

I am not completely certain of the origin of the data file (supplied to me as “Dixson et al. 2014 Raw Data .xlsx”, the space before the extension being [*sic*]). I imagine that it may have been downloaded at some point from an online repository.

The file creation date (File/Properties/Statistics) is 22 September 2016 19:45:10 and the author (File/Properties/Summary) is listed as “Danielle Dixson”. It might be interesting that the file creation date is more than two years after the Dixson et al. article was accepted by *Science*; this could suggest that the analyses in the article were performed using data that were stored in another format, which would raise the question of why that other data file(s) could not have been shared instead of them being copied into a Excel sheet in what must have been a rather labour-intensive and intrinsically error-prone procedure.

### How was the Excel file created?

There would seem to be three main possibilities for how this file was created:

- 1 Contemporaneously with collection of the data, or perhaps in the hours or days following that collection (writing up of handwritten field notes).
- 2 Later export from a statistical package to Excel (it seems likely that some other package was used, judging by the quality of the figures and the use of the Kolmogorov-Smirnov test and “multi-factorial repeated measures ANOVA” (supplement, p. 3); however, neither the article nor the supplemental information makes any mention of how the statistical analyses were conducted).
- 3 Some other manual process.

There is some evidence against the first two of these points.

First, and of course somewhat anecdotally, the file is in fairly good shape, with only a small number of inconsistencies in formatting across, say, each of the 15 sheets that represent the behaviour of a particular species of fish. For example, almost all of the groups of observations start at the same row and column on every “fish sheet”. A couple of exceptions are to be found in (a) the omission of the narrow column with the (somewhat redundant) sample size of 20 at column AA of the sheet “*C. viridis*” and column AI of the sheets “*D. aruanus*”, “*C. biocellata*”, and “*D. trimaculatus*”; (b) the groups starting at cells BO131 and BW131 of the sheet “*H. trimaculatus*”, which should probably be one row lower; and (c) cell M3 of the sheet “*C. raffelsi*”, which is blank when it should presumably contain 0. Overall, though, the file looks like it was put together in quite a short period of time; it does not look like it has been laboriously edited over a period of many weeks. On the other hand, if it had been generated by a software package’s export function, one would expect the

sheets to be totally internally consistent; the small discrepancies just mentioned ought not to occur at all.

Second, the summary statistics are almost all constant numbers, not formulas. (The only exception I found to this is in the sheet “Fish Transect”, where the summary statistics in columns AJ, AK, and AL at lines 113 through 299 and 332 through 408 are formulas.) This would be unusual if the Excel files had been assembled by inputting recently-acquired data; normally a mean would be calculated using the Excel AVERAGE() function over a range of cells, the standard deviation with STDEV(), and the standard error of the mean of a sample of 20 by dividing the standard deviation by SQRT(20). The fact that these numbers are constants is consistent with the idea that the raw data in the sheets were generated using some kind of function based on random numbers, and then all of the formulas in the sheet were converted to numbers with Select All / Copy / Paste Special... / Values.

### Duplicate groups

A total of 33 pairs of groups (same fish collection location, same comparisons) are identical in all 80 numbers (20 fish  $\times$  4 observations) across two (or, in one case, three) species, as shown in Table 1. Furthermore, in 94 cases, all 20 numbers in one column of fish observations are identical across between two and eight species, as shown in Table 2. That is, the same numbers occur in the same order in a column for all four columns in 33 cases, and for one column in 94 cases (e.g., for species A and species B, with fish collection L and comparison C, the observations for fish 1 of each species are the same, as are the observations for 2, 3, ..., through 20). See the examples in Figures Figure 1 and Figure 2.

It does not seem likely that this duplication can have been the result of a natural process. These duplicates account for 348 of the 3,600 total observation runs (first or second test of 20 fish from one location on one comparison), and they almost always occur in the same row and column position on multiple sheets (give or take a difference of one column, in cases where a column is missing from one sheet versus another), with the exception of sets 203 and 204 (cf. Table 2), in which the four and two duplicates, respectively, all occur in the same columns of the same sheet. Of the remaining 3,252 runs, *none* of the sequences of 20 observations are duplicated at any other point in the entire spreadsheet. It is very difficult to think of a way in which these duplicates could have arisen, apart from the per-species sheets having been made by successive copying of a previous sheet with (imperfect) replacement of the numbers.

There are also a number of instances where two sets of observations in the same place on two different sheets are almost identical, except for one or two values. I will just mention one example here. On the sheets “D. trimaculatus” and “C. vagabundus”, the ranges AD28:AD47 and AE28:AE47 are identical (this is set 207 in Table 2), but 19 of the 20 numbers in each of these ranges are also identical in the same place on the sheet “C. cyanea”, with the only differences being in cells AD39 and AE39. The R code that accompanies the present report outputs a list of all pairs of columns (20 fish observations) in the same position across sheets that have 16 or more identical values.

### Unexpected correlations

If the observations of each fish were truly independent, we would not expect there to be a correlation between the 20 values for fish A in the first (etc) position of its spreadsheet and the 20 values for fish B, C, etc. That is, there is no reason to expect that the first fish would be more likely to have a higher score than the 2nd, 3rd, ... 20th fish; we would expect the sequence of observations to be random, even in the presence of an effect. Yet, there is a curious bias in the Excel file towards positive correlations. Specifically, when I compute the 105 correlations between the first columns on each of the 15 sheets, then the 105 correlations for the fourth columns<sup>1</sup>, and repeat this for all 60 experiments on each fish, I end up with 12,600 correlations. Of those, 61.3% are positive, and the mean correlation is about .097. Even if I remove the cases where the columns are identical (and thus give a correlation of 1.0), the mean correlation of the remainder is .080 and 60.6% are positive. There thus seems to be some kind of background noise in the sheet, which would be consistent with the idea that the data values have been systematically altered from a slightly biased baseline.

---

<sup>1</sup> The columns correspond to the count of observations of each fish swimming in the “dangerous” water. We ignore the second and third columns (the “safe” water) as, in principle, the values in each cell should be 24 minus the value in the “dangerous” column, although—as discussed on page 8—this is not always the case, leading to a total number of observations for each fish that is not equal to  $20 \times 24 = 480$ .

Figure 1: Duplicated groups of observations in cells T28 through Y51 in the sheets “C. viridis” (top) and “D. aruanus” (bottom).

	A	B	C	S	T	U	V	W	X	Y
1	Location Captured	Species	n	n	MPA (Votualailai)	Non-MPA (Votualailai)	Non-MPA (Votualailai)	MPA (Votualailai)	Score	Indivd. Average
27										
28	Votua Non-MPA	Chromis viridis	20		22	2	3	21	43	89.58333333
29					21	3	1	23	44	91.66666667
30					24	0	3	21	45	93.75
31					22	2	3	21	43	89.58333333
32					20	4	5	19	39	81.25
33					24	0	3	21	45	93.75
34					20	4	2	22	42	87.5
35					19	5	3	21	40	83.33333333
36					21	3	2	22	43	89.58333333
37					21	3	1	23	44	91.66666667
38					23	1	3	21	44	91.66666667
39					24	0	0	24	48	100
40					21	3	5	19	40	83.33333333
41					23	1	3	21	44	91.66666667
42					21	3	2	22	43	89.58333333
43					21	3	2	22	43	89.58333333
44					21	3	2	22	43	89.58333333
45					24	0	0	24	48	100
46					22	2	3	21	43	89.58333333
47					22	2	4	20	42	87.5
48										
49									Mean	90.20833333
50									SD	4.687804571
51									SE	1.048224969
52										
53										
54	Votualailia MPA	Chromis viridis	20	20	22	2	2	22	44	91.66666667
55					19	5	4	20	39	81.25
56					21	3	3	21	42	87.5
57					22	2	5	19	41	85.41666667
58					23	1	4	20	43	89.58333333

	A	B	C	T	U	V	W	X	Y	
1	Location Captured	Species	n	MPA (Votualailai)	Non-MPA (Votualailai)	Non-MPA (Votualailai)	MPA (Votualailai)	Score	Indivd. Average	
27										
28	Votua Non-MPA	Dascyllus aruanus	20		22	2	3	21	43	89.58333333
29					21	3	1	23	44	91.66666667
30					24	0	3	21	45	93.75
31					22	2	3	21	43	89.58333333
32					20	4	5	19	39	81.25
33					24	0	3	21	45	93.75
34					20	4	2	22	42	87.5
35					19	5	3	21	40	83.33333333
36					21	3	2	22	43	89.58333333
37					21	3	1	23	44	91.66666667
38					23	1	3	21	44	91.66666667
39					24	0	0	24	48	100
40					21	3	5	19	40	83.33333333
41					23	1	3	21	44	91.66666667
42					21	3	2	22	43	89.58333333
43					21	3	2	22	43	89.58333333
44					21	3	2	22	43	89.58333333
45					24	0	0	24	48	100
46					22	2	3	21	43	89.58333333
47					22	2	4	20	42	87.5
48										
49									Mean	90.20833333
50									SD	4.687804571
51									SE	1.048224969
52										
53										
54	Votualailia MPA	Dascyllus aruanus	20		23	1	1	23	46	95.83333333
55					20	4	4	20	40	83.33333333
56					22	2	2	22	44	91.66666667
57					20	4	3	21	41	85.41666667
58					19	5	1	23	42	87.5

Figure 2: Duplicated groups of observations in the sheets “C. vagabundus” (top) and “C. raffelsi” (bottom). For lines 54 through 73 columns AJ and AM are identical, while columns AK and AL are different (although there are quite a few identical cells). For C. raffelsi, all of the pairs in columns AJ/AK and AL/AM sum to 24; for C. vagabundus, AJ/AK sum to values between 19 and 34, while AL/AM sum to values between 19 and 24.

	A	B	C	AJ	AK	AL	AM	AN	AO
1	Location Captured	Species	n	Non-MPA (A. nasuta)	Non-MPA	Non-MPA	Non-MPA (A. nasuta)	Score	Individ. Average
53									
54	Votualailia MPA	Chaetodon vagabundus	20	14	10	10	14	28	58.33333333
55				13	9	8	16	29	60.41666667
56				14	5	7	15	29	60.41666667
57				15	9	8	16	31	64.58333333
58				14	6	5	14	28	58.33333333
59				15	9	9	15	30	62.5
60				14	9	11	13	27	56.25
61				16	10	9	13	29	60.41666667
62				15	12	10	14	29	60.41666667
63				14	11	6	14	28	58.33333333
64				15	13	8	14	29	60.41666667
65				16	18	9	15	31	64.58333333
66				15	9	10	14	29	60.41666667
67				16	9	9	14	30	62.5
68				15	6	7	14	29	60.41666667
69				15	8	9	13	28	58.33333333
70				14	6	5	14	28	58.33333333
71				16	10	8	15	31	64.58333333
72				14	10	11	13	27	56.25
73				15	9	10	14	29	60.41666667
74									
75								Mean	60.3125
76								SD	2.481197569
77								SE	0.554812643
78									

	A	B	C	AJ	AK	AL	AM	AN	AO
1	Location Captured	Species	n	Non-MPA (Coral)	Non-MPA	Non-MPA	Non-MPA (Coral)	Score	Individ. Average
53									
54	Votualailia MPA	Chaetodon raffelsi	20	14	10	10	14	28	58.33333333
55				13	11	8	16	29	60.41666667
56				14	10	9	15	29	60.41666667
57				15	9	8	16	31	64.58333333
58				14	10	10	14	28	58.33333333
59				15	9	9	15	30	62.5
60				14	10	11	13	27	56.25
61				16	8	11	13	29	60.41666667
62				15	9	10	14	29	60.41666667
63				14	10	10	14	28	58.33333333
64				15	9	10	14	29	60.41666667
65				16	8	9	15	31	64.58333333
66				15	9	10	14	29	60.41666667
67				16	8	10	14	30	62.5
68				15	9	10	14	29	60.41666667
69				15	9	11	13	28	58.33333333
70				14	10	10	14	28	58.33333333
71				16	8	9	15	31	64.58333333
72				14	10	11	13	27	56.25
73				15	9	10	14	29	60.41666667
74									
75								Mean	60.3125
76								SD	2.481197569
77								SE	0.554812643
78									

Table 1: List of test groups (pairs of tests, with a total of four columns of data for 20 fish) that are identical in all 80 values.

Set	Sheet name 1	Sheet name 2	First row	First column
101	<i>C. viridis</i>	<i>D. aruanus</i>	28	T
102	<i>C. viridis</i>	<i>D. aruanus</i>	106	AY
103	<i>D. aruanus</i>	<i>P bankanensis</i>	80	BG/BH
104	<i>C. vagabundus</i>	<i>C. raffelsi</i>	54	BP
105	<i>C. vagabundus</i>	<i>C. raffelsi</i>	54	BX
106	<i>C. vagabundus</i>	<i>C. raffelsi</i>	80	AJ
107	<i>A. triostegus</i>	<i>C. striatus</i>	132	BH
107	<i>A. triostegus</i>	<i>C. sordidus</i>	132	BH
108	<i>C. striatus</i>	<i>Scarus sp.</i>	2	T
109	<i>C. striatus</i>	<i>Scarus sp.</i>	28	T
110	<i>C. striatus</i>	<i>Scarus sp.</i>	54	T
111	<i>C. striatus</i>	<i>Scarus sp.</i>	80	T
112	<i>C. striatus</i>	<i>Scarus sp.</i>	106	T
113	<i>C. striatus</i>	<i>Scarus sp.</i>	132	T
114	<i>C. striatus</i>	<i>Scarus sp.</i>	2	AB
115	<i>C. striatus</i>	<i>Scarus sp.</i>	28	AB
116	<i>C. striatus</i>	<i>Scarus sp.</i>	54	AB
117	<i>C. striatus</i>	<i>Scarus sp.</i>	80	AB
118	<i>C. striatus</i>	<i>Scarus sp.</i>	106	AB
119	<i>C. striatus</i>	<i>Scarus sp.</i>	132	AB
120	<i>C. striatus</i>	<i>Scarus sp.</i>	2	AJ
121	<i>C. striatus</i>	<i>Scarus sp.</i>	28	AJ
122	<i>C. striatus</i>	<i>Scarus sp.</i>	54	AJ
123	<i>C. striatus</i>	<i>Scarus sp.</i>	80	AJ
124	<i>C. striatus</i>	<i>Scarus sp.</i>	106	AJ
125	<i>C. striatus</i>	<i>Scarus sp.</i>	132	AJ
126	<i>C. striatus</i>	<i>Scarus sp.</i>	2	AZ
127	<i>C. striatus</i>	<i>Scarus sp.</i>	28	AZ
128	<i>C. striatus</i>	<i>Scarus sp.</i>	54	AZ
129	<i>C. striatus</i>	<i>Scarus sp.</i>	80	AZ
130	<i>C. striatus</i>	<i>Scarus sp.</i>	106	AZ
131	<i>C. striatus</i>	<i>Scarus sp.</i>	132	AZ
132	<i>C. striatus</i>	<i>Scarus sp.</i>	106	BP

Set numbers are arbitrarily assigned labels to denote identical sets of results. Set number 107 occurs twice because three sheets contain copies of the same group.

Table 2: List of identical tests (one column of data for 20 fish), other than those in Table 1.

Set	Sheet name	First row	Column(s)
201	C. viridis	2	AK, AL
201	D. aruanus	2	AK, AL
202	C. viridis	54	AL
202	D. aruanus	54	AL
202	C. vagabundus	54	AM
202	C. raffelsi	54	AM
202	Scarus sp.	54	AM
202	A. triostegus	54	AM
202	C. striatus	54	AM
202	A. angustatus	54	AM
202	H. trimaculatus	54	AM
203	C. viridis	54	T, U
203	C. viridis	80	T, U
203	C. viridis	106	T, U
203	C. viridis	132	T, U
204	C. viridis	54	V, W
204	C. viridis	106	V, W
205	P bankanensis	28	L, O
205	C. biocellata	28	L, O
206	P bankanensis	28	T, W
206	C. cyanea	28	T, W
206	D. trimaculatus	28	T, W
206	C. vagabundus	28	T, W
206	A. triostegus	28	T, W
206	C. striatus	28	T, W
206	Scarus sp.	28	T, W
206	C. sordidus	28	T, W
206	A. angustatus	28	T, W
206	H. trimaculatus	28	T, W
207	D. trimaculatus	28	AD, AE
207	C. vagabundus	28	AD, AE
208	D. trimaculatus	28	L, O
208	C. vagabundus	28	L, O
209	C. vagabundus	2	AJ, AM
209	C. raffelsi	2	AJ, AM
209	A. triostegus	2	AJ, AM
210	C. vagabundus	2	M, N
210	S. spinus	2	M, N
211	C. vagabundus	28	AJ, AM
211	C. raffelsi	28	AJ, AM
211	A. triostegus	28	AJ, AM
212	C. vagabundus	54	AJ
212	C. raffelsi	54	AJ
212	A. triostegus	54	AJ
213	C. vagabundus	132	AM
213	A. triostegus	132	AM
214	A. triostegus	54	BH, BI
214	C. striatus	54	BH, BI
* 106	A. triostegus	80	AJ, AM
215	C. striatus	28	L, O
215	S. spinus	28	L, O
215	Scarus sp.	28	L, O
215	C. sordidus	28	L, O
215	H. trimaculatus	28	L, O

Set numbers are arbitrarily assigned labels to denote identical sets of results.

\* This pair of columns has the same values as in the equivalently numbered complete set in Table 1.

### Errors in the numbers of observations

There are a number of inconsistencies within groups in the “fish sheets” regarding the total number of observations. In principle there should be a total of 480 observations in each of two pairs of columns, corresponding to 24 observations of 20 fish located in one of two channels of water, with each observation taking place five seconds apart. However, this is not the case for 54 of the 900 groups, where one or both of the sets of observations in a test do not sum to 480. These are shown in Table 3 (where a blank cell for one or the other test indicates that the number of observations for that test does indeed sum to the expected value of 480).

The authors do not mention that any number of observations other than 24 was ever performed on each fish, nor do they discuss what the consequences of such variation would be for their results. For the 54 cases with one or both sets of observations that do not sum to 480, the mean number of observations in the two sets is 959.56, with a range of 863 through 1,019 and a standard deviation of 36.97. Yet for the 846 other pairs of tests, exactly 480 observations were carried out in each of the two tests of each fish. It might not be surprising that errors in the numbers of tests were made, but it seems strange that when errors occur they seem to be so substantial.

The existence of numbers of observations over 20 fish that do not sum to 480 implies, of course, that in some cases the number of observations in both flumes for a single fish are different from 24. For example, in the sheet “*C. vagabundus*”, cells AJ65 and AK65 sum to 34, while in the sheet “*H. trimaculatus*”, cells AL58 and AM58 sum to 14, as do cells AL63 and AM63. There seems to be no sensible explanation for this, because even with manual data entry into the Excel file it would seem illogical not to use a formula to calculate the number of observations in the “safe” water by subtracting the number in the “dangerous” water from 24, thus halving the amount of typing to be done. Alternatively, if the idea was to use one number as a control for the other, a formula ought to check that they sum to 24.

An alternative explanation is that the values other than 480 are the result of manual intervention in the Excel file. It seems noteworthy that half (27) of the non-480 totals correspond to cases where there is a duplicate column of data in the set (as shown in Table 2).



Table 3: List of tests where the observations (of 20 fish in area A or B) do not sum to 480.

Sheet name	First row	First column	Total obs.	Test 1	Test 2
C. viridis *	54	AI	915		151+284=435
D. aruanus *	54	AI	899		135+284=419
P bankanensis *	28	T	961	430+51=481	
C. biocellata	2	L	983	433+70=503	
C. biocellata *	28	L	1001	436+74=510	68+423=491
C. cyanea *	28	T	992	430+69=499	61+432=493
C. cyanea	28	AB	957		70+407=477
D. trimaculatus	2	T	969	430+59=489	
D. trimaculatus *	28	T	1019	430+85=515	72+432=504
C. vagabundus	2	T	978	429+69=498	
C. vagabundus *	2	AJ	928	317+139=456	156+316=472
C. vagabundus *	28	L	995	432+67=499	67+429=496
C. vagabundus *	28	T	993	430+67=497	64+432=496
C. vagabundus *	28	AJ	901	338+112=450	107+344=451
C. vagabundus *	54	AJ	936	295+188=483	169+284=453
C. vagabundus	132	AJ	911	332+137=469	132+310=442
A. triostegus	2	T	969	425+64=489	
A. triostegus *	2	AJ	951	317+160=477	158+316=474
A. triostegus *	28	T	1005	430+72=502	71+432=503
A. triostegus	28	AB	971	429+63=492	49+430=479
A. triostegus *	2	AJ	951	317+160=477	158+316=474
A. triostegus *	54	AJ	883	295+151=446	153+284=437
A. triostegus *	80	AJ	965	353+136=489	131+345=476
A. triostegus	132	AJ	863	329+124=453	100+310=410
C. striatus	2	L	948	426+44=470	46+432=478
C. striatus	2	T	963	429+54=483	
C. striatus *	28	L	986	433+72=505	51+430=481
C. striatus	28	T	1013	430+80=510	71+432=503
C. striatus	28	AB	974	431+64=495	87+392=479
C. striatus	54	AJ	919		155+284=439
S. spinus *	2	L	966	425+53=478	60+428=488
S. spinus	2	T	972	403+89=492	
S. spinus	2	AB	970	407+70=477	65+428=493
S. spinus *	28	L	997	433+67=500	67+430=497
S. spinus	28	AB	970	426+64=490	
Scarus sp.	2	L	969	413+68=481	53+435=488
Scarus sp.	2	T	963	429+54=483	
Scarus sp. *	28	L	993	433+65=498	65+430=495
Scarus sp.	28	T	1013	430+80=510	71+432=503
Scarus sp.	28	AB	974	431+64=495	87+392=479
Scarus sp.	54	AJ	919		155+284=439
C. sordidus	2	L	969	428+49=477	64+428=492
C. sordidus	2	T	972	408+84=492	
C. sordidus *	28	L	957	433+54=487	40+430=470
C. sordidus *	28	T	1014	430+79=509	73+432=505
A. angustatus	2	L	945	427+41=468	44+433=477
A. angustatus	2	T	963	437+46=483	
A. angustatus *	28	T	983	430+71=501	50+432=482
A. angustatus *	54	AJ	901		137+284=421
H. trimaculatus	2	L	952	429+45=474	46+432=478
H. trimaculatus	2	T	968	431+57=488	
H. trimaculatus *	28	L	960	433+49=482	48+430=478
H. trimaculatus *	28	T	995	430+69=499	64+432=496
H. trimaculatus *	54	AJ	912		148+284=432

\* These tests correspond to duplicates in Table 2.

## Extra fish

There are some extraneous data points in the “fish sheets”, as shown in Table 4. Specifically, for the four combinations of two locations (Votua non-MPA, row 48; Votualailia non-MPA, row 100) and two comparisons (Acropora/Porites; Acropora/P. demacornis), eight different “fish sheets” contain a 21st record in between one and four groups of tests. In addition, one sheet contains a 21st record for the MPA versus non-MPA tests at Votualailai. At no point in the article or the supplement do the authors mention that they used 21 fish in any of their tests. While only having results for, say, 19 fish might be explained by error (or perhaps death of the fish<sup>2</sup>), it seems unlikely that one might accidentally run too many fish, record their results, and then fail to mention it.

Table 4: List of tests where there are observations of 21 fish.

Sheet name	Row	First column	Values
C. biocellata	100	BG	22 2 4 20
C. cyanea	100	BH	22 2 3 21
D. trimaculatus	100	BG	22 2 4 20
C. vagabundus	48	AZ	21 3 5 19
C. vagabundus	100	AZ	20 4 2 22
C. vagabundus	48	BH	22 2 1 23
C. vagabundus	100	BH	22 2 4 20
C. raffelsi	100	T	23 1 2 22
C. raffelsi	48	AZ	20 4 5 19
C. raffelsi	100	AZ	20 4 5 19
C. raffelsi	48	BH	22 2 1 23
C. raffelsi	100	BH	22 2 4 20
Scarus sp.	100	BH	22 2 4 20
C. sordidus	100	BH	22 2 4 20
H. trimaculatus	100	BH	22 2 4 20

Notably, of the eight records that appear at line 100 for the comparisons between Acropora and P. demacornis (column BG or BH, depending on whether the sample size is missing from column AI, as mentioned earlier), seven have the same four observations in the same order (22, 2, 4, 20).

Investigation of the summary statistics for the 14 groups containing 21 records shows that in each case the mean and standard deviation is based on all 21 data points, but the standard error is equal to the standard deviation divided by the square root of 20 (and not 21). This is consistent with the idea that the 21st records were included inadvertently (perhaps while dragging a set of formulas one line too far), with the mean and SD then being calculated by a formula that included the whole range, and the SE being derived with the same formula (dividing the SD by SQRT(20) rather than a calculation of the size of the range, for example with the ROWS() function) that was used for all of

---

<sup>2</sup> Indeed, it seems remarkable that there appear to have been no loss of fish or data records at any point in the extraordinarily long sequence of tests that were performed. Most researchers who deal with human or animal subjects are familiar with the possibility of subjects dropping out, or of individual measurements being lost or otherwise unusable, for a variety of reasons.

the groups. The remaining formulas in the “Fish Transect” sheet, mentioned earlier, show this pattern; the SE is calculated by dividing the SD by  $\text{SQRT}(30)$ , there being data for 30 transects.

### Some other thoughts

The preceding comments mostly been based on a neutral observation of the data in the context of the article, and are intended to be objectively verifiable by anyone with a modest degree of numeracy and access to the data file. Here are a couple of other notes, which should be read while keeping in mind that I know absolutely nothing about marine ecology.

### How long did all this take?

Each “fish sheet” contains records for ten groups of tests at each of six sites (three MPA, three non-MPA), making 60 groups of tests per species. Each group involved two tests of each of 20 fish, with each test lasting four minutes (two minutes of acclimation time and two of actual testing), so the time spent observing each fish species was  $(60 \times 20 \times 2 \times 4) = 9,600$  minutes, or 160 hours. Multiplied by 15 species, this gives 2,400 hours, *just for the time spent observing the fish*. This does not include time to set up each run, collect the fish and the water, or conduct any of the other experiments. At 10 hours a day and six days a week, this would take a minimum of 40 weeks; with one minute and 12 seconds between fish it would take a year (and have been soul-crushingly repetitive). Only one person (“V. Bonito”) is named in the Acknowledgements, and it is not even clear if they were a research assistant (the current director of Reef Explorer Fiji is named Victor Bonito, cf. <https://www.iucn.org/news/commission-environmental-economic-and-social-policy/202101/coral-restoration-training-fijis-coral-coast>). The number of observations conducted must have been  $(60 \text{ groups} \times 20 \text{ fish} \times 2 \text{ tests} \times 2 \text{ minutes} \times 12 \text{ observations/minute} \times 15 \text{ species}) = 864,000$ , give or take any errors that may have occurred; this is confirmed by an analysis of the data file, where the total number of observations sums to 863,976 (see Table 3 for why it is not equal to 864,000). Yet the effects are so huge that it must have been obvious to the experimenters that they had found a reproducible phenomenon after only a small percentage of the trials had been run. If this had been a clinical trial, it would have been stopped early because the results were so spectacular that it would have been unethical to the patients to continue collecting data. In this case it seems to have been unfair to the experimenters to force them to collect data, day after day. What must it have been like to drag oneself into the lab for day 12 of testing on the 14th fish, when every day until then had been marked by the same astonishingly large effects?

### Lack of variability

Something that stands out from the “fish sheets” is a remarkable lack of variability in the performance of the fish. The position of each fish (in one of two streams) was recorded 24 times in a two-minute period. Every species of fish showed a very similar degree of preference for the “better” (MPA vs non-MPA, or coral-infused vs seaweed-infused) water, but the degree of variation

within a species was also remarkably small. I calculated the minimum and maximum number of visits by the 20 fish of each species to each water option in each of the 3,600 tests (15 species  $\times$  6 catch locations  $\times$  10 water comparisons  $\times$  2 tests  $\times$  2 possibilities<sup>3</sup>) and calculated the size of the range of numbers in each case. As Table 5 shows, this was generally quite small, with 3,496 (97%) of the ranges falling between 3 and 8.

Table 5: Range (maximum minus minimum observation, plus 1) for each of the 3,600 sets of observations of the presence of 20 fish.

Range	Count
1	0
2	0
3	22
4	237
5	908
6	1046
7	794
8	393
9	118
10	35
11	10
12	11
13	15
14	7
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	2
23	2
24	0
25	0

---

<sup>3</sup> The “2 possibilities” part should be redundant as one number should be 24 minus the other, but because of the problem illustrated in Table 3 it seems reasonable to include both here.