# Network models of driver behaviour

Markus Mattsson

University of Helsinki, Traffic Research Unit

# **Abstract**

The way people behave in traffic is not always optimal from the road safety perspective: drivers exceed speed limits, misjudge speeds or distances, tailgate other road users or fail to perceive them. Such behaviours are commonly investigated using self-report-based latent variable models, and conceptualized as reflections of violation- and error-proneness. However, traffic safety work is unlikely to progress by attributing dangerous behaviour to stable properties of individuals, whereas investigating direct relationships between traffic behaviours offers a fruitful way forward. Network models of driver behaviour and background factors influencing behaviour were constructed using a large UK sample of novice drivers. The models show how individual violations, such as speeding, are related to and may contribute to individual errors such as tailgating and braking to avoid an accident. In addition, a network model of the background factors and driver behaviours is reported. Finally, a model predicting crashes based on prior behaviour was built and tested in separate datasets. This contribution helps to bridge a gap between experimental / theoretical studies and self-report-based studies in traffic research: the former have recognized the importance of focusing on relationships between individual driver behaviours, while network analysis offers a way to do so for self-report studies.

### Introduction

It has become something of a truism in human factors research that people contribute to road crashes by either deliberately violating rules or by making unintended errors. In traffic psychology, the perception of the importance of this distinction coincides with the development of a much-used questionnaire instrument, the Driver Behaviour Questionnaire (DBQ). A seminal study hypothesized that the distinction might be due to violations and errors being "mediated by different psychological mechanisms", and interpreted the results of a principal components analysis (PCA) from this point of view. The idea that there exist fundamentally different types of "aberrant behaviour" that need to be targeted by different types of intervention has since appeared in numerous research articles, some of which have been based on the errors / violations dichotomy, some of which have made finer distinctions between different types of errors and violations. Subsequent studies have commonly referred to Reason et al. as having proved the existence of two qualitatively different psychological mechanisms even though the authors of the original study were careful to note that the distinction was hypothetical in nature. The idea is echoed in statements such as: "Errors and violations result from different psychological processes (Reason et al., 1990) and therefore should be treated differently"2; "Since errors and violations result from different psychological processes, they should be treated differently (Reason et al., 1990)", and "As each type of behaviour has a distinct psychological

underpinning (Reason et al., 1990), different interventions are required to reduce their frequency and also associated crash risk"<sup>4</sup>. This idea is herein referred to as the *latent variable view of violations and errors*.

This received wisdom has gone unchallenged in the lack of methodological tools that would enable alternative explanations to be sought. This is problematic, since the different driving behaviours encoded as question items in the DBQ are likely to depend on a multitude of psychological mechanisms<sup>5, 6</sup>, some of which may differ from one person to another. Further, individual violations and errors may well be causally related: for example, exceeding the speed limit (a violation) may cause one to miss observing something (an error). These observations are difficult to accommodate within the latent variable view; thus, this contribution conceptualises driver behaviour in a novel manner as an interacting network of behaviours, emotional reactions and perceptions. Similar network models have been recently employed in research on psychopathology<sup>7, 8</sup>, personality<sup>9</sup>, attitudes<sup>10</sup> and intelligence<sup>11</sup>.

But first, let us look into some of the limitations of the latent variable view, some of which are theoretical and conceptual in nature, some practical and related to the use of statistical methods. Consider the common assertion, cited above, that dimensionality reduction produces insights about psychological mechanisms. However, a mechanistic explanation of a cognitive process involves accounting for the component operations of the process and their interaction<sup>12, 13</sup> while PCA and factor analysis (FA) are – stated simply – procedures for grouping correlated variables. In fact, offering violation- and error-proneness as explanations for why individuals break rules and commit errors amounts to reified circular reasoning<sup>14</sup> insofar as the "pronenesses" refer to nothing but being likely to behave in said manner. Further, the individual DBQ errors and violations are of such a wide variety that it is unclear how they could all be implemented by a single cognitive or psychological mechanism. The errors include missing observing something important, forgetting the road one was travelling, ending up on an often-used route when intending to travel elsewhere etc. Accordingly, it has been argued that these errors are differentially related to top-down and bottom-up attention, updating of a mental model of the driving situation and resolving the conflict between schemas in memory<sup>5</sup>. Similar considerations apply to individual violations.

In addition, the latent variable view involves certain statistical assumptions whose consequences seem unrealistic in practice. First, variation in the observed variables is viewed as due to either the latent variables or measurement error, and no covariance among the observed variables is allowed once these effects are accounted for<sup>15</sup>. The relation between latent and observed variables is often interpreted as causal, and attributing causal power to the latent variables renders the observed variables inert, interchangeable reflections of the latent properties being measured<sup>15</sup>. The idea that observed variables are rendered statistically independent when conditioning on the latent variable(s) is known as the assumption of local independence<sup>15</sup> and it is problematic in at least two cases: 1) when several latent variables prove to influence a given set of observed variables (lack of unidimensionality) and 2) when direct causal connections exist among the observed variables. The problematic nature of the former is concretely reflected in the common practice of representing errors and violations as unweighted or weighted<sup>16</sup> sum scores. Consider two imaginary persons filling in the DBQ: John, known for his quick

temper, answers the three items related to aggressive behaviour with the option "nearly all the time", and does not report performing other violations, thus obtaining the sum score of 21. Bill, on the other hand, known for his careful nature, chooses the option "never" to the aggression-related items and the option "hardly ever" or "occasionally" to the other violation items. As there are many more items related to non-aggressive violations than to aggressive ones, both respondents receive identical scores, even though their behavioural profiles are quite different. The latter question of direct causal connections among observed variables is less often discussed, and it motivates applying network models in in the first place. For instance, driving under the influence of alcohol or driving fast may predispose the driver to other violations and errors, and the consequences of these two violations are likely to be different.

Finally, previous research has indicated equivocal factorial structures for the DBQ. This is evidenced by

- 1) Different factor structures being obtained for the same version of the instrument. For instance, the often-used 27-item questionnaire is thought to reflect two<sup>17, 18</sup>, three<sup>19</sup> or four<sup>20</sup> different psychological processes, with parallel analyses sometimes indicating even more latent variables<sup>21</sup>;
- 2) Typically high cross-loadings of items on factors<sup>5, 22</sup>;
- 3) Complex factor structures needed to adequately fit the data, either by specifying second-order factors<sup>3</sup> or a general factor<sup>20</sup> and
- 4) Failures of the test of measurement equivalence across certain subgroups<sup>4, 5, 22</sup>.

Figure 1 shows schematically the fundamental assumptions of latent variable models. Individual behaviours, such as speeding or misjudging speed, reflect the level of the underlying latent variable and measurement error. Background factors, e.g. enjoying speed, serve to predict the level of the latent variables. Importantly, background factors are not related to individual driver behaviours, which are assumed causally inefficacious. Typically, a dependent variable of interest, such as the number of crashes, is regressed on the sum variables representing violations and errors <sup>16</sup>.

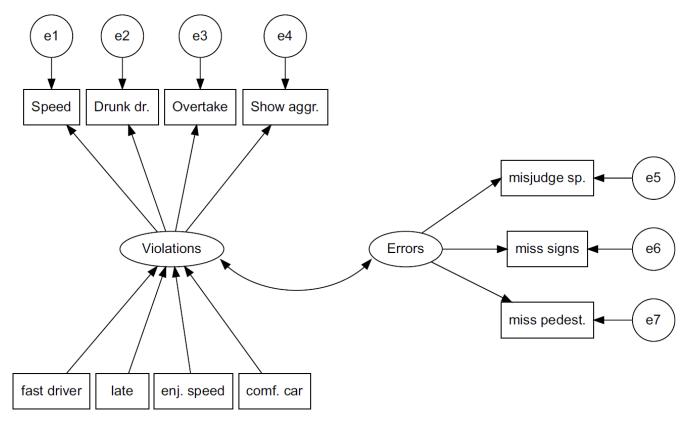


Figure 1 The latent variable view of violations and errors. Rectangles refer to observed variables, circles and ovals to latent variables. The ovals with the letter "e" and a number refer to error variances of the respective observed variables. The error variances (e1-e7) are assumed uncorrelated as per the assumption of conditional independence. Arrows pointing in one direction refer to causal associations, while the two-headed arrow represents a covariance relation. Predictors for errors are not shown.

Network models, such as the hypothetical model shown in Figure 2, are naturally suited to investigating direct associations between behaviours such as those described above. Individual behaviours are also directly affected by background factors. The absence of an edge between two nodes is interpreted as showing that when the effects of all other nodes in the network are controlled, the two nodes are rendered statistically independent<sup>23</sup>. The non-zero edges can be interpreted as potential causal connections<sup>23</sup>, logical relationships of entailment<sup>24</sup> or perhaps semantic relationships of the nodes being close to each other in meaning. Finally, clusters of nodes may reflect the effect of a latent variable influencing all the nodes in the cluster<sup>25</sup>.

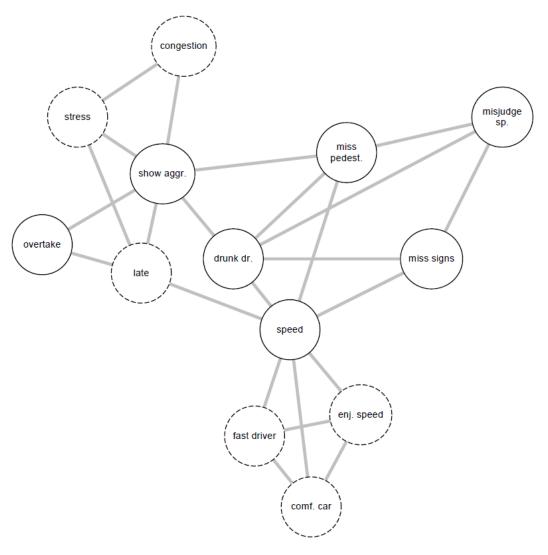


Figure 2 The network view of traffic behaviour. Traffic behaviours encoded in the DBQ are drawn using solid lines, background factors using dashed lines. Pairwise relationship among individual violations and errors, when controlling for the effects of all other variables, are shown as the edges of the network.

This report conceptualizes the behaviour of newly licensed drivers as an interacting network of component behaviours. First, differences between drivers during the first three years post-licensure are represented as the between-person network<sup>26</sup>, which enables assessing direct associations between behaviours and judging the centrality of the behaviours in the network. Second, relationships between driver behaviours and background factors are assessed in a cross-sectional analysis using data collected at six months post-licensure. Importantly, the background factors are directly linked to individual driver behaviours rather than latent variables or sum scores. Third, individual driver behaviours recorded at 6 months post-licensure are used for predicting crashes occurring during the following 2 ½ years. Different predictive models are compared by building the models in one subset of data and testing them in another one.

# Method

#### Data

This study is based on the archival dataset collected in the Cohort II study in 2001 - 2005 on new and novice drivers in the UK<sup>27</sup>. The initial sample size was 20,512 with four waves of data collection taking place after the respondents had passed their driving tests. The numbers of responses and response rates were as follows: 10,064 at 6 months (49 %), 7,450 at 12 months (36 %), 4,189 at 24 months (26 %) and 2,765 at 36 months (26 %). The age distribution of the respondents corresponded to that of the population of newly licensed drivers, with 59 % of respondents under the age of 20 at the first wave of data collection and 76 % under the age of 25. Women, on the other hand, were overrepresented: at the first wave, 64 % of the respondents were females. The dataset has been previously used in several studies  $^{16,18,20,29}$  and is freely available online  $^{28}$ .

The cross-sectional network analyses were performed on data obtained at six months post-licensure. Cases with non-missing data in all variables were included, resulting in a sample size of 8858. The respondents included in this analysis had a mean age of 22.51 years (SD = 7.95), and 64 % of them were female. The between-person network was based on 1173 cases with no missing data. The respondents had a mean age of 24.04 years (SD = 9.62), and 71 % of them were female. The regression analyses were based on cases with no missing values on any of the predictor variables or the number of crashes variable, resulting in N = 1152, 69 % female.

#### Measures

This study uses data collected using the Driving Experience Questionnaire (DEQ) that includes a 39-item version of the DBQ. In addition to the DBQ items, items referred to as "background factors" were used. The network analyses are based on the common 27-item version of the DBQ<sup>3,5,30</sup> together with items related to driving under the influence of alcohol and drugs, using the cell phone while driving and having to brake or swerve to avoid an accident, resulting in 31 DBQ items. The regression analyses were based on all 39 DBQ items to maximize predictive power.

The responses to the DBQ items were recorded on a six-point Likert scale ("never" to "nearly all the time"). The response options for the background factors were as follows. Drivers' self-image was assessed on a seven-point scale anchored to the end points of a continuum (e.g., "inattentive" to "attentive"). The self-perceived improvement needs were recorded on a three-point scale (e.g., "no improvement needed" to "a lot of improvement needed"). Attitudes were assessed on a five-point scale ranging from "Strongly disagree" to "Strongly agree" using the "Learning to drive questionnaire" filled in after the practical driving test, prior to responding to the DEQ. In addition, self-reported age, sex and mileage at six months post-licensure were used as predictors of public road crashes (both major and minor crashes) in the Poisson regression analyses.

### Statistical analyses

### Network analyses

Two types of network models were estimated. The cross-sectional network model was based on the 31 DBQ items together with nine items related to the background factors. The latter were chosen from among candidate variables using a series of exploratory network analyses described in the

Supplementary materials. The between-person network<sup>26, 31</sup> was formed by calculating average scores for each respondent across the four time points to represent the overall pattern of driving behaviours during the first years of learning to drive. The benefit of between-person networks is that they are less susceptible to spurious effects and reporting biases<sup>26</sup> such as mood-congruent recall<sup>32</sup> than cross-sectional networks as different biasing effects are likely to cancel each other out. Descriptive statistics of the DBQ variables at the four time points are shown in Table 1.

In both models, individual questionnaire items correspond to the nodes of the network, while the edges represent partial correlations controlling for all other nodes. The procedure attempts to uncover a graph known as a Gaussian Graphical Model in which each node is independent of the rest given the values of immediately neighbouring nodes as described by the Markov properties<sup>33</sup>. In other words, an edge connecting two nodes indicates their conditional dependence given the other nodes. The edge weights were scaled to the joint maximum value of the two models to ensure the comparability of the results. All network analyses were performed in R<sup>34</sup> using the packages ggraph<sup>35</sup> and bootnet<sup>36</sup>.

Some of the partial correlations are likely to differ from zero because of sampling variation and can be thought of as false positive findings<sup>9</sup>. For this reason, the graphical LASSO<sup>37</sup> was used in estimating the networks based on polychoric correlations. The procedure constrains low values of partial correlations to zero, thus resulting in sparse models<sup>23</sup>. The level of sparsity is determined by the tuning parameter  $\lambda$ , the value of which was chosen based on the Extended Bayesian Information Criterion (EBIC), as this has been shown to work well in retrieving the true network structure<sup>38</sup>, especially if the true model is sparse<sup>23</sup>. The hyperparameter  $\gamma$  used in EBIC model selection was set to the recommended default value of  $0.5^{38}$ . The locations of the nodes were determined using a modified version of the Fruchterman-Reingold algorithm<sup>39</sup> for weighted networks<sup>35</sup>, which places strongly connected nodes that have many edges in common close to one another.

The importance of the individual nodes in the network was assessed by calculating three indices of centrality: strength, betweenness and closeness. Strength centrality refers to the sum of edge weights of the focal node, closeness centrality to the reciprocal of the sum of distances from the focal node to all other nodes and betweenness to the number of shortest paths between two nodes that pass through the focal node. The values of the centrality indices were standardized to ensure comparability between networks and between studies. Further, the generalization of Zhang's local clustering coefficient<sup>40</sup> to signed networks<sup>41</sup> was calculated to represent the redundancy of the focal node. This coefficient was used due to its sensitivity to weak edges<sup>42</sup>. The stability of the centrality indices and the accuracy of edge weight estimates were assessed by bootstrap analyses (Supplementary materials) using the bootnet package<sup>36</sup>. Further, the stability of the indices was quantified by the correlation stability (CS) coefficient, the value of which should preferentially exceed 0.5<sup>36</sup>.

Table 1 DBQ item means and standard deviations at 6 months, 12 months, 24 months and 36 months after licensure. The calculations are based on cases with non-missing values on all DBQ variables on all occasions, resulting in N = 1173.

	Time								
	6 m	onths	12 m	onths	24 m	onths	36 m	onths	Change 6 months -
Variable	Mea	n (sd)	36 months						
Drive away from traffic lights at too high a gear	1.51	(0.69)	1.53	(0.66)	1.56	(0.67)	1.61	(0.68)	0.10
2 Overtake a slow driver on inside	1.30	(0.67)	1.31	(0.66)	1.40	(0.73)	1.46	(0.79)	0.16
3 Attempt to overtake and hadn't noticed signalling right	1.09	(0.32)	1.12	(0.36)	1.13	(0.40)	1.15	(0.40)	0.06
4 Forget where left car in carpark	1.71	(1.06)	1.76	(1.02)	1.84	(1.06)	1.88	(1.06)	0.17
5 Sound horn to indicate annoyance	1.55	(0.87)	1.64	(0.94)	1.72	(0.95)	1.79	(0.95)	0.24
6 Switch on one thing when meant to switch on other	1.58	(0.78)	1.52	(0.72)	1.50	(0.72)	1.49	(0.68)	-0.09
7 Pull out of junction so far that driver has to let you out	1.28	(0.58)	1.30	(0.57)	1.35	(0.66)	1.38	(0.66)	0.10
Realised you have no recollection of road you've been travelling	1.70	(0.91)	1.79	(0.88)	1.86	(0.89)	1.98	(0.94)	0.28
9 Crossed junction knowing lights have turned against you	1.36	(0.65)	1.41	(0.63)	1.49	(0.69)	1.53	(0.72)	0.17
10 Failed to notice people crossing when turned into sidestreet	1.22	(0.46)	1.24	(0.47)	1.28	(0.50)	1.30	(0.51)	0.08
11 Become angered by driver and given chase	1.12	(0.49)	1.11	(0.39)	1.15	(0.47)	1.16	(0.48)	0.04
12 Misread signs and taken wrong turning off roundabout	1.74	(0.80)	1.79	(0.80)	1.82	(0.76)	1.86	(0.77)	0.12
13 Have disregarded speed limit on a residential road	1.59	(0.85)	1.66	(0.90)	1.75	(0.96)	1.81	(0.97)	0.22
L4 When turning left have nearly hit cyclist on inside	1.06	(0.25)	1.05	(0.23)	1.08	(0.29)	1.08	(0.29)	0.02
L5 Used mobile phone without hands free kit	1.34	(0.68)	1.39	(0.73)	1.35	(0.70)	1.45	(0.79)	0.11
16 Stay in motorway lane know will be closed	1.12	(0.37)	1.18	(0.47)	1.24	(0.53)	1.29	(0.60)	0.17
17 When queuing to turn left nearly hit car in front	1.26	(0.52)	1.31	(0.56)	1.34	(0.55)	1.37	(0.58)	0.11
18 Drive when suspect over legal alcohol limit	1.05	(0.23)	1.07	(0.29)	1.08	(0.30)	1.08	(0.32)	0.03
19 Become angered by driver and indicate hostility	1.36	(0.68)	1.46	(0.76)	1.51	(0.78)	1.53	(0.79)	0.17
20 Misjudged speed of oncoming vehicle when overtaking	1.30	(0.55)	1.30	(0.52)	1.31	(0.53)	1.32	(0.55)	0.02
21 Hit something when reversing that hadn't seen	1.16	(0.42)	1.16	(0.40)	1.18	(0.44)	1.18	(0.44)	0.02
22 Raced away from traffic lights to beat other driver	1.46	(0.84)	1.53	(0.87)	1.58	(0.93)	1.60	(0.94)	0.14
Noticed ending up on a different road than intended	1.64	(0.83)	1.67	(0.84)	1.68	(0.81)	1.71	(0.82)	0.07
24 Get into wrong lane when approaching roundabout/junction	1.99	(0.74)	1.95	(0.72)	2.00	(0.70)	1.97	(0.72)	-0.02
25 Drive so close to car that would not be able to stop	1.41	(0.65)	1.44	(0.67)	1.52	(0.72)	1.54	(0.71)	0.13
26 Missed give way signs and narrowly avoided colliding with traffic	1.09	(0.29)	1.10	(0.32)	1.11	(0.33)	1.12	(0.34)	0.03
7 Have disregarded speed limit on motorway	1.61	(0.97)	1.79	(1.10)	1.99	(1.23)	2.12	(1.29)	0.51
28 Failed to check rear-view mirror before manoeuvring	1.59	(0.81)	1.61	(0.76)	1.57	(0.75)	1.60	(0.76)	0.01
29 Brake too quickly on slippery road or steer wrong in skid	1.20	(0.50)	1.20	(0.46)	1.25	(0.49)	1.27	(0.52)	0.07
30 Drove after taking drugs which affected you	1.01	(0.16)	1.02	(0.18)	1.02	(0.17)	1.03	(0.20)	0.02
31 Had to brake or swerve to avoid accident	1.39	(0.57)	1.41	(0.55)	1.46	(0.57)	1.49	(0.62)	0.10

### Regression analyses

The 39 DBQ variables, together with the variables mileage, age and sex, recorded at six months post-licensure, were used as predictors of the number of crashes during the latter waves of data collection in a Poisson regression model with a logarithmic link function. The analysis was based on the idea of maximizing predictive accuracy via minimizing generalized cross-validation error (GCVE, operationalized as the deviation score), as described by Chapman et al.<sup>43</sup>. The minimisation of the GCVE aims at creating models that explain the maximum amount of variance without overfitting the model to data at hand. This is achieved by trading some increase in bias to a reduction in variance <sup>43</sup>.

The data set was randomly split into a training set and a test set (75 / 25 ratio), with  $N_{training} = 864$  and  $N_{testing} = 288$ ). The uneven ratio was chosen to enable a sufficiently large number of cross-validation splits, with initial model fitting and cross-validation taking place within the training set, followed by fitting the same model in the test set using the R-package glmnet<sup>44</sup>. The same penalty, controlled by the

hyperparameter  $\lambda$ , was applied to all the predictors, which were standardized prior to analysis. Self-reported mileage was also log-transformed.

Three Poisson regression models were fit to the training and test data. First, an elastic net model<sup>45</sup>, was formed based on a grid search of optimal values of the hyperparameters  $\alpha$  and  $\lambda$ , the values of which were chosen by minimizing the value of the deviance statistic in a 10-fold cross-validation analysis. The elastic net combines penalties based on squared sums of regression coefficients (ridge) and the sum of their absolute values (lasso); it thus performs variable selection similarly to the lasso, and performs well with correlated predictors similarly to ridge regression<sup>45</sup>. Second, a ridge regression model (regularisation without variable selection) was obtained by taking the cross-validated value of  $\alpha$  and setting the value of  $\alpha$  to zero. Finally, the naive Poisson model with all predictors and no regularisation was fitted to the data.

Model fit was assessed using several descriptive statistics. First, residual variance was quantified using deviance residuals<sup>44</sup> and mean squared error; second, the similarity of the predicted values and the actual values were assessed by calculating their Pearson correlation; third, McFadden's Pseudo R<sup>2 46</sup> was used for assessing how much model fit improved from the null model; finally, the min-max index was calculated as  $MinMax = \sum_i \left(\frac{min(y_i,\hat{y}_i)}{max(y_i,\hat{y}_i)}\right) / n$ .

## Results

## **Network analyses**

The results reported below are based on the maximum number of cases available for the respective analyses. First, the between-person network is shown in Figure 3. It illustrates between-person differences in the connectivity of the driver behaviours during the first three years of learning to drive. Errors are shown as striped blue nodes and violations as striped red nodes. The presence of an edge between two nodes represents their conditional dependence when controlling for all other nodes.

First, in outline, violations and errors occupy different regions of the graph, and their distinction seems a sensible rough description of the data. However, if all violations and all errors were reflections of respective latent variables, we would expect all violations and all errors to be interconnected<sup>23</sup> and for all violations and errors to be independent. In contrast, thematically related violations are clustered together and connected by strong edges, aggression-related nodes (v2, v4, v9) and speeding-related nodes (v5, v11) being a case in point. The edges are interpreted as showing, e.g., that drivers who were more likely than others to exceed speed limits within residential areas (v5) were also more likely to do so on highways (v11). Similarly, nodes related to substance abuse (v8, v12) were connected to each other and few other behaviours, except using the cell phone while driving (v6). Similar considerations apply to errors. For instance, the nodes related to forgetting something (e3, e6) shared a strong edge and were connected to few other nodes except absent-mindedness (v13).

Second, certain violations and errors were connected by relatively strong edges. Notably, drivers who exceeded speed limits within residential areas (v5) were more likely than others to tailgate other drivers (e15). Further, the tailgating drivers were more likely to need to brake or swerve to avoid an accident

(e19). Similarly, crossing junctions against a red light (v3, a violation) was connected by roughly equally strong edges to other violations and errors.

Third, not all thematically related nodes were connected by strong edges, the two items related to perceiving traffic signs (e8, e16) being a case in point. Still, both shared strong edges with other errors; for instance, misreading signs (e8) with getting into a wrong lane (e14). Finally, a methodological note is in order: differences between edge strengths can be interpreted only if their confidence intervals are not excessively wide (the related bootstrap analyses are reported in the Supplementary materials).

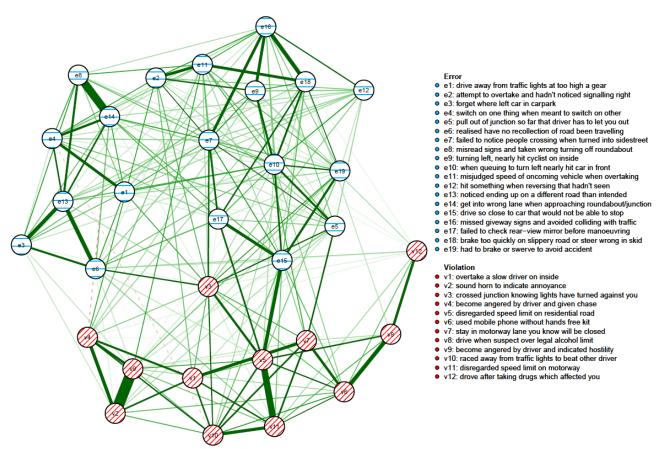


Figure 3 The between-person network model. The colours of the nodes correspond with the errors / violations dichotomy commonly used in DBQ studies. Green edges signify positive associations, dashed red edges negative associations. The wider and the more opaque the edge, the stronger the association.

Interpreting large networks such as those in Figure 3 becomes easier with examining indices of node centrality and clustering. The interpretability of the indices themselves depends on their stability, which can be judged by calculating the correlation stability coefficient performing bootstrap tests, both of which are reported in the Supplementary materials. Only the indices whose CS-coefficient exceeded the recommended value of  $0.5^{36}$  are shown in Figures 4 and 6.

Both strength centrality (associations with immediate neighbours) and closeness centrality (associations with all nodes) of the between-person network indicated the presence of a group of nodes that were especially central in the network. Closeness centrality is perhaps the more revealing in this context, as it capitalizes less on single strong edges. The node with the highest closeness centrality was v5

(speeding within a residential area), followed by nodes v3 (crossing junction on red), e15 (tailgating), e10 (queuing, nearly hit car), e17 (fail to check mirror) and v11 (speeding on motorway); nodes e10, v5, e15 and v3 had also high strength centralities. In addition, nodes e14 (getting into a wrong lane) and e7 (failing to notice pedestrians) had a high strength centrality. In general, the nodes along the path connecting speeding with various errors (v11-v5-v3-e7-e17-e15-e10 or v11-v5-e15 etc.) were central. Figure 4 also shows Zhang's clustering coefficient, according to which certain nodes, most notably v4 (become angered, give chase) and node e3 (forget where left car) contained little unique information.

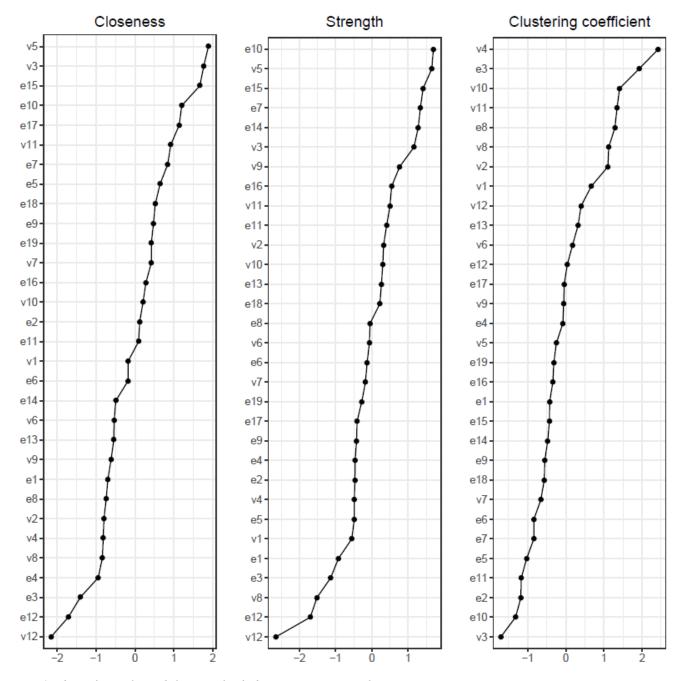


Figure 4 Indices of centrality and clustering for the between-person network.

The cross-sectional network model (Figure 5) shows the DBQ variables in relation to various background factors: attitudes (yellow), self-judged improvement needs (green) and self-image as a driver (lilac). The model is based on data collected at the first time point, 6 months post-licensure. It appears at the first sight quite different from the between-person model, but this is largely because the node placements are now different due the use of the Fruchterman-Reingold algorithm. Similarities between the two models are revealed by examining the connection strengths and the centrality indices (Figure 6). The speeding-related nodes (v5 and v11) were again central together with node e7. Node e16 (missing give way signs) now had a high strength centrality, while in the between-person network the corresponding centrality value was average; on the other hand, the connectivity patterns of the node were similar. The path v11-v5-e15 was present also in this model, even though node e15 was now slightly less central. A prominent difference between the models was that the aggression-related nodes v2, v4 and v9 had quite high values in at least one of the centrality indices in the cross-sectional model.

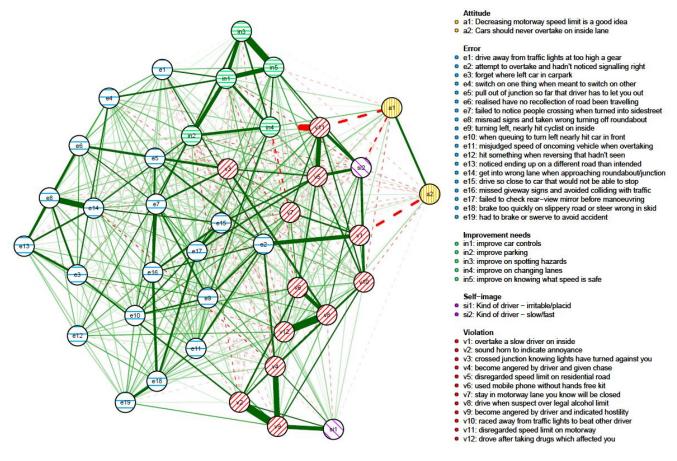


Figure 5 The cross-sectional network model at the first time point (6 months post-licensure). The relationships of the DBQ variables and background variables are shown.

Looking at the background factors, the self-judged improvement needs were strongly interconnected. They were, however, also related to driving behaviours in a revealing pattern. In general, they shared negative associations with different violations. In particular, the judged need for improvement in changing lanes (in4) was related to less speeding (v11) and less pushing into a lane (v7), while the perceived need of improving controlling the car (in1) was associated with having problems with gears (e1) and car controls (e4). Negative attitudes to speeding (a1) and overtaking on the inside (a2) were

related to fewer self-reported behaviours of those kinds (v11 and v1, respectively). On the other hand, perceiving oneself as a fast driver (si2) was positively associated with speeding (v5 and v11) and racing from the lights (v10). The drivers' self-judgment of themselves as irritable (si1) was closely associated with the anger-related nodes, even though the node was quite redundant as judged by the clustering coefficient. Self-perceived need of improvement in hazard-perception (in3) had a similarly high value of the clustering coefficient.

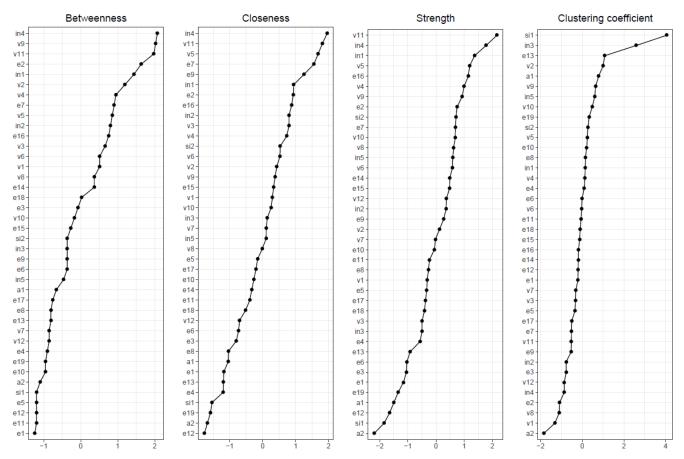


Figure 6 Indices of centrality and clustering for the cross-sectional network.

### **Regression analysis**

The self-reported driving behaviours, age, sex and mileage at the first time point were used in predicting subsequent crashes. As seen in Table 2, the naive Poisson model fit the training data best (as it should), but its performance deteriorated notably in the hold-out data. In fact, the model fit the hold-out data worse than the null model without predictors, offering a dramatic demonstration of how such models sometimes overfit data. The elastic net model fit the training data the worst, but its fit to the hold-out data was almost identical to that of the ridge regression model and is to be preferred due to its parsimony.

Table 2 Regression model fit in the training and test samples

Variable	Elastic net	Ridge	Naive Poisson			
Training sample						
Pearson r	0.246	0.260	0.317			
Mean Square Error	0.625	0.622	0.590			
Min-Max Index	0.121	0.122	0.128			
McFadden Pseudo R <sup>2</sup>	0.030	0.033	0.057			
Deviance	968.4	964.8	924.8			
Null model deviance		1018.1				
Hold-out sample						
Pearson r	0.155	0.160	0.086			
Mean Square Error	0.627	0.626	0.745			
Min-Max Index	0.158	0.157	0.156			
McFadden Pseudo R <sup>2</sup>	0.010	0.011	-0.044			
Deviance	317.9	317.2	348.4			
Null model deviance	323.6					

Table 3 shows the regression weights of the elastic net model that are low because of the dual penalization involved. The driving behaviours included in the model comprise various violations and errors. Variables related to failing to notice pedestrians, problems with car controls and different forms of aggressive behaviour had the highest regression weights. Hitting something when reversing, which is itself a minor crash, also predicted subsequent crashes.

Table 3 Regression weights of predicting accidents based on the elastic net regression analysis						
Variable	Regression weight					
Mileage	0.04					
Gender						
Age						
Drive away from traffic lights at too high a gear	0.07					
Overtake a slow driver on inside	0.04					
Have to confirm youre in right gear						
Attempt to overtake and hadnt noticed signalling right						
Forget where left car in carpark						
Sound horn to indicate annoyance	0.07					
Switch on one thing when meant to switch on other						
Change into wrong gear when driving along						
Pull out of junction so far that driver has to let you out	0.04					
Have used plates to warn drivers you are a new driver						
Realised have no recollection of road been travelling						
Crossed junction knowing lights have turned against you	0.01					
Failed to notice people crossing when turned into sidestreet	0.08					
Become angered by driver and given chase						
Misread signs and taken wrong turning off roundabout						
Drive in either too low or high gear for conditions						
Disregard speed limit on residential road	0.04					
When turning left have nearly hit cyclist on inside						
Used mobile phone without hands free kit						
Stay in motorway lane know will be closed						
When queueing to turn left nearly hit car in front						
Drive when suspect over legal alcohol limit						
Forget to take handbrake off before moving off	0.03					
Become angered by driver and indicate hostility	0.06					
Misjudge speed of oncoming vehicle when overtaking						
Hit something when reversing that hadn't seen	0.06					
Raced away from traffic lights to beat other driver	0.01					
Used a hands free kit						
Selected wrong gear when wanting to go in reverse						
Noticed on different road to destination want to go	-0.01					
Get into wrong lane when approaching roundabout/junction						
Drive so close to car that wouldn't be able to stop						
Forget headlights were on full beam						
Missed give-way signs and barely avoided collision	0.02					
Disregarded speed limit on motorway	0.01					
Failed to check rear-view mirror before manouevring						
Brake too quickly on slippery road or steer wrong in skid						
Drove after taking drugs which think affected you						
Brake/swerve to avoid accident	0.01					

# Discussion

The present contribution put forward three main ideas: 1) Individual violations and errors correlate because direct, possibly causal, relationships exist between perceptions, thoughts, emotions and actions in traffic. In particular, individual violations and errors do not reflect the effects of underlying, in principle unobservable variables, such as violation-proneness and error-proneness. 2) Hypotheses concerning these relationships can be formulated in a data-driven manner by estimating the strengths of associations between pairs of variables while controlling for all other variables in a network model. Similar ideas of network models as hypothesis-generating structures have been expressed in the context of psychopathology research<sup>8, 47</sup>. 3) Background factors such as attitudes, self-image etc. are directly related to individual traffic behaviours such as speeding or drunk driving; similarly, individual traffic behaviours are related to crash risk, and these relationships can be explored by developing criterion-keyed psychometric scales<sup>43</sup>. Finally, to understand different errors and violations, one needs to consider the context in which they take place rather than assuming that they reflect stable traits of the individuals (error-proneness and violation-proneness). The empirical part of the study dealt with the first three questions. Research related to the fourth is also discussed, as the context-sensitivity of behaviour is an important motivation for network models of human behaviour.

The relationships between individual violations and errors were examined by constructing the between-subject network as a representation of individual differences between novice drivers. The analysis showed that drivers who were more likely than others to exceed speed limits within residential areas were also more likely to tailgate the vehicle in front; further, the tailgating drivers were more likely than others to need to brake or swerve to avoid an accident. These associations were interpreted as causal hypotheses: drivers who exceed speed limits will likely catch up with the traffic flow and may as a consequence need to react. Similarly, crossing a junction against a red light (a violation) was associated with failing to notice pedestrians and with other errors. Importantly, such associations between violations and errors are difficult to accommodate within latent variable models of driver behaviour.

Exceeding speed limits within residential areas appeared as the most central node in the between-subject network. Thus, a successful intervention of reducing speeding might affect the other behaviours directly or indirectly linked to it. In fact, network models benefit traffic research precisely in that they encourage thinking of the dynamics of the behaviours. In contrast, latent variable models remain silent in this respect, as they conceptualize individual behaviours as causally passive indicators of the latent variables. Thus, if we take the latent variable view seriously, we can only influence individual behaviours through manipulating the latent variables: whether we want to reduce drunk driving or speeding, we should aim at the drivers' rule-breaking tendencies, because influencing an individual violation has no effect on other behaviours under the latent variable view.

The relationships between background factors and traffic behaviours were examined in the cross-sectional network model based on data collected six months post-licensure. Drivers who perceived themselves in need of improving lane-changing skills were less likely to report speeding and pushing into a lane. Further, the respondents reported behaving according to their attitudes: a preference of decreasing speed limits was associated with less speeding. On the other hand, perceiving oneself as a "fast driver" was associated with more speeding and racing from the lights. In short, individual background factors were related to driver behaviours in an understandable pattern.

More generally, network models show promise for bridging a gap between experimental and theoretical work in traffic research on the one hand and self-report-based research on the other hand. The latter has

commonly assumed the existence of a small number of mutually exclusive psychological mechanisms whose operation is reflected in respective sets of driver behaviours (e.g. violations and errors) that can be represented using latent variables or sum scores. On the other hand, the importance of individual driver behaviours, such as speeding, is recognized in theories of driver motivation<sup>48, 49</sup>, studies that aim at determining reasons for speeding<sup>50-52</sup> and engineering models of accidents<sup>53</sup>. Similarly, an error taxonomy involving action errors, cognitive and decision-making errors, observation errors and information retrieval errors has been proposed<sup>54</sup>, indicating the need to differentiate errors in a fine manner. Network models that focus on individual driving behaviours and their interrelationships offer a novel point of view from which to integrate the results of self-report studies with these lines of research.

Even though it is difficult to explain direct associations between driver behaviours using latent variable models, they pose a challenge to the network view: certain behaviours are more likely to occur together than others; why is this so if there are no latent variables? One can begin to answer by considering relationships between behaviour and environment, as illustrated by the following quote from a study on driver irritation and aggression:

"... Drivers who enjoy a somewhat faster speed than other drivers will more often be obstructed by other traffic, and therefore they will become irritated more often and be more likely to educate other road users. They probably also will become more irritated than other drivers when obstructed, because they want a faster progress"<sup>55</sup>.

In other words, people have characteristics such as emotions, attitudes and personality components<sup>56</sup> that affect their behaviour in traffic, which is not only something to be explained, but also a variable that feeds back into the system of emotions, perceptions and other behaviours. Further, not all behavioural characteristics are equally compatible with each other. For instance, the people described in the quote may be unlikely to make errors related to car controls, which are perhaps related to lack of experience or lack of interest in cars. In technical terms, network models exhibit non-trivial topology<sup>47</sup>. Further, they can – as also demonstrated in the present study – accommodate background variables such as beliefs, and account for why beliefs, feelings and behaviours become aligned<sup>10</sup>. For instance, the drivers described in the quote are perhaps likely to consider speed limits in general too low to avoid cognitive dissonance between their actions and beliefs.

The above example illustrated stable differences between people. However, people do not always behave in a stable manner; rather, their behaviour is context-dependent. The idea has much in common with the cognitive-affective personality system<sup>57</sup> (CAPS) theory developed within personality psychology, which posits if-then rules that describe how someone typically reacts in a certain type of a situation. CAPS is influenced by connectionist models, and characterises behaviours, memories and emotions as differently activated by the situational context and each other. The interconnected elements are described as a network in which activation is sustained by feedback loops. Such situation-specific rules might apply to, e.g., young people driving with their friends vs. with a mother and a baby on board. In the former situation, the driver's repertoire of certain risky behaviours is more highly activated and activation spreads through excitatory links. For example, the threshold of speeding might be lowered, activating an impulse to race other drivers. In short, the network of driving behaviours can be seen as being in two qualitatively different states distinguished by the connection strengths between

the behaviours. As another example, it has been shown that young mothers have an elevated crash risk when driving with an infant passenger compared to driving alone<sup>58</sup>; the dynamics of their behaviour in these contexts are likely to differ.

In addition to presenting network models of driver behaviour, this study involved predicting crashes from individual behaviours. In contrast, previous self-report studies have mainly used latent variables in crash prediction (even though see Wallén-Warner et al.<sup>59</sup> for an exception). The novel contributions of the present study were three-fold. First, crashes were truly predicted from earlier behaviour. Second, predictive models were first fit in a training subsample and then verified in an independent subsample. Third, regularised regression was used to prevent overfitting.

Three predictive models were tested, out of which the elastic net model<sup>45</sup> and the ridge regression model fit the hold-out data roughly equally well, and notably better than the naïve Poisson regression model. The elastic net model has the benefit of being more parsimonious than the ridge regression model, so it is to be preferred, other things equal. In the elastic net model, predictors included displaying anger while driving and driving fast (disregarding speed limits and racing from traffic lights). Further, errors in visual search (failing to notice people, missing signs) and in controlling the car (driving off at wrong gear, forgetting the handbrake) were included in the model. Interestingly, hitting something when reversing, which is itself a minor crash<sup>16</sup>, predicted future crashes. Remarkably, the naive Poisson model including all predictors fit the hold-out subsample worse than the null model with no predictors, offering a dramatic demonstration of the dangers of overfitting.

The present study has its limitations. First, network models are motivated by modelling the *components* of a phenomenon, with a component defined as something having unique causal relations with the rest of the network<sup>56</sup>. The DBQ has been psychometrically developed to maximize reliability, which has resulted in a certain redundancy of the items. In this study, this shows as high values of the clustering coefficient for the nodes related to aggression and speeding, which could in future studies be represented by single nodes. On the other hand, developing a novel questionnaire describing causally related behaviours, thoughts and emotional reactions is another option. Further, the behaviours currently examined are influenced by other road users, which could not be accounted for. For instance, aggressive behaviour is difficult to understand without knowing something about its target. It is naturally possible to take an even more critical perspective and argue that none of the relationships indicate potential causal relationships and driver behaviours should continue to be viewed as being caused by latent variables. Even so, the present study presents a challenge: why are variables that commonly load on different factors strongly correlated?

In addition to the substantive questions, some methodological decisions were problematic. First, polychoric correlations that estimate normally distributed variables underlying ordinal input variables were used, even though the variables were positively skewed. Although the practice is widespread in psychometric network models, it is not optimal<sup>60</sup>. Further, listwise deletion of missing data was performed; less wasteful methods are under development for network models but not yet available<sup>60</sup>. Further, it has been shown that young respondents may give exaggerated responses to questions they find funny, a response bias dubbed "mischievous responding"<sup>61</sup>; this could affect items such as drug use while driving. On the other hand, if such biases are transient in nature, between-person networks are likely a suitable method to use<sup>26</sup>.

Choosing the correct components of driver behaviour is a central issue to be tackled in future network models of driver behaviour. In addition, future studies should aim at developing a network theory of driver behaviour instead of merely applying a novel modelling tool. A recently developed intricate error<sup>54</sup> might provide a good starting point together with factors contributing to such errors. Also, the need for self-report studies will remain in the future even though studies involving instrumented vehicles have become ever more intricate<sup>62</sup>, because physical measurements are underdetermined psychologically<sup>29</sup>: for instance, a sudden acceleration can be due to either racing from the lights or unfamiliarity with car controls.

Further, the context-dependent nature of driving needs to be taken into account in future studies. Drivers may behave differently depending on who they are traveling with; similarly, investigating drivers' developing situation-awareness of when to desist from violating rules has been called for<sup>29</sup>, and network models are an ideal tool for investigating such developmental trajectories<sup>31</sup>. Another fascinating future direction is to consider the effects of the driver's state: being fatigued, intoxicated, stressed, in a hurry or in a strong emotional state could conceivably cause the network of driver behaviours to occupy qualitatively different states. Existing task analyses and models of driving situations<sup>63, 64</sup> are likely to be a good starting point, because individuals may behave in a more-or-less stable manner in a certain type of a situation, but not across situations<sup>57</sup>. Intriguing avenues for future research await those willing to look into the networks of driver behaviour.

# Acknowledgments

I would like to thank Jami Pekkanen and Otto Lappi for constructive feedback on earlier versions of the manuscript and Jami Pekkanen for helpful discussions on performing the regression analyses.

# References

- 1. Reason, J., Manstead, A., Stradling, S., Baxter, J. & Campbell, K. Errors and violations on the roads: a real distinction? *Ergonomics* **33**, 1315-1332 (1990).
- 2. Mesken, J., Lajunen, T. & Summala, H. Interpersonal violations, speeding violations and their relation to accident involvement in Finland. *Ergonomics* **45**, 469-483 (2002).
- 3. Lajunen, T., Parker, D. & Summala, H. The Manchester Driver Behaviour Questionnaire: A cross-cultural study. *Accident Analysis and Prevention* **36**, 231-238 (2004).
- 4. Stephens, A. & Fitzharris, M. Validation of the Driver Behaviour Questionnaire in a representative sample of drivers in Australia. *Accident Analysis & Prevention* **86**, 186-198 (2016).
- 5. Mattsson, M. Investigating the factorial invariance of the 28-item DBQ across genders and age groups: An Exploratory Structural Equation Modeling Study. *Accident Analysis & Prevention* **48**, 379-396 (2012).
- 6. Koppel, S. *et al.* The Driver Behaviour Questionnaire for older drivers: Do errors, violations and lapses change over time? *Accident Analysis & Prevention* **113**, 171-178 (2018).
- 7. Borsboom, D. & Cramer, A. O. Network analysis: an integrative approach to the structure of psychopathology. *Annual review of clinical psychology* **9**, 91-121 (2013).
- 8. Fried, E. I. & Cramer, A. O. Moving forward: challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science* **12**, 999-1020 (2017).
- 9. Costantini, G. *et al.* State of the aRt personality research: A tutorial on network analysis of personality data in R. *Journal of Research in Personality* **54**, 13-29 (2015).

- 10. Dalege, J. *et al.* Toward a formalized account of attitudes: The Causal Attitude Network (CAN) model. *Psychol. Rev.* **123**, 2 (2016).
- 11. Van Der Maas, H. *et al.* A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychol. Rev.* **113**, 842-861 (2006).
- 12. Lappi, O. & Rusanen, A. Turing machines and causal mechanisms in cognitive science. *Causality in the sciences*, 224-239 (2011).
- 13. Bechtel, W. Mechanisms in Cognitive Psychology: What Are the Operations? *Philosophy of Science* **75**, 983-994 (2008).
- 14. Boag, S. Explanation in personality psychology: "Verbal magic" and the five-factor model. *Philosophical Psychology* **24**, 223-243 (2011).
- 15. Borsboom, D., Mellenbergh, G. J. & Van Heerden, J. The theoretical status of latent variables. *Psychol. Rev.* **110**, 203 (2003).
- 16. de Winter, J., Dodou, D. & Stanton, N. A quarter of a century of the DBQ: some supplementary notes on its validity with regard to accidents. *Ergonomics* **58**, 1745-1769 (2015).
- 17. De Winter, J. Small sample sizes, overextraction, and unrealistic expectations: A commentary on M. Mattsson. *Accident Analysis & Prevention* **50**, 776-777 (2013).
- 18. de Winter, J. & Dodou, D. The Driver Behaviour Questionnaire as a predictor of accidents: A meta-analysis. *J. Saf. Res.* **41**, 463-470 (2010).
- 19. Parker, D., Reason, J. T., Manstead, A. S. & Stradling, S. G. Driving errors, driving violations and accident involvement. *Ergonomics* **38**, 1036-1048 (1995).
- 20. Rowe, R., Roman, G. D., McKenna, F. P., Barker, E. & Poulter, D. Measuring errors and violations on the road: A bifactor modeling approach to the Driver Behavior Questionnaire. *Accident Analysis & Prevention* **74**, 118-125 (2015).
- 21. Stanojević, P., Lajunen, T., Jovanović, D., Sârbescu, P. & Kostadinov, S. The driver behaviour questionnaire in South-East Europe countries: Bulgaria, Romania and Serbia. *Transportation Research Part F: Traffic Psychology and Behaviour* **53**, 24-33 (2018).
- 22. Mattsson, M., Lajunen, T., Gormley, M. & Summala, H. Measurement invariance of the Driver Behavior Questionnaire across samples of young drivers from Finland and Ireland. *Accident Analysis & Prevention* **78**, 185-200 (2015).
- 23. Epskamp, S. & Fried, E. I. A tutorial on regularized partial correlation networks. *arXiv preprint arXiv:1607.01367* (2016).
- 24. Kossakowski, J. J. *et al.* The application of a network approach to Health-Related Quality of Life (HRQoL): introducing a new method for assessing HRQoL in healthy adults and cancer patients. *Quality of Life Research* **25**, 781-792 (2016).
- 25. Golino, H. F. & Epskamp, S. Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PloS one* **12**, e0174035 (2017).
- 26. Costantini, G. *et al.* Stability and variability of personality networks. A tutorial on recent developments in network psychometrics. *Personality and Individual Differences* (2017).
- 27. Wells, P., Tong, S., Grayson, G. & Jones, E. Cohort II-A study of learner and new drivers-volume 1-main report, volume 2-questionnaires and data tables. *Road Safety Research Report* (2008).
- 28. Roman, G. D., Poulter, D., Barker, E., McKenna, F. P. & Rowe, R. Novice drivers' individual trajectories of driver behavior over the first three years of driving. *Accident Analysis & Prevention* **82**, 61-69 (2015).
- 29. Transport Research Laboratory. Safety, Security and Investigations Division & Poulter, D. Cohort II: a Study of Learner and Novice Drivers, 2001-2005. (2015).
- 30. Mattsson, M. On testing factorial invariance: A reply to J.C.F. de Winter. *Accident Analysis & Prevention* **63**, 89-93 (2014).
- 31. Epskamp, S., Waldorp, L. J., Mõttus, R. & Borsboom, D. Discovering psychological dynamics: The Gaussian graphical model in cross-sectional and time-series data. *arXiv preprint arXiv:1609.04156* (2016).
- 32. Shiffman, S., Stone, A. A. & Hufford, M. R. Ecological momentary assessment. *Annu.Rev.Clin.Psychol.* **4**, 1-32 (2008).
- 33. Lauritzen, S. L. in *Graphical models* (Clarendon Press, 1996).

- 34. Team, R. C. R: A language and environment for statistical computing. (2013).
- 35. Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D. & Borsboom, D. qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software* **48**, 1-18 (2012).
- 36. Epskamp, S., Borsboom, D. & Fried, E. I. Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, 1-18 (2017).
- 37. Friedman, J., Hastie, T. & Tibshirani, R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**, 432-441 (2008).
- 38. Foygel, R. &Drton, M. *Extended Bayesian information criteria for Gaussian graphical models* (Advances in neural information processing systems, 2010).
- 39. Fruchterman, T. & Reingold, E. Graph drawing by force-directed placement. *Software: Practice and experience* **21**, 1129-1164 (1991).
- 40. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* **4** (2005).
- 41. Costantini, G. & Perugini, M. Generalization of clustering coefficients to signed correlation networks. *PloS one* **9**, e88669 (2014).
- 42. Saramäki, J., Kivelä, M., Onnela, J., Kaski, K. & Kertesz, J. Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E* **75**, 027105 (2007).
- 43. Chapman, B. P., Weiss, A. & Duberstein, P. R. Statistical learning theory for high dimensional prediction: Application to criterion-keyed scale development. *Psychol. Methods* **21**, 603 (2016).
- 44. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw* **33**, 1-22 (2010).
- 45. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**, 301-320 (2005).
- 46. Mittlböck, M. & Schemper, M. Explained variation for logistic regression. Stat. Med. 15, 1987-1997 (1996).
- 47. Borsboom, D. A network theory of mental disorders. World Psychiatry 16, 5-13 (2017).
- 48. Fuller, R. Towards a general theory of driver behaviour. Accident Analysis & Prevention 37, 461-472 (2005).
- 49. Summala, H. in Modelling driver behaviour in automotive environments 189-207 (Springer, 2007).
- 50. Lawton, R., Parker, D., Stradling, S. G. & Manstead, A. S. Self-reported attitude towards speeding and its possible consequences in five different road contexts. *Journal of Community & Applied Social Psychology* **7**, 153-165 (1997).
- 51. Parker, D., Manstead, A. S., Stradling, S. G., Reason, J. T. & Baxter, J. S. Intention to commit driving violations: An application of the theory of planned behavior. *J. Appl. Psychol.* **77**, 94 (1992).
- 52. Warner, H. W. & Åberg, L. Drivers' decision to speed: A study inspired by the theory of planned behavior. *Transportation Research Part F: Traffic Psychology and Behaviour* **9**, 427-433 (2006).
- 53. Abdel-Aty, M. A. & Radwan, A. E. Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention* **32**, 633-642 (2000).
- 54. Stanton, N. A. & Salmon, P. M. Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems. *Saf. Sci.* **47**, 227-237 (2009).
- 55. Björklund, G. M. Driver irritation and aggressive behaviour. *Accident Analysis & Prevention* **40**, 1069-1077 (2008).
- 56. Cramer, A. O. *et al.* Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality* **26**, 414-431 (2012).
- 57. Mischel, W. & Shoda, Y. Reconciling processing dynamics and personality dispositions. *Annu. Rev. Psychol.* **49**, 229-258 (1998).
- 58. Maasalo, I., Lehtonen, E. & Summala, H. Young females at risk while driving with a small child. *Accident Analysis & Prevention* **108**, 321-331 (2017).
- 59. Warner, H. W., Özkan, T., Lajunen, T. & Tzamalouka, G. Cross-cultural comparison of drivers' tendency to commit different aberrant driving behaviours. *Transportation research part F: traffic psychology and behaviour* **14**, 390-399 (2011).
- 60. Epskamp, S. Network psychometrics. (2017).

- 61. Robinson-Cimpian, J. P. Inaccurate estimation of disparities due to mischievous responders: Several suggestions to assess conclusions. *Educational Researcher* **43**, 171-185 (2014).
- 62. Precht, L., Keinath, A. & Krems, J. F. Identifying effects of driving and secondary task demands, passenger presence, and driver characteristics on driving errors and traffic violations—Using naturalistic driving data segments preceding both safety critical events and matched baselines. *Transportation research part F: traffic psychology and behaviour* **51**, 103-144 (2017).
- 63. Fastenmeier, W. & Gstalter, H. Driving task analysis as a tool in traffic safety research and practice. *Saf. Sci.* **45**, 952-979 (2007).
- 64. Oppenheim, I. & Shinar, D. A context-sensitive model of driving behaviour and its implications for invehicle safety systems. *Cognition, Technology & Work* **14**, 261-281 (2012).