

The combined structural equation modeling-model selection approach: a user guide for ecologists and evolutionary biologists

Journal:	<i>Methods in Ecology and Evolution</i>
Manuscript ID	Draft
Manuscript Type:	Review
Date Submitted by the Author:	n/a
Complete List of Authors:	Garrido, Mario; Ben-Gurion University of the Negev - Sede Boqer Campus, Mitrani Department of Desert Ecology; Jacob Blaustein Institutes for Desert Research Hawlena, Hadas; Ben-Gurion University of the Negev - Sede Boqer Campus
Keywords:	Evolutionary Biology, Species interactions < Community Ecology, Interaction < Population Ecology, Diversity < Community Ecology, Food webs < Community Ecology, Statistics, Linear Models < Statistics, Microbial Ecology
Abstract:	<p>1. As part of the efforts to understand the processes shaping natural communities, ecologists and evolutionary biologists have increasingly been taking a top-down approach. Accordingly, they first apply statistical analyses to quantify the associations between and within the measured biotic and abiotic factors and then usually design experiments to test specific hypotheses uncovered by these analyses. In this paper, we discuss an approach combining structural equation modeling (SEM) and model selection methodology that we believe constitutes a satisfactory means to address the statistical challenges that top-down approaches often entail.</p> <p>2. We first introduce the combined SEM-model selection approach and emphasize its advantages. We further provide a practical guide for ecologists and evolutionary biologists interested in applying this combined approach in their research and illustrate its capabilities compared to those of univariate, traditional SEM, and stepwise SEM analyses, using data from a natural multihost-multiparasite system.</p> <p>3. The comparison among the four above approaches illustrates that the combined approach is more sensitive and provides a more thorough investigation of mediation effects than its alternatives.</p> <p>4. By providing this systematic overview of tentative solutions to practical challenges faced, we hope that the combined SEM-model selection approach will become more accessible to those interested in the analysis of complex ecological and evolutionary processes, and its use will become uniform among such studies.</p>

SCHOLARONE™
Manuscripts

For Review Only

1 **The combined structural equation modeling-model selection approach: a user guide for**
2 **ecologists and evolutionary biologists**

3 Running title: SEM-Model selection guide for practitioners

4 Word count: **6951 words**, from Abstract including table and figure legends

5

6 Mario Garrido* and Hadas Hawlena

7 Mitrani Department of Desert Ecology, Jacob Blaustein Institutes for Desert Research, Ben-Gurion University
8 of the Negev, Midreshet Ben-Gurion, Israel

9

10 *Corresponding author name and address:

11 Mario Garrido

12 Mitrani Department of Desert Ecology, Jacob Blaustein Institutes for Desert Research

13 Ben-Gurion University of the Negev, Midreshet Ben-Gurion 84990 ISRAEL

14 Email: gaiarrido@gmail.com; phone: (+972) 08-6596854

15 Abstract

- 16 1. As part of the efforts to understand the processes shaping natural communities,
17 ecologists and evolutionary biologists have increasingly been taking a top-down
18 approach. Accordingly, they first apply statistical analyses to quantify the associations
19 between and within the measured biotic and abiotic factors and then usually design
20 experiments to test specific hypotheses uncovered by these analyses. In this paper, we
21 discuss an approach combining structural equation modeling (SEM) and model
22 selection methodology that we believe constitutes a satisfactory means to address the
23 statistical challenges that top-down approaches often entail.
- 24 2. We first introduce the combined SEM-model selection approach and emphasize its
25 advantages. We further provide a practical guide for ecologists and evolutionary
26 biologists interested in applying this combined approach in their research and
27 illustrate its capabilities compared to those of univariate, traditional SEM, and
28 stepwise SEM analyses, using data from a natural multihost-multiparasite system.
- 29 3. The comparison among the four above approaches illustrates that the combined
30 approach is more sensitive and provides a more thorough investigation of mediation
31 effects than its alternatives.
- 32 4. By providing this systematic overview of tentative solutions to practical challenges
33 faced, we hope that the combined SEM-model selection approach will become more
34 accessible to those interested in the analysis of complex ecological and evolutionary
35 processes, and its use will become uniform among such studies.

36

37

38

39

40 **Keywords**

41 Complex ecological systems, evolutionary studies, model selection, multimodal inference,
42 multivariate statistical analyses, path analysis, review, structural equation modeling, top-
43 down approach

44

45 **Introduction**

46 Traditionally, ecologists and evolutionary biologists have addressed the inherent complexity
47 of natural systems by dividing the target system into its simplest parts as a tool to achieve a
48 workable understanding. However, in nature, interactions between organisms, as well as
49 between organisms and their environment, are mostly nonlinear and interdependent, and are
50 highly determined by indirect effects (e.g., Pilosof *et al.* 2014). To approximate the real
51 scenarios found in nature, in the last three decades, there has been a growing tendency to take
52 a top-down approach. Accordingly, researchers first get an overview by applying statistical
53 analyses to quantify the biological associations between and within the sampled biotic and
54 abiotic factors (e.g., CERoPath project; Herbreteau *et al.* 2011). Then, they move to the
55 bottom of the mechanism by addressing specific hypotheses derived from the field patterns
56 (Pilosof *et al.* 2014).

57 There is no doubt that the top-down approach brings us closer to the real scenarios
58 found in nature than the traditional approach, by simultaneously exploring all components
59 rather than studying them separately and independently. This approach has enhanced our
60 understanding of plant community composition (e.g. Coomes & Allen 2007), population
61 dynamics (e.g. Belinchon *et al.* 2017), host-parasite interactions (Pilosof *et al.* 2014), and
62 some aspects of spatial ecology (e.g. Bunnefeld *et al.* 2011). However, simultaneously
63 considering multiple organisms and environmental conditions is challenging, from both
64 empirical and statistical perspectives. In particular, the large scales and multiple dimensions

65 that the top-down approach entails limit the experimental potential of the studies, making
66 them highly dependent on field patterns determined by statistical analyses.

67 Various statistical approaches can be applied to quantify the associations between
68 multiple biotic and abiotic factors, including (but not limited to) principal components
69 analysis, cluster analysis, discriminant analysis, canonical correlation, canonical
70 correspondence analysis, MANOVA, and structural equation modeling (SEM) (e.g., Grace
71 2006). The overall goal of this paper is to emphasize the advantages of an approach
72 combining SEM and model selection methodology, which we believe constitutes one of the
73 most satisfactory solutions for addressing the statistical challenges entailed by the top-down
74 approaches. While the combined approach is powerful in facilitating inferences regarding
75 multivariate associations (Johnson 2002; Cardon *et al.* 2011), the current lack of consistency
76 in its use increases the likelihood of encountering pitfalls, restricts comparability between
77 studies, and makes the conclusions and interpretations of each study more subjective.

78 We start by introducing the rationale of the SEM and model selection approaches
79 separately and then of the integration between the two; we next discuss the advantages of the
80 combined approach over more traditional methods. We then introduce step-by-step guidelines
81 for practitioners in ecology and evolutionary biology who wish to apply the combined
82 approach to their data (Fig. 1). We finally use data from a multihost-multiparasite system to
83 illustrate the application and the added insights gained from the combined approach when
84 studying complex ecological communities. By providing this systematic overview of possible
85 solutions to practical challenges faced, we hope that the combined SEM-model selection
86 approach will become more accessible to those interested in the analysis of complex
87 ecological and evolutionary processes and its use will become uniform among such studies.

88

89 **The combined SEM-model selection approach**

90 SEM is a broad family of multivariate statistical techniques that use two or more statistical
91 dependencies, subject to causal interpretation, to model simultaneously influences and
92 responses (Grace 2006). In contrast to traditional univariate generalized linear model (GLM)
93 analyses (e.g., Fig. 2A), SEM distinguishes between direct and indirect effects and may
94 include multiple response variables and the associations between them (Grace 2006; Hoyle
95 2012). In addition, it is very flexible under a wide range of conditions—incorporating
96 unobservable data, as well as handling data with missing cells, a diverse array of residual
97 distributions, and different data designs (e.g., hierarchical models and repeated measures
98 design). Its intuitive connection to the way we conceive our study systems makes it a useful
99 technique for ecologists and evolutionary biologists (Grace 2006; Fan *et al.* 2016). The SEM
100 family includes path analysis, confirmatory and exploratory factor analysis, partial least
101 squares path modeling, and latent growth modeling, but from all the family members, path
102 analysis constitutes, by far, the most popular and accessible technique used by ecologists
103 (Kline 2016).

104 Traditionally in SEM, model fit, importance of paths, and hence statistical inference
105 are based on statistics (e.g. χ^2 , RMSEA, Z; Grace 2006; Fan *et al.* 2016), which can be further
106 used for a comparison between competing nested models, in a process analogous to backward
107 elimination in univariate models (stepwise SEM; e.g., Bentler & Bonett 1980). However,
108 while the traditional and stepwise SEM approaches are straightforward and simple (e.g., Fig.
109 2B), they suffer from a dependency of result interpretations on an artificial probability
110 threshold (usually $p < 0.05$) and are highly dependent on the level of model complexity (i.e.,
111 the number of parameters and the number and complexity of interactions) and on the sample
112 size (Bentler & Bonett 1980; Hu & Bentler 1999). The outcome of such fit-dependent
113 comparisons is a single model, which depends on the sequence of parameter or path deletion

114 and ignores model uncertainty (Bentler & Bonett 1980; Hu & Bentler 1999; Whittingham *et*
115 *al.* 2006) .

116 To overcome these drawbacks, SEM models can be complemented by a model
117 selection approach (Burnham & Anderson 2002). In this approach, the researcher develops an
118 *a priori* set of hypotheses (represented as alternative models) and quantifies the databased
119 evidence for and the ranking of each hypothesis. The models that minimize the loss of
120 information about the full reality, as reflected in a lower Akaike information criteria (AIC)
121 index or any other information theory indices (e.g., cAIC, BIC, DIC), receive the highest
122 ranks (Burnham & Anderson 2002). The information theory indices rank the models by
123 balancing between fit and precision, considering model complexity, and avoiding the use of
124 artificial thresholds. In some cases, the data are best explained by a single model, and results
125 can be inferred solely from it. In other cases, information should be inferred from the set of
126 models that best fit the data, and the results are interpreted in light of the relative likelihood
127 of each model, given the data and the set of models (relative weight w_i ; Johnson & Omland
128 2004).

129 These advantages make the combined SEM-model selection approach a powerful way
130 to evaluate alternative causal hypotheses regarding complex interactions between variables.
131 As such, this approach has been deployed across several ecological and evolutionary fields to
132 investigate the relative importance of different mechanisms in shaping natural communities
133 and to generate testable hypotheses about: 1) species interactions, 2) species abundances, 3)
134 species dependencies on abiotic factors, 4) the way that the relative role of these determinants
135 change over time and space, and 5) the relative roles that species play in the community (e.g.,
136 Johnson 2002; Cardon *et al.* 2011; Lootvoet *et al.* 2013; Jing *et al.* 2015).

137

138 **Guidelines for the combined SEM-model selection approach**

139 Below we offer guidelines for the combined SEM-model selection approach (Fig. 1a). Our
140 guidelines are not meant to be exhaustive since many of the concepts are common to all
141 univariate GLM and multivariate methods (e.g., Zuur *et al.* 2009) or are general to either
142 SEM (Grace 2006; Hoyle 2012) or model selection approaches (Burnham & Anderson 2002).
143 Moreover, often the analyses depend on the nature of the data and on the specific study
144 objectives. Rather, we focus on the major steps that are especially important for the combined
145 approach to hopefully avoid pitfalls.

146 1) Formulating the research goal and hypotheses

147 The formulation of the research goal and hypotheses is a crucial step in the combined
148 approach, as it will determine the construction of the saturated models and model set (steps
149 2–3 in Fig. 1). For example, the simplification approach, estimation method, and set of
150 competing models in a study that is driven by a solid hypothesis often will be different from
151 those in a study driven by field observations with no *a priori* hypotheses.

152 2) Constructing the saturated SEM/s

153 a) Simplification

154 After formulating the research goal, it is recommended to construct a saturated or several
155 saturated models (in the case of multiple alternative hypotheses, each refers to a different
156 SEM; Cohen, Einav & Hawlena 2015). Each saturated model defines the range of possible
157 variables and paths between them, constituting the foundations of the set of competing
158 models (step 3). To this end, the choice of the variables and paths to be included therein will
159 greatly affect the inference step (step 4). SEM models may include a combination of multiple
160 dependent variables, mediators, which act as dependent in one equation and independent in
161 others, and independent variables. In some cases, the interactions between variables is also
162 important (e.g., Flatau *et al. in press*) and could be handled, using composites or multigroup
163 approaches (Grace 2006). There is no doubt that including all measured variables and

164 possible paths may bring us closer to real natural scenarios. However, when we are dealing
165 with complex systems, such a detailed approach may be translated into millions of competing
166 models, leading to overfitting and spurious findings (Burnham & Anderson 2002).

167 To simplify the task in a biologically sound way, we offer three, non-mutually
168 exclusive approaches, depending on the goal of the study, the theoretical background, and
169 existing knowledge. First, the desired goal is always to construct one or several alternative
170 hypotheses; each represents a complete SEM. These hypotheses are often based on a theory
171 or on previous findings (Cardon *et al.* 2011; Lootvoet *et al.* 2013). For example, Cohen,
172 Einav and Hawlena (2015) made comparisons between and within three sets of models, each
173 led by a saturated SEM that describes an alternative hypothesis (e.g., interactions in the
174 community are directly determined by parasite A, by parasite B, or are mediated by the host's
175 response) and is followed by its nested derivative models (Cohen, Einav & Hawlena 2015).
176 However, since usually no such *a priori* hypotheses exist, prior knowledge on the
177 relationships between at least some of the variables can instead be incorporated to construct
178 the saturated model (Palomares *et al.* 1998; Johnson 2002, and Step 2a in the case study). To
179 complement the latter approach or in cases with no prior knowledge, the potential causal
180 relationships between some of the variables can be statistically defined. Here too, we
181 recommend using statistical tools from the model selection approach (Messika *et al.* 2017).
182 For example, one can search for the most important factors that best explain the variability in
183 each of the dependent and mediator variables by assessing the sum of w_i over all the models
184 in which the variable of interest appears, given that the number of models that contain each
185 variable is balanced (relative importance; Johnson & Omland 2004). Commonly, variables
186 with a relative importance greater than 70% are considered important and should thus be
187 included in the saturated model (Wagner, Harmon & Seehausen 2014).

188 A parallel reduction in the number of dependent variables can be done statistically by
189 either replacing multiple dependent variables with a principal component (e.g. Mommer *et al.*
190 2006), using unobservable variables, or by formulating an equation that incorporates multiple
191 dependent variables (e.g., the integrated index of female flea reproductive success; Fig. 1B).

192 b) Biological meaning

193 Regardless of the approach one chooses, all paths should have a biological sense. Thus, *a*
194 *priori* analyses or knowledge of at least some of the causal relations between the variables
195 under study are crucial in constructing a manageable saturated path model.

196 c) Statistical considerations

197 Dealing with missing data, multicollinearity, the assumption of parametric tests, and sample
198 size constraints are common concerns in most GLM analyses. However, they are critical in
199 the combined approach since they determine the estimation method (section 2e).

200 Recommendations on the requirements and calculations of the minimum sample size in SEM
201 analyses are found elsewhere (e.g. Fan *et al.* 2016; Kline 2016). For the model selection
202 (steps 3–4 in Fig. 1), the sample size will often determine the choice of information theory
203 indices. For example, the use of AIC is only recommended when $n/p > 40$ (where n is the
204 sample size and p is the number of parameters), and below this value, it is recommended to
205 use its corrected version, AICc (Johnson & Omland 2004). Thus, users of the combined
206 approach should confirm that their sample size meets the requirements of both approaches,
207 and if not, they should enlarge the sampling or reduce the number of parameters in question.

208 Violations of multicollinearity (the inclusion of highly correlated independent
209 variables, i.e., correlation coefficient, $|r| > 0.6$), which can produce spurious results, are
210 common in SEM models since they include both independent and mediator variables that can
211 potentially correlate (Grewal, Cote & Baumgartner 2004). To overcome this problem, it is
212 possible to create a principal component variable from the set of correlated variables (e.g.

213 Cardon *et al.* 2011), replace correlated variables with unobservable variables, or use specific
214 estimation methods that are more robust to deal with multicollinearity (e.g., Consistent Partial
215 Least Squares; Vinzi *et al.* 2010). Similarly, missing data may greatly constrain the combined
216 approach since often the same variables are simultaneously used for multiple paths and since
217 it is only possible to compare models that have the same matrix. The problem can be solved
218 by using a proper estimator (e.g., full information maximum likelihood; Hoyle 2012) or by
219 removing the problematic variables.

220 Results may also be sensitive to violation in the parametric test's assumptions. In
221 SEM, different mediator and dependent variables may have distinct distributions, and
222 therefore, multivariate normality needs to be achieved. Various analytical methods have been
223 developed to check for multivariate normality, depending on factors such as sample size,
224 skewness, and kurtosis (e.g., Mardia's, Henze-Zirkler's and Royston's tests; Farrell, Salibian-
225 Barrera & Naczki 2007). A common factor violating the multivariate normality assumption is
226 the presence of outliers in either the univariate or multivariate dimensions, which can be
227 detected by a variety of techniques (e.g., Cook's and Mahalanobis' distances). When the
228 removal of outliers does not satisfy the multivariate normality assumption, SEM can be
229 adjusted to simultaneously fit a variety of residual distributions, including Poisson and
230 binomial, for count and binary data, respectively (Grace 2006; Hoyle 2012; Kline 2016).
231 Alternatively, normalizing transformations can be used (Hoyle 2012). In cases of high
232 overdispersion, inference can be improved by applying Quasi-AIC (qAIC) as the information
233 theory index (Johnson & Omland 2004).

234 d) Considerations of the nature and structure of the data

235 The choice of a proper estimator (step "e") depends not only on the above statistical
236 considerations but also on the nature and structure of the data. One should define, in advance,
237 which of the variables are categorical, ordinal, or continuous. Moreover, in many cases, the

238 inclusion of unobservable variables could be helpful even in the absence of collinearity or
239 demand for simplification (Grace 2006; Hoyle 2012). These variables, which are abstract and
240 are only inferred from the observed variables, may be particularly valuable in the study of
241 complex natural systems, in which concepts are frequently multifaceted and interactions are
242 nonlinear (e.g., fitness-correlative traits and variables related to ecological niches, climate
243 change, or ecosystem structure and function; Grace & Bollen 2008). Unobservable variables
244 include latent and composite variables. Both are similar in their uses, but the latter variables
245 assume no error variance among the indicators; thus, while the latent factor is extracted from
246 the set of indicators, the composite variable is an exact linear combination of the indicator
247 variables based on given weights (Grace & Bollen 2008; Fan *et al.* 2016). Recent
248 developments in statistical packages also now allow the integration into SEM of other data
249 types, representing basic ecosystem and evolutionary processes and dynamics, including
250 hierarchical linear modeling, time series, repeated measure analyses, and feedback loops
251 (e.g., Meredith & Tisak 1990; Grace 2006).

252 e) Choosing an estimation method

253 The choice of a proper estimation method depends on the statistical properties, nature, and
254 structure of the data (see above), and on the user's software (reviewed in Hoyle 2012;
255 Narayanan 2012). Two families of estimation methods prevail in SEM: the maximum
256 likelihood (ML) and the least squares (LS). For the combined approach, the former family is
257 often used since the calculation of information theory indices has been commonly computed
258 from ML (but see alternatives for least squares methods and for PLS; Yamaoka, Nakagawa &
259 Uno 1978; Vinzi *et al.* 2010).

260 3) Identifying the best models and quantifying their fit to the data

261 a) Constructing the set of competing models

262 As soon as we have a saturated model or models, it is recommended to generate a set of
263 models nested in the saturated ones that will constitute the set of competing models. The
264 main constraint in this step is that all competing models should have the same log likelihood
265 and hence the same metric of the information theory index. This means that a variable that is
266 used as a mediator or as a dependent variable in one model cannot be used as an independent
267 variable in another model (e.g., the fixed association between community composition and
268 host density in the case study; Fig. 2C). However, in most cases, even after accounting for the
269 above constraint and while keeping only paths that are biologically meaningful, there are
270 more models left than the data could reasonably support. Thus, here again, the challenge is to
271 minimize the number of competing models. As in the previous step (2a), the choice of
272 competing models can be based on theoretical background, prior experiences, or logical
273 arguments—choosing to emphasize each time different parts of the saturated model/s. When
274 there is no *a priori* knowledge, “backward elimination” can be helpful, relying on the relative
275 importance weight or effect size of paths (see the case study).

276 b) Comparing the models in a set and assessing the fit of the best ones
277 Once a set of candidate models is specified, they can be compared based on model selection
278 criteria. The choice of the information theory index depends on the sample size, residual
279 distribution, user’s software, and whether the approach is Bayesian or not, as discussed in
280 recent reviews (Burnham & Anderson 2002; Johnson & Omland 2004). The calculated index
281 is then used to rank the competing models and to weigh the relative support for each one
282 (Johnson & Omland 2004). The likelihood of each model is quantified by its w_i , which is
283 calculated as the differences in the information theory index scores between each model and
284 the best model (i.e., the model with the lowest information theory index score), normalized
285 across the set of candidate models to sum to 1 and thus interpreted as a probability. Often

286 there is no single model overwhelmingly supported by the data ($w_i < 0.9$), and thus inference
287 is based on a group of the best models (Johnson & Omland 2004).

288 The models identified as best have a relatively higher fit compared to the other
289 candidate models in the set. Yet, for the inference (step 4), it is important to check whether
290 they sufficiently fit to the data (Hu & Bentler 1999). The choice of the goodness-of-fit test
291 (e.g., Chi-square, RMSEA, CFI, and SRMR) depends on the sample size, residual
292 distribution, number of parameters, and whether there are missing data (Hu & Bentler 1999;
293 Hoyle 2012; Fan *et al.* 2016).

294 4) Multimodel inference

295 From the identification of the best models, one can already infer which hypotheses to reject
296 or accept. To complement the inference process, it is recommended to also quantitatively
297 compare the likelihood of each of the best models by either w_i , evidence ratio (ER), or model
298 likelihood (L) values (Burnham, Anderson & Huyvaert 2011). To assess the strength and
299 importance of specific paths of interest, researchers often examine whether their coefficients
300 are significantly different from zero (Fig. 2B). However, since it is recommended not to mix
301 hypothesis-driven with model selection approaches (Burnham, Anderson & Huyvaert 2011),
302 we propose here to extend the relative importance approach by Burnham and Anderson
303 (2002), namely to search for the most important paths (relative importance greater than 70%)
304 by assessing the sum of the w_i s over all the models in which the path of interest appears. The
305 challenge of such an approach is to achieve a balance in the number of models that contain
306 each path. Regardless of the method of quantifying the path importance, it is recommended to
307 report standardized or unstandardized path coefficients, when the interest is either in
308 comparing paths or in predicting relationships, respectively (see discussion in Grace 2006).
309 These could be done by using a weighted average of parameter estimates either over the
310 model set or just over the best models (Johnson & Omland 2004). Finally, the influence of

311 one variable on another can be decomposed into direct, indirect (calculated as the product of
312 the effect of the independent variable on the mediator/s and of the mediator/s on the
313 dependent variable), and total (the sum of the direct and indirect effects) effects (Grace
314 2006). The decomposition to direct and indirect effects may be especially important in
315 quantifying cascading effects (Grace 2009). We suggest here again following the relative
316 importance approach by assessing the sum of w_i over all the models in which the type of
317 effect (direct or indirect) appears and presenting their weighted, averaged coefficients (see
318 the case study).

319

320 **A case study: factors determining the reproductive success of arthropod vectors**

321 Below we illustrate our proposed step-by-step guidelines for the combined approach with our
322 recently published analysis on the factors determining the reproductive success of arthropod
323 vectors (Fig. 1B; Messika *et al.* 2017). We then assess the usefulness of the combined
324 approach analysis in the exploration of complex ecological communities by comparing the
325 results and insights gained through this approach with those gained through reanalysis of the
326 data using a univariate approach, a traditional SEM, and a stepwise SEM (Fig. 2). In this
327 study, we trapped, marked, and characterized rodents in their natural environment, including
328 sites differing in their proportions of the most dominant rodent species (*Gerbillus andersoni*
329 and *Gerbillus pyramidum*) and their total rodent density. We then sampled their fleas and
330 ticks, randomly collected one female flea from each host, and assessed its current
331 reproductive success and infection status by vector-borne bacteria (*Bartonella* and
332 *Mycoplasma*) under laboratory conditions. The analyses were performed with the software
333 Mplus 7 (Muthén & Muthén 2012) but could be run by any alternative software.

334 **a. Step-by-step analysis using the SEM-model selection combined approach**

335 Step 1: Formulating the research goal and hypotheses

336 The main goal of the research was to reveal the most likely pathways connecting a suite of
337 interrelated factors, ranging from female flea endosymbionts to the community structure of
338 the rodent hosts, and to flea reproductive success. We tested four non-mutually exclusive
339 hypotheses:

340 *H1 (microscale)*: we expected to find neutral or positive associations between infection by
341 vector-borne bacteria and female flea reproductive success.

342 *H2 (vector scale)*: we expected a positive association between female flea reproductive
343 success and their body mass and a negative association with their level of fluctuating
344 asymmetry.

345 *H3 (host scale)*: we expected that female flea reproductive success would be (1) greater on *G.*
346 *pyramidum* and on male hosts; (2) positively associated with host body condition; and (3)
347 negatively associated with the flea and tick burden on the host.

348 *H4 (macroscale)*: we expected that female flea reproductive success would be negatively
349 associated with host community diversity, but positively associated with total host
350 density.

351 The four hypotheses were formulated based on well-supported theories in evolutionary
352 biology (Ewald 1983), general ecology (Møller 1997), and the ecology of host-parasite
353 interactions (Krasnov, Khokhlova & Shenbrot 2002; Seppälä *et al.* 2008) and were
354 complemented by published data on the studied organisms (references and more details in
355 Messika *et al.* 2017).

356 Step 2: Constructing the saturated SEM/s

357 Our data set included 14 potentially independent or mediator variables and 10 dependent
358 variables, all associated with current female reproductive success (Messika *et al.* 2017).
359 Therefore, with no *a priori* analyses and knowledge, we would have had to deal with more
360 than a million competing SEMs, considering all possible relations between and within

361 mediators and dependent variables . To simplify the task in a biologically sound way, we
362 constructed our saturated SEM through a combination of three approaches. The associations
363 between host community diversity, total host density, and the flea probability to encounter
364 one or another host species were predicted based on extensive knowledge of the system's
365 natural history (e.g. Rosenzweig, Abramsky & Brand 1984). In addition, to filter out
366 presumably less important variables, we relied on a relative importance analysis by univariate
367 GLMs and selected the most influential factors (relative importance weight > 0.7) that best
368 explained the variability in female reproductive success (Table 2 in Messika *et al.* 2017).
369 Finally, following previous studies (e.g., Herreras *et al.* 2007), we created an index that
370 mathematically integrates the various components of female reproductive success . We could
371 have alternatively created a latent variable of female reproductive success with the specific
372 measured variables (e.g., offspring mass, number of offspring, and offspring sex ratio) as
373 indicators. However, in our case, this would have required significantly reducing the sample
374 size to adjust for the females who either had no offspring (i.e., no data for the sex ratio or
375 offspring body mass) or had only a single offspring (i.e., no reliable data for sex ratio).

376 As a result, our saturated SEM model included three independent categorical
377 variables, four mediators (three continuous and one categorical), one continuous dependent
378 variable, and a total of 22 links between these variables (Fig. 1S in Messika *et al.* 2017).
379 These sum-up to a total of 30 free parameters and 141 observations with no missing data,
380 with nearly five observations per free parameter. In addition, there was no indication for
381 multicollinearity ($|r| < 0.3$ for all paired correlations) or outliers, and no deviation from
382 multivariate normality. Despite the adherence of our data to the parameter test assumptions,
383 to be most conservative, we used robust maximum likelihood as the estimator.

384 Step 3: Identifying the best models and quantifying their fit to the data

385 We created a set of competing models, all nested in the saturated model, considering that the
386 mediators and dependent variables are fixed and the independent variables and paths are
387 changeable. The set of alternative models was created by omitting in order the weakest paths
388 according to their effect size. The process was done manually to create a new combination of
389 paths and to allow a continuous check that the models and all paths included have ecological
390 sense and, in particular, do not contradict any of our four hypotheses (H1–H4). This approach
391 resulted in 42 competing models (Table S1 in Messika *et al.* 2017). w_i s of the competing
392 models revealed that four SEMs have good support from and fit to the data ($w_i > 10\%$; $\sum w_i =$
393 66% ; CFI > 0.9 and RMSEA 90% CI: 0.000–0.025 for all four models; Fig. 2C).

394 Step 4: Multimodel inference

395 The combined SEM-model selection approach provided support for our overall hypothesis
396 that a suite of interrelated factors, ranging from micro to macroscales, is linked with female
397 flea reproductive success. The four best models indicated positive associations between the
398 index of flea reproductive success and (1) its infection by *Mycoplasma* bacteria (H1,
399 microscale), (2) its body mass (H2, vector scale), (3) the size of the host species (H3, host
400 scale), and (4) the total host density (H4, macroscale) (Fig. 2C). Direct, indirect and total
401 effects of each predictor on the female flea reproductive success were calculated, using the
402 model-averaged standardized path coefficients. Two factors at different scales—the total host
403 density (H4, macroscale) and host species (H3, host scale) —showed the strongest total
404 effects (Table 1). Moreover, this analysis highlights host community diversity as an important
405 predictor, indirectly affecting flea reproductive success (Table 1).

406 **b. A comparison between the results and insights gained by the combined** 407 **approach with those gained by alternative statistical approaches**

408 We compared the results obtained by the SEM-model selection combined approach to those
409 obtained by the univariate approach, the traditional SEM, and the stepwise SEM. The

410 univariate approach included a set of seven independent GLM models, correlating between
411 the seven independent variables (three of which are designated as “mediators” in the other
412 approaches) and the integrated index of reproductive success (Fig. 2A). The best model
413 according to the traditional SEM analysis was based on the significant pathways in the
414 saturated model, whereas in the backward SEM analysis, weak paths were omitted using a
415 backward elimination process supported by Chi-square tests until no further variables could
416 be deleted without a statistically significant loss of fit (Fig. 2B).

417 The most likely pathways that were suggested to connect the tested factors and female
418 reproductive success according to the four alternative approaches are illustrated in Fig. 2. For
419 our particular case study, the traditional and the stepwise SEM approaches generated
420 identical results except for the association between the vector’s asymmetry and body mass,
421 which was only included in the backward SEM (Fig. 2B). As seen, all four approaches
422 detected the direct associations between (i) total host density, (ii) host species, and (iii)
423 female mass and the integrated index of reproductive success (thin arrows in Fig. 2). In
424 addition, the three SEM-related analyses supported a mediation role for the (i) host species
425 and (ii) female mass (grey bold arrows in Figs. 2B–C).

426 Considering the differences between the four approaches, our example highlights a
427 possible limitation of the combined approach, namely, its requirement to retain the same
428 mediators and dependent variables in all competing models. In our case study, for example,
429 in order to keep the total host density as a mediator, we were forced to retain the path
430 between host community diversity and total host density in all competing models despite its
431 low standardized coefficient value. In this particular case, the host community diversity-total
432 density association is supported by previous studies (e.g. Rosenzweig, Abramsky & Brand
433 1984), but in other cases, such constraints may limit the predictive power of the models.

434 On the other hand, our comparison suggests that the combined approach is more
435 sensitive to the weaker (yet possibly important) associations than the three alternative
436 approaches. First, the combined approach supports the direct effects of female asymmetry
437 and *Mycoplasma* infection status on the integrated index of reproductive success (black bold
438 arrows in Fig. 2C). Importantly, the positive association between female infection status and
439 the integrated index of reproductive success was included in three out of the four most
440 supportive models (Fig. 1D in Messika *et al.* 2017). In addition, it was highly important (path
441 relative importance = 58%) and is crucial for both our understanding of flea-host interactions
442 and for flea control programs; thus, it needs to be experimentally verified. Second, the
443 combined approach highlighted mediation effects that were overlooked by the other two
444 SEM-related approaches. These include the indirect effects of the host's species and female
445 flea asymmetry, through effects on the female flea mass, on the integrated index of
446 reproductive success (black bold arrows in Fig. 2C). Notably, the relative importance of
447 female asymmetry's indirect effect through its mass was greater (65%) than its direct effect
448 (26%) on the integrated index of reproductive success.

449 Revealing these additional mediation effects are of special interest to our study goal
450 since they connect factors that represent the hypotheses at different scales and can be crucial
451 when making management decisions (Messika *et al.* 2017). In particular, the understanding
452 that the effects of host community diversity—one of the most important predictors of the
453 bacterial communities in hosts (Gavish *et al.* 2014; Kedem *et al.* 2014)—on flea reproductive
454 success is mediated both by changes in total host density and flea mass would not have been
455 revealed by the alternative approaches (Fig. 2).

456

457 **Acknowledgments**

458 We thank I. Messika and H. Kedem for permitting us to reanalyze the case study. The authors
 459 declare no conflict of interest.

460

461 **Author's contribution**

462 HH conceived the idea and edited the manuscript; MG reviewed the literature and wrote the
 463 manuscript; together they conducted the analyses.

464

465 **Data accessibility**

466 Data will be made available on FigShare.com once manuscript is accepted for publication

467

468 **References**

- 469 Belinchon, R., Harrison, P.J., Mair, L., Varkonyi, G. & Snall, T. (2017) Local epiphyte establishment
 470 and future metapopulation dynamics in landscapes with different spatiotemporal properties.
 471 *Ecology*, **98**, 741-750.
- 472 Bentler, P.M. & Bonett, D.G. (1980) Significance tests and goodness of fit in the analysis of
 473 covariance structures. *Psychological Bulletin*, **88**, 588-606.
- 474 Bunnefeld, N., Borger, L., van Moorter, B., Rolandsen, C.M., Dettki, H., Solberg, E.J. & Ericsson, G.
 475 (2011) A model-driven approach to quantify migration patterns: individual, regional and
 476 yearly differences. *Journal of Animal Ecology*, **80**, 466-476.
- 477 Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference: a practical
 478 information-theoretic approach*. Springer Science & Business Media.
- 479 Burnham, K.P., Anderson, D.R. & Huyvaert, K.P. (2011) AIC model selection and multimodel
 480 inference in behavioral ecology: some background, observations, and comparisons.
 481 *Behavioral Ecology and Sociobiology*, **65**, 23-35.
- 482 Cardon, M., Loot, G., Grenouillet, G. & Blanchet, S. (2011) Host characteristics and environmental
 483 factors differentially drive the burden and pathogenicity of an ectoparasite: a multilevel causal
 484 analysis. *Journal of Animal Ecology*, **80**, 657-667.
- 485 Cohen, C., Einav, M. & Hawlena, H. (2015) Path analyses of cross-sectional and longitudinal data
 486 suggest that variability in natural communities of blood-associated parasites is derived from
 487 host characteristics and not interspecific interactions. *Parasites & Vectors*, **8**, 429-440.
- 488 Coomes, D.A. & Allen, R.B. (2007) Effects of size, competition and altitude on tree growth. *Journal
 489 of Ecology*, **95**, 1084-1097.
- 490 Ewald, P.W. (1983) Host-parasite relations, vectors, and the evolution of disease severity. *Annual
 491 Review of Ecology and Systematics*, **14**, 465-485.
- 492 Fan, Y., Chen, J., Shirkey, G., John, R., Wu, S.R., Park, H. & Shao, C. (2016) Applications of
 493 structural equation modeling (SEM) in ecological studies: an updated review. *Ecological
 494 Processes*, **5**, 19-30.
- 495 Farrell, P.J., Salibian-Barrera, M. & Naczk, K. (2007) On tests for multivariate normality and
 496 associated simulation studies. *Journal of Statistical Computation and Simulation*, **77**, 1065-
 497 1080.
- 498 Flatau, R., Segoli, M., Khokhlova, I.S. & Hawlena, H. (*in press*) The role of *Wolbachia* as a mediator
 499 in shaping the reproductive success of its host. *FEMS microbiology ecology*.

- 500 Gavish, Y., Kedem, H., Messika, I., Cohen, C., Toh, E., Munro, D., Dong, Q., Fuqua, C., Clay, K. &
501 Hawlena, H. (2014) Association of host and microbial species diversity across spatial scales
502 in desert rodent communities. *PLoS One*, **9**, e109677.
- 503 Grace, J.B. (2006) *Structural equation modeling and natural systems*. Cambridge University Press,
504 Cambridge, UK.
- 505 Grace, J.B. (2009) Structural Equation Modeling for Observational Studies. *Journal of Wildlife*
506 *Management*, **72**, 14-22.
- 507 Grace, J.B. & Bollen, K.A. (2008) Representing general theoretical concepts in structural equation
508 models: the role of composite variables. *Environmental and Ecological Statistics*, **15**, 191-
509 213.
- 510 Grace, J.B., Schoolmaster, D.R., Guntenspergen, G.R., Little, A.M., Mitchell, B.R., Miller, K.M. &
511 Schweiger, E.W. (2012) Guidelines for a graph-theoretic implementation of structural
512 equation modeling. *Ecosphere*, **3**, art73.
- 513 Grewal, R., Cote, J.A. & Baumgartner, H. (2004) Multicollinearity and measurement error in
514 structural equation models: Implications for theory testing. *Marketing Science*, **23**, 519-529.
- 515 Herbreteau, V., Jittapalpong, S., Rerkamnuaychoke, W., Chaval, Y., Cosson, J.F. & Morand, S.
516 (2011) *Protocols For Field And Laboratory Rodent Studies*. Kasetsart University Press.
- 517 Herreras, M.V., Montero, F.E., Marcogliese, D.J., Raga, J.A. & Balbuena, J.A. (2007) Phenotypic
518 tradeoffs between egg number and egg size in three parasitic anisakid nematodes. *Oikos*, **116**,
519 1737-1747.
- 520 Hoyle, R.H. (2012) *Handbook of structural equation modeling*. Guilford Press, New York, USA.
- 521 Hu, L. & Bentler, P.M. (1999) Cutoff criteria for fit indexes in covariance structure analysis:
522 Conventional criteria versus new alternatives. *Structural Equation Modeling: A*
523 *Multidisciplinary Journal*, **6**, 1-55.
- 524 Jing, X., Sanders, N.J., Shi, Y., Chu, H., Classen, A.T., Zhao, K., Chen, L., Shi, Y., Jiang, Y. & He,
525 J.S. (2015) The links between ecosystem multifunctionality and above- and belowground
526 biodiversity are mediated by climate. *Nature Communications*, **6**, 8159.
- 527 Johnson, J.B. (2002) Divergent life histories among populations of the fish *Brachyrrhaphis*
528 *rhabdophora*: detecting putative agents of selection by candidate model analysis. *Oikos*, **96**,
529 82-91.
- 530 Johnson, J.B. & Omland, K.S. (2004) Model selection in ecology and evolution. *Trends in Ecology &*
531 *Evolution*, **19**, 101-108.
- 532 Kedem, H., Cohen, C., Messika, I., Einav, M., Pilosof, S. & Hawlena, H. (2014) Multiple effects of
533 host-species diversity on coexisting host-specific and host-opportunistic microbes. *Ecology*,
534 **95**, 1173-1183.
- 535 Kline, R.B. (2016) *Principles and Practice of Structural Equation Modeling*, 4th edn. The Guilford
536 Press, New York, USA.
- 537 Krasnov, B.R., Khokhlova, I.S. & Shenbrot, G.I. (2002) The effect of host density on ectoparasite
538 distribution: An example of a rodent parasitized by fleas. *Ecology*, **83**, 164-175.
- 539 Lootvoet, A., Blanchet, S., Gevrey, M., Buisson, L., Tudesque, L., Loot, G. & Mayhew, P. (2013)
540 Patterns and processes of alternative host use in a generalist parasite: insights from a natural
541 host-parasite interaction. *Functional Ecology*, **27**, 1403-1414.
- 542 Meredith, W. & Tisak, J. (1990) Latent curve analysis. *Psychometrika*, **55**, 107-122.
- 543 Messika, I., Garrido, M., Kedem, H., China, V., Gavish, Y., Dong, Q., Fuqua, C., Clay, K. &
544 Hawlena, H. (2017) From endosymbionts to host communities: factors determining the
545 reproductive success of arthropod vectors. *Oecologia*, **184**, 859-871.
- 546 Møller, A.P. (1997) Developmental stability and fitness: a review. *The American naturalist*, **149**, 916-
547 932.
- 548 Mommer, L., Lenssen, J.P.M., Huber, H., Visser, E.J.W. & De Kroon, H. (2006) Ecophysiological
549 determinants of plant performance under flooding: a comparative study of seven plant
550 families. *Journal of Ecology*, **94**, 1117-1129.
- 551 Muthén, B.O. & Muthén, L.K. (2012) *Mplus Version 7: User's guide*. Muthén & Muthén, Los
552 Angeles, CA.
- 553 Narayanan, A. (2012) A Review of Eight Software Packages for Structural Equation Modeling. *The*
554 *American Statistician*, **66**, 129-138.

- 555 Palomares, F., Ferraras, P., Travaini, A. & Delibes, M. (1998) Coexistence between Iberian lynx and
556 Egyptian mongooses: estimating interaction strength by structural equation modelling and
557 testing by an observational study. *Journal of Animal Ecology*, **67**, 967-978.
- 558 Pilosof, S., Fortuna, M.A., Cosson, J.F., Galan, M., Kittipong, C., Ribas, A., Segal, E., Krasnov, B.R.,
559 Morand, S. & Bascompte, J. (2014) Host-parasite network structure is associated with
560 community-level immunogenetic diversity. *Nature Communications*, **5**, 5172.
- 561 Rosenzweig, M.L., Abramsky, Z. & Brand, S. (1984) Estimating Species Interactions in
562 Heterogeneous Environments. *Oikos*, **43**, 329-340.
- 563 Seppälä, O., Liljeroos, K., Karvonen, A. & Jokela, J. (2008) Host condition as a constraint for parasite
564 reproduction. *Oikos*, **117**, 749-753.
- 565 Vinzi, V.E., Chin, W.W., Henseler, J. & Wang, H. (2010) *Handbook of Partial Least Squares*.
566 Springer-Verlag, Berlin Heidelberg, Germany.
- 567 Wagner, C.E., Harmon, L.J. & Seehausen, O. (2014) Cichlid species-area relationships are shaped by
568 adaptive radiations that scale with area. *Ecology Letters*, **17**, 583-592.
- 569 Whittingham, M.J., Stephens, P.A., Bradbury, R.B. & Freckleton, R.P. (2006) Why do we still use
570 stepwise modelling in ecology and behaviour? *Journal of Animal Ecology*, **75**, 1182-1189.
- 571 Yamaoka, K., Nakagawa, T. & Uno, T. (1978) Application of Akaike's information criteria (AIC) in
572 the evaluation of linear pharmacokinetic equations. *Journal of Pharmacokinetics and*
573 *Biopharmaceutics*, **6**, 165-175.
- 574 Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. & Smith, G.M. (2009) *Mixed effects models and*
575 *extensions in ecology with R*. Springer Science, NY.

576

577

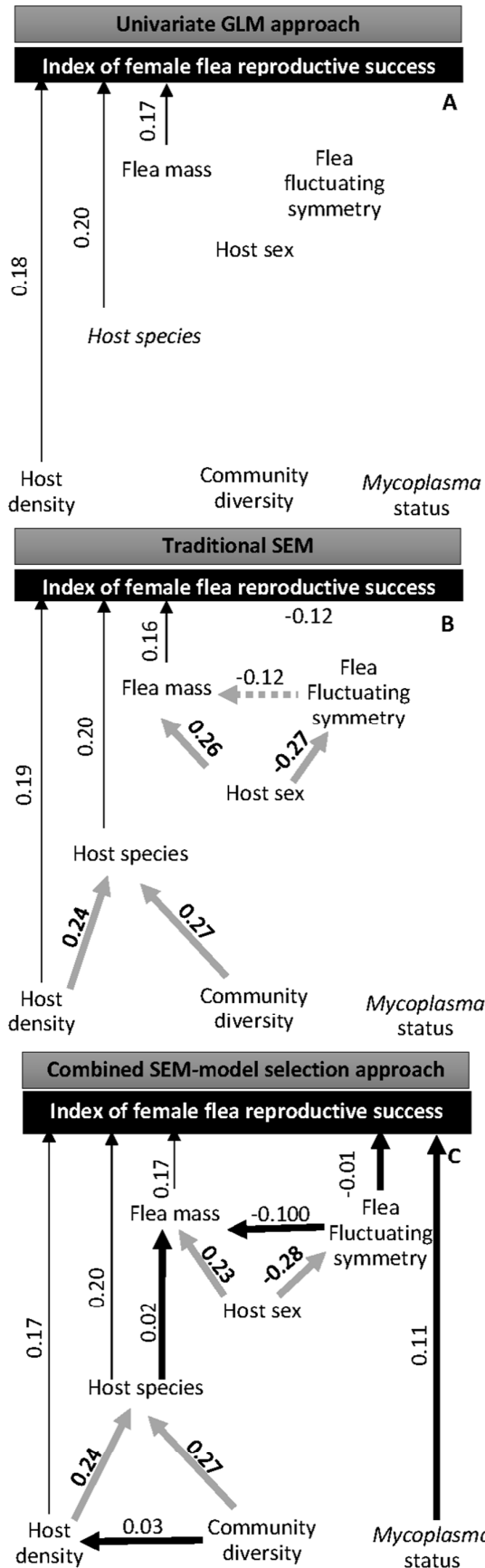
578 **Tables and Figures**

579 **Table 1.** Summary of weighted model-averaged standardized path coefficients for the direct,
 580 sum of all indirect effects, and total effects of each predictor variable on the integrated index
 581 of female reproductive success. Standardized path coefficients were weighted by the weight
 582 (w_i) of the best models.

	Direct	Indirect	Total
Community diversity	0.02	0.06	0.06
Rodent host density	0.2	0.05	0.2
Rodent host species	0.2	0.004	0.2
Rodent host sex	0.02	0.05	0.05
Female flea mass	0.2	0.0	0.2
Female flea FA	-0.01	0.0	-0.01
<i>Mycoplasma</i> infection status	0.1	0.0	0.1

	A	B
	1. Formulating the research goal and hypotheses	To reveal the most likely pathways that may connect a suite of interrelated factors to the reproductive success of fleas (H1–H4 in the case study)
SEM	2. Constructing the saturated SEM/s	The saturated model is described in Fig. S1 in Messika et al., 2017
	a) Simplification	We focused on 8 of the 24 measured variables based on previous knowledge and their relative importance, and by creating an index, integrating the various correlatives of reproductive success
	b) Biological meaning	All paths are consistent with the natural history of rodents and fleas and consider the four hypotheses
	c) Statistical considerations	~Five cases per free parameter; no outliers, missing data, or collinearity; multivariate normality
	d) Considerations of the nature and structure of the data	Variables include three categorical independent variables, one categorical and three continuous mediators, and a continuous dependent variable. There are no unobservable variables or hierarchical or repeated measures structure
	e) Choice of an estimation method	Robust maximum likelihood estimation
Model Selection	3 Identifying the best models and quantifying their fit to the data	Comparing the 42 candidate models, nested within a single saturated model
	f) Constructing the set of competing models	The set of competing models was created by dropping in order their weakest paths according to their effect size while keeping the biological logic and considering the four hypotheses
	g) Comparing the models in a set and assessing the fit of the best ones	Four best models with $w_i > 10\%$; $\sum w_i = 66\%$ (Table S1 in Messika et al., 2017); all have a decent fit to the data ($CFI > 0.9$ and $RMSEA$ 90% CI: 0.000–0.025)
	4 Multimodel inference	Support of the overall hypothesis that a suite of interrelated factors is linked with the flea reproductive success, including partial support in all specific H1–H4 hypotheses; highlight the indirect effects of community diversity and host density on flea reproductive success (Fig. 2C; Table 1).

583 **Figure 1.** Recommended steps in the combined SEM-model selection approach (A) as
 584 exemplified by our flea-rodent case study (B). SEM steps (a–e) are modified from Figure 1 in
 585 Grace *et al.* (2012).



View Only

587 **Figure 2.** Results of four alternative statistical approaches used to explore the effects of a set
588 of interrelated factors on flea reproductive success. Arrows represent directional relationships
589 between variables. (A) Univariate GLMs, only significant paths are included. (B) A
590 traditional SEM, in which only significant paths construct the best SEM and a backward
591 SEM. The two approaches generated identical results in our case study, except for the
592 association between the vector's asymmetry and its body mass, which was only included in
593 the backward SEM (dashed arrow). (C) A combined SEM-model selection approach, in
594 which paths are present only if they are constrained in all competing models (i.e., community
595 diversity on host density) or appear in at least one of the best models ($w_i > 10\%$), and
596 standardized path coefficients are averaged among the best models. Thin arrows represent
597 paths that are highlighted as important by all approaches; bold grey arrows represent paths
598 that are highlighted as important in all three SEM-based approaches, and bold black arrows
599 represent paths exclusively detected by the combined approach. Numbers near arrows
600 indicate standardized path coefficient values.

Dr. Hadas Hawlena
Mitrani Department of Desert Ecology
Jacob Blaustein Institutes for Desert Research

July 18, 2018

Prof. Rob Freckleton, Editor-in-Chief, *Methods in Ecology and Evolution*
University of Sheffield, UK

Dear Professor Freckleton,

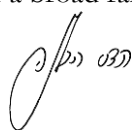
We hereby submit the manuscript “The combined structural equation modeling-model selection approach: a user guide for ecologists and evolutionary biologists” by Garrido and Hawlena to be considered for publication as a review in *Methods in Ecology and Evolution*.

The motivation to compile this review and supplement it with a practical guide was the direct result of our research group’s cumulative frustration over the past seven years caused by spending a remarkable amount of time in figuring out the correct ways to combine the structural equation modeling (SEM) and the model selection approaches (e.g. Cohen, Einav & Hawlena 2015; Garrido *et al.* 2016; Messika *et al.* 2017; Flatau *et al. in press*). Specifically, while we became aware of the power and potential of the combined approach for exploring ecological and evolutionary hypotheses, we realized that it requires an overwhelming process of decision-making and entails pitfalls that can lead to incorrect biological conclusions. This is partly because both the model selection and the SEM approaches originated in different disciplines and only gradually, and independently, became popular among ecologists and evolutionary biologists. Accordingly, the relevant literature is difficult to follow and complicates the integration of the two approaches for the end-users, so that neither approach, either separately or in combination, is routinely used in the fields of Ecology or Evolution. Furthermore, among those papers that used the combined approach, we found some inconsistencies, and it was challenging for us to locate a statistician who is not only an expert in both fields but can also communicate with ecologists and evolutionary biologists.

To maximize the uptake of the combined SEM-model selection approach by the ecological and evolutionary biologist communities and disseminate its use among practitioners, we have demonstrated its use with our data on a multihost-multiparasite system. While the results of part of these analyses were recently published (Messika *et al.* 2017), here we use the rodent-flea data for a different propose, namely to illustrate, step-by-step, the rationale in the decision-making process. This background information does not appear in the Messika *et al.* paper. To further illustrate the potential breadth of data interference, we added several analyses that are absent in the Messika *et al.* paper (e.g., the decomposition of paths into direct, indirect, and total effects and the quantification of their relative weight and importance). Finally, to illustrate the added insights gained from the combined approach, we reanalyzed the data, using a univariate approach, a traditional SEM, and a stepwise SEM.

Our review provides a single source that summarizes tentative solutions to practical challenges faced and, as such, will make the combined SEM-model selection approach more accessible to those interested in the analysis of complex ecological and evolutionary processes. This is especially important in the current era when there is a tendency to shift the focus from the study of specific isolated components in a system to the way that these components integrate, and will hopefully guide the investigation of questions that were not previously tractable for a broad range of ecologists and evolutionary biologists.

Sincerely, Hadas Hawlena



References

- Cohen, C., Einav, M. & Hawlena, H. (2015) Path analyses of cross-sectional and longitudinal data suggest that variability in natural communities of blood-associated parasites is derived from host characteristics and not interspecific interactions. *Parasites & Vectors*, **8**, 429-440.
- Flatau, R., Segoli, M., Khokhlova, I.S. & Hawlena, H. (*in press*) The role of *Wolbachia* as a mediator in shaping the reproductive success of its host. *FEMS microbiology ecology*.
- Garrido, M., Adler, V.H., Pnini, M., Abramsky, Z., Krasnov, B.R., Gutman, R., Kronfeld-Schor, N. & Hawlena, H. (2016) Time budget, oxygen consumption and body mass responses to parasites in juvenile and adult wild rodents. *Parasites & Vectors*, **9**, 120.
- Messika, I., Garrido, M., Kedem, H., China, V., Gavish, Y., Dong, Q., Fuqua, C., Clay, K. & Hawlena, H. (2017) From endosymbionts to host communities: factors determining the reproductive success of arthropod vectors. *Oecologia*, **184**, 859-871.