

Calculating nominal group statistics in collaboration studies

DANIEL B. WRIGHT

University of Sussex, Brighton, England

In many areas of psychology researchers compare the output of pairs of people with people working individually. This is done by calculating estimates for nominal groups, the output of two individuals if they had worked together. The way this is often done is by creating a single set of pairs either randomly or based on their location in a data file. This paper shows that this approach introduces unnecessary error. Two alternatives are developed and described. The first calculates statistics for all permissible sets of pairs. Unfortunately the number of sets is too large for modern computers for moderate sample sizes. The second alternative calculates statistics on all possible pairs. Several simulations are reported which show that both methods provide good estimates for the mean and trimmed mean. However, the all pairs procedure provides a biased estimate of the variance. Based on simulations, an adjustment is recommended for estimating the variance. Functions in S-Plus/R are provided in an appendix and are available from the author's Web page along with updates and alternatives (www.sussex.ac.uk/users/danw/s-plus/ngstats.htm).

Are two heads better than one? This question is at the heart of much research on collaboration in many areas of psychology (e.g., Diehl & Stroebe, 1987). Applicable areas include organizational, forensic, cognitive, educational, and social psychology. Should people work together or should they work separately?

While this research question and the statistical techniques described below are relevant to many areas of psychology, for illustrative purposes this paper will concentrate on memory research. In recent years many memory researchers have explored whether groups can recall more than individuals (e.g., Andersson & Rönnerberg, 1995, 1996; Basden, Basden, & Henry, 2000; Finlay, Hitch, & Meudell, 2000; Thompson, 2002; Meudell, Hitch, & Boyle, 1995; Weldon & Bellinger, 1997; Weldon, Blair, & Huebsch, 2000; Wright & Klumpp, 2004). In a typical study, participants are shown some stimuli and then are asked to recall as many of the stimuli as they can. They either recall in pairs (or larger groups) or individually. The finding is that the mean amount recalled is higher for the pairs than for the individuals. This is not surprising because there are two people recalling in the pairs, but only one for those recalling individually. To account for this, researchers calculate the amount recalled for *nominal groups*. Nominal groups are individuals who recall separately but whose scores are combined as if they were in a group.

The recall for a nominal group is the number of stimuli recalled by the two individuals, but only counting items that both people recall as one item. Nominal groups tend to recall more than actual groups, a finding known as *collaborative inhibition*. Thus, two heads working together

are better than one, but not as good as two heads working separately. This is an important finding for both cognitive and social psychological theories of memory, and for real world problems like students preparing for an exam together and eyewitnesses speaking with each other (Wright, Mathews, & Skagerberg, 2005).

The nominal groups can be formed in different ways. Sometimes it is not clear from articles how the groups were constructed and speaking with some of these authors they have said they grouped individuals who happened to be next to each other in the data file. In other articles it is explicitly stated that "random combinations of participants who recalled alone" were used (Finlay et al., 2000, p. 1558). I will refer to these approaches as arbitrary, each choosing one particular set of nominal groups from all the possible sets. The focus of this paper is alternatives to this arbitrary approach. In other studies people have participants arrive in groups, and then are assigned groups either to a recall-together condition or to a recall-separately condition (for example, Weldon & Bellinger, 1997). In these studies, the researcher can argue that these are intact groups before the random assignment and can treat them as such. Unfortunately, the authors of many of these papers also conduct statistics on the individual data assuming that they are independent rather than using multi-level modeling which is arguably more appropriate if they really do consider these individuals to be in their groups (Wright, 1998). Thus, it is not clear how they are treating the individuals within these nominal groups.

The remainder of this paper is divided into five parts. First, problems using the arbitrary approach to nominal

D. B. Wright, danw@sussex.ac.uk

Table 1
Data to Illustrate the Different Methods of Assigning
Participants to Nominal Groups and Calculating Nominal
Group Recall

Participant	Item 1	Item 2	Item 3	Item 4	Item 5	Participant total
1	1	1	0	1	0	3
2	0	0	1	0	1	2
3	1	1	0	1	0	3
4	1	1	1	1	0	4
5	0	1	0	0	1	2
6	0	1	1	0	1	3
Item total	3	5	3	3	3	

group construction are identified. Second, a simple example is presented to illustrate the different methods which will be compared. The hypothetical data set contains only six participants and five items to recall. The numbers are small for illustrative purposes. Increasing the number of items to recall creates no difficulties. Calculating the group's total only requires calculating the dot product of the individuals' responses within each group. Increasing the sample size *does* create computational difficulties, though conceptually the problem is not difficult. In the third section the computational details are discussed. In the fourth section some simulations are conducted and the results reported. The final section provides recommendations and offers some extensions.

The aim of forming nominal groups is to provide the best estimates of central tendency and spread that would have been expected from a sample of size n allocated into groups. The most common measures of central tendency and spread are the mean and variance, respectively. The focus here is on these, although some of the simulations also use robust estimators.

COLLABORATIVE MEMORY AND NOMINAL GROUPS

Consider the following hypothetical study. Suppose eight participants are recruited to take part in a collaborative memory study where they are presented with the names of ten fruits: apple, apricot, banana, cherry, lemon, lime, grape, orange, peach, and pear. Four people are allocated to a collaboration condition and four to a control condition. The four people in the collaboration condition are grouped into two actual pairs. Within the pair, the people attempt to recall the items that they were previously presented. The pair receives a score for the number of items recalled. Suppose one pair recalls five items: apple, banana, grape, orange, and pear.

Participants in the control condition recall individually. Suppose the first two control participants each recalled four items:

- Participant 1: apple, banana, lemon, and lime,
- Participant 2: apple, banana, orange, and pear.

These people are combined into a nominal group. While separately each scores less than the actual group, com-

bined they recalled: apple, banana, lemon, lime, orange, and pear. Thus, they recalled six items, one more than the actual group. If each participant is given a 0 for each fruit ($j = 1, \dots, 10$) that they recall and a 1 for fruits that they do not recall, then the total for the nominal group of participant 1 ($P1$) and participant 2 ($P2$) is $10 - \sum P1_j P2_j$ (or 10 minus the dot product of each participant's vector of responses). In words, this is the total number possible minus the number of fruits that neither participant recalled.

The problem with the arbitrary method is that assignment to nominal groups is based on only one possible grouping, sometimes just based on when the participants arrived for the study or their location in the data file. The choice can make a difference. Consider two other participants:

- Participant 3: apple, cherry, lemon, and lime,
- Participant 4: banana, orange, peach, and pear.

The nominal group ($P3, P4$) has a total of eight, so the set $\{(P1, P2)(P3, P4)\}$ has mean of seven fruits recalled. If the pairs ($P1, P3$) and ($P2, P4$) are used then the nominal group totals are five and five, a mean of five. The difference between 70% recalled and 50% recalled is substantial and could affect interpretation. Simulations show that the size of this difference can occur. It is clear that some alternative methods need to be developed.

The number of unique pairs that can be made from n participants is C_2^n (i.e., n choose 2), so with 10 participants the total number of pairs is $C_2^{10} = 45$. In a study with n participants in the control condition, $n/2$ nominal groups are created. With $n = 10$, this would be five nominal pairs. The total number of sets of five pairs that can be drawn from 45 is $C_5^{45} = 1,221,759$. In general, this total becomes large quickly. However, not all sets of pairs are permissible because each participant can only be used once in each set. The number of permissible sets is: $(n-1)(n-3) \dots (n - [n-1])$ (proof for this in Appendix A). For $n = 10$ this is only 945. If the nominal groups are created arbitrarily, then any of these 945 sets of 5 pairs could be chosen. When a researcher compares experimental groups with nominal groups, their results will be determined by this choice, which of the 945 sets has been chosen. The main idea of this paper is that this needlessly adds error into the analysis and should be avoided.

Two alternatives are developed and compared. Other alternatives also exist. The first is to pick all the permissible sets of pairs and calculate the relevant statistics. In combinatoric terminology, this is taking all set partitions of the group where each set has only two elements. This takes away the arbitrariness of the traditional approach, but it does have some computational difficulties. The second alternative is to calculate statistics for all the pairs of people in the control group. In the collection of all permissible sets, each pair is used the same number of times. For some of the statistics, like the mean, it is unnecessary to group the pairs into sets. The researcher can simply calculate statistics on the C_2^n pairs, though as will be shown, the standard error of the mean is biased so some caution is needed.

Table 2
The Number of Items Recalled for Each of the 15 Nominal Groups

Participant	1	2	3	4	5
2	5				
3	2	5			
4	2	4	2		
5	4	4	4	4	
6	4	3	4	3	3

Note—Statistics: mean = 3.53, variance = 0.98.

An Example

Table 1 shows the data for six participants who were shown five items and then individually asked to recall these. In a collaboration condition people would have been placed into actual pairs. Here the interest is with the control condition and it is assumed that each person’s responses are independent of each other’s responses. If the person recalled the item there is a 0 in the cell; if the person failed to recall the item there is a 1 in the cell. With $n = 6$, there are $C_2^6 = 15$ pairs and $(6-1)(6-3)(6-5) = 15$ permissible sets of pairs. These are shown in Tables 2 and 3 with some descriptive statistics.

Table 2 shows the mean of all pairs: 3.53. This is the same as the mean of all sets, shown in Table 3. Thus, if only the mean is desired, then calculating the mean of all pairs can be done, rather than using the more complex and time consuming procedure used for Table 3. If using the arbitrary method, any of the 15 sets in Table 3 could have been chosen. Depending on the choice, the mean control recall could be as low as 60% (3.00 items) or as high as 87% (4.33 items). While scientists are used to observing large differences between different samples of data (i.e., due to sampling error), these estimates are from the *same* data. Arbitrary choosing how to construct nominal groups is analogous to calculating a bootstrap estimate with only one replication. While using any single randomly chosen set will provide an unbiased estimate, it will have a large amount of variability associated with it.

Further, note that the variances in the final column of Table 3 differ considerably. Given that t is inversely proportional to the standard deviation, shifts in the standard deviations/ variances often have a large effect on inference about means (Wright, 2006).

Computational Details

Computational aspects of the different methods for calculating nominal group performance are described here. The procedures for the two alternatives are called `ngallsets` and `ngallpairs` for nominal groups for all sets and for all pairs, respectively. The procedures have been written in S-Plus 6 and runs without any modification in R.2.2.0 (other than changing `stdev` to `sd` in one part). The code is listed in Appendix B and documented so that others can easily adapt it to their needs and their software. Other software could have also been used. Software designed more for mathematics has more inbuilt probability and set functions. S-Plus was used here because it is relatively common in academic institutions, it has a

freeware version (R), and it has many statistical functions available which can be used if the user desires.

Nominal groups for all permissible sets of pairs. A permissible set of pairs is one which includes each individual once and only once. It is the set of all partitions of the $(1..n)$ of where each set has two elements. The number of permissible sets of pairs increases with the sample size as: $(n-1)(n-3) \dots (n - [n-1])$. There are two problems calculating this set. The first is that for moderate sample sizes this number becomes too large for modern computers. The second problem is that the conceptually simplest methods for calculating this require first calculating an even larger number of sets, which makes them impractical for even small samples. I will first describe the conceptually simplest methods.

One method to calculate all permissible sets is to calculate all combinations of the C_2^n pairs and then save the sets that are permissible. This would mean just making sure that no person is used more than once. However, the number of combinations to check is massive for the typical sample sizes used in psychology studies. This is: $C_{n/2}^x$, where $x = C_2^n$. For $n = 20$ this is about 1.3×10^{16} , which makes calculations impractical. An alternative is to calculate all permutations of $1..n$ and treat the first and second people as a pair, the third and fourth as a pair, etc. These sets satisfy the condition that no person is used more than once, but this method will still include many impermissible sets because there will be duplicates. The subset that are permissible include only those sets where $i < j$ for all pairs (i,j) and that the first item in every pair is less than the first item in the next pair. The problem with this method is that there are $n!$ permutations of $1..n$. With $n = 20$ this is approximately 2.4×10^{18} . For the typical sample sizes, these methods exceed computational capabilities.

A search was done of software packages and code to see if this could be calculated without having to calculate

Table 3
All the Sets of Permissible Pairs From the Data in Table 1, the Nominal Group Recall for Each of the Three Groups With the Set, the Mean and Variance for the Set, and Some Summary Descriptive Statistics for All the Sets

Set	Recall1	Recall2	Recall3	Mean	Variance
12 34 56	5	2	3	3.33	5.43
12 35 46	5	4	3	4.00	1.00
12 36 45	5	4	4	4.33	0.11
13 24 56	2	4	3	3.00	1.00
13 25 46	2	4	3	3.00	1.00
13 26 45	2	3	4	3.00	1.00
14 23 56	2	5	3	3.33	5.43
14 25 36	2	4	4	3.33	1.77
14 26 35	2	3	4	3.00	1.00
15 23 46	4	5	3	4.00	1.00
15 24 36	4	4	4	4.00	0.00
15 26 34	4	3	2	3.00	1.00
16 23 45	4	5	4	4.33	0.11
16 24 35	4	4	4	4.00	0.00
16 25 34	4	4	2	3.33	1.77

Note—Statistics for means: $M = 3.53$, $SD = 0.52$, $\min = 3$, $\max = 4.33$; statistics for variances: $M = 1.44$, $SD = 1.71$, $\min = 0$, $\max = 5.43$.

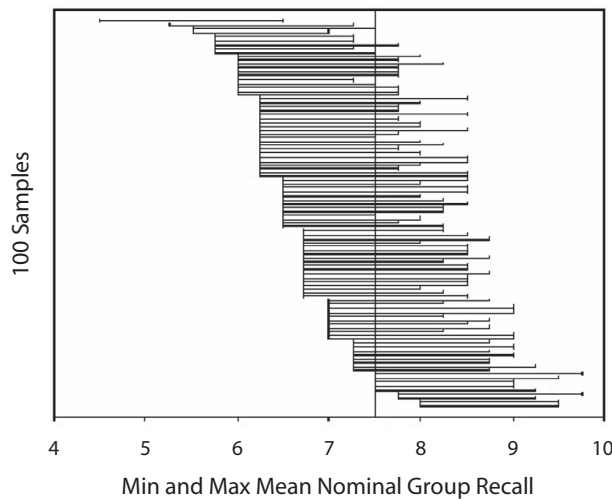


Figure 1. The range of mean nominal group recall for 100 samples with $p = .50$ and $k = 10$ that could be chosen using the traditional approach. The range is due to the arbitrary nature of how nominal groups are chosen is approximately the size due to sampling variability.

a much larger set. Maple appears to have a function that does this, but the source code could not be located. Instead, an alternative method was devised to calculate the permissible sets without having to calculate some larger set first. The technique is based on the proof given in Appendix A. The permissible sets can be calculated by first calculating the set for when $n = 2$. This is $\{(1,2)\}$. Then you merge this with the next 2 numbers so that you have $\{(3,4),(1,2)\}$. This is a permissible set when $n = 4$. Next you switch one of the new numbers with each of the existing numbers. If you switch the “3” this yields $\{(1,4),(3,2)\}$ and $\{(2,4),(1,3)\}$. The three permissible sets when $n = 4$ are: $\{(3,4),(1,2)\}$, $\{(1,4),(3,2)\}$, and $\{(2,4),(1,3)\}$. This is repeated so that one of the items in the next pair, (5, 6), is switch with each of the four numbers in each of the three existing sets. Including the original permissible set of $n = 4$ with (5,6) as the first pair, this yields 15. This procedure is repeated for $n = 8, n = 10$, etc. The computation for this is done in the function `permiss` in the Appendix B. It can be run in R 2.2.0 without modification, although for the statistics produced `stdev` needs to be changed for `sd`, for the standard deviation.

When n is small this method is possible, but as n increases the number of permissible sets gets large. For example, with $n = 30$, which is only 15 pairs, the number of permissible sets of pairs is approximately 6.19×10^{15} . While this is much less than $n!$ or all possible sets of pairs, as the sample size increases this number quickly exceeds computational capabilities. In a typical study a researcher might wish to compare collaborative pairs and control pairs using a t test with an ability to reject a large effect size (0.8 sd difference) at $\alpha = .05$ with an 80% likelihood. To do this, 25 pairs are needed in each condition so an $n = 50$ is needed. For $n = 50$ there are 5.8×10^{31} permissible sets of pairs. If the researcher wished to detect Cohen’s (1988) medium effect size (0.5 sd difference)

this requires 63 pairs. This produces 1.3×10^{105} permissible sets of pairs. Clearly performing calculations on all permissible sets, as in Table 3, is not possible even with moderate sample sizes.

Nominal groups for all pairs. Every pair occurs the same number of times as each other in the collection of all permissible sets. Therefore, an alternative would be to calculate statistics on all pairs. If there are n people, the number of pairs is C_2^n . While this increases with the sample size (for example, $C_2^{20} = 190$, $C_2^{30} = 435$, and $C_2^{100} = 4950$), the increase is not large enough to cause any particular concerns. The calculations are simple. All possible combinations (i,j) are created and those which have $i < j$ are included. The amount of items free recalled by the nominal group (i,j) is based on the dot product of the i^{th} and the j^{th} participants’ scores. These C_2^n scores are stored in a vector and statistics computed.

Simulations Comparing Methods

Three sets of simulations are conducted. The first is simply to show that the arbitrary choice of sets is problematic. The second is to compare the two alternatives, `ngallsets` and `ngallpairs`. Because of computational difficulties with moderate ns , this will be done with a small n of 8, varying the number of items (5 and 10) and the individual accuracy rate (75%, 50%, and 25%) in the first simulation and then allowing accuracy rates to vary by individual person in the second simulation. The third set of simulations examines `ngallpairs` for different sample sizes. The first set of simulations is done with only 100 replications because that is all that is necessary to demonstrate the inadequacy of the arbitrary approach. For the second sets of simulations there are 1000 replications for each condition and the mean, the 20% trimmed mean (20% from each tail), and variance are estimated for each trial. Quantile values (minimum, 2.5%, 25%, 50%, 75%, 97.5%, and maximum) and the mean are calculated for these 1000 replications for the three statistics. There are many advantages using the quantiles to understand the distribution (Yu, Lu & Stander, 2003). The final set of simulations systematically examines the effects of probability and sample size on estimates of mean and variance. The functions in the appendix were used with some minor additions for printing output and running multiple replications.

For a fixed probability of recalling an item, the predicted, or true, values can be calculated. If the probability for someone recalling an item is p , and this is the same for both people, then the probability of either member of pair recalling the item is $1 - (1 - p)^2$ which will be denoted as p' . If the probabilities of recalling each item are the same, then the mean should be kp' where k is the number of items to be recalled. The variance for the binomial distribution is also known: $kp'(1 - p')$.

Examining the arbitrary approach. Two simulations were done to demonstrate that the arbitrary method produces too wide a set of estimates for any given sample for the method to be used. Both simulations produced 100 samples of size eight and sampled all the sets (105 for each sample) that could have been used to create the set

of nominal pairs (more replications could be taken, but this is unnecessary for showing the problems with this method). For the first simulation, the probability of someone accurately recalling an item was fixed at 50% and there were ten items. The minimum and maximum for each sample's data are plotted in Figure 1. The predicted mean for $p = .50$ ($p' = .75$) and $k = 10$ is 7.50 items recalled. This is shown by the vertical line in Figure 1. The mean range for the different samples is 1.73 items which shows how even after the data have been collected, the choice of how nominal groups are constructed can make a sizeable difference.

The next simulation shows that having different levels of ability increases the unreliability of the arbitrary method of determining nominal groups. The simulation was repeated with people having different probabilities for recalling the items. For simplicity half the people had a probability of .30 to recall each item and half had a probability of .70 to recall each item. Items are assumed not to vary in difficulty. For 25% of the permissible pairs $p' = .51$ (2 low recall people), for 25% $p' = .91$ (2 high recall people), and for half $p' = .79$ (1 low recall person and 1 high recall person), which results in a weighted mean of 0.75 and therefore a predicted mean of 7.50 items recalled. Because p' is higher for the mixed pairs than for the mean of the other 2 groups, the more mixed pairs the higher the likely nominal group mean. Thus, having different probabilities of accurate recall theoretically should increase the variability of the estimates. The results for 100 samples are shown in Figure 2 and they show that the variability does increase. The mean range is now 1.91 items. If the discrepancy is increased so that half of the sample have $p = .10$ and half have $p = .90$, the mean range for 100 samples becomes 3.46 items.

The aim of these first two simulations was to show that arbitrarily choosing one set of pairs introduces additional error that can make a difference. An obvious solution is that instead of arbitrarily choosing one set from within the range of each line shown in Figures 1 and 2 the researcher should to take some measure of central tendency of all the sets. This is the method used in the next set of simulations with `ngallsets`.

Calculating all sets of pairs versus all pairs. The computational demands of `ngallsets` are large, but it theoretically is optimal for calculating all possible statistics because each permissible set is used. Therefore, the estimates from this can be used as the best estimates, being affected only by sampling error in the simulation. Further, for fixed accuracy rates the true nominal group performance in the population can be easily calculated so the values can be compared.

The third simulation uses three different fixed accuracy rates: $p = .75$, $.50$, and $.25$, with both five and ten items. A thousand replications are used for each condition. The results are shown in Table 4. The mean values are the same for the two methods and differ from the true means only by sampling variability (and rounding). The 20% trimmed means are also fairly close to the true means, but the two methods do produce different results. When p' gets closer to either the maximum or minimum possible, the estimates

from the pairs method for the trimmed mean are slightly more extreme (the difference in means is about 0.1 of an item). The variance estimates show a large discrepancy between the sets and pairs methods. The sets method is approximating the true value. However, the pairs estimate is consistently underestimating the true value. If this value is used, unadjusted, to calculate the standard error, then the estimates will appear too precise. More comparisons with the experimental group will be statistically significant than is appropriate. Importantly, the underestimation appears to increase as p gets smaller (or to be more precise, as p' moves away from .50). This will be investigated in the next set of simulations.

The simulation was repeated, for illustrative purposes, with the individual differences used above: half the sample with a probability of .30 of recalling an item and half with a probability of .70 of recalling an item. Ten items were used. The results are shown in the bottom of Table 4. The change to having two different recall probabilities does not substantially alter the mean and trimmed mean estimates. The variances are larger than the comparable condition with a fixed probability of .50, as expected. If it is assumed that `ngallsets` provides approximately the correct values, then `ngallpairs` again underestimates the variance.

Given the computational difficulties using all sets of pairs, it is worth looking closer at the all pairs statistics. While the mean and trimmed mean estimates are near the true values, the estimates for the variance are below the true value. This is important because this will affect the estimated standard error making any estimate appear more precise than it actually is. This is examined in the next simulation.

Increasing sample size. A fifth simulation was conducted with the sample size varied with the all pairs function (`ngallpairs`). Table 5 shows the results for control conditions with $n = 8$ (using data from Table 4),

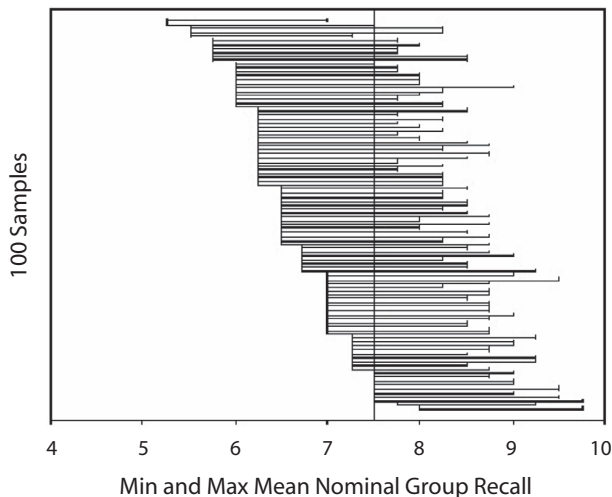


Figure 2. The range of mean nominal group recall for 100 samples when half the sample have $p = .30$ and half have $p = .70$ of recalling each item ($k = 10$) using the traditional approach. The range of scores is even larger than when p is constant.

Table 4
Comparing Using All Sets of Pairs and All Pairs for Estimating Nominal Group Statistics

Condition	Methods	Stats	Min	2.5%	25%	Median	75%	97.5%	Max	Mean	True Value
$p = .75$ $k = 5$	Sets	mean	3.82	4.25	4.57	4.71	4.82	4.96	5.00	4.69	4.688
		trimmed	3.83	4.25	4.61	4.74	4.86	5.00	5.00	4.71	
		variance	0.00	0.04	0.15	0.24	0.37	0.73	1.22	0.28	0.293
	Pairs	mean	3.82	4.25	4.57	4.71	4.82	4.96	5.00	4.69	4.688
		trimmed	3.94	4.33	4.72	4.89	5.00	5.00	5.00	4.82	
		variance	0.00	0.04	0.15	0.23	0.33	0.62	1.04	0.26	0.293
$p = .50$ $k = 5$	Sets	mean	2.39	2.86	3.50	3.79	4.07	4.50	4.71	3.75	3.750
		trimmed	2.39	2.84	3.49	3.78	4.07	4.49	4.72	3.76	
		variance	0.23	0.36	0.64	0.84	1.12	1.92	3.30	0.92	0.938
	Pairs	mean	2.39	2.86	3.50	3.79	4.07	4.50	4.71	3.75	3.750
		trimmed	2.33	2.89	3.50	3.83	4.17	4.61	4.83	3.82	
		variance	0.21	0.33	0.56	0.72	0.94	1.57	2.69	0.79	0.938
$p = .25$ $k = 5$	Sets	mean	0.50	1.14	1.82	2.21	2.50	3.11	3.82	2.16	2.188
		trimmed	0.50	1.15	1.84	2.20	2.53	3.13	3.83	2.18	
		variance	0.21	0.42	0.81	1.10	1.49	2.67	4.68	1.21	1.231
	Pairs	mean	0.50	1.14	1.82	2.21	2.50	3.11	3.82	2.16	2.188
		trimmed	0.44	1.00	1.78	2.17	2.56	3.17	4.06	2.16	
		variance	0.19	0.34	0.67	0.89	1.21	2.13	3.66	0.98	1.231
$p = .75$ $k = 10$	Sets	mean	8.18	8.79	9.25	9.43	9.57	9.82	9.64	9.39	9.375
		trimmed	8.20	8.78	9.25	9.42	9.60	9.83	10.00	9.40	
		variance	0.04	0.17	0.36	0.51	0.69	1.23	1.98	0.56	0.586
	Pairs	mean	8.18	8.79	9.25	9.43	9.57	9.82	9.96	9.39	9.375
		trimmed	8.22	8.78	9.39	9.56	9.72	10.00	10.00	9.51	
		variance	0.04	0.15	0.33	0.48	0.63	1.10	1.66	0.51	0.586
$p = .50$ $k = 10$	Sets	mean	5.11	6.25	7.14	7.54	7.93	8.57	9.00	7.51	7.500
		trimmed	5.11	6.25	7.14	7.55	7.93	8.58	9.02	7.52	
		variance	0.31	0.69	1.30	1.72	2.27	3.80	6.47	1.86	1.875
	Pairs	mean	5.11	6.25	7.14	7.54	7.93	8.57	9.00	7.51	7.500
		trimmed	5.28	6.33	7.17	7.61	8.00	8.67	9.11	7.56	
		variance	0.30	0.67	1.15	1.48	1.90	3.13	5.15	1.59	1.875
$p = .25$ $k = 10$	Sets	mean	1.71	2.86	3.86	4.32	4.82	5.79	6.68	4.35	4.375
		trimmed	1.75	2.88	3.87	4.34	4.85	5.78	6.73	4.36	
		variance	0.41	0.76	1.56	2.22	3.00	5.30	9.86	2.41	2.461
	Pairs	mean	1.71	2.86	3.86	4.32	4.82	5.79	6.78	4.35	4.375
		trimmed	1.67	2.83	3.83	4.33	4.83	5.78	6.72	4.34	
		variance	0.40	0.66	1.29	1.80	2.53	4.18	7.76	1.95	2.461
$p = .30, .70$ $k = 10$	Sets	mean	5.25	6.46	7.18	7.54	7.89	8.50	8.93	7.53	7.500
		trimmed	5.22	6.46	7.20	7.55	7.91	8.48	8.95	7.54	
		variance	0.83	1.64	2.96	3.97	5.05	7.73	9.73	4.14	
	Pairs	mean	5.25	6.46	7.18	7.54	7.89	8.50	8.93	7.53	7.500
		trimmed	5.44	6.56	7.44	7.78	8.17	8.83	9.33	7.78	
		variance	0.76	1.40	2.46	3.27	4.17	6.40	8.08	3.41	

Note—Trimmed mean is a 20% trim. No true value is estimated for the true trimmed mean. Sample size is 8 for all simulations. Each condition is based on 1,000 replications.

12, 16, 32, 64, and 128. The time taken to run these increases slightly with the sample size, but even much larger samples can easily be run. These were run with $p = .50$ and $k = 10$, so have a predicted mean of 7.50 items and a predicted variance of 1.875. Statistics for the means, trimmed means, and variances are shown in Table 5. The means and trimmed means provide good estimates for the true mean. While there is variability in these estimates it is due to the data varying, rather than the choice of nominal groups. The variance estimates remain problematic as they underestimate the true value. However, increasing the sampling size improves the estimates. Recall from Table 4 that the underestimation was greatest when p' was further away from .50.

A final simulation was conducted to investigate how the sample size and probability affect the variance estimates for `ngallpairs`. Samples sizes from 4 to 120 (by increments of 4) were tested with $p = .30, .40, .50, .60, .70, .80, \text{ and } .90$. Figure 3 shows the mean variances for 100 replicates for each of these conditions. The horizontal lines show the predicted variances for the different probability levels. As the sample size increases the observed variances approach the predicted variance. While the absolute size of the deviation between observed and predicted variances is largest for low probabilities (where the predicted variance is high), the ratio of the observed and predicted variances is less variable with respect to probability. A model for estimating the true variance that

Table 5
Simulations for Different Sample Sizes (With 10 Items and $p = .50$) Using the All Pairs
Method for Calculating Nominal Group Statistics

Sample size	Stats	Min	2.5%	25%	Median	75%	97.5%	Max	Mean
$n = 8$	mean	5.11	6.25	7.14	7.54	7.93	8.57	9.00	7.51
	trimmed	5.28	6.33	7.17	7.61	8.00	8.67	9.11	7.56
	variance	0.30	0.67	1.15	1.48	1.90	3.13	5.15	1.59
$n = 12$	mean	6.09	6.55	7.18	7.52	7.83	8.39	8.80	7.51
	trimmed	6.10	6.58	7.23	7.58	7.93	8.48	8.95	7.57
	variance	0.65	0.91	1.32	1.60	2.00	3.00	4.65	1.69
$n = 16$	mean	6.13	6.69	7.23	7.54	7.80	8.26	8.73	7.51
	trimmed	6.13	6.74	7.28	7.60	7.89	8.40	8.78	7.58
	variance	0.74	1.02	1.39	1.69	2.00	2.82	3.78	1.73
$n = 32$	mean	6.59	6.97	7.32	7.52	7.70	8.05	8.44	7.51
	trimmed	6.64	7.01	7.36	7.59	7.81	8.14	8.53	7.58
	variance	1.14	1.27	1.58	1.79	2.00	2.50	3.05	1.81
$n = 64$	mean	6.83	7.10	7.37	7.51	7.64	7.87	8.11	7.50
	trimmed	6.88	7.14	7.42	7.57	7.73	7.97	8.19	7.57
	variance	1.29	1.43	1.69	1.84	1.99	2.34	2.63	1.85
$n = 128$	mean	7.00	7.20	7.41	7.50	7.60	7.77	7.97	7.50
	trimmed	7.05	7.24	7.46	7.58	7.69	7.87	8.08	7.57
	variance	1.45	1.57	1.75	1.85	1.97	2.20	2.42	1.86

Note—Trimmed mean is a 20% trim. All estimates based on 1,000 replications.

works fairly well is: $Var + x Var/n$, where Var is the observed variance, n is the number of people in the control condition, and x is a constant between 1 and 2. Providing n is above 10, values of about 1.5 are good, but if you wish to be conservative a value of 2 can be used. When there are more than 20 groups the value of this constant makes little difference. Other methods for adjustment can be used, the important point being that the absolute value of the adjustment should increase with the estimated variance and decrease with the sample size.

In summary, the final set of simulations show that the variance estimates from `ngallpairs` are biased. They are lower than they should be which would mean that the Type 1 error rate would be higher than α if an adjustment

were not made. The bias is related mostly to the sample size, and slightly to the probability. An adjustment based just on the sample size is proposed. This is based on simulation studies.

Usage and Extensions

The main recommendation from this paper is that arbitrarily assigning participants to nominal groups, should not be used. Figures 1 and 2 show that the variability due to the arbitrary nature of how the nominal groups are chosen is large, often around two items on a ten item list. This is on the order of sampling variation of data in the $n = 8$ groups shown in these Figures. The difference is large enough to either create or negate many of the differences

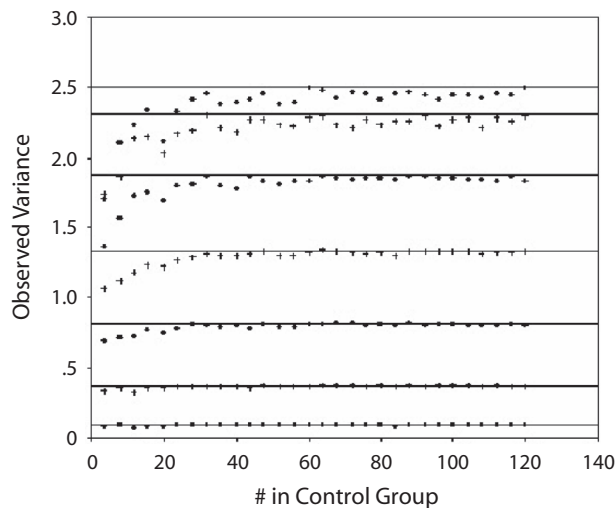


Figure 3. The observed variance for all pairs for different probabilities and sample sizes. The horizontal lines are for the predicted, or true, variances for each probability level. The observed values approach the true values as n increases.

between control conditions and experimental conditions reported in the literature. It may be worth re-examining some of these studies to make sure their conclusions were not due to the method of determining nominal groups.

The second recommendation is that sampling all permissible pairs produces similar results to sampling all permissible sets of pairs except for underestimating the variance. However, this estimate can be adjusted. Given that to achieve satisfactory power for even moderate effect sizes sampling all permissible sets of pairs is computationally impractical, the current recommendation is to use the procedure where all permissible pairs are sampled (`ngallpairs`) with the adjusted variance estimate unless the sample sizes are small. In these cases both procedures can be used. If computational power increases in the future then sampling all permissible sets of pairs has conceptual advantages and has advantages for estimating more complex statistics. Further, as discussed below alternatives may exist for sampling from all permissible pairs.

Suppose a researcher conducts a study with 40 participants, 20 in a control group and 20 in an experimental group. People in the experimenter group are allocated to 10 pairs. Supposed the mean and variance of this group are 5 and 2, respectively. The standard error of mean for the actual group is: $\sqrt{(2/10)}$. With 20 people in the control group `ngallpairs` should be used. Let the mean and variance values provided be 7 and 3. First the variance is adjusted so that the new estimate is $3 + 1.5(3)/10 = 3.45$. The standard error of the mean is: $\sqrt{(3.45/10)}$. Depending on how the researcher wishes to combine the standard errors, the researcher can then conduct a *t* test on these values.

The concentration here has been on the way the typical collaborative memory study is conducted, with an even number of people being allocated into a control condition and then allocated into nominal pairs with the task to recall items. Variants of these designs are briefly considered.

For an odd number of people, the S-Plus function `ngallpairs` works fine for creating all permissible pairs. For all permissible sets one person will be missing from each set. The sets for $n-1$ can be created and then the additional person's scores switched for each of the other people to create a new set. This only requires minor amendments to `ngallpairs` but given that these studies are usually conducted in controlled environments, it is best to keep the sample size as a multiple of the individual group size.

Groups with more than two people may be of particular interest for a researcher's theory. To create all permissible groups further matrices simply need to be multiplied together in `ngallpairs` function. With a two person group the product of the scores of the two people are taken for each trial. With more than two people the product is still taken. If the individual scores are not as simple as the binary recall/not recall that occurs in most memory studies, then some decisions have to be made on how to combine the scores for both nominal and actual groups. Suppose the score is on an attainment scale from 0-10. It may be that researchers want to take the highest of the members of the nominal group, or in other cases it may be more prudent to combine scores in some other way.

For example, if the outcome is number of sales, then the researcher might sum the scores for all nominal group members. The choice depends on the particular application and how scores are calculated for actual groups. The appropriate function can be included where the dot product is calculated within the current function.

There are other methods that could estimate nominal group statistics that have been considered. These involve taking random samples of one of three different populations. The first is of all permissible sets of pairs. However, this population would need to be created. Simply stopping the function `permiss` after a certain number of sets were created could produce an unrepresentative sample. Further, no complete set is produced within `permiss` until after all sets for a sample of size $n-2$ have been created. It would be simpler to sample from either all the possible pairs or all sequences of the numbers 1 thru n . Sampling from all possible pairs would be problematic because almost all of these sets are not permissible. Sampling from all $n!$ sequences is more promising; although almost all of these are duplications this may not create bias in the final statistics. Given the problems with the method used in most research to calculate nominal groups, it is important to examine several alternatives.

CONCLUSION

In summary, researchers should not arbitrarily assign people to nominal groups and should carefully consider how to create their sets of nominal groups. Existing research that has used arbitrarily created nominal groups should be viewed cautiously. While the estimates are unbiased, they include unnecessary error. With small sample sizes researchers can sample all permissible sets using the S-Plus function `ngallsets` in Appendix B. With larger sample sizes `ngallpairs` should be used with the adjusted variance estimate.

AUTHOR NOTE

Correspondence regarding this article may be sent to Daniel B. Wright, Psychology Department, University of Sussex, BN1 9QH, UK (e-mail: danw@sussex.ac.uk).

REFERENCES

- ANDERSSON, J. & RÖNNBERG, J. (1995). Recall suffers from collaboration: Joint recall suffers from collaboration: Joint recall effects of friendship and task complexity. *Applied Cognitive Psychology*, **9**, 199-211.
- ANDERSSON, J. & RÖNNBERG, J. (1996). Collaboration and memory: Effects of dyadic retrieval on different memory tasks. *Applied Cognitive Psychology*, **10**, 171-181.
- BASDEN, B. H., BASDEN, D. R., & HENRY, S. (2000). Costs and benefits of collaborative remembering. *Applied Cognitive Psychology*, **14**, 497-507.
- COHEN, J. (1988). A power primer. *Psychological Bulletin*, **112**, 155-159.
- DIEHL, M., & STROEBE, W. (1987). Productivity loss in brainstorming sets: Toward solution for a riddle. *Journal of Personality & Social Psychology*, **53**, 487-509.
- FINLAY, F., HITCH, G. J., & MEUDELL, P. R. (2000). Mutual inhibition in collaborative recall: evidence for a retrieval-based account. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 1556-1567.

- MEUDELL, P. R., HITCH, G. J., & BOYLE, M. M. (1995). Collaboration in recall: Do pairs of people cross cue each other to produce new memories? *Quarterly Journal of Experimental Psychology*, **48A**, 141-152.
- THOMPSON, R. (2002). Are two heads better than one? *Psychologist*, **15**, 616-619.
- WELDON, M. S. & BELLINGER, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 1160-1175.
- WELDON, M. S., BLAIR, C., & HUEBSCH, D. (2000). Group remembering: Does social loafing underlie collaborative inhibition? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 1568-1577.
- WRIGHT, D. B. (1998) Modelling clustered data in autobiographical memory research: The multilevel approach. *Applied Cognitive Psychology*, **12**, 339-357.
- WRIGHT, D. B. (2006). The art of statistics: A survey of modern techniques. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of research in educational psychology* (2nd ed., pp. 879-901). Mahwah, NJ: Erlbaum.
- WRIGHT, D. B. & KLUMPP, A. (2004). Collaborative inhibition is due to the product, not the process, of recalling in groups. *Psychonomic Bulletin & Review*, **11**, 1080-1083.
- WRIGHT, D. B., MATHEWS, S., A. & SKAGERBERG, E. M. (2005). Social recognition memory: The effect of other people's responses for previously seen and unseen items. *Journal of Experimental Psychology: Applied*, **17**, 200-209.
- YU, K., LU, Z., & STANDER, J. (2003). Quantile regression: Applications and current research areas. *The Statistician*, **52**, 331-350.

APPENDIX A

Proof that the total number of unique sets of pairs from a set of size n , where n is even and positive, is: $(n - 1)(n - 3) \dots (n - [n - 1])$

When $n = 2$, the set $\{(AB)\}$ is the only set (and $2 - 1 = 1$). When $n = 4$, the sets $\{(AB),(CD)\}$, $\{(AC),(BD)\}$, and $\{(AD),(BC)\}$ are the only sets (and $(4 - 1)(4 - 3) = 3$). Assume that $(n - 1)(n - 3) \dots (n - [n - 1])$ produces the correct number for size n . To show that this holds generally it will be shown, given this assumption, that the number of sets for $n + 2$ is: $((n + 2 - 1)((n + 2) - 3) \dots ((n + 2) - [(n + 2) - 1])$. This can be simplified to $(n + 1)(n - 1) \dots (n - [n - 1])$.

Let the collection of sets for n be denoted X_{ij} where i denotes the individual pair for each of the j sets. i goes from 1 to $n/2$ and j goes from 1 to $(n - 1)(n - 3) \dots (n - [n - 1])$. Suppose the 2 new elements are A and B and consider any of the permissible sets X_j . One new set is $\{(AB), X_j\}$. Further, element A can be switched with any of the original n elements to create a new permissible. Given that the remaining $n/2 - 1$ pairs cannot all be part of any other X_j , (otherwise they would have been the exact same set of $n/2$ pairs), this is a unique set. Therefore, the number of sets that can be produced from each X_j is $(n + 1)$. Thus, the total number of sets for $n + 2$ is: $(n + 1)(n - 1)(n - 3) \dots (n - [n - 1])$.

This approach can also be used to show that the method in `permiss` creates all permissible pairs. The method above produces unique sets that are all permissible, and the appropriate number of sets, and therefore it produces all the permissible sets.

APPENDIX B

S-Plus Functions: `ngallpairs`, `permiss`, `ngscores`, `setpairs`, and `ngallpairs`

```
"ngallsets" <- function(nomg,...)
# Calculates all permissible sets of nominal pairs and runs
# some basic descriptive statistics.
# Requires a rectangular matrix with an even number of rows.
# The data file should have 1 row for each participant.
# 0 should be for recalling the item, 1 if not.
# Calculates all possible sets of pairs.
# All permissible sets is a big number so probably only worth using
# up to about n=14, and even then it will be slow.
# Calls "permiss", "ngscores", and "setpairs".
# For loops are necessary in many places in these because the values
# used in the calculations in the ith step are dependent on the ith-1
step.
{n <- dim(nomg)[1]}
if (n > 10)
  print("This may take awhile and may not run on all computers")
lsets <- permiss(n)
ngscores(lsets, nomg, n)
"permiss" <- function(n)
# Creates all permissible sets of pairs for nominal groups.
# This function works by creating an initial pair for n=2 (i.e.,
# (1,2)) and then joining that set with the next 2 numbers (3,4)
# and then switching the 4 with 1 to create a new set and then
# with 2 to create another new set. The functions combines these
# sets, and then repeats this for the next pair, and so on.
```

```

# Calls "setpairs".
{ init <- c(1,2)
  dim(init) <- c(1,2)
  currmat <- init
  permmat <- init
  cn <- 2
  x <- seq(1,(n-2)/2)
  sp <- 1
  for (i in x)
  {
    # let cn be for the previous loop
    # this combines the 2 new numbers
    currmat <- data.frame(rep((cn+2),sp),rep(cn+1,sp),currmat)
    initmat <- currmat
    for (j in 1:cn)
    {
      tempmat <- initmat # This is where numbers are switched
      tempmat[,2] <- initmat[,j+2]
      tempmat[,j + 2] <- initmat[,2]
      currmat <- rbind(currmat,tempmat)
    }
    cn <- cn + 2
    sp <- setpairs(cn)
  }
  return(currmat)}
"setpairs" <- function(n,...)
# Calculates the number of permissible sets for a given n.
{
  setpairs <- 1
  for (i in 1:(n/2-1)) setpairs <- setpairs*(n+1 - 2*i)
  if (n < 4) setpairs <- 1
  return(setpairs)
}
"ngscores" <- function(lsets,nomg,n,...)
# This calculates the nominal group recall for each pair and
# also provides the code for descriptive statistics. These can
# be changed to suit the users' needs.
{
  nomgmat <- {}
  items <- dim(nomg)[2]
  allgroups <- {}
  sp <- setpairs(n)
  for (i in 1:sp)
  {recallv <- {}
    for (j in seq(1,n-1,2))
    {recall <- items - sum(nomg[lsets[i,j],]*nomg[lsets[i,j+1],])
      recallv <- c(recallv,recall)
      allgroups <- c(allgroups,recall)}
    recallv <- c(lsets[i,], recallv, mean(recallv), var(recallv))
    nomgmat <- c(nomgmat, recallv)
  }
  dim(nomgmat) <- c(3*n/2 + 2, sp)
  nomgmat <- t(nomgmat)
  print(nomgmat) # This can be useful to see, but can be deleted.
  dim(allgroups) <- c(n/2, sp)
  allgroups <- [t(allgroups)
  mv <- vector("numeric," sp)
  varv <- vector("numeric," sp)
  for (i in 1:sp)
  {
    mv[i] <- mean(allgroups[i,]) # Choose your favorite stats
    varv[i] <- var(allgroups[i,]) # from huge list of S-Plus ones.
  }
  mstats <- c(mean(mv),stdev(mv),range(mv),quantile(mv,c(.025,.975)))
  # For R change stdev to sd

```

```
varstats <- c(mean(varv),stdev(varv),range(varv),quantile(varv,c(.025,.975)))
x <- c(mstats,varstats)
dim(x) <- c(6,2) # Change these dimensions depending on the stats used
return(x)
}
"ngallpairs" <- function(nomg,...)
{
# Sample all nominal pairs, rather than possible sets.
# Requires the matrix entered to be rectangular with even vertical
# length. It should have 1 row for each participant.
# 0 should be for recall the item, 1 if not.
items <- dim(nomg)[2]
pairmat <- items - nomg%*%t(nomg)
vect <- pairmat[lower.tri(pairmat)]
# Add whichever stats you want
x <- mean(vect)
variance <- var(vect)
vecstat <- c(x,variance)
# If you want to use the adjusted variance estimate use:
# variance <- variance + 1.5*variance/n
return(vecstat)}
```

(Manuscript received April 4, 2005
revision accepted for publication May 1, 2006.)