Imperfect Mappings of Qualitative Structure to Quantitative Scales

Maas van Steenbergen [1,*]

[1] *Faculty of Behavioural and Social Sciences, Methodology & Statistics, Utrecht University, the Netherlands*

Correspondence*:
Corresponding Author
m.vansteenbergen@uu.nl

> "*Eh bien!*"–exclaimed Walras characteristically–"this difficulty is not insurmountable. Let us suppose that this measure exists, and we shall be able to give an exact and mathematical account of it". [. . .] In view of the fact that theoretical science is a living organism, it would not be exaggerating to say that this attitude is tantamount to planning a fish hatchery in a moist flower bed.
> – *Nicholas Georgescu-Roegen*

> "I think the foundations of measurements [. . .] have a lot of implications about the way you do and actually think about measurement. There is a great deal of feedback from work on foundations on the actual practice, unlike a lot of other fields of mathematics where work on foundations is divorced from the actual practice of other disciplines"
> – *Amon Tversky*

1 INTRODUCTION

Measurement in psychology is much more controversial than measurement in the physical sciences. Variables in psychology often consider attitudes, strength of sensation, ability, or personality traits. These variables are often said to not be directly observable. Researchers rely on an observer, the participant, who estimates values of psychological variables.

o introduce our topic, we make a number of assumptions about the nature of psychological constructs that are studied using psychology. It is important to make these assumptions explicit to increase the ability to reject their tenets if they turn out not to hold (10). We will use these assumptions to introduce the topic and embed the study in the literature.

We assume, in alignment with Krantz, Luce, Suppes, and Tversky (henceforth referred to as KLST), that an attribute is a quantity if it is or can be proven to be measurable using empirical methods, by criteria set out by the representational theory of measurement(8). This is presently the most completely developed theory of measurement within psychology. This means that we accept both the representation theorem and the uniqueness theorem in the context of measurement. The representation theorem asserts the following. Imagine that a set, together with one or more relations on that set, follows certain axioms so that it can represent qualitative attributes of an object in a quantitative manner. If this is the case, it can preserve the properties of the qualitative attribute after the assignment of numbers. In other words, a *homomorphic* mapping of the qualitative relational structure into a quantitative representation can be constructed.

Suppose that the transformation function $\phi(a)$ is a homomorphism and the representation theorem holds for the mapping of qualitative ordering $\succ$ to quantitative ordering $>$. If $\phi(a)$ then maps a qualitative attribute of set A into $\mathbb{R}$, it should maintain the structure of the qualitative attribute in its numerical representation: if and only if a is qualitatively greater than b, it is numerically greater than b. The uniqueness theorem specifies the permissible homomorphic transformations of A that lead to the same structure of numerical relations. E.g., both Fahrenheit or Celsius measurements are legitimate homomorphic mappings that maintain the qualitative structure of temperature, differing in zero-point and scale. Note that these aspects of measurement do not concern modelling or statistical analysis, but are a *necessary condition* for those.

We will assume a version called “representation minimalism” that rids it of most of its epistemological basis so that it can serve as a common ground for discussion about measurement, recognizing that most of the critical literature about measurement takes place either within or in discussion of this framework (18). We therefore do not make firm claims if a realist, an operationalist, or a representationalist ontology should be matched to the formal aspects of the framework. We do, however, choose to use representationalist language, as this is used in the original source and it is the most well-known and fully realized. It should be noted that choosing a different framework means that the formalism might change. If this change is subtle. In some frameworks, the current experiment might not make a whole lot of sense.

KLST is considered the definitive characterisation of a representational measurement theory, but its origins lie much further in history. It is said to originate from Russels. This is the source of the textbook definition of representationalism in psychology. That definition of measurement within psychology stems from Stevens: “the assignment of numerals according to rule” (15). However, for various reasons, both philosophers of science and measurement theorists, whatever their inclinations, have moved on from this characterization (11). Unsurprisingly, then, the definition of measurement in psychology is heavily debated. For our purposes, we present a very rough separation into three categories. The first category consists of a group that actively considers and thinks about measurement, and does not rule out quantitative psychology in some adjusted form. This can be subdivided into different groups. E.g., there is a school that claims that quantifiability can be empirically settled by using techniques such as Conjoint Measurement Theory either based on a representationalist or classical groundings (9, 8, 11). Another group of philosophers say that the Rasch model is already similar to conjoint measurement (3). On the other side, there are others who argue that psychological processes are not quantifiable at all (17), or that that the question itself rests on conceptual confusion (7, 16). The third category concerns those scientific psychologists, for whatever reason, that do use quantitative psychology but refrain from making epistemological commitments or assume that every measured construct is measurable. We consider only those cases that can be described through representation minimalism. In other words, we will consider it a settled manner that the attribute in question has a quantitative structure and that this can be proven if the empirical relations between objects or attributes permit measurability. This excludes any approach that is inconsistent with this basis. This means that only the first category of theories will be considered. While representation minimalism can ameliorate some controversy, measurement in psychology is too hotly debated to avoid all.

For the current project, we focus our attention on cases where an attribute of interest is pricipially continuously measurable. Therefore, a continuous, homomorphic mapping of the qualitative structure towards a quantitative representation can be constructed. In these cases, there are practical limitations in the construction of such a mapping. These complications should be familiar to the working psychologist. It is impossible to measure a psychological attribute such as happiness as one would measure height or temperature. Height and temperature measurements are relatively easy to independently verify between different measurement instruments and people. The area of interest of psychologists involves subjectivity, such as with attitudes, psychological state, and feelings. These are only directly accessible to the subject, and this makes it more difficult to verify the structural relationship between psychological phenomena than it is for physical phenomena. Usually, relatively crude measures are used to make an estimate of .

Let X, Y & Z by any three values of the mapping. Then the projection of the quantity to an ordinal scale, which we will refer to as P, holds to the following conditions:

Let A be a set and $\succeq$ be a binary relation on A. $< A, \geq >$ is a weak order iff, for all a, b, axiom 1 and 2 are met:

- if $X \succeq Y$ & $Y \succeq Z$, then $X \succeq Z$ (transitivity).
- Either $X \succeq Y \parallel Y \succeq X$ (connectedness).
- For the numerical representation of binary relations on A, if $X \succeq Y$ & $Y \succeq X$, then $X = Z$ (antisymmetry).

The score is only dependent on rank-order, and does not support concatenation operations. A score in P can only indicate that a score is higher or lower than another score in P. Based on the measure only, beyond that ordering, we know nothing about the real value of P (assuming a scaling S).

As for the principially measurable attribute Q we will assume that the following additional characteristics will hold (above the ones for the proje), combining Michell's distillation of Krantz, Tukey, Suppes, and Luce, but framed in the formalism of representational measurement (8):

- $X \oplus (X \oplus Y) = (X \oplus Y) \oplus Z$ (associativity);
- $X \oplus Y = Y \oplus X$ (commutativity);
- $X \succeq Y$ iff $X \oplus Z \succeq Y \oplus Z$ (monotonicity);
- if $X \succ Y$ then there exists a Z such that $X = Y \oplus Z$ (solvability);
- $X \oplus Y \succ X$ (positivity).
- there exists a number n such that $nX \succeq Y$ (where $1X = 1$ and $(n \oplus 1)X = nX \oplus X$) (Archimeadean condition).

This means essentially that a value q in terms of another value r in Q always sustain ratios. Every scale is homomorphic to the qualitative attribute that is being measured, and scales are thus homomorphic to each other. The last axiom (the archimedean condition) is added to ensure that the set of possible scores is finite (ratios cannot be infinite). E.g., say that we have developed a standard measure for happiness: H. Then we can say that any value x is written in terms of H iff the finite ratio $\frac{x}{HCS}$ holds [1].

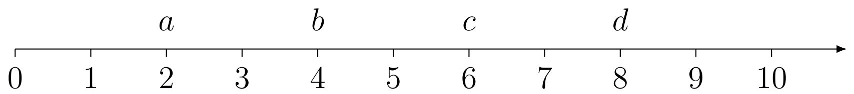
By pulling the scores away from the ordinal measurement instrument, we can formalize their relationship. This allows us to provide an unambiguous classification mechanism of scores in Q to scores in P[2]Assume that instrument P has n elements. Then we can define a series of $n-1$ 'pegs' R as $\{a, b, c \ldots z\}$. These pegs are the threshold of q where a score p_1 in P jumps to another classification p_2 in P on the basis of a score q in Q. We then define a score p in P for each q in Q as follows, assuming we have an n-level measurement scale for P:

[1] We agree with the point by Franz that it is misleading to give physical magnitudes as examples when discussing psychological phenomena (7). Therefore, we choose to use psychological examples when we are thinking about psychological phenomena. This adds an extra face validity check: does discussing psychological attributes in this manner make sense?

[2] This lack of ambiguity is only a result of this formalization. We may find that the relationship of the 'real' variable to its ordinal estimate is not quite so straightforward. We will discuss this later on.

$$\begin{cases} 1 & q \leq a \\ 2 & a \leq q \leq b \\ 3 & b \leq q \leq c \; . \\ \dots & \dots \\ n & z \leq q \end{cases}$$

This relationship can also be visualized on a number line. Consider a five-point instrument is used to measure a quantitative psychological variable. At a particular point in time t, we assume that the projection on the real number line is divided into four segments of equal length and one element which goes on infinitely. It should be noted that the scaling is arbitrary, but it has been set here to multiples of two for convenience.



,

The relationship between P and Q in this case can then be described with the following step function:

$$\begin{cases} 1 & q \leq 2 \\ 2 & 2 \leq q \leq 4 \\ 3 & 4 \leq q \leq 6 \; . \\ 4 & 6 \leq q \leq 8 \\ 5 & 8 \leq q \end{cases}$$

This construct can be said to be equally spaced. It is equally spaced because the distance each ordinal

1.1 Between-person variation

It is often assumed that aggregations of individual measures rescind the limitations of different scaling because these measurement differences cancel out when enough samples are drawn. The origins of this assertion are from Knapp. This idea can be seen as follows: because

1.2 Within-person variation

When studying within-person mappings, we can assume that those psychological processes part of Q are time-dependent. These measures are shaped by different forces and the previous states of the construct (13). Their values are continuous and are related to each other in a structured manner (2). We also assume that their values are differentiable over time, changing smoothly. Their values can increase or increase very quickly, but not instantaneously. This, they are modeled best using the rate of change of the variable, through differential equations (12).

Making a, b, c, and d functions of time, step function F(x, t) becomes:

$$\begin{cases} 1 & q \leq a(t) \\ 2 & a(t) \leq q \leq b(t) \\ 3 & b(t) \leq q \leq c(t) \\ \dots & \dots \\ n & z(t) \leq q \end{cases} .$$

This means that each threshold becomes a function, with time as input. That is, for each time t, a .

1.3 Using the framework for simulation studies

If you want to make claims about the consequences of different structural deficiencies of a framework, it can be helpful to run a simulation of the effect of those deficiencies to show the consequences of those effects to researchers.

1.4

We used the Julia language, and in particular the 'DynamicalSystems.jl', 'RecurrenceAnalysis.jl', and 'Statistics.jl' packages to implement the toy model and run the recurrence analyses (1, 5, 6). Analyses were run on a personal computer. Full information about dependencies and version numbers can be found in a machine-readable format in the Manifest.toml file in the Github-repository. Instructions for running the analysis through a sandboxed project environment identical to our system can be found on the main page of this repository.

2 DISCUSSION

2.1 Measurability

We should note that we are skeptical in regarding many psychological attributes as quantitative in the first place. One of the arguments for this is that it is unclear how an attribute would be decomposable. A stick that is one meter in length can be broken up, so that you are left with two sticks of half a meter in size. The 'length' attribute is the same, irrespective of the other qualities of the object. Similarly, a pound of butter can be split in two so that you have two bricks of butter that are half a pound in size. With psychological attributes, this kind of reasoning is quite difficult to imagine if you take into account what Georgescu-Roegen calls the qualitative residual that is left after . Take the concept of happiness. This, and other psychological attributes, are often imagined to be a derived measurement of sorts: a measurement that is defined as a cluster of more specific attributes. If one were to decompose happiness into two parts of equal size, there is always some kind of combination of attributes. You would need to make [3] For example, say you would need to balance eudaimonia and euphoria. For happiness to decrease equally, you would need to keep the balance of the two. This kind of quantitative decomposition can be easily imagined to be sustained to an infinite regress, resulting in having to balance or track an infinite number of decomposed items that should all be in some kind of relationship with each other .

It must also be said that we do feel some sympathy towards the viewpoint by Sijtsma(14).

[3] for a great

An hereto unnoticed (as far as I know) strong resemblance to the discussion from economists about the status of utility measures, giving extra credence to the charges that are often levied by social scientists of this school. For example, there is no connection to the work of the Romanian mathematician Georgescu-Roegen made in , but his

Note that this is also a problem for ordinal scaling.

There is a distinction between unobservable and observable properties made in psychology that pops up often, such as in the discussion of latent variables. This distinction is not without criticism. As Burgos notes, the distinction is a (4)

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

FUNDING

No external funding was used for this project.

ACKNOWLEDGMENTS

I acknowledge the work of my thesis supervisors, who introduced me to the method and left me free to pursue the project as I imagined it in all its weird and shifting shapes. The great help of the Julia community was also appreciated, as they have been pushing me forward where I got stuck and took the time to respond to my stupid questions. Finally, I would like to acknowledge the feedback and conversations between me and my thesis group, who have been working through my text and made sure that it is easy to follow and well-written. Special thanks to Giuliana Orizzonte, Daniel Anadria, dr. Derksen, dr. Grelli and dr. Bringmann for fruitful discussions and feedback about my topic. Lastly, I would like to thank my girlfriend, family, and friends for the mental support throughout.

DATA AVAILABILITY STATEMENT

The code, additional material, and generated data for this study can be found on GitHub.

REFERENCES

1 .Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM review* 59, 65–98

2 .Boker, S. M. (2002). Consequences of Continuity: The Hunt for Intrinsic Properties within Parameters of Dynamics in Psychological Processes. *Multivariate Behavioral Research* 37, 405–422. doi:10.1207/S15327906MBR3703_5

3 .Borsboom, D. and Mellenbergh, G. J. (2004). Why Psychometrics is Not Pathological: A Comment on Michell. *Theory & Psychology* 14, 105–120. doi:10.1177/0959354304040200

4 .Burgos, J. E. (2021). The Real Problem with Hypothetical Constructs. *Perspectives on Behavior Science* 44, 683–704. doi:10.1007/s40614-021-00311-0

5 .Datseris, G. (2018). DynamicalSystems.jl: A Julia software library for chaos and nonlinear dynamics. *Journal of Open Source Software* 3, 598. doi:10.21105/joss.00598

6 .Datseris, G. and Parlitz, U. (2022). *Nonlinear Dynamics: A Concise Introduction Interlaced with Code* (Cham, Switzerland: Springer Nature). doi:10.1007/978-3-030-91032-7

7 .Franz, D. J. (2022). “Are psychological attributes quantitative?” is not an empirical question: Conceptual confusions in the measurement debate. *Theory & Psychology* 32, 131–150. doi:10.1177/09593543211045340

8 .Krantz, D. H. and Suppes, P. (1971). *Foundations of Measurement* (New York : Academic Press)

9 .Luce, R. D. and Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology* 1, 1–27. doi:10.1016/0022-2496(64)90015-X

10 .Meehl, P. E. (2004). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Applied and Preventive Psychology* 11, 1. doi:10.1016/j.appsy.2004.02.001

11 .Michell, J. (1999). *Measurement in Psychology: A Critical History of a Methodological Concept*. Ideas in Context (Cambridge: Cambridge University Press). doi:10.1017/CBO9780511490040

12 .Molenaar, P. C. and Campbell, C. G. (2009). The New Person-Specific Paradigm in Psychology. *Current Directions in Psychological Science* 18, 112–117. doi:10.1111/j.1467-8721.2009.01619.x

13 .Olthof, M., Hasselman, F., and Lichtwarck-Aschoff, A. (2020). Complexity in psychological self-ratings: Implications for research and practice. *BMC Medicine* 18, 317. doi:10.1186/s12916-020-01727-2

14 .Sijtsma, K. (2012). Psychological measurement between physics and statistics. *Theory & Psychology* 22, 786–809. doi:10.1177/0959354312454353

15 .Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In *Handbook of Experimental Psychology* (Oxford, England: Wiley). 1–49

16 .Tafreshi, D. (2022). Sense and nonsense in psychological measurement: A case of problem and method passing one another by. *Theory & Psychology* 32, 158–163. doi:10.1177/09593543211049371

17 .Trendler, G. (2019). Conjoint measurement undone. *Theory & Psychology* 29, 100–128. doi:10.1177/0959354318788729

18 .Vessonen, E. (2021). Representation in measurement. *European Journal for Philosophy of Science* 11, 76. doi:10.1007/s13194-021-00365-6

FIGURES

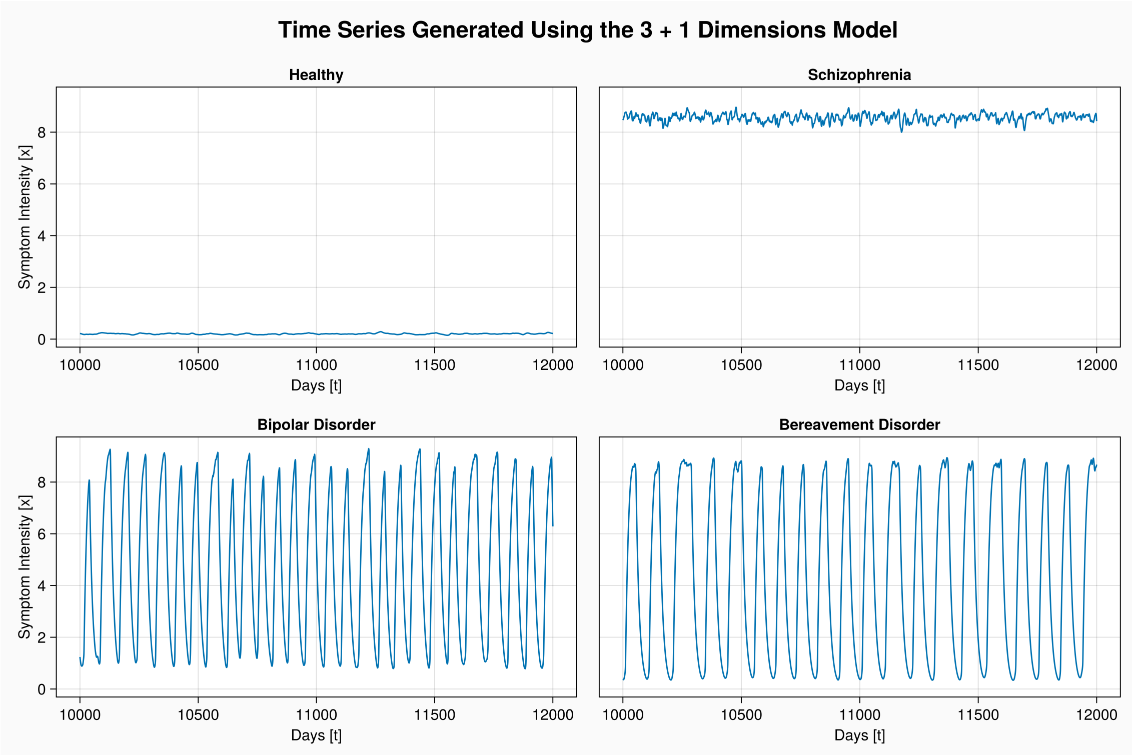


Figure 1. A section of the time series created using the coupled differential equations and parameter settings specified in section 2.1.1. This is the intact data, before degradation takes place.

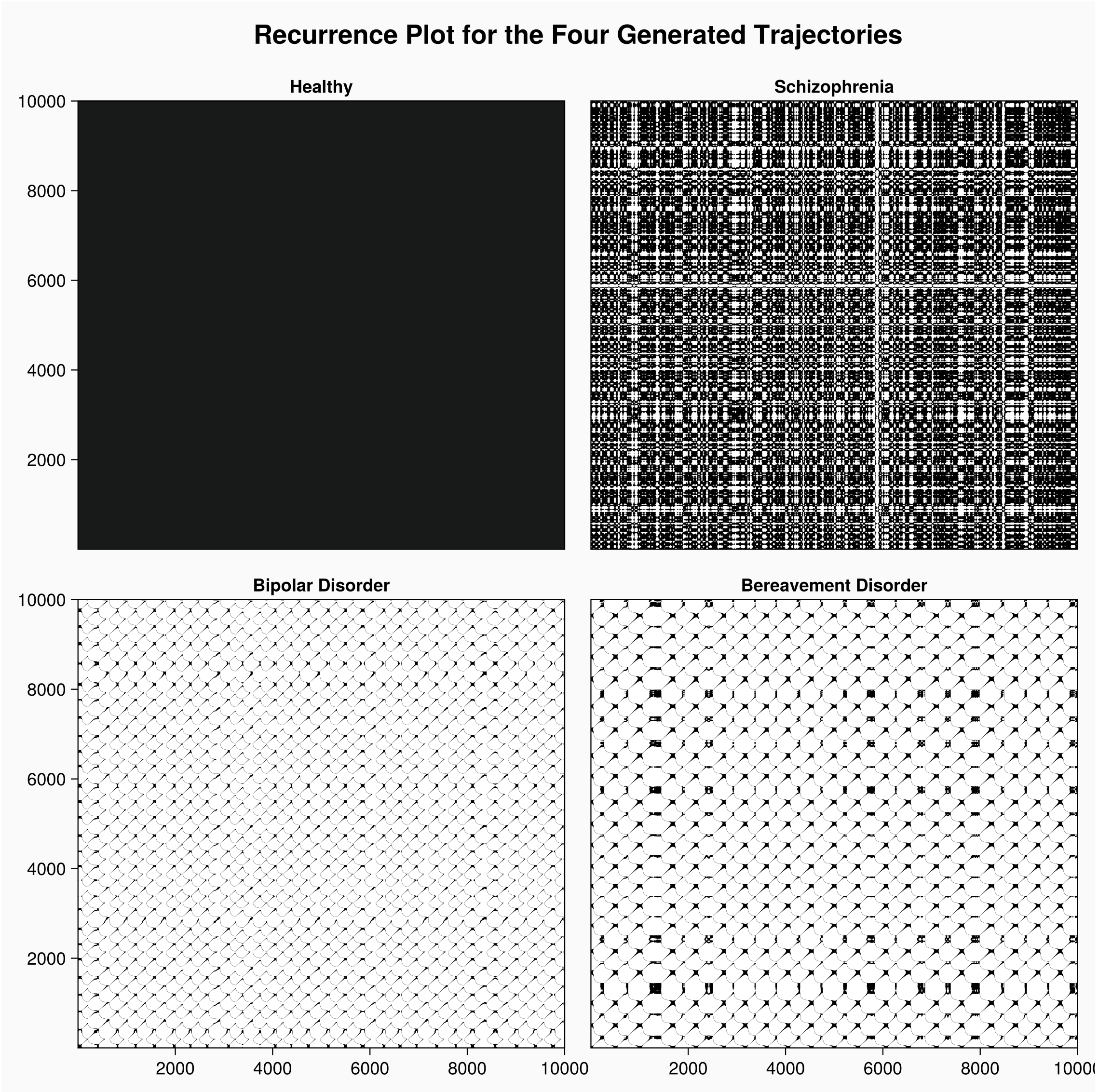


Figure 2. Recurrence plot for the four time series generated using the coupled differential equations and parameter settings specified in section 2.1.1. A point recurs when it is within the recurrence threshold of another point. Recurrent points are black, non-recurrent points are white. The axes represent time points, each location on the matrix represents a combination of time points. The recurrence threshold is set at 0.2 for illustration purpose. Note that the plot for the 'healthy' trajectory is completely black: this is because every point in the plot falls within the recurrence threshold. Also note the black 'boxes' where the bottom two trajectories are stagnant.

TABLES

Parameter	S_{max}	R_s	λ_s	τ_x	P	R_b	λ_b	L	τ_y	S	α	β	τ_z	λ_d	τ_f
Healthy	10	1	0.1	14	10	1.04	0.05	0.2	14	4	0.5	0.5	1	1	720
Schizophrenia	10	1	0.1	14	10	0.904	0.05	0.2	14	4	0.5	0.5	1	1	720
Bipolar	10	1	0.1	14	10	1.04	0.05	1.01	14	10	0.5	0.5	1	1	720
Bereavement	10	1	0.1	14	10	1	0.05	0.6	14	4.5	0.5	0.5	1	1	720

Table 1. The parameter settings used as initial parameter settings for the coupled differential equations specified in paragraph 2.1.1