Supplementary Information

Methods

Paper selection

We used the 2012 Web of Science impact factors to select the highest ranked ecology-themed journals that published studies with an observational component, excluding journals devoted to reviews, meta-analyses, laboratory, cellular, or experimental studies. To select a representative sample of recent ecology studies, we downloaded the metadata for all papers published in the selected journals (Table S1) between 2004 and 2014. Our study involved six different observers (those reviewing the papers to extract the observational scales), each of whom was given a randomly selected batch of 500 titles. A separate set of 20 randomly selected papers was also given to each observer, which formed the basis for calibrating observers' interpretation and extraction of scale-related information, as well as estimating between observer variance.

Table 1. The selected journals and their 2012 impact factors.

Journal	Impact Factor
Ecology Letters	17.95
Ecological Monographs	8.09
Frontiers In Ecology And The Environment	7.62
Global Ecology And Biogeography	7.22
Global Change Biology	6.91
Diversity And Distributions	6.12
Methods In Ecology And Evolution	5.92
Proceedings Of The Royal Society B-biological Sciences	5.68
Journal Of Ecology	5.43
Ecology	5.17
Ecography	5.12
Journal Of Biogeography	4.86
Functional Ecology	4.86
Journal Of Animal Ecology	4.84
Journal Of Applied Ecology	4.74
American Naturalist	4.55
Conservation Biology	4.36
Ecological Applications	3.81
Biological Conservation	3.79

Journal	Impact Factor
Biogeosciences	3.75
Bulletin Of The American Museum Of Natural History	3.48
Biology Letters	3.35
Oikos	3.32
Behavioral Ecology	3.22
Ecosystems	3.17
Advances In Ecological Research	3.08
Oecologia	3.01
Landscape Ecology	2.90
Agriculture Ecosystems & Environment	2.86
Ecological Economics	2.85

Estimating observational scales

Each observer first reviewed the papers in the calibration set, and then commenced reviewing papers in their individual random draws, beginning at the top of the list and then proceeding until at least 20 eligible papers describing ecological observations were reviewed. In cases where the reviewed papers used observations that were described in another publication, we reviewed those source papers in order to extract the observational dimensions. We excluded papers that were opinion or perspectives pieces (unless they presented or used existing observational data), theoretical studies based on generated data, or those which were entirely based on experimental manipulations. We left out the latter category because our intent was to evaluate the domains for observations of natural systems, and we wanted to avoid the bias that would be imposed by the relatively narrow spatial and temporal scales of experiments (*1, 2*).

To record the spatio-temporal dimensions of ecological observations, we considered the observational scales to be those at which a total measurement of the feature(s) of interest was made, rather than the scales that might be represented by the observations. As an example, if a paper reported data on plant species diversity collected using 10 1 m^2 sub-plots randomly placed within a larger plot of 100 m^2, of which 10 were placed, we recorded 1 m^2 as the spatial resolution and 100 m^2 as the total extent of the observation. While the 10 sub-plots might be representative of

diversity at larger scales (e.g. the particular plant community), that relationship depends both on sample design (number of replicates, placement, grain/resolution) and the properties of the ecosystem (*3–5*). Most studies also did not quantify the broader scales represented by their observations, nor did they report dimensions that would allow these to be independently estimated, such as the maximum distance between sample plots, which could have been used to estimate the spatial extent represented by the sample (*4*).

Similarly, we did not try to quantify the extent to which values of individual observations (e.g. a single plot) of an ecological feature might be correlated with spatially or temporally adjacent values. Autocorrelation is common to most ecological phenomena (*6, 7*), and it means that an observation is descriptive of the properties within a broader area or time around the point at which it is taken. For example, soil properties are typically measured by analysis of soil cores collected using an augur. The two-dimensional resolution of this measurement is equal to the area of the augur, but the measured soil properties are likely to be very similar, if not identical, to those found in the soil within a radius of at least a few to several hundreds of meters around the core location. However, the distance and strength of this correlation can vary substantially according to the ecological features being observed, and such correlation lengths were reported in almost no studies. Attempting to estimate these lengths, and add them to actual observations scales, would have added further uncertainty to estimates. Moreover, this concern only applies to spatially and temporally discontinuous (or discontiguous) observations, and since many of observations are collected using techniques that permit continuous measurements (e.g. weather data over time or satellite images over space), we elected to use a consistent standard for estimating dimensions.

In assessing spatial scales, our analysis only considered evaluated the Cartesian plane. We did not consider the z, or depth, dimension, which for certain aspects of ecology, this measure is of greater value (e.g. depth profiles in marine ecology), although in some cases (paleoecology) values extracted from the z-dimension provided a temporal measure that was used to calculated both the

3

interval and the duration of the observation.

A fuller description of the project methodology is contained within the answers to the list of questions below. This set of Frequently Asked Questions (FAQ) was provided to each observer for initial study and reference, in order to ensure methodological consistency (see next section).

General:

Q1. ***What are the general inclusion/exclusion criteria for studies?*** Studies should be excluded from this analysis if they are: 1) opinion/perspectives pieces; 2) model-only studies, particularly theoretical models, which are not developed or tested against observed data; 3) if they are experimental manipulations (but if there is a study that has a mix of observational and experimental, record the observational treatments and exclude the manipulated treatments).

Q2. ***What are the standard categories to be used for defining Study type?*** Define study type according to the following categories: Remote sensing, passive/automated data collection, other geographic data (e.g. non-remotely sensed GIS data), field/direct observation, or paleo-reconstruction (tree rings, charcoal cores, etc).

Q3. ***What happens when the study draws on a separately published dataset as a key part of the methods?*** Track down the study describing the paper, and then record the DOI of that paper/those papers.

Q4. ***What is the best unique identifier of a study I am reviewing?*** The DOI!

Q5. ***What do I record for a time or space scale when it is not clearly reported in the paper, or when I am unsure? For example, in a paleo-ecological study looking at historical charcoal deposition, sediment cores were extracted from lakes, which the authors report as the number of samples. However, it is unclear how many sediment cores were drawn from each lake, and it is these which should be the number of spatial replicates.*** For these

4

78   sorts of issues, we record that the scale in question is uncertain, and then your best estimate

79   of the measure (e.g. you might assume that only 1 core was made per lake).

Temporal scales:

Q6. ***What is sampling duration, and how do we record it?*** How long an individual observation of an individual point in space took to make. Sampling duration multiplied by the number of repeats observations is used to calculate study duration. Often this value will not be reported, so you will have to use your best judgement, based on your knowledge of ecological methods, to approximate the duration. For example, for a field based method with intensive plot methods, if you can't estimate a plausible duration, assign a token 1 day. For remote sensing observations you can assume one second (the observations are effectively instantaneous).

Q7. ***What is sampling interval, and how do we record it?*** Sampling interval is the time that elapsed between repeat observations of the same point in space or individual organism. In many cases, observations will only be made one time–list a value of 0 for these.

Q8. ***What is study duration, and how do we calculate it?*** Study duration is the integral of sampling duration, or the time spent making one observation of the phenomenon in question. To clarify, duration is the total time spent sampling/observing a single point in space–not the span of time between first and last sample, nor the integral of time spent in observing all spatial replicates.

Q9. ***Should study duration be the total time spent sampling all sites or the amount of time spent sampling per site (e.g. for 5 minute point counts of birds at 10 sites each repeated twice, should we enter 100 minutes (5 minutes X 2 repeats X 10 sites) or 10 minutes (5 minutes X 2 repeats) for duration)?*** As stated above, duration is the total spent observing a single point in space, so in this case that would be 10 minutes (then converted to days, so 10 / (60 * 24))).

5

Q10. ***How do we record study duration when there are no repeat observations?*** In these cases, study duration is equal to sampling duration.

Q11. ***How do I record sampling interval in cases where the interval is inconsistent? For example, in a study were observations were observations were repeated in 1979, 1980, 1981, 1984, 2007, 2009?*** Find the time between each successive period, and then take the average of that (remember to convert to days!). If there are two or more sets of unevenly spaced days for each site/plot/measurement being taken in the study, then find the average interval for each, and average the averages.

Q12. ***How do you determine the time between samples for paleo-reconstructions?*** Use the minimum estimate for dating precision as the estimate of time between samples (e.g. 50 years in the study of European charcoal deposits DOI:10.1111/geb.12090).

Q13. ***How do you determine the sample duration (our third time category) for paleo-reconstructions?*** Similar to the previous question, the sampling duration is also the same as the minimum estimate of dating precision. The logic behind this is that in such cases, where a sediment or tree core or similar measurement is being collected, this effectively represents a continuous "observation", and the value associated with the minimum (or other reported) interval is typically an average (or another summary statistic like the maximum) of the amount accumulated.

Q14. ***What about sampling intervals and durations for instrument-collected observations?*** These are often similar in essence to the paleo-reconstruction case. Take the minimum temperature or daily rainfall recorded at a weather station, which require constant observation over 24 hours to report. In such cases, the sampling interval and sampling duration are both 24 hours. On the other hand, automated logging systems will provide a series of high frequency observations that are collected instantaneously. In these cases the sampling duration should be

$\leq$second, and the sampling interval should be the period between successive instantaneous

measurements.

Q15. ***How do you handle the time between samples for a case where repeated samples are taken***

during a season, across several seasons (e.g."we performed repeat bird counts every 10

days between March and June of 2005, 2006, and 2008")? Since the sampling is focused

on seasons, and presumably some season-specific phenomenon (e.g. breeding behavior), the

reported values should be pegged to the season, not averaged across the full study span (the

start and end dates of the study). So in this example, it would be 10 days.

Spatial scales:

Q16. ***What is sampling resolution, and how do I record it?*** This is the finest scale at which a

complete measurement of every unit of the quantity of interest is recorded. For example, if

the measurement in question is a tree stem count, the resolution is determined by the size of

the plot used to record every tree stem. Taking this example further, let's say a study reports

a plot size of 100 x 100 m, but then goes on to report that they counted stems within a single

1 m wide transect within this larger plot. In this case, the plot resolution is in fact 100 m x

1 m, or 100 m^2 (sampling resolution should be reported in m^2). In another example, if the

reported plot size was 20 x 20 m, but the authors in fact only measured a random selection

of, say, grass stems on which they counted aphids within those plots, then use an estimate of

the area of the grass stem as the sampling resolution (*8*).

Q17. ***What is study extent, and how do I record it?*** Broadly defined, study extent is the sampling

resolution multiplied by the number of spatial replicates, divided by 10,000 to convert to

hectares. For studies in which the spatial replicates are not spatially contiguous (as with most

field-based studies), this means plot resolution (see Q16) multiplied by number of plots. For

spatially contiguous studies (e.g. those based on remote sensing imagery), it should be the

total area covered by the imagery, i.e. pixel resolution multiplied by the number of pixels. However, only record the area analyzed by the authors. If they used a sub-sample of pixels, extent is the number of those pixels a multiplied by pixel resolution.

Q18. ***How do you determine sampling resolution for paleo-reconstructions and other approaches where a sampling method is presumed to draw from a larger area (e.g. mammal traps, mist nets, etc)?*** For sampling resolution, estimate the size of the sample actually taken, rather than the assumed catchment/shed area of the sample (e.g. the area of the corer used to take a sediment sample, rather than an estimate of the area that that sample is assumed to draw from), and then indicate that the plot resolution was uncertain. Related to this, you may also have to estimate the number of samples collected, as exemplified in a charcoal study of Europe where the number of lakes sampled was provided, not the number of cores per lake (*9*).

Q19. ***What about studies that sample individual organisms?*** If the study is making a total count of all organisms (let's say a mammal species) within a fixed plot size, or even a variable plot size from which an average plot size (and thus sampling resolution) can be estimated, then follow the procedures described in Q16-17. However, if the individual animals are the unit of measure (either because a sub-sample of them is being made within a defined plot, or because the observation is not contingent on being located within a plot (maybe a blood sample or body weight is being recorded, for example), then simply estimate two-dimensional area occupied by the animal, the number of animals sampled, and then calculate area from that. Occasionally individual animals might be recorded, but within the context of some natural feature, such as a nesting site where the survival of individual chicks is the measurement of interest (*?*). In this case, an estimate of the nest area provides the sampling resolution. Another possible case is where individual animals are being tracked using radio or GPS collars. In these cases, use the estimate of the collared/tagged animal's body area, and then

treat the number of reported locational observations as the number of spatial replicates, and then use those two values to calculate sampled area. The total GPS points collected from all individuals can be used if we can safely assume that each individual sampled did not cover the same exact location. If the number of GPS points is not given, the number of fixes can be estimated from the duration during which individuals were collared and the recording interval.

Calibration and consistency

Most studies did not explicitly report values for all four of the assessed scales, and thus interpretation and judgement had to be applied to develop reasonable estimates for their values. The FAQ provided the protocol we followed, and was initially developed following a consultative process prior to the commencement of review. We conducted an iterative process of calibration to ensure consistency and reliability of estimates. First, we used the calibration set to calculate between-observer variability with respect to paper selection/rejection and the estimation of scales. Based on this, the lead author reviewed individual records in each observer's calibration set, flagged values were the estimation procedure departed from the protocol, and returned these to observers for re-estimation, without providing an estimate of the actual value. Instead, the relevant section of the protocol was highlighted, and further explanation and clarifying discussion undertaken as needed. Protocol language was adjusted for clarity during this process, and new items added to cover circumstances that had not been addressed by the initial version. The variability measures (see Results) were re-calculated after each iteration.

To ensure consistency within the main analysis, the lead author also reviewed each observer's results from their individual draw of papers and flagged values that appeared to deviate from the protocol for re-review by the observer. Re-reviewed values were re-inspected, and in some cases a secondary review of particular papers was undertaken to cross-check the estimated scales.

Estimating observational density

In presenting results (Fig. 1 and 2 in main text), we log-transformed (base 10) the observational scales because of the large range in values. To reveal the densies of observations within two dimensions (Fig. 2 main text), we used the splancs package (*10*) of R (*11*) to calculate a kernel density estimate of the log-transformed values, using a bandwidth of 1 onto a 0.1 resolution image to provide a smoothed result that served to more effectively highlight domains in which ecological observations are concentrated. Bandwidths of varying resolutions were tested (see Supplementary Results section).

Sensitivity of results to observer and estimation error

To test the impact of between-observer inconsistencies on results, as well as the possibility of error by individual observers made in estimating the scales of actual observations scale, we conducted a sensitivity analysis. For each of the four dimensions, we performed a 1000-iteration resample of values, adding to each record a random amount of error at each iteration. The error values were constrained to fall within a maximum percentage of the record's value, which was determined by the coefficient of variation between observers' estimates for the calibration set. We then combined the set of 1000 perturbed observations, log-transformed them, recalculated the kernel density estimates for the comparisons of the two resolution and extent dimensions (i.e. Fig. 2A and B in main text), and assessed the degree to which the domains of observational concentration differed from those in our main analysis (Fig. S2).

Results

Variability and consistency between observers

We assessed variability and consistency between observers along multiple dimensions. First, we assessed how reliablly observers agreed with respect to selecting or rejecting papers for review, using the R's Agreement package (*12*) to calculate Fleiss' kappa statistic (*13*), which was 0.72 (z = 12.5, p<0.000), indicating substantial and significant agreement between observers (*14*).

Second, we calculated the intra-class correlation coefficient (*15*) to assess agreement between observers regarding the number of ecological observations that could be extracted from each paper (multiple ecological observations were reported in many studies, and we listed observations as separate records if they varied on one or more dimensions). The coefficient, calculated using the IRR package (*16*) of R, was 0.70 (F = 15.4, p<0.0001, 95% confidence interval = 0.54 - 0.85)

Finally, we calculated the coefficient of variation (CV) between observers' estimates of scales for each dimension, first across all six observers' mean scale estimates, and then as the average CV among observers' estimates of each individual record (Table S2).

Table S2: Estimates of variability calculated from each observer's estimates of the spatial and temporal scales of ecological observations reported within a common set of the 20 papers used for calibration. Variability is expressed as the coefficient of variation (CV; standard deviation divided by mean multiplied by 100) between each observer's overall mean, and as the mean CV of observers' estimates for individual records.

	Spatial		Temporal	
Value	Resolution	Extent	Interval	Duration
CV of overall mean	54	82	54	110
Mean of record-wise CV	58	68	99	124

Sensitivity analysis

We used the maximum uncertainty values for each dimension from the inter-observer variability analysis (Table S2) to determine the bounds of the random perturbations applied to each record in the sensitivity analysis. Kernel density estimates applied to the merged results of this analysis for the contrasts between spatial resolution and temporal intervals, and to spatial extent and temporal duration, suggest that the primary features observed in Figures 2A and 2B (main text) are largely unaltered by the addition of uncertainty to the estimates. The largest difference is in the greater density of second to sub-second duration observations across all spatial extents.
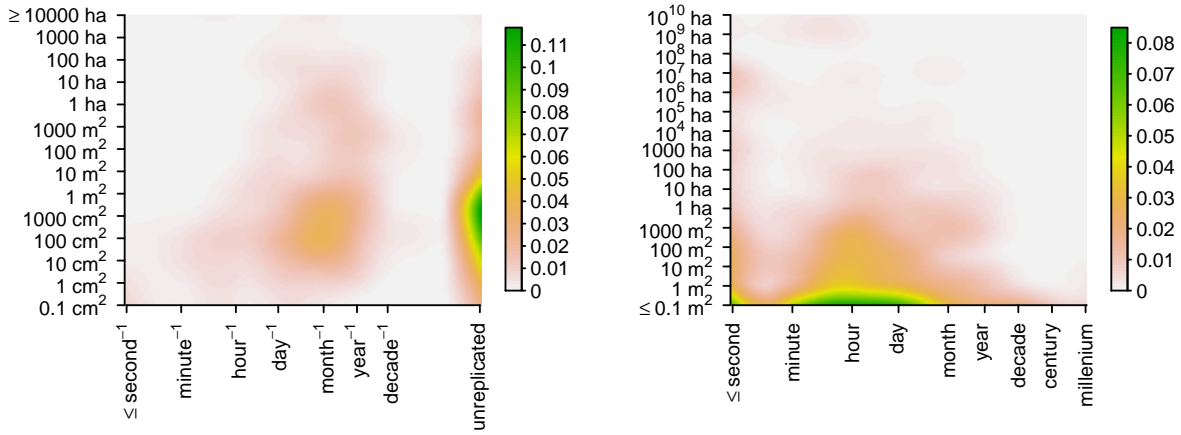
Figure S1: Two-dimensional kernel density estimates of observational densities within the domains defined by sampling interval and spatial resolution (left) and temporal duration and spatial extent (right). Density estimates were applied to the log-transformed values of each observational dimension after being randomly perturbed within uncertainty bounds drawn from an assessment of between-observer variability. Density estimates are rescaled to represent percentages.

Figure S2 indicates the effect that varying bandwidth has on the appearance of the kernel density estimates, as estimated for the same two contrasts shown in Figure S1. the smallest bandwidth (0.4) tested (Fig. S2 top) shows that the primary observational concentrations described in the main text are evident but harder to discern. Most difficult to discern is the oblong concentration of observations that in Fig. 1A (main text) is bounded on the lower right at monthly to yearly intervals and 100-1000 m^2 resolutions and on the upper left by near-daily to monthly observations and 0.1-10 ha resolutions. With the smaller these appears as two separate patches (Fig. S2 top left), but with 0.7 bandwidth applied become coherent (Fig. S3 middle left).
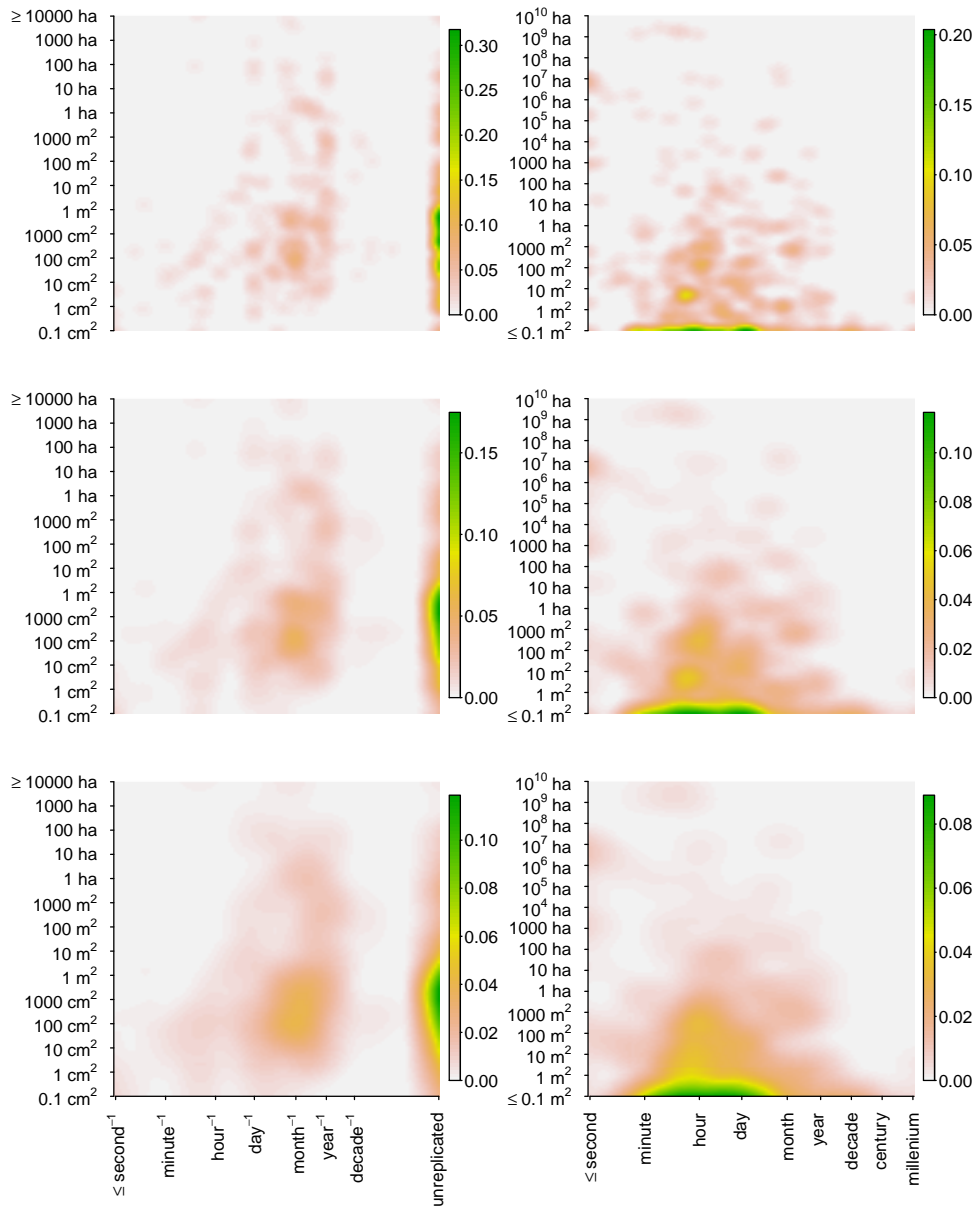
Figure S2: Two-dimensional kernel density estimates of observational densities within the domains defined by sampling interval and spatial resolution (left column) and temporal duration and spatial extent (right column), applied to log-transformed values of each observational dimension. Rows indicate the effects of selecting different bandwidths: 0.4 (top row); 0.7 (middle row); 1 (bottom row).

References and Notes

1. P. Kareiva, M. Andersen, *Community ecology* (Springer, 1988), pp. 35–50.

2. D. Tilman, C. Balzer, J. Hill, B. L. Befort, *Proceedings of the National Academy of Sciences* (2011).

3. A. J. Underwood, *Experiments in Ecology: Their Logical Design and Interpretation Using Analysis of Variance* (Cambridge University Press, 1997).

4. M. W. Palmer, P. S. White, *American Naturalist* pp. 717–740 (1994).

5. Y. Cao, D. D. Williams, D. P. Larsen, *Ecological Monographs* **72**, 41 (2002).

6. P. Legendre, *Ecology* **74**, 1659 (1993).

7. S. L. Collins, F. Micheli, L. Hartt, *Oikos* **91**, 285 (2000).

8. V. Gagic, *et al.*, *Oecologia* **170**, 1099 (2012).

9. C. Molinari, *et al.*, *Global Ecology and Biogeography* **22**, 1248 (2013).

10. B. S. Rowlingson, P. J. Diggle, *Computers & Geosciences* **19**, 627 (1993).

11. R Development Core Team, *R: A Language and Environment for Statistical Computing* (Vienna, Austria, 2011). ISBN 3-900051-07-0.

12. Y. Yu, L. Lin, *Agreement: Statistical Tools for Measuring Agreement* (2012). R package version 0.8-1.

13. J. L. Fleiss, *Psychological Bulletin* **76**, 378 (1971).

14. J. R. Landis, G. G. Koch, *Biometrics* **33**, 159 (1977).

266   15. J. J. Bartko, *Psychological Reports* **19**, 3 (1966).

267   16. M. Gamer, J. Lemon, I. F. P. S. <puspendra.pusp22@gmail.com>, *irr: Various Coefficients*

268   *of Interrater Reliability and Agreement* (2012). R package version 0.84.