ECMs can be estimated from cross-sectional measurements of a dynamical system, as long as those measurements capture the equilibrium position of each process. In SEM terminology, we can think about this as proposing a particular type of *measurement model*, which we depict in panels (a) and (b) of Figure 5. We imagine that there is some latent process $\boldsymbol{X}_t$ which evolves over time within an individual, and that when we take measurements of this process at a certain point in time $\boldsymbol{Y}_t$, we capture the long-run tendency or mean of this process. Formally, we have that

$$\boldsymbol{Y}_{it} = \boldsymbol{\mu}_i, \tag{12}$$

where i indexes the person. We can think about this as a reflective measurement model where the measurements are akin to a weighted sum (i.e., average) of some previous values of the latent process stretching back in time.

This is a somewhat different way of conceptualizing single-time-point measurements of a dynamical system than is typically found in the psychological methods literature, which typically assumes that cross-sectional measurements of a dynamical system consist of *snapshots*, that is, direct measurements of the position or value of the process at the time of the measurement ($\boldsymbol{X}_t$), as depicted in Figure 5(b). This model of psychological measurement is present in almost all discussions of the *ergodicity* problem (Hamaker, 2012; Molenaar, 2004) in psychology.[9] Essentially, the literature on the ergodicity problem shows that, when there are inter-individual differences in the underlying dynamic process (such as individual differences in the intercepts or $\boldsymbol{c}_i$ parameters in our data-generating model from Equation (7)), then the means and (co-)variances of snapshot measurements collected cross-sectionally will differ from the means and (co-)variances of data collected repeatedly over time *and* from the means and (co-)variances of the equilibrium positions between individuals. In other words, if measurements represent snapshots, then—even for the basic dynamical system considered in the current paper—cross-sectional statistical dependencies will not be equivalent to either the equilibrium dependencies or the moment-to-moment dependencies between processes in the dynamical systems model, with the latter shown and discussed in depth by, amongst others, Hamaker (2012, 2022); Molenaar (2004) and Schuurman (2023).

Clearly, this challenging feature of studying processes with individual differences is relevant for the estimation of ECMs, since it implies that the ECM cannot directly be estimated from snapshot data using the simple multivariate SEM defined in the previous section. The basic intuition for this is that, when the data represent equilibrium positions, then variation in the observed variables represents the variation across individuals in their equilibrium positions, represented by $\boldsymbol{\Sigma}_\mu$. This variation can, in turn, be attributed to the fixed parameter values in Equation (10) and the variance between individuals in the time-constant forces acting on the system, $\boldsymbol{\Sigma}_c$. However, when observations are snapshots, then additional variation in these data come from the variance *around* the equilibrium positions at a certain point in time. Formally, we have that

$$\underbrace{\boldsymbol{Y}_{it}}_{\text{Observation}} = \underbrace{\boldsymbol{\mu}_i}_{\text{Trait / Equilibrium}} + \underbrace{\boldsymbol{s}_{it}}_{\text{State}}, \tag{13}$$

which states that snapshot measurements at a particular point in time ($\boldsymbol{Y}_{it}$) can be considered as a sum of

[9] An ergodic process is one for which the structure of variation captured between individuals at a single time point (also referred to as interindividual variation) is equivalent or equal to the structure of variation which would be captured over time within a single individual (also referred to as intraindividual variation) (Molenaar, 2004). The ergodicity problem can be broadly understood as the problem of making inferences about non-ergodic processes from single-time-point observations across individuals. The dynamic process under consideration in the current manuscript, defined in Equation (7), is *nonergodic*.

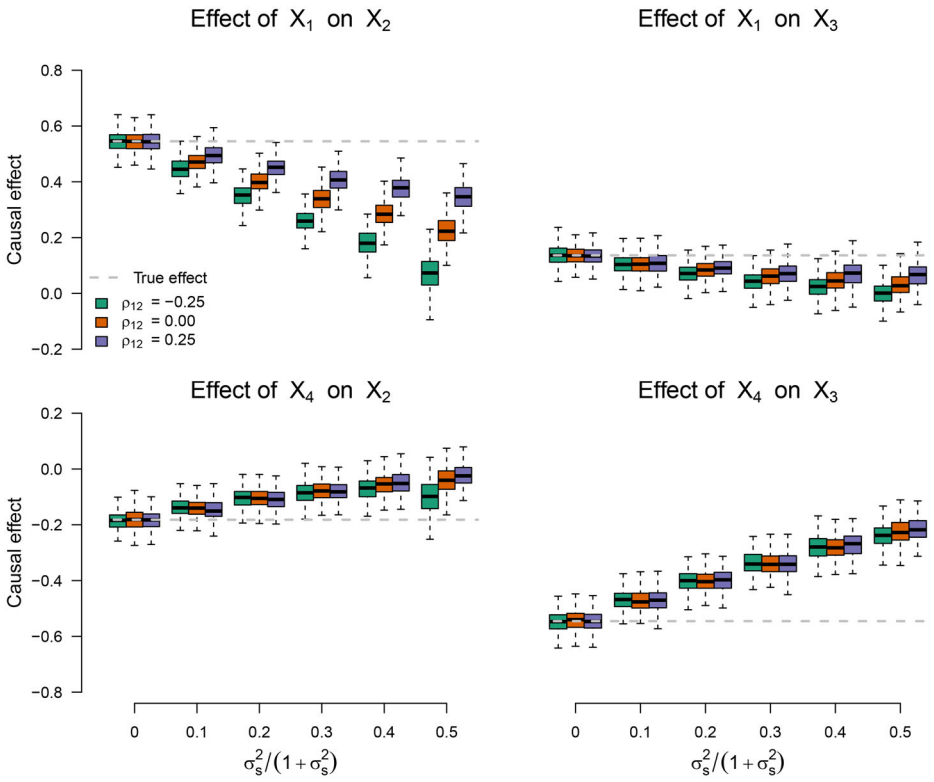

Figure 6. Estimates of the effect of the intervention $press(X_1 = 1)$ on X_2 (top left) and X_3 (top right), and $press(X_4 = 1)$ on X_2 (bottom left) and X_3 (bottom right) across different state variances (given as a proportion of the total variance) and residual correlations between X_1 and X_2. As the state variances increase, the causal effects generally become attenuated. For negative residual correlation between X_1 and X_2 (green), the attenuation is stronger, while it is weaker for a positive one (purple). Dashed gray lines indicate the true causal effect.

the person-specific equilibrium ($\boldsymbol{\mu}_i$) and a person's deviation from their equilibrium ($\boldsymbol{s}_{it}$). Hamaker and Wichers (2017) refer to the $\boldsymbol{\mu}_i$ term as a *trait*, representing the stable mean values of the processes $\boldsymbol{X}$ for a given individual and so equivalent to the equilibrium in a linear dynamical system, while referring to $\boldsymbol{s}_{it}$ as the *state* value. This expression makes it explicit that the variance in our data in the snapshot scenario comes both from the variance between individuals $\boldsymbol{\Sigma}_{\boldsymbol{\mu}}$ and the variance within individuals around their equilibrium, that is, the variance of the state $\boldsymbol{\Sigma}_{\mathbf{s}}$. Critically, failing to separate these two sources of variance means that we cannot use snapshot data to estimate the equilibrium parameters: Estimating a SEM model from the covariance matrix of the snapshot data $\boldsymbol{\Sigma}_{\mathbf{Y}}$ instead of the covariance matrix of the equilibria $\boldsymbol{\Sigma}_{\boldsymbol{\mu}}$ will not recover the ECM.

While this shows that using snapshot data to estimate equilibrium parameters is problematic in theory, the extent to which the estimates are biased in practice depends on the variance-covariance matrix of the states. To assess this, we conduct a small simulation study. Specifically, we simulate $n = 2000$ observations from our example ECM while varying the state variances (assumed equal for all states, $\sigma_s^2 = \text{diag}(\boldsymbol{\Sigma}_{\mathbf{s}})$, and, consistent with the data-generating model defined in Equation 7, equal for all individuals). In particular, setting the covariance matrix of the equilibrium positions to be diagonal with variances 1, we vary the proportion of the total variance that is attributable to the state variance—denoted $\sigma_s^2/(1+\sigma_s^2)$—from 0 to 0.50 in increments of 0.10. A value of 0 indicates that we observe the equilibrium positions directly, while a value of 0.50 indicates that half of the total variance is due to the state variance. In essence, all conditions except the 0 value represent different versions of a "snapshot" measurement, with only the amount of variance attributable to the state varied.

Figure 6 shows the results of the simulation study for the press intervention effect of forcing $X_1 = 1$ on X_2 (top left) and X_3 (top right), and of $X_4 = 1$ on X_2 (bottom left) and X_3 (bottom right). Focusing on the case in which the states are uncorrelated, that is, $\boldsymbol{\Sigma}_s$ is a diagonal matrix (indicated by $\rho_{12} = 0$ in the figure), we find that the causal effect estimates are attenuated, that is, pulled toward zero with increasingly large state

variance. The situation becomes more complicated when we allow for unobserved common causes between the deviations of the equilibrium positions, which results in correlations between them. We simulate two such scenarios: One in which the correlation between X_1 and X_2 is $\rho_{12} = 0.25$ (purple) and one in which it is $\rho_{12} = -0.25$ (green). Naturally, this has the strongest effect on the estimate of the causal effect of X_1 on X_2. A positive correlation increases the causal effect, which works against the attenuation effect from the increasing state variance. A negative correlation on the other hand decreases the causal effect. The effect is similar, but much less pronounced, for the other causal effects.

States and traits in psychological measurement

The simulation study above showed that we obtain biased estimates of the ECM parameters if we use snapshot measurements, capturing the current value of the process at a single point in time, rather than equilibrium measurements, capturing the long-run average of the process. As such, if ECMs are to be of potential use for psychological research, the crucial question is: To what degree should psychological measurements be considered to capture snapshots, equilibrium positions, or something in-between? To gain insight into this question, we can make use of the extensive literature on psychological measurement, in particular the literature on latent state-trait (LST) modeling (Hertzog & Nesselroade, 1987; Steyer et al., 1989, 1999, 2015).

In the psychological measurement literature, a distinction is often made between measurement instruments which aim to capture trait aspects of a psychological process versus those which aim to capture state aspect of that process. Typically, a psychological trait is in this context defined as a stable-over-time or long run characteristic of the process, while a state is considered to vary over time. For example state anxiety could be conceptualized as the degree of anxiety that an individual is feeling at a given moment in time in reaction to a stressful event, while trait anxiety would be the tendency to experience anxiety in reaction to stressful events (Speilberger et al., 1983). In LST research, traits are often specifically defined as long-run mean values of psychological processes, with the trait component of a measurement tool identified by taking the mean of repeatedly administered measurements over time (Steyer et al., 1989). As such, in the context of the dynamical systems model we study in the current paper, the trait aspect of our process $\mathbf{X}$ can be considered to be equivalent to the equilibrium of that process, $\boldsymbol{\mu}$.

When we consider the types of questions typically used in cross-sectional psychological research, we can see that researchers often aim to capture long-run summaries rather than snapshots of the target process. For example, experience sampling studies which aim to study *anxiety* (Bringmann et al., 2016; Rowland & Wenzel, 2020) typically ask participants the degree to which they currently feel anxious multiple times a day, that is, at the moment they receive the measurement prompt. In that situation, it is clear that the researchers aim to capture snapshot measurements of some time-varying psychological process. In contrast, a cross-sectional measurement instrument such as the trait part of the State-Trait Anxiety Inventory (Speilberger et al., 1983) asks participants about how anxious they *generally* feel, as distinct from how they feel at that moment in time. Similar comparisons can be made with depressed mood, where experience sampling studies typically query depressed mood levels at the moment of measurement (Bringmann et al., 2016; Rowland & Wenzel, 2020); cross-sectional studies based on the Beck Depression Index or DSM-criteria ask participants to assess their depressed mood over the past week, two weeks, or the past 30 days (Alegria et al., 2007; Beck et al., 1987; Fried et al., 2016; Kendler et al., 2018); while still other cross-sectional studies assess stable levels of depressed mood using measures designed to assess trait positive and negative affect (Clark & Watson, 1991; Watson et al., 1994). This already gives some indication that the (perhaps implicit) measurement model being used by empirical researchers in cross-sectional settings may be closer to that of equilibrium measurements than that of snapshot measurements.

Of course, prior research has also shown that people's responses to such measurements can be influenced by contextual effects, such as one's current state (Augustine & Larsen, 2012; Barrett, 1997; Bower, 1981; Brose et al., 2013; Leertouwer et al., 2021), which means that, even if we attempt to measure the equilibrium or trait, we are likely not able to do so perfectly. For example, measurements of general life satisfaction appear to be influenced by the mood of the participant, and potentially even the weather on the day of measurement (Schwarz & Clore, 1983). However, as we saw in the simulation study above, the degree to which this is a problem in practice depends on the degree to which measurements deviate from the equilibrium values.

From the LST literature, we know that designing an item to ask about long-term general tendencies does appear to have positive effect on our ability to capture stable features of the process of interest. Braun et al. (2021) showed that, when items were phrased to explicitly probe general trait-like levels of self-esteem and depressive symptoms, around 70% of the variance in these items could be attributed to trait variability ($67.5 - 68.5\%$ for depression, $71 - 72.6\%$ for self-esteem), in contrast to state-like measurements, for which $50 - 60\%$ of the variance was due to traits. In a similar vein, Eid and Diener (2004) showed that items which queried participants *general* affect intensity showed a higher proportion of trait variance ($60 - 83\%$) than items which queried affect frequency over a shorter time frame ($46 - 83\%$) or current mood levels ($33 - 47\%$).

These studies show that careful design of items could realistically yield measurements in which upwards of 70 to 80% of the variability is due to true variation in stable between-person differences. Based on this, we can say that trait-like measurement instruments could be conceptualized as something in-between an equilibrium measurement and a snapshot measurement, with many instruments tools skewing closer to the former than the latter. Furthermore, a number of studies have aimed to characterize the amount of trait variance in different psychological measurement tools. For example, Eid and Diener (2004) showed that, in their sample of undergraduate students, the satisfaction with life scale (SWLS; Diener et al., 1985) exhibited $74 - 80\%$ trait variance; the Rosenberg Self-Esteem Scale (Rosenberg, 1965) $93 - 96\%$; the Life Orientation Scale (LOT; Scheier & Carver, 1985) $75 - 85\%$; and the Eysenck Personality Inventory (EPI; Eysenck, 1968) neurotisicm and extraversion scales $90 - 92\%$ and $88 - 93\%$ trait variances, respectively.

Equilibrium models from non-equilibrium cross-sectional data

Having prior knowledge about the degree of trait variance in a measurement instrument already allows us to get a clearer idea of the amount of bias that we might expect when fitting a statistical model which assumes a perfect equilibrium measurement model. For instance, for our example system, if 70 to 80% of the variance in our measurements is attributable to variance in the equilibrium positions, and so, correspondingly, 20 to 30% attributable to variance in the state, then, consulting Figure 6, we would expect only a small amount of bias to be present. In principle, however, if prior research has established how much trait variability we can expect in an instrument, then we can use this information to relax the measurement assumption stated earlier in the current paper. That is, we can use this information to, in principle, estimate the ECM from non-equilibrium measurements.

We can see how such a correction would work by returning to our running example. Suppose that we measure variables X_1 through X_4 with a questionnaire designed to capture the long-run value of the process, and that we only take measurements at a single moment in time. Further suppose that previous research is available which shows that 70% of the variance in these items is due to the trait, that is, equilibrium variability across people, and we wish to estimate the equilibrium dependencies between these variables. In order to do this, we first need to expand our SEM model, defined in Equation (10), with a *measurement component* which relates the observed variables $\boldsymbol{Y}$ to their equilibria $\boldsymbol{\mu}$, now considered latent variables. We write

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i + \boldsymbol{\omega}_i, \tag{14}$$

where $\boldsymbol{\omega}_i$ is (a vector of) deviations from the equilibrium value for each individual. In SEM terms, we have a typical measurement model consisting of a latent variable $\boldsymbol{\mu}$, an observed variable $\boldsymbol{Y}$, and what would usually be referred to as a measurement error term $\boldsymbol{\omega}$. Note that we make no distinction between variance due to errors in the measurement (for instance, mistakes in filling out the survey) and variance due to the state values, referred to $\boldsymbol{s}_{it}$ in Equation (13) (Steyer and colleagues note that LST studies often decompose measurement instruments into trait, state, and measurement error variances, see Steyer et al. 1989).

We assume without loss of generality that the variances of the observed variables $\boldsymbol{Y}$ are equal to one. To be able to estimate this model from data, we use our prior knowledge about the measurement instrument and fix the measurement error variances σ^2_ω to 0.30, and assume that the measurement errors are uncorrelated with each other. We further constrain the latent variable variances to 0.70 (see Appendix C.1 for how this can be done in standard SEM software). With these constraints in place, the model is statistically identified, and the ECM parameters can be estimated in much the same way as above.

To illustrate this, we conduct a small simulation study. Specifically, we simulate $n = 2000$ observations whose variance are 70% due to the trait and 30% due to the state, which in the model is now considered as

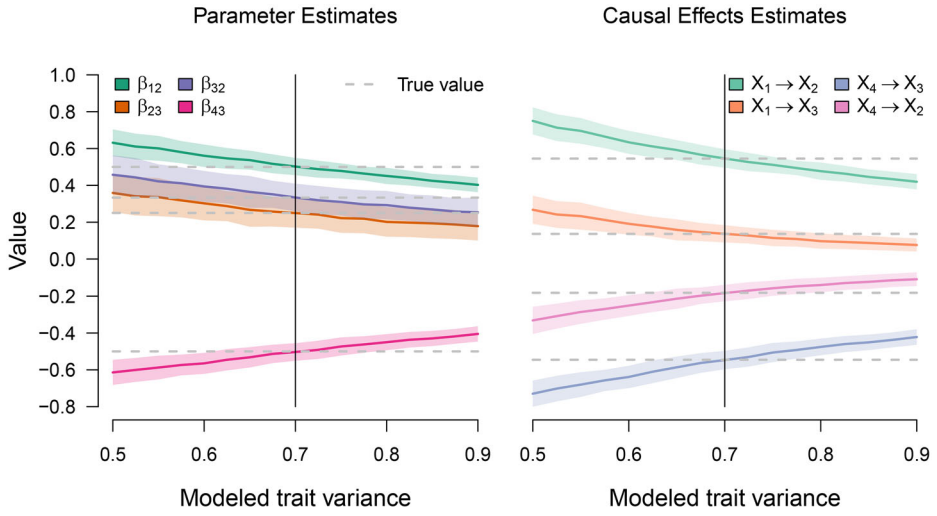

Figure 7. Mean (solid lines) and standard deviation (shaded areas) of parameter estimates (left) and causal effect estimates (right) with $n = 2000$ observations as a function of the modeled trait variance, assuming a true trait variance of 0.70 (indicated by the vertical black line). True values are shown as dashed grey lines.

error variance. We estimate the ECM by specifying the modeled trait variance to be between 50% and 90%. The left panel in Figure 7 shows the values of the estimated parameters while the right panel shows the estimates of selected causal effects. If we fix the trait variance in the model to be equal to the true trait variance in the measurements (0.70), we obtain unbiased estimates, as indicated by the vertical lines in Figure 7. As we would expect, we not only obtain unbiased estimates of the ECM parameters (left panel), but also of the causal effects (right panel). In other words, if we have sufficient knowledge about the measurement instrument employed, we can incorporate this information into our SEM model, thereby correcting for the unwanted variance component and recovering the equilibrium model parameters and causal effects.

In practice, of course, researchers may be uncertain about the degree of trait variation in their measurements. As Figure 7 shows, when we *under*-specify the true trait variance (e.g., by setting the modeled trait proportion to 0.50), then the estimated parameters and causal effects are inflated (larger in absolute value). This is because, if we assume that the measurement error is smaller than it really is, then the parameter estimates have to account for the larger remaining (co)variance. Correspondingly, *over*-specifying the true trait variance means that the estimates are shrunk toward zero because more of the (co)variance is attributed to measurement error and the parameter estimates thus do not have to be as large. In Appendix C.2 we show that additional bias may result when the measurement errors are correlated but we fail to specify this in the measurement model.

This simulation shows that when snapshot measurements of a process are available, researchers can still *potentially* estimate the ECM from data if they have sufficient prior knowledge about the degree of state vs trait variance present in a measurement instrument. When this is available, a measurement model can be specified implementing these values as fixed parameters, and the ECM model can again in principle by estimated using standard SEM methods. In this example, we can see that the degree of bias introduced through misspecifying the modeled trait variance is not too large. Of course, the larger the literature we can rely on, the more confident we can be in specifying the trait variance. In practice, if researchers wish to apply this approach, they should identify a range of plausible values for the trait variance percentage, and vary the fixed model variance as a sensitivity analysis to check the robustness of their conclusions to this type of model misspecification. In principle, this approach could be extended to include modeling of *measurement error variance* in addition to state and trait variance if researchers have access to multiple indicators of their latent variables of interest (Bollen, 1989; Hertzog & Nesselroade, 1987; Steyer et al., 2015).

Equilibrium causal models in practice

So far, we have focused on introducing the basic ideas behind Equilibrium Causal Models and showed that statistical modeling tools already familiar to psychological researchers can—in certain idealized situations—be applied to cross-sectional data to yield a

model that is informative about the long-term effects of interventions in linear dynamical systems.

One implication of our analysis above is that, when studying the type of linear system considered here, an ECM will contain *cycles* whenever there are feedback relationships in the underlying dynamical system. In our example, there exists a feedback relationship between the variables X_2 and X_3, and in the dynamical systems representation, this cyclic relationship can be *unrolled* in time, resulting in a representation that is *acyclic*: X_2 has a causal effect on X_3 at time point t, while X_3 has a causal effect on X_2 at time point $t + 1$. In the corresponding ECM there is no notion of time ordering any more, since the model describes causal relations between equilibrium positions. However, the feedback relationship between X_2 and X_3 carries over, resulting in a cycle. Thus, despite the fact that ECM is a model for cross-sectional data, we have an intuitive interpretation of cycles in this model: Cycles exist whenever the underlying dynamical system has feedback relationships. In the context of psychological research, we may expect the dynamics underlying psychological processes to consist of many such feedback loops, and so it is an advantage of ECMs that they provide a clear connection between cross-sectional statistical dependencies and feedback relations in dynamical systems.

While in the previous section we outlined the considerations that researchers must make in regards to the nature of psychological measurements, there remain a number of other challenges that are likely to arise if researchers wish to use ECMs in empirical practice. These challenges relate to practical difficulties of estimating cyclic SEM models, the problem of unobserved confounding, and the issue of learning unknown causal structures from empirical data. After outlining these challenges, we discuss how the modern causal discovery literature may help address them.

Challenges in estimating ECMs

While ECMs are a promising tool to bridge dynamical systems thinking and cross-sectional data analysis, there are at least four key challenges with applying them in practice. First, since we expect that psychological processes are typically characterized by a number of feedback relationships, we expect ECMs of psychological processes to include cyclic relationships. Unfortunately however, cyclic relationships are typically more challenging to estimate from data. The vast majority of both SEM and causal modeling applications are typically limited to considering only acyclic causal relations (DAGs; Pearl, 2009; Ryan et al., 2022). In the language of SEMs, most models are *recursive* (i.e., acyclic) rather than *nonrecursive* (i.e., cyclic). The reason for considering only acyclic causal models is partly because of practical concerns, since they generally have more convenient properties than cyclic causal models (for details, see Lauritzen et al., 1990; Bongers et al., 2021; Spirtes et al., 1995). In the SEM literature, it is well known that a necessary condition for estimating $\mathbf{\Phi}$ relates to the *invertibility* of the matrix matrix $\mathbf{I} - \mathbf{\Phi}$ with $(\mathbf{I} - \mathbf{\Phi})^{-1}$ appearing in the expression for the model-implied covariance matrix of a SEM. If $\mathbf{\Phi}$ is triangular, as is the case for acyclic models, then this condition is always met, regardless of the parameter values. This is not the case for cyclic models, where a necessary condition for this to hold relates to the values of eigenvalues of $\mathbf{\Phi}$; for instance, this will hold if the eigenvalues are smaller than one in absolute value, $|\lambda| < 1$.[10] Recall that this is in fact the same condition we needed for our linear dynamical system to be stable. Usually, this means that we require that the feedback relations in the dynamical system are not too strong such that, when iterating the system ($\mathbf{X}_{t+1} = \mathbf{\Phi}\mathbf{X}_t + \epsilon$), it reaches equilibrium (Rothenhäusler et al., 2015). From a structural standpoint, cyclic models may be statistically identified in, for instance, the presence of sufficient instrumental variables; variables that are direct causes of only one variable in the model, as in our running example the variables X_1 and X_4.

Second, even if the true parameters $\mathbf{\Phi}$ meet this condition, the model itself may not be *statistically identified*. A model is *not identified* when there are (at least) two distinct sets of parameter values that, given the same data, give rise to the same likelihood. It is known that, if we estimate only directed relationships in the structural model, that is, regressing observed variables directly on one another without any measurement model, then every acyclic model is identified. However, not all cyclic models are identified, with the *rank condition* providing a necessary and sufficient condition for a model to be identified (for details, see Bollen, 1989, pp. 98–103). So far, we have considered an example system which is relatively *sparse*: There is a feedback relationship between X_2 and X_3, but X_1 and X_4 only act as unique causes of X_2 and X_3, respectively. This model is statistically identified, but

[10]Strictly speaking, the condition is that the eigenvalues of $D\mathbf{\Phi}D^{-1}$ are smaller than 1 in absolute value, where D is a diagonal invertible "stabilization" matrix (for details, see Bongers et al., 2022, Corollary 4.22). The freedom to choose D significantly enlarges the class of stable linear dynamical systems that equilibrate to the ECM. Equation (8) in Corollary 4.23 of Bongers et al. (2022) provides a condition on $\mathbf{\Phi}$ for when a suitable "stabilization" matrix D can be found. We thank an anonymous reviewer for this clarification.

if we were to add more relationships, such as additional feedback loops $X_1 \leftrightarrows X_3$, $X_2 \leftrightarrows X_4$, and $X_1 \leftrightarrows X_4$, this would no longer be the case. The practical implication of this is that, even if we know the structure of the ECM and have observed all relevant variables, we may in some cases not be able to estimate it from data using standard approaches.

Third, unobserved confounding can obscure the estimation of causal effects. For example, we may find a statistical dependency between X_1 and X_2, yet this dependency may be explained by the common cause X_3, which we failed to include in our model (Dablander & van Bork, 2021; Pearl, 2009; Peters et al., 2017; Rohrer, 2018). Unobserved confounding is the bane of causal inference from observational data and likely the de facto situation in psychology and the social sciences more broadly. Thus, researchers who wish to interpret the SEM models they fit to empirical data as ECMs, that is, interpret estimated statistical relationships as causal effects, should be very cautious.

Fourth, while we have so far assumed that we know the *structure* of the ECM, that is, we know (or have strong theoretical expectations about) the causal relations between variables, this is generally not the case in practice. For example, we used knowledge of the ECM structure previously by specifying the specific directed relationships which should be estimated. In the causal modeling literature, this is sometimes referred to as *causal inference*: Using knowledge about the causal system at hand in order to estimate the causal effects of interest (Peters et al., 2017). The theoretical understanding of psychological phenomena may in many situations be too weak to confidently assert which variables cause (or do not cause) which other variables. In the majority of cases, the key difficulty is in correctly specifying the *structure* of the causal model, rather than just estimating the causal effects given a particular structure. This is known as *causal discovery* in the literature (Peters et al., 2017).

To illustrate the challenges of estimating ECMs from cross-sectional data in practice, Appendix D provides an empirical application of an ECM estimation approach. Recall that in the simulated examples above we have shown that an ECM estimated from equilibrium data yields the same inferences about (equilibrium) causal relations and interventions as would be obtained if the true data-generating dynamic system parameters were known. One may be tempted to infer from this that, given a time-series dataset, one would expect to obtain the same inferences about causal relations by fitting an appropriate VAR(1) model to data and deriving the implied equilibrium relations or by simply fitting the appropriate equilibrium model to the person-means of the time-series. As we have discussed above, and as we show in the empirical example, this will not generally be the case. Model identifiability, unobserved confounding, model misspecification, and measurement issues can all cause inconsistencies between models fit to equilibrium and time-series data, respectively.

In sum, while ECMs provide us with an understanding of how, hypothetically, cross-sectional data could yield insights into (cyclic) dynamic causal relations, estimating ECMs from data using SEM-based confirmatory modeling approaches familiar to social science researchers faces considerable challenges. However, as we will outline in the next section, some of these difficulties may be directly addressed by using practical tools and approaches developed in the modern causal discovery literature. As we will see, by using different research designs, and by leveraging information from different sources, these tools can in principle allow us to overcome these challenges, enabling us to learn cyclic ECMs from data.

Cyclic causal discovery

As outlined above, in practice we generally do not know the structure of the ECM, that is, which variables have direct causal effects on which other variables. When the structure of the ECM is unknown, the inference problem changes from *estimating* known (or at least hypothesized) causal relations to *discovering* the causal model itself (Peters et al., 2017; Spirtes et al., 2000). The field of causal discovery is an active area of research that has made great progress in developing tools to help researchers learn the structure of causal relations between variables from data (for recent overviews, see e.g., Eberhardt, 2017; Heinze-Deml et al., 2018; Zhang et al., 2017). As we have outlined in the introduction of this paper, in many psychological contexts researchers are specifically interested in studying systems which are characterized by reciprocal feedback relations (Borsboom, 2017; Borsboom & Cramer, 2013; Haslbeck et al., 2022), which specifically necessitates the use of *cyclic* causal discovery methods.

There are many challenges associated with learning causal relations from data (for a list, see e.g., Spirtes & Zhang, 2018). One fundamental problem of causal discovery lies in the fact that many causal models are compatible with the same set of data (MacCallum et al., 1993; Raykov & Marcoulides, 2001; Verma & Pearl, 1990). In SEM terms, we would say that there are many models which are *statistically equivalent* (Bollen, 1989; Ryan et al., 2022). For example,

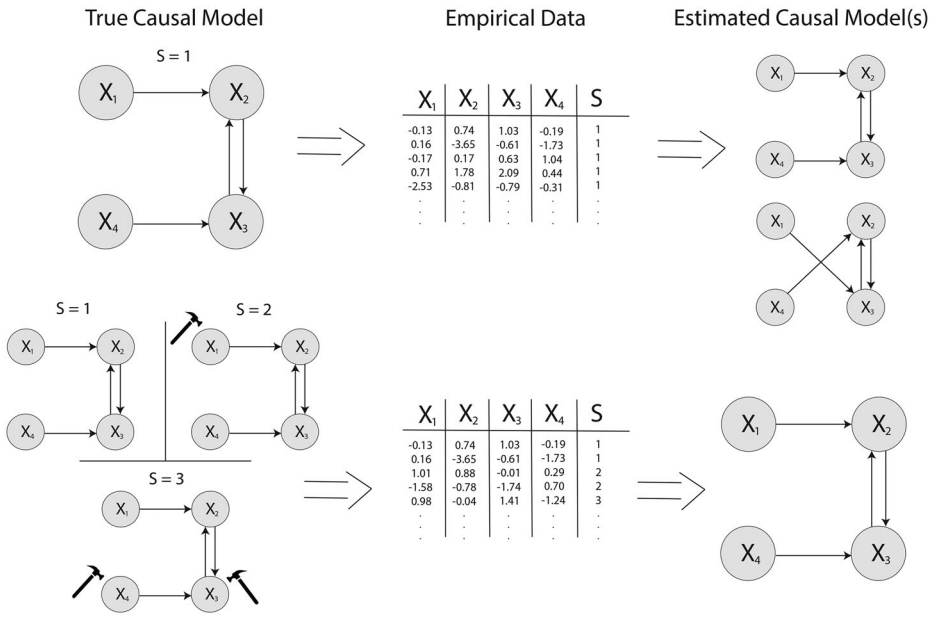

Figure 8. Top: True causal model (left) observed in one setting ($S = 1$) giving rise to empirical observations (middle). Applying a causal discovery method on only these data results in two causal models that both reproduce the statistical dependencies in the data, that is, in an equivalence class of estimated causal models (right). Bottom: Same except that the causal model is observed in three settings ($S \in \{1, 2, 3\}$). Utilizing this additional information allows one to rule out statistically equivalent models and arrive at the correct causal model (right).

suppose you observe a correlation between variables X and Y. Even if we are willing to make the simplifying assumption that there are no unobserved (latent) confounding variables, the causal models compatible with this observation are $X \rightarrow Y$, $X \leftarrow Y$, and $X \rightleftarrows Y$. Using our running $p = 4$ variable example, the top row in Figure 8 further illustrates this equivalence problem. The rightmost panel shows that there are two causal graphs that are consistent with the empirical data gathered from observing the true causal system on the left. In other words, observing only these data cannot allow us to rule out any of the estimated graphs—all of them exhibit the same set of conditional independencies found in the data.

That the causal graph often cannot be uniquely identified from data is a challenge for all causal discovery methods. So-called *constraint-based* causal discovery methods, which aim to estimate causal graphs from patterns of statistical (in)dependence in a single observational dataset, typically output an *equivalence class* of causal graphs consistent with the data at hand.[11] Although these can still be quite informative for researchers, they may be challenging to interpret in practice. For treatments of how constraint-based cyclic causal discovery methods can be applied and interpreted in psychological settings, we refer readers to Park et al. (2023) and Kossakowski et al. (2021).

A major recent insight in the causal discovery literature is that using data from different contexts can improve our ability to recover causal models (Mooij et al., 2020; Peters et al., 2016). Although a *context* can be broadly defined, the most straightforward example is when we have a mix of data from observational settings and settings where some intervention is applied to the system.[12] This is illustrated in the bottom row in Figure 8, where we now not only observe the causal system in the context $S = 1$, where no intervention takes place, but also in the contexts $S = 2$ and $S = 3$, where interventions on X_1 and $\{X_3, X_4\}$ occur, respectively. Using observations from these different contexts allows us to completely recover the true causal graph, as shown in the rightmost panel. The intuition behind this is that the set of graphs that are interventionally equivalent (i.e., imply the same statistical dependencies when we intervene on the system) is generally much smaller than the set of graphs that are observationally equivalent. Returning to our two variable example involving only X and Y, if we have data where X has been intervened on and Y

[11] Currently existing cyclic causal discovery methods do not directly return a set of equivalent graphs. Instead, they return a Partial Ancestral Graph which encodes ancestral relations and from which one can derive the Markov equivalent set of directed cyclic graphs, for example by brute force enumeration.

[12] Different contexts can also refer to observing the system at different points in time (Rothenhäusler et al., 2015, for an empirical example).

changes as a result, we know that there must be a path $X \rightarrow Y$, excluding all models where such a path does not exist.

Using modern causal discovery methods that draw on data from different contexts can potentially address the challenges outlined above. Many of these methods allow for the discovery of cyclic causal relations even in the presence of unobserved confounding (for an overview, see e.g., Mooij et al., 2020). Some of them also identify the causal structure exactly, rather than an equivalence class, as well as returning causal effects estimates. The *Backshift* (Rothenhäusler et al., 2015) method seems particularly promising as it fulfills all these requirements. Backshift assumes that shift interventions are applied in the different contexts, but the intervention targets themselves need not be known, which is important for psychological research where interventions are likely to often be "fat-hand" in nature, that is, targeting multiple variables at once (Eronen, 2020). Backshift further assumes linear relationships similar to the example system which we have used throughout the current paper. Note that since we assume that the causal structure is the same across the different settings, the data need not be collected from the same set of individuals. However, if the individuals come from different populations for which the underlying causal structure is in fact different, Backshift's performance will be negatively impacted.

The main disadvantages of Backshift are a) that it is limited to considering shift interventions, and b) that it requires at least three different settings, where one setting can be entirely observational. The first assumption can potentially be relaxed by switching to conceptually similar multiple-context causal discovery methods which can handle a larger variety of interventions, such as JCI-FCI (Mooij et al., 2020). However, as many psychological interventions are likely to be "soft" in nature, meaning that the value of the variable is not completely determined by the intervention, as in a shift intervention (Campbell, 2007; Eronen, 2020), we consider this to be a potentially acceptable tradeoff for the Backshift method in the context of psychological research, especially as it handles linear systems and outputs interpretable weighted causal graphs, unlike more general but related methods. The second disadvantage of the Backshift method, that data be drawn from multiple (intervention-based) settings, is more fundamental, and shared across all such so-called *invariance-based* methods. We provide a small simulation study in Appendix E that assesses the performance of Backshift for our example system. Our results show that the number of settings or contexts available to researchers is more important than the sample size per setting. This implies that researchers who are interested in estimating ECMs, should try to maximize the settings in which the system is observed rather than the sample size per se. This would constitute a shift in how data is currently collected in psychology, away from a focus on simply collecting more data in an absolute sense, and toward a focus of observing the system in different contexts. In Appendix F we provide an empirical example (Blanken et al. using data from 2019), of how the Backshift method can be applied, and what the output of this method is, when data from multiple contexts is available.

Discussion

In this paper, we introduced the concept of Equilibrium Causal Models to the psychological literature. While Equilibrium Causal Models have been studied before (Bongers et al., 2022; Dash, 2005; Iwasaki & Simon, 1994; Spirtes, 1995; Strotz & Wold, 1960; Weinberger, 2020, 2021)—forming the target of inference in many cyclic causal discovery methods (Bongers et al., 2021; Lacerda et al., 2012; Mooij et al., 2013; 2020; Mooij & Claassen, 2020; Richardson, 1996; Rothenhäusler et al., 2015)—they are virtually unknown in psychology. Using the example of a linear dynamical system, we showed that ECMs provide a missing link that connects cross-sectional data analysis with dynamical systems modeling. We demonstrated that ECMs (a) can yield insights into the long-term effects of different interventions and (b) can, under certain conditions, be estimated from cross-sectional data. Focusing on linear systems, we showed how ECMs can be estimated using standard SEM software when the structure of causal relations is known. In case this structure is not known, as is generally the case in practice, we also showed how ECMs can be estimated using modern causal discovery techniques.

Our analysis of ECMs made three key simplifying assumptions. First, we limited our analysis to the case of linear dynamical systems that are stable in both the observational and intervention setting. Linearity implies (assuming that all eigenvalues are nonzero) that the system has a single global equilibrium. This rules out the possibility of multiple equilibria, including the fact that a small intervention can have a large effect on the dynamics of the system (Dablander et al., 2023; van der Maas et al., 2020). ECMs and causal discovery methods for more general (nonlinear) systems—including systems whose equilibrium

depends on the initial condition—are an active area of research (Bongers et al., 2022; Mooij et al., 2013). Second, we assumed that individuals exhibit only limited heterogeneity, expressed in our model as differing with respect to the intercepts but not the lagged relationships. In psychological settings, this assumption may be overly strict, and further research is needed to investigate to what extent this assumption can be relaxed. Third, we assumed that psychological measurements either yield equilibrium positions directly, or that sufficient knowledge about the properties of measurement instruments was available. We showed that this knowledge could potentially be gleaned from the psychological measurement literature on latent-state-trait variance decompositions, and demonstrated how this can be used to correct for deviations from the equilibrium using standard SEM software. In practice, however, this information may not be available for all measurement instruments and may also differ for different populations. Furthermore, if measurement instruments yield correlated errors, then knowledge of these correlations may be required to fully correct for imperfect equilibrium measurements. This highlights the need to build on the existing literature with detailed and thorough studies on psychological measurement.

The ECM perspective put forward in this paper has a number of implications for empirical researchers. First, ECMs connect cross-sectional data analysis to dynamical systems modeling, providing a new perspective on the ergodicity debate in psychological research. Specifically, if the assumptions outlined above hold, then between-person data can be informative about causal relations present in within-person processes. Importantly, ECMs can include cyclic causal relations and—at least for the linear case discussed here—give them a straightforward interpretation: A cyclic relationship exists in the ECM if there is a feedback relationship in the underlying dynamical system. Of course, for more complicated dynamical systems, the mapping between equilibrium and dynamic causal dependencies may be less straightforward (Blom et al., 2020; Dash, 2005; Dash & Druzdzel, 2001; Weinberger, 2023). However, pursuing equilibrium causal dependencies may be a fruitful avenue for research, both for its own end and as a way of constraining the space of possible dynamical models which may underlie those equilibrium relationships.

Second, researchers interested in estimating ECMs can connect to a large array of tools developed in the field of causal discovery. While there are important statistical and conceptual challenges to causal discovery (see e.g., Eronen, 2020; Spirtes & Zhang, 2018), we have seen that utilizing observations from multiple contexts can substantially improve performance (Mooij et al., 2020). This draws from the fact that causal relations are relations that should be *invariant* across settings (Bühlmann, 2020), suggesting a potential shift in the way psychological data is collected. Specifically, rather than solely focusing on increasing sample size, researchers may wish to increase the number of settings in which the psychological system is observed. Several causal discovery methods, including the one we have focused on in this paper, do not require precise knowledge of which variables were intervened on, nor do they assume that all causally relevant variables have been observed. These constitute important advances that psychological researchers can benefit from in practice. For causal discovery methods to become more widely applied in psychology, however, more extensive research must be directed into several issues. Initially, we may need to develop or adapt existing methods to deal with the presence of measurement error (Blom et al., 2018; Saeed et al., 2020; Zhang et al., 2017). Following the development of fit-for-purpose methods, we need studies investigating how well these methods work for settings common to psychology, varying effect sizes, the density of the causal graph, the number of nodes, the sample size, the extent and type of measurement error, and the type and target of interventions in a more systematic manner. Once certain methods have shown promise in simulation, they can be applied to empirical data. The results of these analyses can then be probed for sensibility—do certain causal relations make sense given the existing literature? Would experts on particular systems agree with the estimated relationships? If the face validity of the results has been sufficiently established, these methods can be used to make predictions about the outcome of particular interventions, resulting in the strongest test of their usefulness. Some initial promising work has been done on cyclic causal models in psychology (Kossakowski et al., 2019; 2021), but we believe that there are many more research avenues and opportunities for the psychological research community to pursue.

Throughout this paper we have assumed that psychological systems can be fruitfully described using the language of dynamical systems theory and causal modeling. However, some concepts in dynamical systems theory and causal modeling may not map onto psychological systems in an obvious way. Although the

equilibrium of a system of differential equations can be clearly defined, the notion of a resting state for a psychological process is less clear. For example, one's stress might remain constant within a week, but may be changing when viewed on a monthly timescale, and so the notion of an equilibrium and long-term versus short-term effects are inherently tied up with the timescale under consideration. We defined causal effects on an abstract level using the concepts of press, pulse, and shift interventions, but these may not map neatly onto real-world interventions. We should remember to view all models, including the ECMs we describe in the current paper, as simplifications and abstractions of real world systems. Ultimately, their practical utility must be assessed by testing whether the causal predictions they yield are accurate.

Conclusion

Equilibrium Causal Models can yield insights into the long-term effects of different interventions and can, under certain conditions, be estimated from cross-sectional data. They help us understand how cross-sectional data can be used to learn about within-person processes. Causal discovery methods can be used to estimate these models, but they require observations from the system under different settings. This would represent a shift in psychological data collection. There is ample room for future research—both conceptual and empirical—to help establish Equilibrium Causal Models as a valuable tool in the psychologist's toolbox.

Article information

Conflict of Interest Disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical Principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: This work was partially supported by a Consolidator grant (nr.865468) from the European Research Council.

Role of the Funders/Sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

References

Alegria, M., Jackson, J. S., Kessler, R. C., & Takeuchi, D. (2007). *Collaborative psychiatric epidemiology surveys (CPES), 2001–2003 [United States]*. Inter-university Consortium for Political and Social Research.

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 359–388. https://doi.org/10.1080/10705511.2017.1406803

Assaad, C. K., Devijver, E., & Gaussier, E. (2022). Survey and evaluation of causal discovery methods for time series. *Journal of Artificial Intelligence Research*, *73*, 767–819. https://doi.org/10.1613/jair.1.13428

Augustine, A. A., & Larsen, R. J. (2012). Is a trait really the mean of states? Similarities and differences between traditional and aggregate assessments of personality. *Journal of Individual Differences*, *33*(3), 131–137. https://doi.org/10.1027/1614-0001/a000083

Barrett, L. F. (1997). The relationships among momentary emotion experiences, personality descriptions, and retrospective ratings of emotion. *Personality and Social Psychology Bulletin*, *23*(10), 1100–1110. https://doi.org/10.1177/01461672972310010

Beck, A. T., Steer, R. A., Brown, G. K., et al. (1987). *Beck depression inventory*. Harcourt Brace Jovanovich New York.

Bender, E. A., Case, T. J., & Gilpin, M. E. (1984). Perturbation experiments in community ecology: Theory and practice. *Ecology*, *65*(1), 1–13. https://doi.org/10.2307/1939452

Blanken, T. F., Van Der Zweerde, T., Van Straten, A., Van Someren, E. J., Borsboom, D., & Lancee, J. (2019). Introducing network intervention analysis to investigate sequential, symptom-specific treatment effects: A demonstration in co-occurring insomnia and depression. *Psychotherapy and Psychosomatics*, *88*(1), 52–54. https://doi.org/10.1159/000495045

Blom, T., Bongers, S., & Mooij, J. M. (2020). Beyond structural causal models: Causal constraints models. In *Uncertainty in Artificial Intelligence* (pp. 585–594). PMLR.

Blom, T., Klimovskaia, A., Magliacane, S., & Mooij, J. M. (2018). An upper bound for random measurement error in causal discovery. *arXiv preprint* (arXiv:1810.07973).

Bollen, K. A. (1989). *Structural equations with latent variables* (Vol. 210). John Wiley & Sons.
Bongers, S., Blom, T., & Mooij, J. (2022). Causal modeling of dynamical systems. *arXiv preprint* (arXiv:1803.08784).
Bongers, S., Forré, P., Peters, J., & Mooij, J. M. (2021). Foundations of structural causal models with cycles and latent variables. *The Annals of Statistics*, *49*(5), 2885–2915. https://doi.org/10.1214/21-AOS2064
Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13. https://doi.org/10.1002/wps.20375
Borsboom, D., & Cramer, A. O. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*(1), 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608
Bos, F. M., Snippe, E., de Vos, S., Hartmann, J. A., Simons, C. J., van der Krieke, L., de Jonge, P., & Wichers, M. (2017). Can we jump from cross-sectional to dynamic interpretations of networks implications for the network perspective in psychiatry. *Psychotherapy and Psychosomatics*, *86*(3), 175–177. https://doi.org/10.1159/000453583
Bower, G. H. (1981). Mood and memory. *The American Psychologist*, *36*(2), 129–148. https://doi.org/10.1037/0003-066x.36.2.129
Braun, L., Göllner, R., Rieger, S., Trautwein, U., & Spengler, M. (2021). How state and trait versions of self-esteem and depressive symptoms affect their interplay: A longitudinal experimental investigation. *Journal of Personality and Social Psychology*, *120*(1), 206–225. https://doi.org/10.1037/pspp0000295
Briganti, G., Scutari, M., & McNally, R. J. (2022). A tutorial on Bayesian networks for psychopathology researchers. *Psychological Methods*, *28*(4), 947–961. https://doi.org/10.1037/met0000479.
Bringmann, L. F., Pe, M. L., Vissers, N., Ceulemans, E., Borsboom, D., Vanpaemel, W., Tuerlinckx, F., & Kuppens, P. (2016). Assessing temporal emotion dynamics using networks. *Assessment*, *23*(4), 425–435. https://doi.org/10.1177/1073191116645909
Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., & Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PloS One*, *8*(4), e60188. https://doi.org/10.1371/journal.pone.0060188
Brose, A., Lindenberger, U., & Schmiedek, F. (2013). Affective states contribute to trait reports of affective well-being. *Emotion (Washington, D.C.)*, *13*(5), 940–948. https://doi.org/10.1037/a0032401
Bühlmann, P. (2020). Invariance, causality and robustness. *Statistical Science*, *35*(3), 404–426.
Campbell, J. (2007). An Interventionist Approach to Causation in Psychology. In A. Gopnik and L. Schulz (eds.), *Causal Learning: Psychology and Computation*. Oxford: Oxford University Press.
Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, *100*(3), 316–336. https://doi.org/10.1037/0021-843x.100.3.316
Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*(11), 671–684. https://doi.org/10.1037/h0043943
Dablander, F. (2020). A gentle introduction to dynamical systems theory.
Dablander, F., Pichler, A., Cika, A., & Bacilieri, A. (2023). Anticipating critical transitions in psychological systems using early warning signals: Theoretical and practical considerations. *Psychological Methods*, *28*(4), 765–790. https://doi.org/10.1037/met0000450
Dablander, F., & van Bork, R. (2021). Causal inference. In A. Isvoranu, S. Epskamp, & D. Borsboom (Eds.), *Network psychometrics* (Chapter 12, pp. 190–210). Oxford University Press.
Dash, D. (2005). *Restructuring dynamic causal systems in equilibrium* [Paper presentation]. Aistats. Citeseer.
Dash, D., & Druzdzel, M. (2001). Caveats for causal reasoning with equilibrium models. In European conference on symbolic and quantitative approaches to reasoning and uncertainty (pp. 192–203). Springer.
Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13
Diener, E., Northcott, R., Zyphur, M. J., & West, S. G. (2022). Beyond experiments. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *17*(4), 1101–1119. https://doi.org/10.1177/17456916211037670
Driver, C. C., Oud, J. H., & Voelkle, M. C. (2017). Continuous time structural equation modeling with r package ctsem. *Journal of Statistical Software*, *77*(5), 1–35. https://doi.org/10.18637/jss.v077.i05
Eberhardt, F. (2017). Introduction to the foundations of causal discovery. *International Journal of Data Science and Analytics*, *3*(2), 81–91. https://doi.org/10.1007/s41060-016-0038-6
Eberhardt, F., & Scheines, R. (2007). Interventions and causal inference. *Philosophy of Science*, *74*(5), 981–995. https://doi.org/10.1086/525638
Eid, M., & Diener, E. (2004). Global judgments of subjective well-being: Situational variability and long-term stability. *Social Indicators Research*, *65*(3), 245–277. https://doi.org/10.1023/B:SOCI.0000003801.89195.bc
Epskamp, S. (2020). Psychometric network models from time-series and panel data. *Psychometrika*, *85*(1), 206–231. https://doi.org/10.1007/s11336-020-09697-3
Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The Gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, *53*(4), 453–480. https://doi.org/10.1080/00273171.2018.1454823
Eronen, M. I. (2020). Causal discovery and the problem of psychological interventions. *New Ideas in Psychology*, *59*, 100785. https://doi.org/10.1016/j.newideapsych.2020.100785
Eysenck, H. (1968). *The manual of the Eysenck-personality inventory*. Educational and Industrial Testing Service.
Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(27), E6106–E6115. https://doi.org/10.1073/pnas.1711978115
Foster, E. M. (2010). Causal inference and developmental psychology. *Developmental Psychology*, *46*(6), 1454–1480. https://doi.org/10.1037/a0020204
Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., & Borsboom, D. (2016). What are "good" depression

symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders*, *189*, 314–320. https://doi.org/10.1016/j.jad.2015.09.005

Gische, C., & Voelkle, M. C. (2022). Beyond the mean: A flexible framework for studying causal effects using linear models. *Psychometrika*, *87*(3), 868–901. https://doi.org/10.1007/s11336-021-09811-z

Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *15*(5), 1243–1255. https://doi.org/10.1177/1745691620921521

Hamaker, E. (2022). The curious case of the cross-sectional correlation. *Multivariate Behavioral Research*, *59*(6), 1111–1122.

Hamaker, E. L. (2012). Why researchers should think "within-person": A paradigmatic rationale. In M. R. Mehl, & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). The Guilford Press.

Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, *26*(1), 10–15. https://doi.org/10.1177/0963721416666518

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. (2005). Statistical modeling of the individual: Rationale and application of multivariate stationary time series analysis. *Multivariate Behavioral Research*, *40*(2), 207–233. https://doi.org/10.1207/s15327906mbr4002_3

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. https://doi.org/10.1037/a0038889

Hamilton, J. D. (1994). *Time series analysis* (Vol. 2). Princeton university press Princeton.

Haslbeck, J., Ryan, O., Robinaugh, D. J., Waldorp, L. J., & Borsboom, D. (2022). Modeling psychopathology: From data models to formal theories. *Psychological Methods*, *27*(6), 930–957. https://doi.org/10.1037/met0000303

Heinze-Deml, C., Maathuis, M. H., & Meinshausen, N. (2018). Causal structure learning. *Annual Review of Statistics and Its Application*, *5*(1), 371–391. https://doi.org/10.1146/annurev-statistics-031017-100630

Hertzog, C., & Nesselroade, J. R. (1987). Beyond autoregressive models: Some implications of the trait-state distinction for the structural modeling of developmental change. *Child Development*, *58*(1), 93–109. pageshttps://doi.org/10.2307/1130294

Hofmann, S. G., Curtiss, J. E., & Hayes, S. C. (2020). Beyond linear mediation: Toward a dynamic network approach to study treatment processes. *Clinical Psychology Review*, *76*, 101824. https://doi.org/10.1016/j.cpr.2020.101824

Hyttinen, A., Eberhardt, F., & Hoyer, P. O. (2012). Learning linear cyclic causal models with latent variables. *The Journal of Machine Learning Research*, *13*(1), 3387–3439.

Iwasaki, Y., & Simon, H. A. (1994). Causality and model abstraction. *Artificial Intelligence*, *67*(1), 143–194. https://doi.org/10.1016/0004-3702(94)90014-0

Kendler, K. S., Aggen, S. H., Flint, J., Borsboom, D., & Fried, E. I. (2018). The centrality of DSM and non-DSM depressive symptoms in Han Chinese women with major depression. *Journal of Affective Disorders*, *227*, 739–744. https://doi.org/10.1016/j.jad.2017.11.032

Kossakowski, J. J., Waldorp, L. J., & van der Maas, H. L. (2021). The search for causality: A comparison of different techniques for causal inference graphs. *Psychological Methods*, *26*(6), 719–742. https://doi.org/10.1037/met0000390

Kossakowski, J., Oudheusden, L. J., McNally, R. J., Riemann, B. C., Waldorp, L., & van der Maas, H. L. (2019). Introducing the causal graph approach to psychopathology: An illustration in patients with obsessive-compulsive disorder. https://doi.org/10.31234/osf.io/ed2v5

Kuiper, R. M., & Ryan, O. (2018). Drawing conclusions from cross-lagged relationships: Re-considering the role of the time-interval. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(5), 809–823. https://doi.org/10.1080/10705511.2018.1431046

Lacerda, G., Spirtes, P. L., Ramsey, J., & Hoyer, P. O. (2012). Discovering cyclic causal models by independent components analysis. *arXiv preprint* (arXiv:1206.3273).

Lauritzen, S. L., Dawid, A. P., Larsen, B. N., & Leimer, H.-G. (1990). Independence properties of directed Markov fields. *Networks*, *20*(5), 491–505. https://doi.org/10.1002/net.3230200503

Leertouwer, I., Schuurman, N. K., & Vermunt, J. (2021). Are retrospective assessments means of people's experiences? Accounting for interpersonal and intrapersonal variability when comparing retrospective assessment data to ecological momentary assessment data. *Journal for Person-Oriented Research*, *8*(2), 52.

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, *114*(1), 185–199. https://doi.org/10.1037/0033-2909.114.1.185

McNeish, D., & MacKinnon, D. P. (2025). Intensive longitudinal mediation in mplus. *Psychological Methods*, *30*(2), 393–415. https://doi.org/10.1037/met0000536

Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *72*(4), 417–473. https://doi.org/10.1111/j.1467-9868.2010.00740.x

Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*(4), 201–218.

Mooij, J. M., & Claassen, T. (2020). *Constraint-based causal discovery using partial ancestral graphs in the presence of cycles* [Paper presentation]. In *Conference on Uncertainty in Artificial Intelligence* (pp. 1159–1168). PMLR.

Mooij, J. M., Janzing, D., & Schölkopf, B. (2013). *From ordinary differential equations to structural causal models: The deterministic case* [Paper presentation]. In *Uncertainty in Artificial Intelligence* (pp. 440–449).

Mooij, J. M., Magliacane, S., & Claassen, T. (2020). Joint causal inference from multiple contexts. *Journal of Machine Learning Research*, *21*(99), 1–108.

Park, K., Waldorp, L. J., & Ryan, O. (2024). Discovering cyclic causal models in psychological research.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.

Peters, J., Bühlmann, P., & Meinshausen, N. (2016). Causal inference by using invariant prediction: Identification and

confidence intervals. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *78*(5), 947–1012. https://doi.org/10.1111/rssb.12167

Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference: Foundations and learning algorithms*. MIT press.

Pfister, N., Bühlmann, P., & Peters, J. (2019). Invariant causal prediction for sequential data. *Journal of the American Statistical Association*, *114*(527), 1264–1276. https://doi.org/10.1080/01621459.2018.1491403

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Raykov, T., & Marcoulides, G. A. (2001). Can there be infinitely many models equivalent to a given covariance structure model? *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(1), 142–149. https://doi.org/10.1207/S15328007SEM0801_8

Richardson, T. (1996). A discovery algorithm for directed cyclic graphs. In *Proceedings of the twelfth international conference on uncertainty in Artificial Intelligence* (pp. 454–461). Morgan Kaufmann Publishers Inc.

Robinaugh, D. J., Haslbeck, J. M., Ryan, O., Fried, E. I., & Waldorp, L. J. (2021). Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *16*(4), 725–743. https://doi.org/10.1177/1745691620974697

Robinaugh, D. J., Hoekstra, R. H., Toner, E. R., & Borsboom, D. (2020). The network approach to psychopathology: A review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine*, *50*(3), 353–366. https://doi.org/10.1017/S0033291719003404

Robinaugh, D., Haslbeck, J., Waldorp, L., Kossakowski, J., Fried, E. I., Millner, A., McNally, R. J., van Nes, E. H., Scheffer, M., Kendler, K. S., & Borsboom, D. (2024). Advancing the network theory of mental disorders: A computational model of panic disorder. *Psychological Review*, *131*(6), 1482–1508.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, *1*(1), 27–42. https://doi.org/10.1177/2515245917745629

Rosenberg, M. (1965). Rosenberg self-esteem scale (rse). *Acceptance and Commitment Therapy. Measures Package*, *61*(52), 18.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rothenhäusler, D., Heinze, C., Peters, J., & Meinshausen, N. (2015). BACKSHIFT: Learning causal cyclic graphs from unknown shift interventions. In Advances in *neural information processing systems* (pp. 1513–1521).

Rowland, Z., & Wenzel, M. (2020). Mindfulness and affect-network density: Does mindfulness facilitate disengagement from affective experiences in daily life? *Mindfulness*, *11*(5), 1253–1266. https://doi.org/10.1007/s12671-020-01335-4

Ryan, O., & Hamaker, E. L. (2022). Time to intervene: A continuous-time approach to network analysis and centrality. *Psychometrika*, *87*(1), 214–252. https://doi.org/10.1007/s11336-021-09767-0

Ryan, O., Bringmann, L. F., & Schuurman, N. K. (2022). The challenge of generating causal hypotheses using network models. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*(6), 953–970. https://doi.org/10.1080/10705511.2022.2056039

Ryan, O., Kuiper, R. M., & Hamaker, E. L. (2018). A continuous-time approach to intensive longitudinal data: What, why, and how? In *Continuous time modeling in the behavioral and related sciences* (pp. 27–54). Springer.

Saeed, B., Belyaeva, A., Wang, Y., & Uhler, C. (2020). Anchored causal inference in the presence of measurement error. In J. Peters, & D. Sontag (Eds.), *Proceedings of the 36th conference on uncertainty in Artificial Intelligence (UAI)*. Vol. 124 of *Proceedings of Machine Learning Research* (pp. 619–628). PMLR.

Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology*, *4*(3), 219–247. https://doi.org/10.1037/0278-6133.4.3.219

Scherer, E. A., Metcalf, S. A., Whicker, C. L., Bartels, S. M., Grabinski, M., Kim, S. J., Sweeney, M. A., Lemley, S. M., Lavoie, H., Xie, H., Bissett, P. G., Dallery, J., Kiernan, M., Lowe, M. R., Onken, L., Prochaska, J. J., Stoeckel, L. E., Poldrack, R. A., MacKinnon, D. P., & Marsch, L. A. (2022). Momentary influences on self-regulation in two populations with health risk behaviors: Adults who smoke and adults who are overweight and have binge-eating disorder. *Frontiers in Digital Health*, *4*, 798895. https://doi.org/10.3389/fdgth.2022.798895

Schmittmann, V. D., Cramer, A. O., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology*, *31*(1), 43–53. https://doi.org/10.1016/j.newideapsych.2011.02.007

Schölkopf, B., & von Kügelgen, J. (2022). From statistical to causal learning. *arXiv preprint* (arXiv:2204.00607).

Scholkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., & Bengio, Y. (2021). Toward causal representation learning. *Proceedings of the IEEE*, *109*(5), 612–634. https://doi.org/10.1109/JPROC.2021.3058954

Schuurman, N. K. (2023). A "within/between problem" primer: About (not) separating within-person variance and between-person variance in psychology.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, *45*(3), 513–523. https://doi.org/10.1037/0022-3514.45.3.513

Shamsollahi, A., Zyphur, M. J., & Ozkok, O. (2022). Long-run effects in dynamic systems: New tools for cross-lagged panel models. *Organizational Research Methods*, *25*(3), 435–458. https://doi.org/10.1177/1094428121993228

Speilberger, C. D., Gorsuch, R., Lushene, R., Vagg, P., & Jacobs, G. (1983). *Manual for the state-trait anxiety inventory*. Consulting Psychologists.

Spirtes, P. (1995). Directed cyclic graphical representations of feedback models. In *Proceedings of the eleventh conference on uncertainty in Artificial Intelligence* (pp. 491–498). Morgan Kaufmann Publishers Inc.

Spirtes, P. and Zhang, K. (2018). Search for causal models. In M. Maathuis, M. Drton, S. Lauritzen, & M.

Wainwright (Eds.), *Handbook of graphical models* (Chapter 18, pp. 439–470). CRC Press.
Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search*. MIT Press.
Spirtes, P., Meek, C., Richardson, T. (1995). Causal inference in the presence of latent variables and selection bias. In *Proceedings of the eleventh conference on uncertainty in Artificial Intelligence* (pp. 499–506).
Steyer, R., Majcen, A.-M., Schwenkmezger, P., & Buchner, A. (1989). A latent state-trait anxiety model and its application to determine consistency and specificity coefficients. *Anxiety Research*, *1*(4), 281–299. https://doi.org/10.1080/08917778908248726
Steyer, R., Mayer, A., Geiser, C., & Cole, D. A. (2015). A theory of states and traits—Revised. *Annual Review of Clinical Psychology*, *11*(1), 71–98. https://doi.org/10.1146/annurev-clinpsy-032813-153719
Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state–trait theory and research in personality and individual differences. *European Journal of Personality*, *13*(5), 389–408. https://doi.org/10.1002/(SICI)1099-0984(199909/10)13:5<389::AID-PER361>3.0.CO;2-A
Strogatz, S. H. (2014). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering (studies in nonlinearity)* (2rd ed.). Westview Press.
Strotz, R. H., & Wold, H. O. (1960). Recursive vs. nonrecursive systems: An attempt at synthesis (part I of a triptych on causal chain systems). *Econometrica*, *28*(2), 417–427. https://doi.org/10.2307/1907731
van der Maas, H. L., Dalege, J., & Waldorp, L. (2020). The polarization within and across individuals: The hierarchical Ising opinion model. *Journal of Complex Networks*, *8*(2), cnaa010. https://doi.org/10.1093/comnet/cnaa010
van der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. https://doi.org/10.1037/0033-295X.113.4.842
van Rooij, I. (2022). Psychological models and their distractors. *Nature Reviews Psychology*, *1*(3), 127–128. https://doi.org/10.1038/s44159-022-00031-5
Vanhasbroeck, N., Ariens, S., Tuerlinckx, F., & Loossens, T. (2021). Computational models for affect dynamics. In Christian E. Waugh, Peter Kuppens, *Affect dynamics* (pp. 213–260). Springer.
Verma, T., Pearl, J. (1990). Equivalence and synthesis of causal models. In *Proceedings of the sixth annual conference on uncertainty in Artificial Intelligence* (pp. 255–270).
von Klipstein, L., Borsboom, D., & Arntz, A. (2021). The exploratory value of cross-sectional partial correlation networks: Predicting relationships between change trajectories in borderline personality disorder. *PloS One*, *16*(7), e0254496. https://doi.org/10.1371/journal.pone.0254496
Watson, D., & Clark, L. A., et al. (1994). *Manual for the positive and negative affect schedule-expanded form* [Unpublished manuscript]. University of Iowa.
Weinberger, N. (2020). Near-decomposability and the time-scale relativity of causal representations. *Philosophy of Science*, *87*(5), 841–856. https://doi.org/10.1086/710760
Weinberger, N. (2023). Intervening and letting go: On the adequacy of equilibrium causal models. *Erkenntnis*, *88*(6), 2467–2491. https://doi.org/10.1007/s10670-021-00463-0
Wichers, M. (2014). The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *Psychological Medicine*, *44*(7), 1349–1360. https://doi.org/10.1017/S0033291713001979
Zhang, K., Gong, M., Ramsey, J., Batmanghelich, K., Spirtes, P., & Glymour, C. (2017). Causal discovery in the presence of measurement error: Identifiability conditions. *arXiv preprint* (arXiv:1706.03768).

Appendix A: Interventions in linear dynamical systems

In this section, we formally define the interventions considered in the main text and the effects those interventions have on different properties of the system. Recall that the equilibrium positions are given by

$$\boldsymbol{\mu} = (\mathbf{I} - \boldsymbol{\Phi})^{-1}\boldsymbol{c}, \tag{15}$$

where $\boldsymbol{\Phi}$ is the matrix of lagged effects in the underlying dynamical system and $\boldsymbol{c}$ are the time-invariant intercepts. Press and shift interventions modify Equation (15), reflecting their long-term effects, while pulse interventions do not cause long-lasting changes in the equilibrium positions. We turn to them first.

A.1. Pulse interventions

Pulse interventions represent surgical interventions whereby the target variable X at a certain point in time $t = \tau$ is forced to obtain a particular value. In the causal modeling literature this operation would be referred to as a *do-intervention* at a single point in time (Pearl, 2009). Consider a pulse intervention which acts by setting the variable $X_{j,\tau} := z$ at a particular point in time τ. Let $\mathbf{Z}_\tau$ represent a $p \times 1$ vector with values $Z_{i,\tau} = X_{i,\tau}$ and $Z_{j,\tau} = z$, again at a particular point in time τ. The effect of the pulse intervention on the values of other variables in the system at the next time point can be expressed as

$$\mathbb{E}[\mathbf{X}_{\tau+1}] = \mathbf{c} + \boldsymbol{\Phi}\mathbf{Z}_\tau. \tag{16}$$

To compute the effect of the same pulse intervention at a longer timescale $t + s$ we can simply take the appropriate power of $\boldsymbol{\Phi}$, yielding

$$\mathbb{E}[\mathbf{X}_{\tau+s}] = \mathbf{c} + (\boldsymbol{\Phi})^s\mathbf{Z}_\tau. \tag{17}$$

As outlined in the main text, if the system is stationary, then $\boldsymbol{\Phi}$ has eigenvalues $|\lambda| < 1$. This implies that as $s \to \infty$, we have that $(\boldsymbol{\Phi})^s \to 0$, meaning that the effect of the pulse intervention eventually disappears from the system at a long enough timescale. We can also see that the pulse intervention does not alter the lagged relationships $\boldsymbol{\Phi}$, and that the expression for the equilibrium position does not depend on $\mathbf{X}_\tau$. As such, we can say that the pulse intervention has no long-term effect on the system.

A.2. Press intervention

Press interventions represent surgical interventions in our system whereby the target variable is made independent of its causes and is forced to obtain a constant value over a window of time. As such, press interventions have a natural interpretation in terms of the *do-operator*, essentially being a do-operation applied at every moment in time. Formally, a press intervention replaces the causal effects on X with a constant a. Consider a press intervention which acts on the variable X_k such that $X_k := a$. Let $\boldsymbol{P}_k$ be a $p \times p$ matrix with zeros on the off-diagonals and the k^{th} diagonal, and ones as the other diagonal elements. The equilibrium positions which are produced as a result of the press intervention can be expressed as

$$\boldsymbol{\mu} = (\mathbf{I} - \boldsymbol{P}_k\boldsymbol{\Phi})^{-1}(\boldsymbol{P}_k\boldsymbol{c} + \boldsymbol{a}_k), \tag{18}$$

where $\boldsymbol{a}_k$ is a $p \times 1$ column vector with k^{th} element a, the value that the intervened-on variable is set to by intervention, and the other elements of $\boldsymbol{a}_k$ are zero. The term $\boldsymbol{P}_k\boldsymbol{c}$ represents setting the time-invariant incoming forces acting on X_k, represented by the intercept, to zero and adding $\boldsymbol{a}_k$ enforces that the causal effect on X_k is the constant a. Focusing on the left term, note that pre-multiplying the parameter matrix Φ by $\boldsymbol{P}_k$ cuts all incoming ties to X_k by setting the k^{th} row of $\boldsymbol{\Phi}$ to zero.

Press interventions can change the stability of the system, potentially making a previously stable system unstable, since the eigenvalues of $\boldsymbol{\Phi}$ will typically not be equal to the eigenvalues of the intervened system $\boldsymbol{P}_k\boldsymbol{\Phi}$. Hyttinen et al. (2012) use the term 'asymptotic stability' to refer to systems which are both stable in the observational setting and under any press intervention.

A.3. Shift intervention

In contrast to press interventions, shift interventions represent 'soft' interventions in the system. Unlike with press interventions, we do not cut off all incoming ties to the target variable, and so we do not force the variable to obtain a specific value. Instead, a shift intervention can be considered as an intervention on the time-invariant causal forces acting on the system—an intervention on the intercept terms directly. In terms of the model equations, the shift intervention involves adding a constant s to the intercept term $\boldsymbol{c}$. Let $\boldsymbol{s}$ represent the p-dimensional column vector of shift values: If we apply a shift intervention to X_k, then $\boldsymbol{s}$ has the k^{th} element equal to s and all other elements zero. The equilibrium positions produced by a shift intervention can be expressed as

$$\boldsymbol{\mu} = (\mathbf{I} - \boldsymbol{\Phi})^{-1}(\boldsymbol{c} + \boldsymbol{s}) \tag{19}$$

From this expression it is clear that the shift intervention has no effect on the stability of the system, but does change the equilibrium positions.

Appendix B: Canonical model form and model equivalence

In this section, we study the relationships between the underlying dynamical system and the corresponding Equilibrium Causal Model, drawing heavily on the results by Hyttinen et al. (2012), which we generalize to shift interventions. Hyttinen et al. (2012) consider the discovery of cyclic causal models from equilibrium data, assuming an underlying dynamic process of the form

$$\mathbf{X}_{i,t} = \mathbf{c}_i + \boldsymbol{\Phi}\mathbf{X}_{i,t-1}, \tag{20}$$

where $\mathbf{c} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\boldsymbol{c}})$ represents a time-invariant perturbation to the system, equivalent to the intercept terms in the model given in the main text. This represents a special case of the linear dynamical systems model discussed in the current paper, where the innovation or time-varying perturbations $\boldsymbol{\epsilon}_t$ are omitted, and where the average intercept value $\boldsymbol{\mu}_{\boldsymbol{c}}$ is assumed equal to be zero, largely for the sake of notational simplicity. The former restriction ensures that at some time point τ sufficiently long after the initial time point $t = 0$ the process is at equilibrium, $\mathbf{X}_{i,\tau} = \mathbf{X}_{i,\tau+1} = \ldots = \boldsymbol{\mu}_i$, essentially meaning that at some time point the equilibrium positions can be directly observed by sampling the position of the system at time τ or later. In the main text, we allow for time-varying perturbations $\boldsymbol{\epsilon}_t$ but also assume that it is possible to observe the equilibrium positions $\boldsymbol{\mu}_i$ directly or inferring it by taking the mean over a suitable sequence of observations of $\mathbf{X}_{i,t}$.

The models considered by Hyttinen et al. (2012) and in the current paper yield the same expression for the equilibrium positions of the process

$$\boldsymbol{\mu}_i = \mathbf{c}_i + \boldsymbol{\Phi}\boldsymbol{\mu}_i, \tag{21}$$

which can be re-arranged as

$$\boldsymbol{\mu}_i = (\boldsymbol{I} - \boldsymbol{\Phi})^{-1}\boldsymbol{c}_i. \tag{22}$$

This equivalence means that we can directly apply several derivations made by Hyttinen et al. (2012) in the current context, applying the additional assumption that we can directly observe the equilibrium positions of our system. The remainder of this appendix is structured as follows. In Section B.1, we discuss how the fact that we cannot assess the auto-regressive effect of a variable X_k from equilibrium data changes the parameters we can estimate. In Section B.2, we show that the equilibrium causal model and the underlying dynamic system produce the same observational data; in Section B.3, we show that they produce the same observations under press interventions; and in Section B.4, we show that they produce the same observations under shift interventions.

B.1. Marginalization of self-cycles and the canonical form

One of the key issues with studying equilibrium models of dynamical systems is that, when the system contains self-loops or auto-regressive effects ($\phi_{ii} \neq 0$), these self-loops cannot be recovered from equilibrium data. Hyttinen et al. (2012) show that any such system which is *stable* can be re-written into an *equivalent equilibrium model* in which the self-loops are standardized or marginalized out. Hyttinen et al. (2012) refers to this as the *canonical form* of the model. Let $\boldsymbol{U}$ be a $p \times p$ matrix with zero off-diagonal elements and diagonal elements

$$U_{ii} = \frac{\phi_{ii}}{1 - \phi_{ii}}. \tag{23}$$

The marginalized or canonical form direct effects ($\tilde{\boldsymbol{\Phi}}$), intercept covariance-matrix ($\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{c}}$), and mean intercept vector ($\tilde{\boldsymbol{\mu}}_{\boldsymbol{c}}$) can be expressed in terms of the original matrices as

$$\tilde{\boldsymbol{\Phi}} = \boldsymbol{\Phi} - \boldsymbol{U}(\mathbf{I} - \boldsymbol{\Phi}) \tag{24}$$

$$\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{c}} = (\mathbf{I} + \boldsymbol{U})\boldsymbol{\Sigma}_{\boldsymbol{c}}(\mathbf{I} + \boldsymbol{U})^T \tag{25}$$

$$\tilde{\boldsymbol{\mu}}_{\boldsymbol{c}} = (\mathbf{I} + \boldsymbol{U})\boldsymbol{\mu}_{\boldsymbol{c}}, \tag{26}$$

where $\tilde{\boldsymbol{\Phi}}$ is a matrix with zero elements on the diagonal (no self-cycles), but which otherwise retains the same structure (non-zero off-diagonal elements) as $\boldsymbol{\Phi}$. The off-diagonal elements of $\tilde{\boldsymbol{\Phi}}$ will differ in numeric value from the off-diagonal elements in $\boldsymbol{\Phi}$, with every element ϕ_{ij} re-scaled by auto-regressive effects of the effect variable ϕ_{ii} such that

$$\begin{aligned}\tilde{\phi}_{ij} &= \phi_{ij} + \phi_{ij}\frac{\phi_{ii}}{(1-\phi_{ii})} \\ &= \frac{\phi_{ij}(1-\phi_{ii}) + \phi_{ij}\phi_{ii}}{(1-\phi_{ii})} \\ &= \frac{\phi_{ij}}{(1-\phi_{ii})}.\end{aligned}$$

The parameters of $\tilde{\boldsymbol{\Phi}}$ can be interpreted as cumulative direct effects over a long timescale, in contrast to the lag-1 direct effects of the original matrix $\boldsymbol{\Phi}$. This interpretation becomes clear if we consider how we would express such an effect in a VAR setting. There, we could consider a cumulative effect of X_i on X_j as the sum of the lag-1 direct effect of ϕ_{ij} and longer lag indirects through the self-loop ϕ_{ii}. We could express this as

$$\begin{aligned}&\phi_{ij} + \phi_{ij}\phi_{ii} + \phi_{ij}\phi_{ii}^2 + \ldots \\ &= \phi_{ij}\left(\sum_{T=0}^{\infty}\phi_{ii}^T\right) \\ &= \frac{\phi_{ij}}{(1-\phi_{ii})}.\end{aligned}$$

where the last line simplification comes from the assumption of asymptotic stability, and can be understood to be valid whenever $|\phi_{ii}| < 1$, as we would expect in a stable system. Notably, while $\tilde{\boldsymbol{\Phi}}$ retains the same structure as $\boldsymbol{\Phi}$, many informative properties of the system are lost. For example, the rank order of the parameters in $\tilde{\boldsymbol{\Phi}}$ may differ from the rank order of cross-lagged effects in $\boldsymbol{\Phi}$ (a property of interest in a number of applications of lagged regression models; Hamaker et al., 2015; Kuiper & Ryan, 2018). The absence of auto-regressive parameters also means that in principle we are unable to recover properties such as the stability of individual variables, or the speed of return to equilibrium of specific trajectories.

Equations (24) and (25) are derived directly by Hyttinen et al. (2012), while Equation (26) follows as an implication of the derivations shown in their Appendix E. We demonstrate the validity of these expressions below. The matrices $\tilde{\boldsymbol{\Phi}}$, $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{c}}$, and $\tilde{\boldsymbol{\mu}}_{\boldsymbol{c}}$ define a model for the equilibrium positions of the dynamical system

$$\boldsymbol{\mu}_i = \tilde{\mathbf{c}}_i + \tilde{\boldsymbol{\Phi}}\boldsymbol{\mu}_i, \tag{27}$$

with $\tilde{\mathbf{c}}_i \sim \mathcal{N}(\tilde{\boldsymbol{\mu}}_{\boldsymbol{c}}, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{c}})$. It turns out that, even though the model matrix $\tilde{\boldsymbol{\Phi}}$ does not contain any information regarding the auto-regressive effects, the canonical model as a whole is actually equivalent to the original dynamical systems model in two ways. First, it yields equivalent means and covariances of equilibrium data in the observational setting (as we will show in Section B.2), and second, it yields equivalent predictions about the effects of press and shift interventions on the equilibrium positions (as we will show in Sections B.3 and B.4). Crucially, since the model does not require the estimation of the auto-regressive effects, the canonical model can be estimated using only equilibrium position data and standard SEM techniques. The implication of this is that, in principle, it is possible to learn a model which allows inferences about equilibrium positions and the causal effects of interventions on equilibrium positions using only data about those equilibrium positions.

B.2. Observational equivalence

Here we show that the underlying dynamical system and the equilibrium causal model in canonical form produce the same equilibrium data in an observational setting. Let $\boldsymbol{M}_{\boldsymbol{\mu}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\mu}}$ denote the mean vector and covariance matrix of equilibrium data in the observational setting. Standard expressions for $\boldsymbol{M}_{\boldsymbol{\mu}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\mu}}$ are known from the study of structural equation models (Bollen, 1989).

Proposition 1. The observational means $\boldsymbol{M}_{\boldsymbol{\mu}}$ are equivalent under the original and the canonical model, that is, we have that

$$(\mathbf{I} - \tilde{\boldsymbol{\Phi}})^{-1}(\tilde{\boldsymbol{\mu}}_{\boldsymbol{c}}) \overset{!}{=} (\mathbf{I} - \boldsymbol{\Phi})^{-1}(\boldsymbol{\mu}_{\boldsymbol{c}}). \tag{28}$$

Proof. Recall that $\tilde{\boldsymbol{\Phi}} = \boldsymbol{\Phi} - \mathbf{U}(\mathbf{I} - \boldsymbol{\Phi})$ and $\tilde{\boldsymbol{\mu}}_c = (\mathbf{I} + \mathbf{U})\boldsymbol{\mu}_c$. We show that substituting the latter into Equation 28 and pre-multiplying by $(\mathbf{I} - \tilde{\boldsymbol{\Phi}})$ results in the former, thereby proving the proposition. We write

$$\tilde{\boldsymbol{\mu}}_c = (\mathbf{I} - \tilde{\boldsymbol{\Phi}})(\mathbf{I} - \boldsymbol{\Phi})^{-1}(\boldsymbol{\mu}_c) \tag{29}$$

$$= (\mathbf{I} - \boldsymbol{\Phi} + \mathbf{U}(\mathbf{I} - \boldsymbol{\Phi}))(\mathbf{I} - \boldsymbol{\Phi})^{-1}(\boldsymbol{\mu}_c) \tag{30}$$

$$= \left((\mathbf{I} - \boldsymbol{\Phi})(\mathbf{I} - \boldsymbol{\Phi})^{-1} + \mathbf{U}(\mathbf{I} - \boldsymbol{\Phi})(\mathbf{I} - \boldsymbol{\Phi})^{-1}\right)(\boldsymbol{\mu}_c) \tag{31}$$

$$= (\mathbf{I} + \mathbf{U})\boldsymbol{\mu}_c, \tag{32}$$

which concludes the proof.

Proposition 2. The observational covariances are equivalent under the original and the canonical model, that is, we have that

$$(\mathbf{I} - \boldsymbol{\Phi})^{-1}\boldsymbol{\Sigma}_c(\mathbf{I} - \boldsymbol{\Phi})^{-T} \overset{!}{=} (\mathbf{I} - \tilde{\boldsymbol{\Phi}})^{-1}\tilde{\boldsymbol{\Sigma}}_c(\mathbf{I} - \tilde{\boldsymbol{\Phi}})^{-T}. \tag{33}$$

Proof. From the derivation above and Equation (29) we have that

$$(\mathbf{I} - \tilde{\boldsymbol{\Phi}})(\mathbf{I} - \boldsymbol{\Phi})^{-1} = (\mathbf{I} + \mathbf{U}) \tag{34}$$

$$(\mathbf{I} - \boldsymbol{\Phi})^{-1} = (\mathbf{I} - \tilde{\boldsymbol{\Phi}})^{-1}(\mathbf{I} + \mathbf{U}) \tag{35}$$

$$(\mathbf{I} - \boldsymbol{\Phi})^{-T} = (\mathbf{I} + \mathbf{U})^T(\mathbf{I} - \tilde{\boldsymbol{\Phi}})^{-T}. \tag{36}$$

Recall that $\tilde{\boldsymbol{\Sigma}}_c = (\mathbf{I}+\mathbf{U})\boldsymbol{\Sigma}_c(\mathbf{I}+\mathbf{U})^T$. Using this and the above we write

$$(\mathbf{I}-\boldsymbol{\Phi})^{-1}\boldsymbol{\Sigma}_c(\mathbf{I}-\boldsymbol{\Phi})^{-T} = (\mathbf{I}-\tilde{\boldsymbol{\Phi}})^{-1}(\mathbf{I}+\mathbf{U})(\boldsymbol{\Sigma}_c)(\mathbf{I}+\mathbf{U})^T(\mathbf{I}-\tilde{\boldsymbol{\Phi}})^{-T} \tag{37}$$

$$= (\mathbf{I}-\tilde{\boldsymbol{\Phi}})^{-1}\tilde{\boldsymbol{\Sigma}}_c(\mathbf{I}-\tilde{\boldsymbol{\Phi}})^{-T}, \tag{38}$$

which concludes the proof.

The implication of observational equivalence is that, given data on the means and covariance matrix of equilibrium positions, and knowledge of the structure of the model, we can in principle estimate the canonical model from observational equilibrium data using standard SEM software. If the structure of the model is not known, one can use causal discovery methods, as discussed in the main text. Irrespective of the estimation method, and even though the parameter values differ between the two models, the structure is the same, and we know exactly how the original lagged parameters map onto the parameters we estimate from equilibrium data. This is interesting, but it does not stop there. As we will see in the next two sections, these models are not only observationally equivalent but also yield the same predictions about the effect of press and shift interventions.

B.3. Equivalence under press interventions

Recall the effect of a press intervention on the equilibrium positions as formalized in Equation (18)

$$\boldsymbol{\mu} = (\mathbf{I}-\boldsymbol{P}_k\boldsymbol{\Phi})^{-1}(\boldsymbol{P}_k\boldsymbol{c}+\boldsymbol{s}), \tag{39}$$

where $\boldsymbol{P}_k$ is a $p \times p$ matrix with zeros on the off-diagonals and the k^{th} diagonal, and ones as the other diagonal elements (such that pre-multiplying by $\boldsymbol{P}_k$ yields a matrix with the k^{th} row equal to zero and other elements unchanged), and $\boldsymbol{a}_k$ is a p column vector with k^{th} element equal to a, the value that the intervened-on variable is set to by intervention.

Proposition 3. The canonical and the original dynamical system yield the same inferences about the effects of press interventions, that is, we have that

$$(\mathbf{I}-\boldsymbol{P}_k\boldsymbol{\Phi})^{-1}(\boldsymbol{P}_k\boldsymbol{c}+\boldsymbol{s}) \overset{!}{=} (\mathbf{I}-\boldsymbol{P}_k\tilde{\boldsymbol{\Phi}})^{-1}(\boldsymbol{P}_k\tilde{\boldsymbol{c}}+\boldsymbol{s}). \tag{40}$$

Proof. We begin by showing that

$$(\mathbf{I}-\boldsymbol{P}_k\tilde{\boldsymbol{\Phi}})(\mathbf{I}-\boldsymbol{P}_k\boldsymbol{\Phi})^{-1} \overset{!}{=} \mathbf{I}+\mathbf{U}\boldsymbol{P}_k. \tag{41}$$

To see this, recall that $\tilde{\boldsymbol{\Phi}} = \boldsymbol{\Phi} - \mathbf{U}(\mathbf{I}-\boldsymbol{\Phi})$ and substitute

$$(\mathbf{I}-\boldsymbol{P}_k\tilde{\boldsymbol{\Phi}})(\mathbf{I}-\boldsymbol{P}_k\boldsymbol{\Phi})^{-1} = (\mathbf{I}-\boldsymbol{P}_k(\boldsymbol{\Phi}-\mathbf{U}(\mathbf{I}-\boldsymbol{\Phi}))(\mathbf{I}-\boldsymbol{P}_k\boldsymbol{\Phi})^{-1} \tag{42}$$

$$= (\mathbf{I}-\boldsymbol{P}_k\boldsymbol{\Phi}+\boldsymbol{P}_k\mathbf{U}(\mathbf{I}-\boldsymbol{\Phi}))(\mathbf{I}-\boldsymbol{P}_k\boldsymbol{\Phi})^{-1} \tag{43}$$

$$= (\mathbf{I}-\boldsymbol{P}_k\boldsymbol{\Phi})(\mathbf{I}-\boldsymbol{P}_k\boldsymbol{\Phi})^{-1} + (\boldsymbol{P}_k\mathbf{U}(\mathbf{I}-\boldsymbol{\Phi}))(\mathbf{I}-\boldsymbol{P}_k\boldsymbol{\Phi})^{-1} \tag{44}$$

$$= \mathbf{I} + (\mathbf{U}\boldsymbol{P}_k(\mathbf{I}-\boldsymbol{\Phi}))(\mathbf{I}-\boldsymbol{P}_k\boldsymbol{\Phi})^{-1}, \tag{45}$$

where the last steps follows because $\mathbf{U}$ is diagonal and hence multiplication commutes. Notice that because $\boldsymbol{P}_k$ sets the k^{th} row to zero, it follows that $\boldsymbol{P}_k = \boldsymbol{P}_k^2$. We can thus write $\boldsymbol{P}_k(\mathbf{I}-\boldsymbol{\Phi}) = \boldsymbol{P}_k(\mathbf{I}-\boldsymbol{P}_k\boldsymbol{\Phi})$, which plugged into (45) yields

$$(\mathbf{I}-\boldsymbol{P}_k\tilde{\boldsymbol{\Phi}})(\mathbf{I}-\boldsymbol{P}_k\boldsymbol{\Phi})^{-1} = \mathbf{I}+\mathbf{U}\boldsymbol{P}_k(\mathbf{I}-\boldsymbol{P}_k\boldsymbol{\Phi})(\mathbf{I}-\boldsymbol{P}_k\boldsymbol{\Phi})^{-1} \tag{46}$$

$$= \mathbf{I}+\mathbf{U}\boldsymbol{P}_k. \tag{47}$$

With this in hand, we left-multiply Equation (40) by $(\mathbf{I}-\boldsymbol{P}_k\tilde{\boldsymbol{\Phi}})$ and substitute, yielding

$$(\mathbf{I}+\mathbf{U}\boldsymbol{P}_k)(\boldsymbol{P}_k\boldsymbol{c}+\boldsymbol{s}) \overset{!}{=} (\boldsymbol{P}_k\tilde{\boldsymbol{c}}+\boldsymbol{s}). \tag{48}$$

Recall that $\tilde{\boldsymbol{c}} = (\mathbf{I}+\mathbf{U})\boldsymbol{c}$. We expand and substitute, writing

$$(\mathbf{I}+\mathbf{U}\boldsymbol{P}_k)\boldsymbol{P}_k\boldsymbol{c} + (\mathbf{I}+\mathbf{U}\boldsymbol{P}_k)\boldsymbol{s} = \boldsymbol{P}_k(\mathbf{I}+\mathbf{U})\boldsymbol{c}+\boldsymbol{s} \tag{49}$$

$$\boldsymbol{P}_k(\mathbf{I}+\mathbf{U}\boldsymbol{P}_k)\boldsymbol{c} + (\mathbf{I}+\mathbf{U}\boldsymbol{P}_k)\boldsymbol{s} = \boldsymbol{P}_k(\mathbf{I}+\mathbf{U})\boldsymbol{c}+\boldsymbol{s} \tag{50}$$

$$(\mathbf{I}+\mathbf{U}\boldsymbol{P}_k)\boldsymbol{s} = \boldsymbol{s} \tag{51}$$

$$\boldsymbol{s} = \boldsymbol{s}. \tag{52}$$

The last step follows because $(\mathbf{I}+\mathbf{U}\boldsymbol{P}_k)$ yields a diagonal matrix with the k^{th} diagonal equal to 1 and all other diagonals i equal to $1+U_{ii}$, and $\boldsymbol{s}$ is a $p \times 1$ column vector with the k^{th} element equal to s and all other elements zero. Multiplying thus yields $\boldsymbol{s}$, which concludes the proof. Another way to prove equivalence under press interventions is to show that the equilibrium means and covariances are the same in the interventional setting. For a proof of this, see Appendix E in Hyttinen et al. (2012).

B.4. Equivalence under shift intervention

Recall the effect of a shift intervention on the equilibrium positions as formalized in Equation (19)

$$\boldsymbol{\mu} = (\mathbf{I}-\boldsymbol{\Phi})^{-1}(\boldsymbol{c}+\boldsymbol{s}), \tag{53}$$

where $\boldsymbol{s}$ is a $p \times 1$ column vector.

Proposition 4. If we consider only shift values s which are defined with respect to a linear function of the population standard deviations of the intercept, then the canonical and the original dynamical system yield the same inferences about the effects of shift interventions. This condition becomes clearer shortly. Formally, we have that

$$(\mathbf{I}-\boldsymbol{\Phi})^{-1}(\boldsymbol{c}+\boldsymbol{s}) \overset{!}{=} (\mathbf{I}-\tilde{\boldsymbol{\Phi}})^{-1}(\tilde{\boldsymbol{c}}+\tilde{\boldsymbol{s}}), \tag{54}$$

where $\boldsymbol{s}$ and $\tilde{\boldsymbol{s}}$ are $p \times 1$ column vectors. Note that if we apply a shift intervention to X_k—respecting the linearity condition mentioned above—then $\boldsymbol{s}$ and $\tilde{\boldsymbol{s}}$ have elements $s_k = q \times \sigma_k + z$ and $\tilde{s_k} = q \times \tilde{\sigma}_k + z$, respectively, where q and z are arbitrary constants and σ_k and $\tilde{\sigma}_k$ represent the population standard deviations of the k^{th} intercept term in the observational and canonical setting, respectively.

Proof. It is easy to show that $(\mathbf{I}-\tilde{\boldsymbol{\Phi}})(\mathbf{I}-\boldsymbol{\Phi})^{-1} = \mathbf{I}+\mathbf{U}$. Further, recall that $\tilde{\boldsymbol{c}} = (\mathbf{I}+\mathbf{U})\boldsymbol{c}$. From this and Equation (54) it follows that

$$(\mathbf{I}-\tilde{\boldsymbol{\Phi}})(\mathbf{I}-\boldsymbol{\Phi})^{-1}(\boldsymbol{c}+\boldsymbol{s}) = \tilde{\boldsymbol{c}}+\tilde{\boldsymbol{s}} \tag{55}$$

$$(\mathbf{I}+\mathbf{U})(\boldsymbol{c}+\boldsymbol{s}) = (\mathbf{I}+\mathbf{U})\boldsymbol{c}+\tilde{\boldsymbol{s}} \tag{56}$$

$$(\mathbf{I}+\mathbf{U})\boldsymbol{s} = \tilde{\boldsymbol{s}}. \tag{57}$$

Equation (57) gives a condition on $\boldsymbol{s}$ and $\tilde{\boldsymbol{s}}$ under which the original and canonical model are equivalent under shift interventions. Without lack of generality, note that if we

intervene on X_k, then all elements of $\boldsymbol{s}$ and $\tilde{\boldsymbol{s}}$ are zero except for the k^{th} element, for which we have

$$(1 + u_k)s_k = \tilde{s_k}, \tag{58}$$

where u_k denotes the k^{th} diagonal element of $\mathbf{U}$. Recall that $\mathbf{U}$ and hence u_k are a function of the auto-regressive effects in the original model, as defined in Equation (23). In the canonical model, we do not have direct access to the auto-regressive effects and thus $\mathbf{U}$, and choosing $\tilde{s_k}$ such that it is equal to $(1 + u_k)s_k$ is impossible to do directly.

We can, however, achieve this implicitly by defining $\tilde{s_k}$ with respect to $\tilde{\sigma}_k$. To see this, note that a very similar relation as in Equation (58) holds with regards to the population standard deviations σ_k and $\tilde{\sigma}_k$ of the original and canonical model, respectively. To see this, recall that $\tilde{\boldsymbol{\Sigma}}_c = (\mathbf{I} + \mathbf{U})\boldsymbol{\Sigma_c}(\mathbf{I} + \mathbf{U})$. Let σ_k and $\tilde{\sigma}_k$ denote the k^{th} diagonal element of $\boldsymbol{\Sigma_c}$ and $\tilde{\boldsymbol{\Sigma}}_c$, respectively. It follows that

$$(1 + u_k)^2\sigma_k^2 = \tilde{\sigma}_k^2 \tag{59}$$

$$(1 + u_k)\sigma_k = \tilde{\sigma}_k. \tag{60}$$

If we define $\tilde{s_k}$ to be a *linear* function of $\tilde{\sigma}_k$, denoted as f, we can—using Equation (58)—write

$$(1 + u_k)s_k = f(\tilde{\sigma}_k) \tag{61}$$

$$(1 + u_k)s_k = f((1 + u_k)\sigma_k) \tag{62}$$

$$(1 + u_k)s_k = (1 + u_k)f(\sigma_k) \tag{63}$$

$$s_k = q \times \sigma_k + z. \tag{64}$$

This shows that, if we define the shift interventions $\boldsymbol{s}$ and $\tilde{\boldsymbol{s}}$ on any variable X_k in terms of the same linear function of σ_k and $\tilde{\sigma}_k$, respectively, then the canonical and the original dynamical system are equivalent under shift interventions.

Appendix C: Latent state trait and residual correlation

In this section, we provide additional information regarding the specification of fixed measurement models and the effects that misspecification of this measurement model has in practice.

C.1. Estimating ECMs with fixed trait variance

In the main text we stated that prior knowledge can be used to fix the measurement error variances and latent variable variances to constants. Using SEM terminology, we have the following structural equation and measurement equation

$$\boldsymbol{\mu} = \tilde{\boldsymbol{\mu}}_{\boldsymbol{c}} + \tilde{\boldsymbol{\Phi}}\boldsymbol{\mu} + \boldsymbol{\zeta} \tag{65}$$

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}, \tag{66}$$

which yields a model-implied covariance matrix given by the standard SEM expression

$$\boldsymbol{\Sigma}_Y = (\mathbf{I} - \tilde{\boldsymbol{\Phi}})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \tilde{\boldsymbol{\Phi}})^T + \boldsymbol{\Theta}, \tag{67}$$

where $\boldsymbol{\Theta}$ is the variance-covariance of the measurement errors $\boldsymbol{\epsilon}$ and $\boldsymbol{\Psi}$ represents the variance-covariance matrix of the latent variables or equivalently the structural error terms ζ: the diagonal elements of $\boldsymbol{\Psi}$ represent the *variances* of exogenous latent variables and the *residual variances* of endogenous variables.

Suppose that based on prior research the proportion of variance in a measurement instrument which can be attribute to the *trait*, that is, equilibrium variance across individuals, is known, and the rest of the variance is attributable to an independent variance term. Using the measurement equation above we could write

$$var(\boldsymbol{Y}) = var(\boldsymbol{\mu}) + var(\boldsymbol{\epsilon}), \tag{68}$$

and statements about the proportion of variance attributable to a trait can then be interpreted as a statement about knowing the magnitude of $var(\boldsymbol{\mu})$ relative to $var(\boldsymbol{\epsilon})$. Suppose without loss of generality that the observed variables $\boldsymbol{Y}$ are standardized, such that $var(\boldsymbol{Y}) = 1$ for all variables, and that we know that 70% of the variance of each observed variable is attributable to variance in the equilibrium and 30% to other sources. In order to use this information in parameter estimation, we need to impose parameter constraints on $\boldsymbol{\Theta}$ and $\boldsymbol{\Psi}$ such that $var(\boldsymbol{\epsilon}) = 0.30$ and $var(\boldsymbol{\mu}) = 0.70$. In the former case, this is straightforward, since the variance of the measurement error is parameterized directly by $\boldsymbol{\Theta}$, meaning that we can set $\theta_{ii} = 0.30\,\forall i$. However, $var(\boldsymbol{\mu})$ is parameterized by ψ_{ii} alone only in the case of an exogenous variable. For endogenous variables, $var(\mu_j)$ is a function of its predictors, their variances and covariances, and the residual variance term ψ_{jj}, and so imposing the appropriate parameter constraint on ψ_{jj} is less straightforward.

To see this, consider our example ECM discussed in the main text, which has weights matrix

$$\tilde{\boldsymbol{\Phi}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ a & 0 & b & 0 \\ 0 & c & 0 & d \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

An expression for the variance of μ_2 can be found by using the structural equation

$$\begin{aligned} \mu_2 &= a\mu_1 + b\mu_3 + \epsilon_2 \\ var(\mu_2) &= var(a\mu_1 + b\mu_3 + \epsilon_2) \\ var(\mu_2) &= var(a\mu_1) + var(b\mu_3) + cov(a\mu_1, b\mu_3) + var(\epsilon_2) \\ var(\mu_2) &= a^2var(\mu_1) + b^2var(\mu_3) + ab \times cov(\mu_1, \mu_3) + \psi_{22}, \end{aligned}$$

which we can see is dependent on the variances of μ_1 and μ_3, and the covariance between μ_1 and μ_3. The variance of μ_3 yields a similar expression, dependent both on $var(\mu_2)$ and the covariance between μ_2 and μ_4, while variance of μ_1 and μ_4 simply reduce to ψ_{11} and ψ_{44}, respectively. This means that we can fix ψ_{11} and ψ_{44} to take on values of 0.70, but imposing the same constraint on ψ_{22} and ψ_{33} would not yield the desired total variance. As such, our goal is to impose parameter constraints on ψ_{22} and ψ_{33} such that the expression on the right hand side of the above equation adds up to the known total trait variance value of 0.70. To do this we must solve for an expression for ψ_{22} which does not involve ψ_{33}, and vice versa. In the current example, this yields

$$\psi_{22} = \frac{a^2\psi_{11} + a^2bc\psi_{11} - 0.70 + b^2 0.70 + bc0.70 - b^3c0.70}{1 + bc} \tag{69}$$

$$\psi_{33} = \frac{d^2\psi_{44} + bcd^2\psi_4 4 - 0.70 + bc0.70 + c^2 0.70 - bc^3 0.70}{1 + bc}. \tag{70}$$

These parameter constraints can be implemented directly in standard SEM software. R code showing how this can be done is available from https://github.com/fdabl/Equilibrium-Causal-Models.

C.2. Fixed trait variance and residual correlation

In the main text, we studied the effect that a residual correlation between the variables Y_1 and Y_2 has on the causal effects estimates between all variables. Specifically, we considered unmodeled correlations between the state components of the measurement s. In the standard SEM notation used in the previous section, this concerns a correlation between the measurement error variances $\theta_{12} \neq 0$. In the main text we also, we studied the effects that misspecifying the measurement model has on the parameter and causal effects estimates. Here, we combine these two issues. In particular, we mirror the simulation study described earlier, except that we now always correctly specify the modeled trait variance correctly. For example, if $\sigma_s^2/(1+\sigma_s^2) = 0.50$, then we assume that we have correctly specified the modeled trait variance as being as high as the state variance in the estimation routine described above. We again study a residual correlation between Y_1 and Y_4 $\rho_{12} \in [-0.25, 0, 0.25]$, but we assume that we do not correctly specify this residual correlation in the measurement model.

Figure 9 shows the estimated causal effects across the relative proportion of trait and state variance, residual correlation, and causal effects. In contrast to Figure 6, were we did not correctly specify the measurement model, we now see that the causal effects are estimated without bias in case of zero residual correlation (orange). In contrast, a negative residual correlation between Y_1 and Y_4 that is not correctly specified in the measurement model leads to an attenuation of the effect estimates (green), most strongly for the causal effect $X_1 \rightarrow X_2$. Correspondingly, a positive residual correlation yields to an overestimate of the causal effects (purple).

Appendix D: Illustrating challenges of estimating ECMs

In the main text we demonstrated that, given assumptions regarding the qualitative dynamics being studied, unobserved confounding, and correct model specification, equilibrium causal effects can be estimated from equilibrium data of a dynamical system. In empirical practice, of course, many of these assumptions are likely not to hold.

Here, using an open-access time-series dataset (McNeish & MacKinnon, 2025) we investigate whether we obtain equivalent estimates of equilibrium causal relations by (a) estimating a lagged regression model and then deriving the equilibrium relations or (b) obtaining estimates of the equilibrium positions and then estimating the equilibrium model directly from that data.

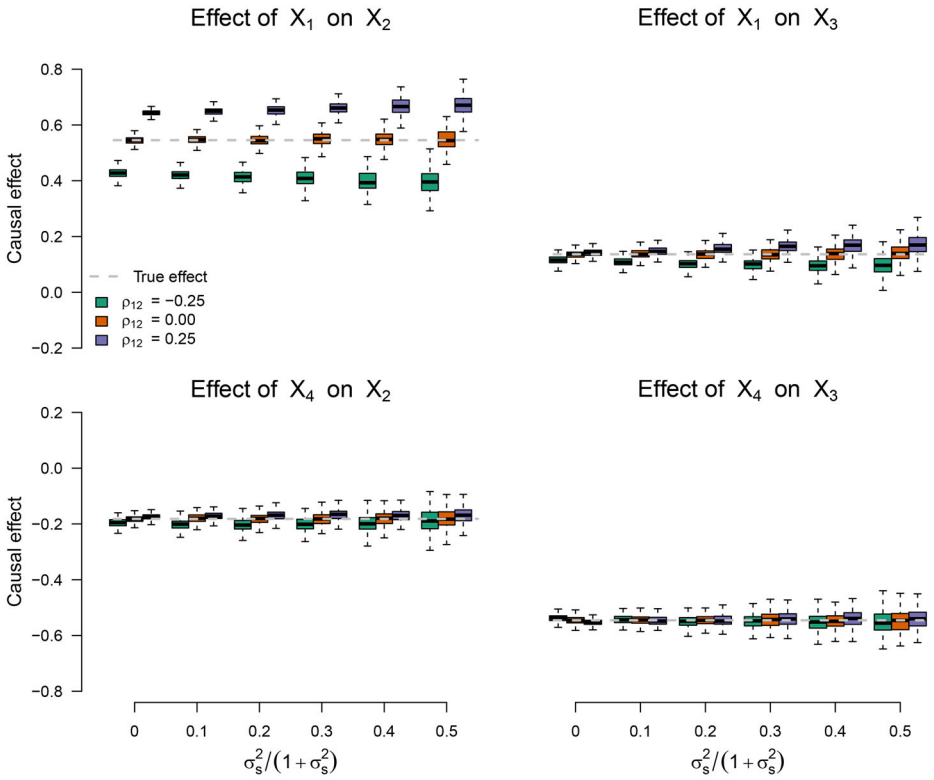

Figure 9. Estimates of the effect of $press(X_1 = 1)$ on X_2 (top left) and X_3 (top right), and $press(X_4 = 1)$ on X_2 (bottom left) and X_3 (bottom right) across different state variances (given as a proportion of the total variance) and residual correlations between X_1 and X_2. Dashed gray lines indicate the true causal effect.

D.1. Data and context

To perform this analysis, we will make use of an open-source synthetic empirical time-series dataset published by McNeish and MacKinnon (2025) and available open access (see https://osf.io/yk3je/). All code to reproduce this analysis is available from https://github.com/fdabl/Equilibrium-Causal-Models. McNeish and MacKinnon (2025) generate these data based on the distributional characteristics of data from an empirical intensive longitudinal study (Scherer et al., 2022) studying the dynamic relations between adherence to a health-behavior intervention (hereafter *Adh*), self-reported measures of perseverance (hereafter *Pers*) and the number of steps taken (hereafter *Steps*). The dataset consists of 35 repeated measures of each variable for 50 individuals. This dataset is suitable for use as an illustration for three reasons. First, the research goal of the original study, and the work of McNeish and MacKinnon (2025), focuses on investigating a hypothesized dynamic *mediation* structure where current *Adh* has a direct effect on future *Steps*, as well as an indirect effect through its effect on intermediate values of *Pers* (*Adh* → *Pers* → *Steps*). The models used to investigate this structure are linear. This structure of dynamic relations would imply a linear ECM which is non-recursive, and therefore identifiable using standard SEM methods. Second, the features of the dataset, in terms of number of measurements within- and between-person, is somewhat typical for what is often encountered in psychological research. Third, although synthetic in nature, the dataset was simulated based on characteristics of a true empirical dataset, and therefore does not impose that any of the idealizing assumptions we make in our simulation study in the main text.

D.2. Estimation procedure

We estimate an equilibrium model using two approaches. First, we fit a multilevel VAR(1) model to the time series data, estimating only lag-1 autoregressive effects and lag-1 cross-lagged effects which impose the mediation structure described above ($Steps_t$ regressed on $Pers_{t-1}$ and Adh_{t-1}, with $Pers_t$ regressed on Adh_{t-1}). The model is estimated using the Bayesian DSEM module in Mplus (Asparouhov et al., 2018), using the stationary mediation script supplied by McNeish and MacKinnon (2025) adapted to estimate only lag-1 cross-lagged effects (available from our online code repository). From this model we extract a matrix of point-estimates for the lagged fixed effects, $\hat{\Phi}_{lag}$. During estimation, we also extract 1000 posterior samples of $\hat{\Phi}_{lag}$. By applying the transformation described in Equation (4) to these point estimates and posterior samples, we obtain point estimates and 95% credible intervals for the *model-implied* equilibrium relations ($\tilde{\Phi}$).

Second, we obtain direct estimates of the equilibrium relations $\tilde{\Phi}$ by using the within-person sample means of each variable as an *estimate* of their equilibrium positions $\hat{\mu}_i$. With this dataset of sample means, we estimate a simple SEM mediation model (μ_{Steps} regressed on μ_{Pers} and μ_{Adh}; μ_{Pers} regressed on μ_{Adh}). The regression weights matrix of this model gives us a direct estimate of the equilibrium relations.

D.3. Results and discussion

Table 2 shows the point estimates and 95% credible intervals obtained using the above approach. In the first column we see the estimates of the model-implied ECM relations; we see positive estimates for all relations, although we see a relatively wide CI encompassing zero for the *Adh* → *Steps* relation. In the second column we see the equilibrium relations directly estimated using the sample-means as stand-ins for the equilibrium positions. We see that the point estimates are again all positive, and relatively similar in magnitude to those in the first column. We see, here, however, that the credible intervals are noticeably wider than in the first column, and all encompass zero. This is in part to be expected as, by taking the means across the 50 individual time series before model estimation, we essentially reduce the size of our dataset to just 50 data points. We also see that, for the *Pers* → *Steps*, the point estimates obtained from both approaches are quite different; although not a formal test in any way, the fact that the CIs obtained from both approaches do not overlap can be taken to mean that both approaches would lead to relatively different conclusions about this particular equilibrium relationship.

Table 2. Point estimates and 95% confidence intervals in parenthesis for the ECM model implied by estimated multilevel VAR(1) model (left column) and directly estimated from the estimated person-specific equilibrium positions (right column).

	Model-implied ECM	Directly estimated ECM
Adh → *Pers*	1.26 (0.97, 1.59)	1.32 (−2.02, 4.66)
Adh → *Steps*	2.01 (−1.52, 5.71)	1.56 (−15.28, 18.40)
Pers → *Steps*	3.62 (2.81, 4.49)	0.72 (−0.67, 2.10)

This exercises demonstrates that, while the equivalence between empirical ECMs and equilibrium relations implied by dynamic relations has been shown to hold under a set of assumptions described in the main text, there is no guarantee that those assumptions will hold in practice. When the true system is unlikely to be satisfactorily modeled by a simple non-recursive linear model such as this, or when we cannot assume the nonexistence of unobserved confounding, then we should also not expect this equivalence to hold. This exercise also illustrates the practical difficulty of testing whether ECM estimates are equivalent to those obtained from a time-series model in a typical psychological setting. Specifically, we saw that this equivalence could only be directly investigated when the implied ECM is identifiable (typically, with few or no feedback relations), and even then, such tests will likely run into issues of statistical power: Collecting longitudinal data is relatively difficult, and so intensive longitudinal datasets in psychology tend to have relatively few participants. When pre-processing this data to obtain sample means (i.e., average across time for each participant), we ended up with a much smaller dataset. As discussed in the main text, however, we do not suggest that researchers estimate ECMs in the way we have done here, unless strong evidence is provided that the necessary assumptions hold. Instead, we suggest that novel developments in the causal discovery literature can be utilized to estimate ECMs on psychological data.

Appendix E: Causal discovery simulation

In this section, we briefly describe the Backshift method and assess its performance in a simulation study using our running $p = 4$ variable example.

E.1. Backshift

Backshift assumes that the observations are equilibrium positions and that for a particular environment j they can be modeled as

$$\mathbf{x}_j = \mathbf{B}\mathbf{x}_j + \boldsymbol{c}_j + \mathbf{e}_j, \tag{71}$$

where the random shift intervention $\boldsymbol{c}_j$ has covariance $\Sigma_{\boldsymbol{c},j}$ and the matrix $\mathbf{B}$ and the error distribution of $\mathbf{e}_j$ are assumed to be identical across environments. While the covariance matrix of the errors $\Sigma_{\mathbf{e},j}$ does not have to be diagonal, $\Sigma_{\boldsymbol{c},j}$ is assumed to be diagonal. This implies that interventions at different variables are assumed to be uncorrelated. Backshift further assumes that the intervention shift $\boldsymbol{c}_j$ and the noise $\mathbf{e}_j$ are uncorrelated. Interventions can differ in their strength, and this is encoded in the variance of $\boldsymbol{c}_j$: a higher variance implies a stronger intervention strength. Of note is that the location (or target) of the intervention (i.e., the indices of $\boldsymbol{c}_j$ who have non-zero values) and their strength can be estimated from data (for details, see Rothenhäusler et al., 2015). R code showing an example of using Backshift is available from E.

E.2. Simulation study

To assess the performance of Backshift we simulate from our $p = 4$ variable example model we have used throughout. Specifically, we generate equilibrium data varying the sample size $n = [250, 500, 1000, 2500, 5000]$; the number of variables the shift intervention targets $t = [1, 2, 3]$; the number of settings $s = [3, 4, ..., 9, 10]$; the strength of the shift intervention $m = [0.50, 1, 2]$ (Rothenhäusler et al. formalized as in 2015); and the extent of unobserved confounding, parameterized as residual correlation $r = [0, 0.50]$—we randomly set each off-diagonal element to either 0.50 or −0.50. We assess the estimation error defined as the (element-wise) average absolute difference between the estimated $\hat{\mathbf{\Phi}}$ and the true parameters $\mathbf{\Phi}$.

E.3. Simulation results

Figure 10 shows the estimation error across different intervention strengths, number of targets, and residual confounding. We find that residual confounding leads to a

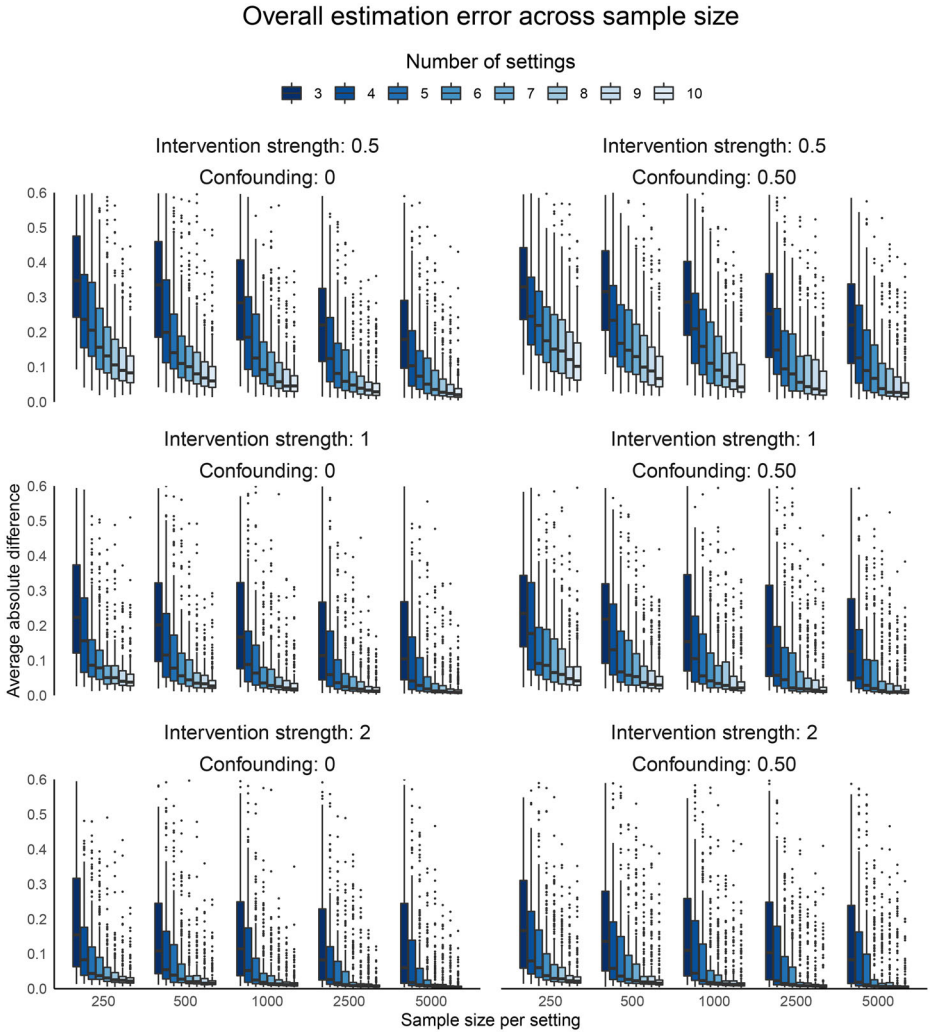

Figure 10. Average absolute difference between estimated and true parameters across sample sizes n and number of settings s for $p = 4$, $t = 3$ number of targets, an effect size of $m = 0.50$ (top), $m = 1$ (middle), and $m = 2$ (bottom) for no (left) or $r = 0.50$ (right) residual confounding. Results are ordered with increasing number of settings from $s = 3$ (dark blue) to $s = 10$ (white).

slightly worse performance and that the number of settings are more important than the sample size per setting. Figure 10 further shows that increasing the intervention strength and especially the number of targets yields to significant reductions in the estimation error. Focusing on the middle left panel, we observe that the estimation error decreases with increasing sample size per setting as well as with increasing the number of settings. The number of settings has a stronger beneficial effect than the sample size: for example, increasing the sample size from $n = 250$ to $n = 500$ with $s = 3$ settings decreases the estimation error from 0.28 (0.29) to only 0.26 (0.27) for $r = 0$ and $r = 0.50$ residual correlation, respectively. Fixing the sample size per setting at $n = 250$ and increasing the number of settings from $s = 3$ to $s = 4$, on the other hand, decreasing the estimation error from 0.28 (0.29) to 0.21 (0.21) for $r = 0$ and $r = 0.50$, respectively, a 25% (28%) reduction rather than just a 7% (7%) reduction when increasing the sample size but keeping the number of settings fixed. Increasing the intervention strength (going from the top to the bottom panels) decreases the estimation error

Figure 10 shows results in which the intervention in each setting is targeting $t = 3$ variables. The top panels in Figure 11 show the estimation error as a function of the number of targets and no (left) and $r = 0.50$ (right) residual correlation across intervention strengths of $m = 0.50$ (top), $m = 1$ (middle), and $m = 2$ (bottom), fixing the sample size at $n = 500$. We find that the estimation error decreases quite strongly with the number of targets in each setting. For example, with $s = 4$ environments and an intervention strength of $m = 1$, the estimation error in the case of two targets $t = 2$ is about almost three times as high (0.20) compared to four targets $t = 4$ (0.07) in case of no residual correlation. For $r = 0.50$, a similar pattern holds (0.30 vs. 0.09). Note that these results are robust to changing the intervention target.

Appendix F: Empirical example of cyclic causal discovery

We illustrate how empirical researchers may utilize backshift to estimate ECMs by reanalyzing the data presented in

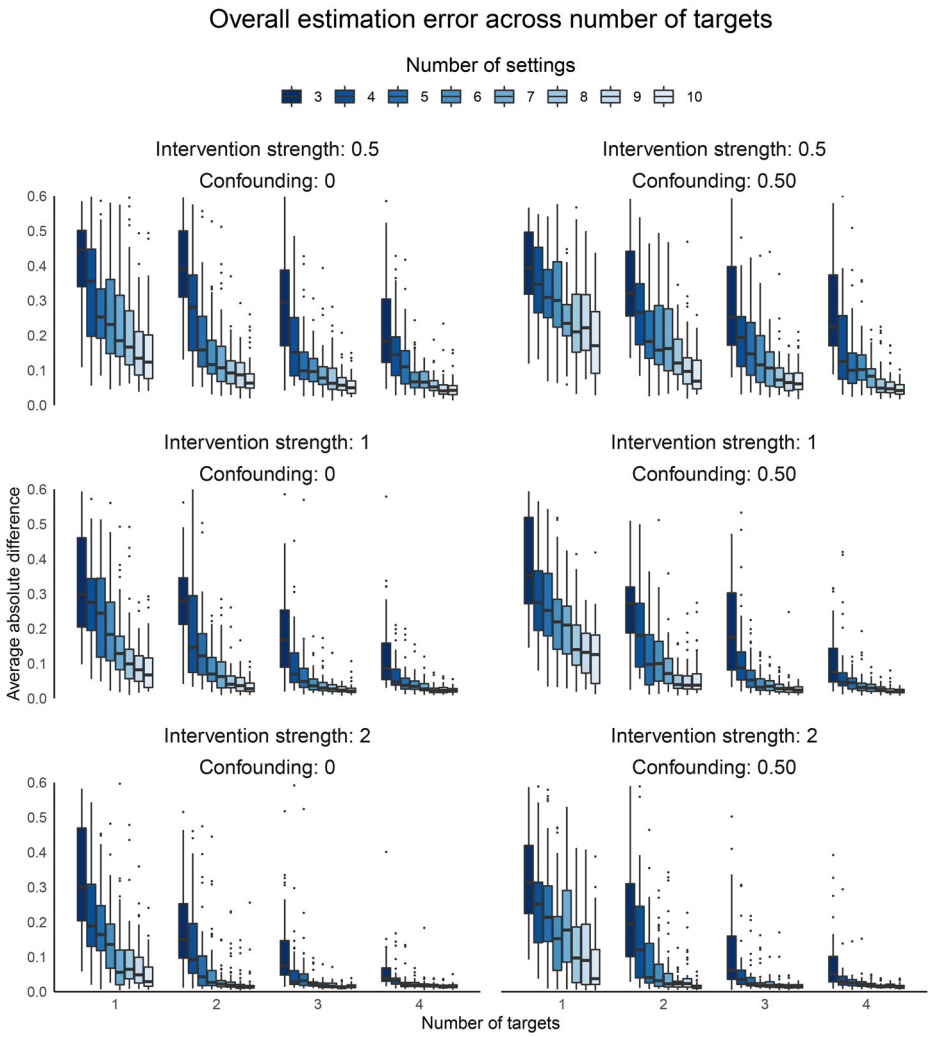

Figure 11. Average absolute difference between estimated and true parameters across across number of targets t and number of settings s for $p = 4$, $n = 500$, an effect size of $m = 0.50$ (top), $m = 1$ (middle), and $m = 2$ (bottom) for no (left) or $r = 0.50$ (right) residual confounding. Results are ordered with increasing number of settings from $s = 3$ (dark blue) to $s = 10$ (white).

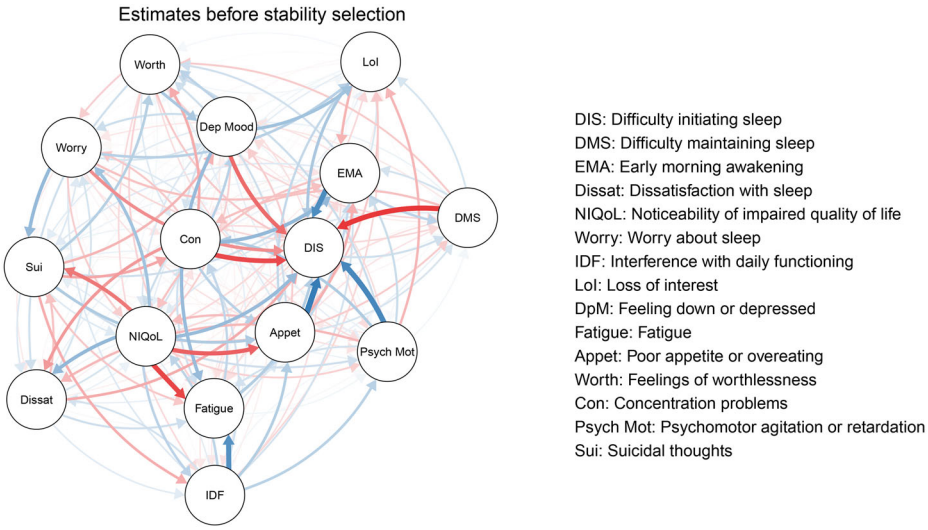

Figure 12. Causal effect estimates from backshift before stability selection. No causal effect remains after stability selection, see main text for details.

Blanken et al. (2019). In a randomized control trial, one group of patients received no treatment ($n = 52$) while another group received cognitive-behavioral therapy for insomnia for five weeks ($n = 52$). Symptoms of insomnia and depression were assessed each week for ten weeks (before treatment T0-T1, during treatment T2-T6, and after treatment T7-T9). Although there are a number of causal discovery methods which can use longitudinal data directly (Assaad et al., 2022; Pfister et al., 2019), for the purpose of illustration in the current paper, we treat this data as though it were cross-sectional in nature. We do this by averaging the measures across all ten weeks for the control group and across the two pretreatment and three post-treatment weeks for the treatment group. We do not use the measures during treatment because they may, conceptually and empirically, not constitute equilibrium observations. The left graph in Figure 12 shows the backshift estimate, yielding a fully connected graph with some effects being much stronger than others. R code to reproduce this analysis is provided in the reproducibility archive of the current paper, though data must be requested from the authors of the original paper (Blanken et al., 2019).

In the cyclic structure shown in Figure 12 we can for example see that *difficulties initiating sleep* is an effect of several variables, being negatively influenced by for example *concentration problems*, *difficulties maintaining sleep*, and *depressed mood*. It is important to note, however, that all of the point estimates in this graph are quite unstable; estimating the stability of each edge using the stability selection procedure described by Meinshausen and Bühlmann (2010) results in no edge being estimate consistently, and so, the point estimates shown in Figure 12 are not robust. They are very likely misleading and should not be interpreted. Significance testing based on bootstrapping also indicates that no edge can be assumed to be significantly different from 0. This is likely in large part due to the comparatively small sample size ($N = 158$ total observations spread across three groups), though it may also be due to the differences between the settings (i.e., the intervention strengths) not being large enough to inform causal effect estimates, or because the averaging procedure obscured any differences. This illustration does however serve the show the *potential* of methods like backshift in helping uncover cyclic causal depencies, though to fully realize this potential may require different data collection and study design strategies than what we commonly see in current network-based approaches in psychology.