NTNU

Institutt for psykologi

Mal for sensorveiledning

Emnekode	Psy3121
Emnenavn	Forskning og prosjektutvikling
Emneansvarlig/oppgavegiver	Robert Biegler
Kvalitetssikret av	Gerit Pfuhl
Semester, år	Vår 2025
Vurderingsform, lengde	**Oppgave**
Tillatte hjelpemidler	A – alle hjelpemidler tillatt

Emnets læringsutbyttebeskivelser angitt I kunnskaper, ferdigheter og generell kompetanse. (Henvisning med lenke til emnesiden på NTNUs nettsider er tilstrekkelig)	Emne - Forskning og prosjektutvikling - PSY3121 - NTNU
Pensum	Goodman, S. and Greenland, S. (2007). Assessing the unreliability of the medical literature: A response to "Why most published research findings are false". Johns Hopkins University, Department of Biostatistics. Available: http://www.bepress.com/jhubiostat/paper135. Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, 2(8), 696-701. Lenharo, M. 2023 Consciousness: The future of an embattled field. Meltzoff, J, and Cooper, H. (2018). Critical Thinking about Research. American Psychological Association, Washington DC. Simmons, J. P., Nelson L. D., and Simonsohn, U. (2011). False-Positive Psychology: Flexibility in Data Collection and Analysis Allows Presenting anything as Significant. *Psychological Science*, 2011 (Nov), Vol 22(11), 1359-1366
Eventuelle formelle krav til besvarelsen	Ingen
Hvordan de ulike oppgavene i eksamenssettet er vektlagt	

NTNU

Institutt for psykologi

Oppgavetekst:

Review the fake paper below as if for a scientific journal.

In **part 1**, list the problems you find, referencing line numbers when you report a problem.

If you think something has been omitted that should be there, and there is a specific line where it should be, also provide the line number.

If you think something has been omitted that should be there, and you can only say that it should be somewhere in introduction, methods, data analysis, results, discussion, or the data file, then reference these sections of the paper, instead of a line number.

If you comment on specific variables in the included data, use the variable name in the columns of the CSV file.

In **part 2**, summarise the *most important* flaws and recommend improvements. There are fatal flaws, which can either be ethical problems so severe that the research should not be done at all, or flaws in design so severe that they prevent meaningful inferences regarding the hypotheses without further experiments and that therefore preclude publication of the paper even after major revision. List all of those and explain them.

Also list the non-fatal flaws in analysis and interpretation that most affect the conclusions that can be drawn. These would be things that can be fixed without further experiments, but that absolutely must be fixed to make sound inferences possible, and that the reader can't resolve without further information.

Short report on the association between choice of musical instrument and the projected death toll of a World Domination Attempt

Furly Loosvelt, Patience Fish, and Comfort The Unfortunate Cripes

Institute for the Comparative Study of Conquest, Perfectly Normal University of Ruritania

Corresponding author: c.t.u.cripes@ac.rur

Abstract

We allowed people with SCGPD access to a music room stocked with a wide variety of musical instruments. We found that the death toll expected if their Word domination Attempts had succeeded was predicted by sound amplitude summed over frequencies, but not by the number of parts of the chosen instrument. We propose a range of security measures to deal with the threat of SCGPD.

NTNU

Institutt for psykologi

Introduction

Schizotypal Creative Genius Personality Discorder (SCGPD; Mcguire, 2013), colloquially known as Mad Scientist Syndrome, exclusively affects natural scientists and engineers who excel in their field. Known risk factors are high capacity for sustained attention, high social dominance orientation, high creativity, and intense dislike of mimes, raspberries, and the colour mauve (Toldervy & Hooton, 2011). Onset of the condition amplifies whichever of these traits already deviate more than 2.3 standard deviations from the norm (Peacock & Smedley, 2017). This amplification of personality traits was followed, in 89% of known cases, by a World Domination Attempt within 71 to 2917 days. The remaining 11% had undergone trait amplification within no more than 389 days before data collection, well below the median time to World Domination of 953 days, (Sixsmith, Delight & Killjoy, 2016). A follow-up study showed World Domination Attempts in two thirds of this remaining population (Delight & Sixsmith, 2019).[1]

Research into a further potential risk factor has been neglected so far. Ever since the index case of Captain Nemo (Verne, 1870), there has been speculation regarding an association between playing the organ and SCGPD. While only 1 in 100 000 people play the organ, 6 out of 229 known sufferers of SCGPD, corresponding to a rate of 2620 in 100 000, have played the organ while preparing for or during a World Domination Attempt. This overrepresentation of organists among people with SCGPD raises the question of what aspect of playing organ makes someone vulnerable to SCGPD? Is it something to do with the sheer volume of sound, with the frequency range of sounds that can be generated at any one time, the variety of sounds the instrument can produce, or the number of parts of the instrument? And should organs be made more widely accessible to detect people vulnerable to SCGPD, or dismantled and outlawed, to prevent conversion from excellence in a field to SCGPD?

All World Domination Attempts associated with SCGPD having failed so far has led some to argue that the condition is not a serious problem (Pangloss & Allswell, 2020). However, all but the initial 17 of the 301 known attempts were interrupted by nerd sniping, asking for an explanation of the scheme, ideally in the tone of a fascinated but bewildered postgraduate. This extremely effective technique was discovered by Rawbone (1989) during the 1987 Sarcastic Fringehead Incident, and subsequently extensively validated (Fish & Pye, 2001). It is a subject of ongoing research why the schemes and doomsday machines of people with SCGPD reliably perplex and baffle the most experienced investigators and scientists, yet any doomsday machines are easily disabled by competent janitors while the would-be tyrant is busy explaining the brilliance and inevitability of their scheme (Cuppage, 2007; Hooton & Toldervy, 2022; Sparkling, 2019).

[1] Sadly, Philomena Killjoy was among the number who tried to take over the world and her brilliant research career in psychology was tragically cut short.

However, a board of 28 healthy experts estimated that, if any of the known World Domination Attempts had succeeded, the likely death tolls would have ranged from 18±7 to 3.7±1.3 million. Interestingly, an independent assessment by 13 people with SCGPD gave estimates with much greater mutual agreement (Intraclass correlation = 0.473 for people with SCGPD versus intraclass correlation = 0.361 for the healthy experts; Wilde & Dry, 2024). In this study, we gave people with SCGPD opportunity to practice a variety of musical instruments and used features of these instruments to predict the expected death tolls obtained by Wilde and Dry (2024).

Methods

We provided 187 people with SCGPD who are confined in a maximum-security installation with unrestricted access, during the hours from 07:00 to 23:00, to a music room that was stocked with telenkas, didgeridoos, bongo drums, saxophones, Scottish pipes, violins, Celtic harps, hurdy-gurdies, electronic keyboards, and a 4000 pipe organ. We recorded the time spent playing each instrument. Over the 101 days of the study, all patients tried out instruments. We included the 148 who played for at least three hours per week through the whole study period. All participants retained broadly human body plans. Age ranged from 15 to 92, 99 were male, 47 female, 2 other. We do not report the age and sex of individuals as their high public profile and modest number would make it possible to identify at least some individuals.

We considered a patient to have chosen an instrument when the patient played that instrument for at least twice as long as the next favourite. Data collection was terminated after 101 days because four of the organists tried to take over the world using the organ. Because the facility lacks enough janitors with the requisite security clearance to provide coverage at all hours, these attempts failed only because the patients questioned each other's methodology and because their independent modifications of the organ interfered with each other and the instrument melted down. Access to the music room has been terminated due to the radiation hazard. The room has been entombed in a concrete sarcophagus.

Data analysis

We summed sound amplitudes over 41 band pass filters with both a width and spacing of a quarter octave, ranging from 20 Hz to 12800 Hz. We then summed amplitudes either over all play during the last four weeks of data collection, or only over the most passionate play, as measured by force sensors built into the keyboards. The latter turned out to be more informative and is reported here.

We record the number of parts that make up an instrument, to test whether complexity has an effect. Because the number of parts in the organ strongly skews the distribution, in the analysis, we use the logarithm of the number of parts.

We carried out a multiple regression with sum of amplitudes across frequencies and logarithm of number of parts as predictors, and the predicted death tolls obtained from Wilde & Dry (2024) as outcome. We treat participants as independent because people with SCGPD have never been known to collaborate. They may attract minions, but they do not work with colleagues (de Autosuficiente & Einzelgänger, 2023).

Results

Ordered by popularity, 41 patients played the didgeridoo, 32 the saxophone, 21 the bongo drums, 16 the electronic keyboard, 15 the violin, 9 the Scottish pipes, 8 the organ, 7 the Celtic harp, 3 the telenka, and 1 the hurdy-gurdy.

We found a significant regression ($F(2, 145) = 42.432$, $p < 0.001$, $R^2 = 0.369$). The regression equation is:

Log10_Death_average = 2.005 + 0.047 Log_N_Parts + 0.002 Freq_amplitude

Contrary to our expectations, the number of parts had no effect on the predicted number of deaths resulting from a successful World Domination Attempt ($t(147) = 0.745$, $p = 0.457$, confidence intervals -0.077 to 0.171; Figure 1).

However, sound amplitude summed over frequency does significantly relate to the expected death toll ($t(147) = 6.608$, $p < 0.001$, confidence intervals 0.002 to 0.003; Figure 2).

NTNU

Institutt for psykologi

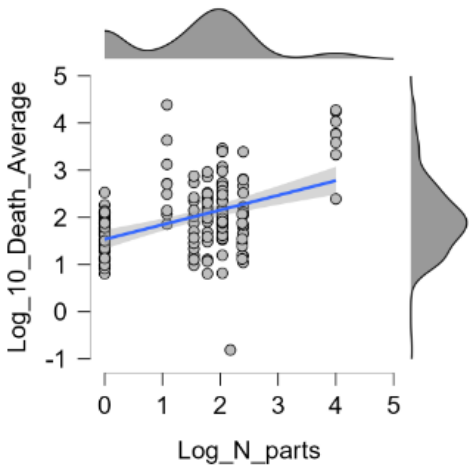


Figure 1. The logarithm of average expected deaths as a function of the logarithm of the number of parts in each musical instrument.

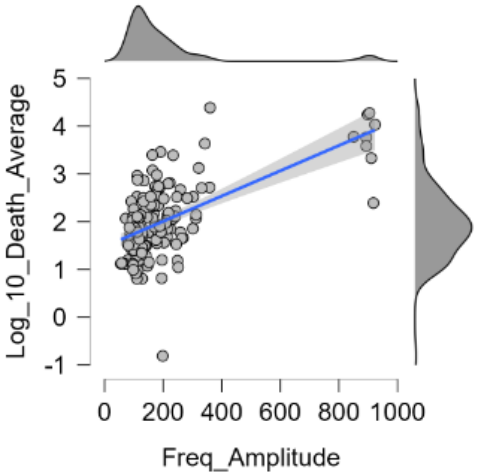


Figure 2. The logarithm of average expected deaths as a function of sound amplitude summed over frequencies.

Discussion

Our results confirm that playing the organ exacerbates the symptoms of SCGPD and increases the likelihood of World Domination Attempts. This has important implications for public safety. We recommend that aspiring organists are subject to strict scrutiny, that anyone working in natural science or engineering who wants to play the organ is screened for risk factors, and that sites where organs are installed are strictly surveilled so that unauthorised users can be apprehended and a thorough background check initiated.

Two further points must be addressed. First, would people with SCGPD not simply refrain from playing the organ? Our observations indicate that playing the organ is a compulsion, as much as an irresistible urge to explain makes people with SCGPD vulnerable to nerd sniping, despite the technique having been in the public domain for decades (Fish & Pye, 2001; Rawbone, 1989). Second, surveillance of existing organs would miss people with SCGPD building their own. Given that organ pipes may be built from a wide variety of materials, using many different manufacturing processes, it is hardly practical to monitor them all. However, only those who have already undergone transition would go to the trouble of building their own organ. Our proposals aim to make transition less likely.

Despite people with SCGPD' penchant for extremely complex plans, complexity has no effect on choice of musical instrument. What matters is how loud the instrument is. Further research should look into whether there is a broader connection, possibly including raves. The possible influence of rhythm is a neglected research area. Some anecdotal reports suggest a possible connection with 1980s German Electronic Dance Music, but it is unclear whether the connection is positive or negative.

References

de Autosuficiente, A. F. & Einzelgänger, Q. E. D. (2023). SCGPD is incompatible with collaborative effort. *Quarterly Review of Organisational Psychology*, 26(3), 412 – 431.

Cuppage, F. (2007). Why Practice Beats Theory. *Applied Epistemology Quarterly*, 28(3), 43 – 51.

Fish, P. and Pye, G. (2001). Nerd Sniping: a Systematic Review and Meta-Analysis. *Annals of the Society*, 114(1), 9 – 17.

Hooton, Q. and Toldervy T. (2022). The Curse of Knowledge in the Analysis of World Domination schemes. *Annual Review of World Domination*, 14(1), 204 – 217.

Mcguire, S. (2013). Laughter at the Academy. In J. J. Adams (ed), *The Mad Scientist's Guide to World Domination*, pp. 38 - 50. Tom Doherty Associates, New York, NY.

Pangloss, D. and Allswell, R. (2020). Calm Down, It's Not so Bad. *Journal of Theoretical Sociology*, 44(2), 89 – 97.

Peacock, P. P. and Smedley, S. O. (2017). Amplification of Deviations Marks the Onset of SCGPD. *Proceedings of the Society for the Delightfully Unhinged*, 13(13), 13 – 26.

Rawbone, U. (1989). Nerd Sniping: an Effective Method for Disrupting Attempts to Take Over the World. *Annals of the Society*, 102(7), 1003 – 1032.

Sixsmith, A. Q. R. J., Delight, B. E. and Killjoy, P. (2016). From Transition to SCGPD to World Domination Attempt in 71 Days. *Journal of Inadvisable Studies*, 10(8), 1 - 8.

Sparkling, W. (2019). On the Surprising Effectiveness of Practical Knowledge. *Journal of Experimental Sociology*, 53(2), 324 – 340.

Toldervy, T. and Hooton, Q. (2011). Risk Factors for Mad Science. *Journal of Inadvisable Studies*, 5(1), 13-22.

Verne, J. (1870). Vingt Mille Lieues sous les mers. Magasin d'éducation et de récréation.

Wilde, C. and Dry, V. (2024). Disturbing or Disturbed Judgement? Predicting the Cost of World Domination.

Data file:

Log_N_parts	Freq_Amplitude	Attention_z-score	Social_Dominance_z-score	Creativity_z-score	Mimes_z-score	Raspberries_z-score	Mauve_z-score	Death_1	Death_2	Dea_3
0	55,37	2,7	1,48	3,25	-1,64	2,72	3,81	4	8	3
0	110,35	0,67	1,31	4,71	0,23	2,48	3,59	3	3	1
1,5441	125,71	1,6	2,58	3,57	-3,04	0,58	2,81	135	31	7
2,3802	69,23	3,51	1,28	3,75	1,66	1,66	2,09	34	8	2
2,3979	258,54	3,85	3,96	3,73	2,54	1,32	1,17	295	364	15
1,5441	126,51	2,99	2,70	2,66	0,38	-0,89	0,43	33	21	14
2,3979	222,63	1,15	4,06	2,41	3,34	-0,69	3,13	465	190	46

NTNU

Institutt for psykologi

1,0792	311,76	3,62	2,93	3,1	-1,23	1,29	-0,38	109	84	45	21
1,7782	105,27	3,32	1,77	1,99	2,91	-0,44	2,73	8	10	14	7
1,0792	301,40	2,58	2,91	3,28	3,68	2,31	4,94	136	15	180	41
1,7782	138,41	2,5	2,21	1,18	1,97	0,57	1,17	28	108	44	16
2,3979	302,24	2,87	3,79	3,57	0,23	3,03	2,89	117	504	181	182
2,0374	212,84	5,52	2,36	4,92	3,37	0,93	2,13	66	24	2	174
0	111,36	1,62	2,50	3,43	-0,7	2,46	1,81	13	27	16	439
4,0012	893,46	4,07	5,92	2,62	5,14	3,94	0,38	2104	13888	3547	4654
2,0374	154,72	2,05	3,32	4,14	2,19	-0,78	2,66	350	10	692	285
0	92,01	2,24	1,70	4,06	-0,81	4,8	0,67	8	4	34	2
1,7782	146,31	1,82	1,54	2,08	1,36	2,27	4,35	2	19	5	20
2,0374	174,87	2,11	3,51	3,1	3,4	1,71	2,71	67	297	965	959
1,7782	132,35	2,73	3,46	3,12	1,58	2,34	2,73	354	119	35	120
0	100,86	0,47	1,70	2,51	2,08	-1,09	2,42	8	4	28	2
0	98,31	1,66	1,36	3,58	-0,24	-0,09	1,01	6	1	12	16
2,3979	267,77	2,52	2,31	2,31	1,64	-2,55	-0,99	36	46	22	31
0	118,47	4,33	2,84	3,39	-0,6	-2,83	0,81	224	92	14	30
2,0374	244,77	3,29	2,17	1,46	1,89	-1,66	3,85	29	50	39	44
2,0374	191,24	4,61	4,80	2,15	-0,44	1,42	2,9	465	1242	4504	514
1,5441	95,93	5,22	1,29	1,11	0,38	2,21	2,11	9	3	57	1
2,0374	183,26	3,3	3,12	2,15	-0,36	2,5	2,65	83	267	50	1023
2,0374	208,82	1,35	2,50	2,06	2,53	-0,42	0,77	166	28	42	72
1,7782	101,79	2,33	2,51	3,46	0,63	3,37	-3,34	32	73	27	59
0	151,56	2,69	1,28	2,09	2,46	-0,53	0,62	54	2	7	16
2,3979	216,97	2,12	3,65	1,42	0,46	2,04	-1,24	712	196	333	189
1,7782	107,57	1,92	3,34	5,39	1,9	0,68	2,26	52	453	48	113
2,0374	195,99	4,6	1,63	4,31	3,24	1,14	1,76	10	6	20	34
1,7782	142,22	2,34	1,93	3,31	0,61	3,29	0,58	14	47	87	19
1,0792	282,09	3,48	4,05	2,53	0,13	1,24	4,56	68	182	436	90
0	106,32	1,22	1,21	2,24	3,64	0,34	-0,58	7	0	10	7
2,0374	217,98	2,9	2,29	3,64	2,92	-0,31	3,89	47	8	1359	15
2,3802	132,61	2,65	1,97	2,87	0,65	-2,21	2,61	18	9	14	8
2,0374	178,85	2,05	2,56	2,91	2,87	0,65	2,16	91	20	9	21
2,0374	164,33	3,62	2,83	4,14	3,84	2,19	0,47	102	49	486	60
0	109,61	4,41	1,67	5,09	2,07	0,51	0,26	12	51	5	0
1,7782	95,53	3,5	1,41	4,95	-0,97	-0,52	-1,71	1	14	36	26
0	107,08	1,77	2,79	3,24	0,05	1,1	0,55	130	196	10	29
0	118,84	4,93	1,79	3,2	3,34	0,26	4,23	123	18	37	61
0	95,42	2	1,11	2,46	-1,08	1,6	4,18	55	5	11	18
2,0374	175,90	1,37	2,25	1,59	3,65	3,43	0,42	99	4	44	12
2,3979	213,05	3,52	2,99	4,21	-0,58	4,94	2,85	530	159	23	86
2,0374	176,83	1,1	2,60	2,69	2,01	1,9	1,54	64	187	33	33
2,3802	87,61	1,35	2,77	2,86	0,25	2,35	2,86	23	71	146	17
0	57,83	1,77	1,28	4,38	-1,68	-2,58	2,56	4	11	17	32
2,0374	203,01	3,04	1,89	3,76	1,21	4,65	2,54	24	68	25	16
1,7782	113,53	4,15	2,02	1,95	-0,12	1,32	5,3	13	9	172	17
1,5441	157,53	1,67	3,40	3,49	6,69	2,79	-0,28	229	194	42	152
2,3802	101,09	1,35	2,08	3,6	-2,62	-1,96	1,36	20	54	10	59
1,5441	112,25	1,66	4,38	2,32	4,27	1,76	2,24	1215	462	449	361
1,5441	106,66	1,51	2,24	4,04	1,44	-0,91	7,1	22	29	40	17
1,7782	169,81	1,9	3,35	3,54	0,85	2,48	1,34	144	346	711	32

NTNU

Institutt for psykologi

2,3802	79,83	1,47	2,00	3,28	-3,09	-0,4	2,02	51	33	7	27
0	140,72	2,9	1,58	5,58	0,53	-1,14	-1,6	16	40	15	5
2,0374	178,34	3,02	4,12	1,88	0,85	0,66	2,1	4175	659	165	569
0	126,68	2,35	1,44	4,68	0,91	-1,01	-3,35	19	1	10	4
2,0374	194,91	1,68	1,33	4,68	-1,53	4,39	1,87	9	4	13	9
2,3802	79,79	4,31	1,94	2,13	-0,99	0,78	1,74	3	5	31	8
4,0012	923,82	2,91	5,86	2,92	2,49	1,89	1,38	9740	2228	16585	32993
0	100,55	3,18	2,04	3,01	0,79	0,7	4,46	36	28	99	22
0	97,05	2,56	2,95	1,2	2,6	-0,44	0,37	223	587	65	38
1,7782	122,53	2,03	2,57	3,04	2,33	0,02	1,13	45	21	180	221
0	105,18	2,48	2,47	3,95	0,69	-0,74	6,48	19	153	44	178
0	108,75	3,38	2,71	4,02	0,8	2,59	2,3	31	49	38	6
0	94,49	4,46	2,80	4,73	3,15	1,73	0,3	247	250	112	56
1,7782	110,50	2,92	3,60	3,7	0,88	-0,5	2,42	170	120	39	100
2,0374	187,29	4,83	3,48	3,98	1,24	2,91	2,99	540	59	470	1022
1,5441	104,64	3,65	1,19	3,02	3,03	1,11	1,44	9	2	1	41
1,5441	129,76	4,21	2,06	2,39	1,29	1,23	4,53	8	51	15	44
2,3979	248,78	2,29	1,63	3,2	-0,99	2,94	-0,37	8	5	49	20
1,5441	112,23	2,17	2,28	4,62	2,32	1,9	7,01	87	126	12	25
0	157,46	4,4	3,49	4,84	-2,88	3,44	2,16	56	195	1347	69
1,7782	134,33	2,66	4,23	1,14	2	1,82	0,08	91	1459	171	409
0	114,41	3,22	2,24	2,15	2,06	0,43	1,38	55	13	29	3
2,3979	241,24	3,18	2,52	3,95	3,55	3,31	-0,86	114	34	552	23
2,0374	169,49	2,12	3,10	1,36	2,09	0,69	2,13	40	861	50	24
2,0374	212,41	4,89	2,90	2,24	0,39	1,02	1,11	39	289	21	61
2,0374	163,18	2,55	4,78	1,66	3,26	-0,91	2	799	1254	1234	416
2,0374	167,86	1,15	3,66	3,38	4,25	1,06	3,92	144	54	227	61
2,0374	170,39	3,38	2,51	4,01	1	1,71	-0,25	24	32	13	24
1,0792	342,50	0,53	5,08	4,64	2,22	-0,42	-0,59	563	4323	1026	1233
1,7782	150,14	2,25	3,22	5,35	0,1	2,1	-0,26	342	149	221	24
4,0012	898,33	2,13	6,37	1,38	0,44	1,63	3,26	14092	8167	7308	3609
1,7782	91,33	2,25	3,42	2,07	1,82	0,98	0,76	59	488	20	81
0	123,64	3,76	2,88	2,85	0,76	-2,96	1,28	428	314	35	154
0	128,51	5,09	1,87	3,73	4,29	2,79	1,5	16	135	19	19
0	84,61	3,78	2,87	3,28	4,17	3,25	-1,9	61	100	393	64
0	98,07	2,42	1,26	2,55	2,47	-0,88	5,57	4	14	9	18
0	111,67	3,08	1,78	4,04	1,86	-1,24	-0,79	21	467	5	54
2,0374	211,16	5,63	2,35	2,55	2,4	-0,67	-0,43	86	103	100	63
2,3802	72,26	3,1	2,56	3,48	2,46	-1,07	2,62	434	72	20	134
2,3979	251,57	2,22	1,61	1,71	0,44	-0,35	-1,61	11	12	5	10
1,5441	93,90	1,11	1,65	5,01	2,36	3,49	4,53	14	5	1	4
4,0012	904,77	3,38	6,21	2,35	5,86	1,55	2,5	31424	36253	21928	30961
4,0012	893,97	0,49	5,25	3,06	2,6	2,25	1,95	11368	2844	1503	910
2,3979	231,72	3,92	2,64	2,57	4,19	1,16	2,62	181	100	25	187
2,3979	224,57	1,88	2,48	3,38	1,15	3,54	1,73	13	119	118	43
0	81,31	4,28	2,09	2,13	2,15	-1,85	3,82	13	3	4	37
1,0792	312,61	2,3	2,99	2,69	-2,97	-0,53	-0,33	99	198	227	45
1,0792	360,33	3,2	6,32	2,62	5,92	-1,59	1,3	18201	37585	3933	18608
1,0792	358,51	2,96	3,92	3,7	1,75	5,65	2,65	147	654	597	233
1,0792	321,41	2,76	4,62	3,11	-1,36	3,83	2,67	1144	1005	1540	923
1,0792	331,37	5,04	3,57	2,82	3,33	0,62	-0,46	344	269	2725	286

NTNU

Institutt for psykologi

2,0374	147,89	1,99	2,59	3,25	2,34	5,67	0,81	38	58	63	25
1,7782	140,65	3,15	3,33	1,98	1,33	-1,07	3,31	86	288	165	137
2,0374	172,70	4,24	3,97	3,32	1,79	0,87	0,49	1207	1065	1194	102
1,7782	141,45	2,31	2,74	2,04	3,83	-1,33	3,33	57	19	99	309
1,7782	128,21	2,8	1,24	4,1	-1,34	3,27	3,06	28	6	3	4
2,3979	263,25	-2,12	2,08	1,35	-0,2	4,41	2,59	160	27	16	12
2,0374	192,39	3,96	2,84	4,93	0,7	1,79	1,88	127	127	62	86
0	118,47	1,77	2,28	2,29	0,6	0,28	1,38	10	57	62	13
1,7782	111,37	1,02	4,26	1,36	0,4	1,71	4,13	621	784	369	798
2,3979	242,68	4,53	3,09	4,25	0,91	4,91	1,63	52	124	31	102
0	90,63	3,15	2,53	4,07	1,07	1,17	0,05	44	20	9	57
4,0012	917,86	-0,91	3,58	2,24	1,43	1,81	1,61	162	254	89	39
4,0012	850,75	3,16	5,34	4	2,03	0,33	3,44	15009	6310	19176	2115
1,7782	89,97	1,92	1,32	2,66	1,54	-0,05	-0,04	2	2	35	26
1,5441	118,78	3,91	2,89	2,19	0,84	3,67	0,51	48	360	180	94
0	87,56	3,81	2,36	3,14	2,32	3,43	0,49	145	29	23	44
2,3979	258,93	3,63	2,59	6,58	2,8	4,71	3,13	92	41	39	50
1,7782	117,39	4,09	2,35	2,05	-0,27	1,19	1,27	57	10	75	21
0	116,61	3,54	2,76	3,55	0,46	4,5	4	138	53	205	80
2,1761	198,69	5,32	-0,56	3,5	0,16	-2,97	1,65	0	0	1	0
2,0374	213,86	1,86	2,95	4,06	-0,1	3,16	3,92	13	76	58	48
2,0374	163,68	3	4,53	3,17	1,63	5,2	0,77	833	311	126	717
2,3979	246,84	2,24	3,30	3,07	0,82	-1,28	-0,16	209	1628	172	151
2,0374	201,87	0,78	3,67	2,09	3,15	-1,2	0,65	182	157	122	347
4,0012	910,60	2,49	4,96	3,71	-1,71	3,32	0,02	396	2045	1860	1039
0	128,37	3,13	2,18	3,77	5,02	1,07	2,02	147	26	193	84
0	105,00	2,21	2,70	2,04	2,71	2,08	3,33	115	55	76	8
2,3979	234,08	2,12	4,59	2,2	3,62	-0,34	-1,92	1580	9419	846	691
1,5441	184,11	2,76	3,73	2,42	2,33	-1,33	3,07	357	430	91	380
0	145,58	2,71	3,05	3,54	0,51	1,2	1,06	174	148	87	212
0	113,79	4,36	1,30	2,31	2,5	0,7	2,03	2	9	1	2
2,0374	188,56	2,02	2,94	6,06	2	3,92	0,64	43	45	25	138
1,5441	135,08	2,56	2,19	2,14	-0,27	-0,53	-2,99	18	5	12	29
2,0374	148,93	0,87	2,96	2,51	2,37	-2,24	-0,94	126	95	32	17
1,5441	117,56	2,89	1,66	3,95	-0,36	-1,02	-0,83	7	35	32	25
1,5441	133,05	2,72	3,26	3,13	1,18	0,8	-0,2	262	77	319	481
0	99,25	2,69	3,07	3,87	2,33	-1,16	3,46	898	127	210	176
2,0374	179,74	1,49	3,85	2,05	-0,4	0,88	1,82	661	577	107	130
0	121,05	1,84	2,22	2,15	1,59	-1,88	1,76	12	28	164	24

Sensurveiledning:

Scoring: I will define difficulty by the proportion of students who notice a problem. Importance is rated as 1 = something an attentive reader can work out but shouldn't

have to; 2 = something that the authors would have to fix in a revision; 3 = something the authors would have to fix in a revision and that seriously affects the interpretation of the results; 4 = a fatal flaw, either something that can't be fixed by reanalysis or reporting missing but available data because, for example, the authors have to collect more data to add a missing control, or an ethical problem so serious that the research should not be done at all. Then I will give points using this formula:

Points = importance/(proportion who noticed)$^{0.3}$

Students noticing flaws I have missed will be treated the same, and these additional flaws will have to be added to these instructions for examiners in case there is a complaint.

A false positive, reporting a flaw that isn't there, loses 2 points.

Because I can't predict the difficulty, I can't know in advance the distribution of points and I can't yet assign point ranges to grades.

I will identify the exam that I think should just barely get a C, based on part 1, use that to establish the lower limit of C. The highest point score establishes the upper limit of A. I will then divide the range in between into five equal intervals, three for the grades, two for borderline ranges. I will divide the range from the lower limit of C to 0 into six equal intervals. When a student's score falls into a borderline interval, the grade given to part 2, where students identify the most important flaws and suggest improvements, determines whether the student gets the grade above or below.

Lines 9 – 13, importance = 1: "SCGPD"

Giving only the abbreviation in the abstract. Spell out what that means.

Line 16, importance = 2: Mcguire (2013) is a science fiction story. It is not even a pretend scientific publication or diagnostic manual that defines the fictional condition of SCGPD.

Line 17, importance = 2: "exclusively affects natural scientists and engineers ".

That needs a definition of what counts as natural science and engineering. Also, the footnote says that Philomena Killjoy, a psychologist, tried to take over the world, and psychology is generally considered a social science, even though studying the behaviour of other species (ethology and animal cognition) would generally be considered natural science. So the meaning of the terms is not established clearly enough to make this claim without either more explanation or a reference.

Line 18, importance = 2: "who excel in their field."

How is excellence defined? There should be a reference or an explanation.

Line 23, importance = 2: That World Domination Attempt is capitalised suggests that this is an accepted and validated term for some well-defined concept. If so, there should be a reference. If not, there should be an explanation of what constitutes a World Domination Attempt.

Line 27, importance = 1: “(Delight & Sixsmith, 2019)”

Not in the references

Line 30, importance = 2: Where does the 1 in 100 000 come from? Is that world-wide? Organs are most commonly installed in churches, so if this number comes from a country in which Christianity dominated, that rate is likely to be higher than where Christians are rare.

Line 31, importance = 2: “6 out of 229 known sufferers “.

Where do those numbers come from? Are those 229 everyone since the index case, or those alive now?

Line 31, importance = 1: are 229 affected people enough for inclusion in a diagnostic manual? Is SCGPD a recognised diagnostic category?

Line 33, importance = : “overrepresentation of organists”

Not backed up by statistics, and there is not enough information for the reader to do the statistics. What is the numerator for that 1 in 100 000 rate?

Line 34, importance = 2: “what aspect of playing organ makes someone vulnerable to SCGPD?”

Hypothesis 1, claims that playing the organ causes vulnerability to SCGPD, but NOT that SCGPD causes an interest in playing the organ.

Even if there is a statistical association, why should playing the organ make people vulnerable to SCGPD, rather than SCGPD making people interested in (some aspect of) playing the organ, or SCGPD and playing the organ both being consequences of a common cause? This needs better justification.

Lines 34 – 36, importance = 2: “Is it something to do with the sheer volume of sound, with the frequency range of sounds that can be generated at any one time”

These variables have been collapsed into a single index in the analysis without any explanation why.

Lines 36 - 37, importance = 2: “variety of sounds the instrument can produce”

This measure disappears from the analysis, without explanation.

Lines 37 – 38, importance = 2: “should organs be made more widely accessible to detect people vulnerable to SCGPD”

Hypothesis 2, claims that SCGPD causes people to be interested in playing the organ, without justification. It would be acceptable if the authors made clear they were trying to cover all the possibilities, but that was not done here.

Lines 37 – 39, importance = 2: “should organs be (…) dismantled and outlawed, to prevent conversion from excellence in a field to SCGPD?”

Hypothesis 3, claims that playing the organ causes SCGPD, without justification. This differs from hypothesis 1 in that here playing the organ is assumed to be a sufficient cause, while hypothesis 1 makes no such commitment. It would be acceptable to just list these hypotheses without justification if the authors made clear they were trying to cover all the possibilities, but that was not done here.

Line 54, importance = 1: “18±7 to c. 3.7±1.3 million”.

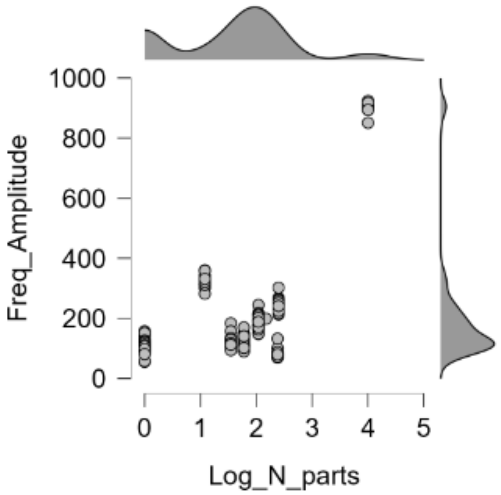
That presumably means 18±7 individuals, rather than 18±7 million individuals. That should be expressed more clearly.

Lines 54 - 57, importance = 2: “18±7 to c. 3.7±1.3 million”.

Where are the corresponding numbers for people with SCGPD?

Lines 65 - 67, importance = 2: “stocked with telenkas, didgeridoos, bongo drums, saxophones, Scottish pipes, violins, Celtic harps, hurdy-gurdies, electronic keyboards, and a 4000 pipe organ.”

How were those instruments chosen? In lines 36 - 39, the authors mention four parameters, importance = : "the sheer volume of sound, with the frequency range of sounds that can be generated at any one time, the variety of sounds the instrument can produce, or the number of parts of the instrument". Were the instruments chosen to cover a good range of that parameter space, perhaps approximately even spacing, and extending close to the corners of that space? The plot on the right suggests not. Or were instruments chosen according to personal preferences by the experimenters, cost, sponsorship, requests by participants, or some other criterion?



Line 67, importance = 3: “We recorded the time spent playing each instrument.”

How? Data were collected in a maximum-security installation, so the study may have tapped into standard surveillance, but that still leaves an ethical issue. Was there free and informed consent? This seems like the kind of study where participants knowing there is a study could distort the data, and if everything is recorded already for security purposes, there is a case for asking for consent retrospectively. However, there is no mention of consent, either in advance or after the fact, and no discussion of how free consent can be in a prison.

Line 67, importance = 2: “We recorded the time spent playing each instrument.”

More information needed. Is that the time spent making sound, or the time between picking up and putting down the instrument, provided it was used to make some sound for some minimum amount of time?

Lines 63 – 69, importance = 4: “We provided 187 people with SCGPD who are confined in a maximum-security installation with unrestricted access, during the hours from 07:00 to 23:00, to a music room ... We included the 148 who played for at least three hours per weeks through the whole study period. “

16 hours/day gives 112 playing hours per week. 148 people playing for at least 3h/week means at least 444 playing hours. Given that there is only one music room, and only one organ, this has two consequences:

1) It is possible for demand for the organ to outstrip the available hours. Instrument choices cannot be regarded as independent. The number of other instruments is not specified, so the same problem may occur there. The hypothesis has nothing to say about what happens when people with SCGPD must coordinate their choices. In the absence of any relevant hypothesis, one must expect that anything could happen. As stated, the hypothesis can only be tested if the data from each individual are independent of the data of other individuals. That is not true. This is a fatal flaw in the design of the experiment. The claim that because people with SCGPD never collaborate, all the data can be treated as independent is true only for half the data, the expected death toll, but not for instrument choice.
2) Unless there are soundproof booths, which should have been mentioned, it is unlikely that the choice of music is independent, which might also affect choice of instrument.

Lines 69 – 70, importance = 2: “All participants retained broadly human body plans.”

Needs more explanation. How much variation is covered by “broadly human”?

Line 70, importance = 3: “Age ranged from 15 to 92”

So at least one participant is not an adult, raising more concern about consent. How was consent regarding adolescents handled?

Lines 74 – 75, importance = 2: “We considered a patient to have chosen an instrument when the patient played that instrument for at least twice as long as the next favourite.”

As described, that condition would be fulfilled the moment someone is considered to have tried any of the instruments, before touching any other. For the condition to make sense, there must be a minimum time period over which it is applied. What is it?

Further, did anyone ever change their mind and switch to a different instrument after that condition was first met? If so, how was that handled?

Lines 76 – 82, importance = 4: “four of the organists tried to take over the world using the organ. Because the facility lacks enough janitors with the requisite security clearance to provide coverage at all hours, these attempts failed only because the patients questioned each other’s methodology and because their independent modifications of the organ interfered with each other and the instrument melted down. Access to the music room has been terminated due to the radiation hazard. The room has been entombed in a concrete sarcophagus.”

Given that the authors speculate in the introduction that playing the organ might cause SCGPD, it is a serious ethical oversight that not enough janitors were available for round the clock coverage, and apparently no nerd snipers at all.

Further, there is no mention of when the organ became a radiation hazard, how much participants or staff may have been exposed, and whether anything is being done to treat the effects of exposure. Or any account of how an organ became a radiation

hazard serious enough that it needs to be treated like the melted-down reactor at Chernobyl.

Lines 85 – 87, importance = 2: “We then averaged amplitudes either over all play during the last four weeks of data collection, or only over the most passionate play, as measured by force sensors built into the keyboards.”

What is the criterion for defining passionate play? Pressure on the keys so many standard deviations above the mean for that person, or the group mean, or the X% of time with the greatest force, or what?

Lines 85 – 87, importance = 2: “We then averaged amplitudes either over all play during the last four weeks of data collection, or only over the most passionate play, as measured by force sensors built into the keyboards.”

Pressure on the keys works for keyboards. What criterion is being applied to other instruments?

Lines 87 – 88, importance = 3: “The latter turned out to be more informative and is reported here.”

“More informative” likely refers to greater effect size and more extreme p-value. However, the phrasing implies that this selection was made only after the experiment was planned and the data had been collected, and so constitutes the kind of experimenter degrees of freedom that raise the rate of false positives, as shown by Simmons et al (2011; Undisclosed flexibility allows presenting anything as significant). This is a form of p-hacking. The closest this can come to being legitimate is if the authors report all the data and perform a multiverse analysis (Silberzahn et al. 2018; Many analysts, one Data Set, importance = : How variations in analytic choices affect results).

Importance = 3: The data file contains only one variable related to the sum of amplitudes, presumably the version considered more informative. It should contain all variants of the measure that were analysed, not only the one the authors chose to report.

Lines 89 – 91, importance = 2: “We record the number of parts that make up an instrument, to test whether complexity has an effect.”

How does that work for the electronic keyboards? Are the electronics one part, is each board a part, each transistor? Are there external loudspeakers, and if yes, are they included in the part count? If amplifiers are separate, are they included?

Lines 89 – 91, importance = 2: “We record the number of parts that make up an instrument, to test whether complexity has an effect.”

Why was number of parts chosen as a measure of complexity? Is it the best measure, or the most convenient, or some combination?

Line 89, importance = 2: “We carried out a multiple regression”

Multiple regression can only use the *average* expected death toll as outcome. That analytic approach ignores that this average comes with a variance. An analysis that can take this variance into account is a General Linear Model.

Lines 92 – 93, importance = 2: Why was number of parts chosen as a measure of complexity? Is it the best measure, or the most convenient, or some combination?

Lines 93 – 94, importance = 2: “the predicted death tolls obtained from Wilde & Dry (2024) as outcome.”

Which predicted death tolls? The previous reference to Wilde & Dry mentions two predictions, one from healthy experts, one from people with SCGPD. A look at the data file shows that the estimates must come from people with SCGPD, importance = : there are nine estimates, rather than 28, and the average correlation is that reported for the latter group.

Lines 93 – 94, importance = 2: “the predicted death tolls obtained from Wilde & Dry (2024) as outcome.”

How valid are those predicted death tolls? Intraclass correlation was reported, but that measures reliability, not validity.

Rating the importance of this omission is a bit tricky. If this were a real paper, with real references, we could look up Wilde & Dry (2024). A strict application of my criteria would mean the lowest level of importance in that an attentive reader can find or infer the missing information, but the authors really should not impose that work on the reader. However, the validity of a measurement is important enough that I rate this at the next level up, something that definitely would need to be fixed in a revision.

Lines 68 and 96 – 98, importance = 1: “41 patients played the didgeridoo, 32 the saxophone, 21 the bongo drums, 16 the electronic keyboard, 15 the violin, 9 the Scottish pipes, 8 the organ, 7 the Celtic harp, 3 the telenka, and 1 the hurdy-gurdy.”

That adds up to 149 people. Line 68 says 148. The data file has only two telenka players.

Lines 101 – 102, importance = 2: “Contrary to our expectations, the number of parts had no effect on the predicted number of deaths resulting from a successful World Domination Attempt ($t(147) = 1.56$, $p > 0.1$, confidence intervals -0.03 to 0.26).“

The numbers indicate merely the lack of a significant effect, not the absence of an effect. Bayesian statistics or equivalence testing would be necessary to provide positive evidence of the absence of an effect.

Lines 103 – 105, importance = 1: “confidence intervals”

What is the range of the confidence intervals? 95% is fairly conventional, but still arbitrary. This should be made clear.

Lines 108 – 109, importance = 2: “Our results confirm that playing the organ exacerbates the symptoms of SCGPD and increases the likelihood of World Domination Attempts.”

That is another causal hypothesis, but differs from all three mentioned in the introduction. It most resembles hypothesis 3, that playing the organ causes SCGPD, but it makes a broader claim of symptoms getting worse over an unspecified time span, rather than causing transition with its accompanying amplification of some personality traits.

Lines 108 – 109, importance = 2: “Our results confirm that playing the organ exacerbates the symptoms of SCGPD and increases the likelihood of World Domination Attempts.”

The results confirm no such thing. The methods and analysis cannot distinguish the various causal relationships.

Lines 110 – 114, importance = 2: “We recommend that aspiring organists are subject to strict scrutiny, that anyone working in natural science or engineering who wants to play the organ is screened for risk factors, and that sites where organs are installed are strictly surveilled so that unauthorised users can be apprehended and a thorough background check initiated.”

Any such policy recommendation should be based on effect sizes in reported in a format that can inform estimation of costs and benefits, perhaps number needed to treat. There is no such report.

Lines 116 – 117, importance = 1: “Our observations indicate that playing the organ is a compulsion”

What observations would these be? There are no relevant data in this report.

Lines 122 – 123, importance = 2: “However, only those who have already undergone transition would go to the trouble of building their own organ.”

Claims confirmation of hypothesis 2, without data to back it up.

Lines 125 – 126, importance = 2: “complexity has no effect on choice of musical instrument”

Repeats the error in lines 101 – 102, Claiming the absence of an effect when the data only support a lack of a significant effect. These are not the same thing.

Lines 128 – 130, importance = 1: “Some anecdotal reports suggest a possible connection with 1980s German Electronic Dance Music”

What anecdotal reports? If they are published, there should be a reference. If this is a personal communication, that should be reported.

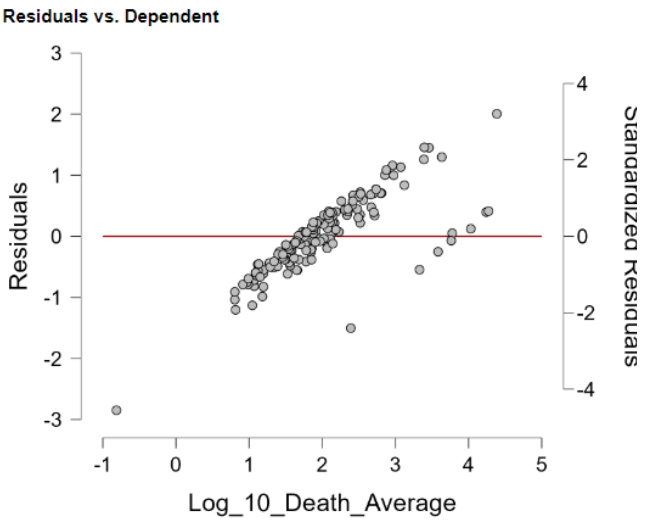
Methods

Importance = 2: The introduction mentions that onset of SCGPD is defined by an amplification of personality traits, and that this is usually followed by a World Domination Attempt. It is implied that a World Domination Attempt is inevitable, and confinement in a maximum security installation suggests that some crime was committed, possibly the attempt itself. But it is not actually stated clearly whether all those taking part in the experiment have actually attempted to take over the world.

NTNU

Institutt for psykologi

Data analysis and Results, importance = 3:



There is no mention of whether assumption checks were made. Plotting the residuals against the dependent variable shows a set of at least eight data points roughly falling on a separate line with the same slope as the other data points. (These represent the Scottish bagpipes, which I gave a systematically higher death toll.) That suggests an analysis that can take into account that not all instruments are the same. Again, a multiple regression won't do the job.

Importance = 3: There is no justification for why interaction terms were excluded from the regression. (The bagpipes create an interaction, but multiple regression is not sensitive to it being the bagpipes because it treats all data points the same, you can't tell it that they fall into different categories.)

Importance = 2: The introduction states that Wilde & Dry (2024) generated *two* sets of predictions, one from healthy experts, one from people with SCGPD. The data file contains only one set, apparently that from people with SCGPD. Why was the other data set excluded and this one chosen? Presumably because of the better intraclass correlation, but that should be made clear.

Importance = 2: The data file contains data that are not being analysed, apparently the six traits that are known risk factors for SCGPD. Why were these not analysed?

Importance = 3: Once Social_Dominance_z-score is included in the analysis, Freq_Amplitude no longer has any significant effect. That is because, for convenience, I generated the fake data by letting Freq_Amplitude and Log_N_Parts jointly determine Social_Dominance_z-score, which then determined the 13 individual death toll predictions, which were aggregated into Lod_10_Death_Average. In other words, Social_Dominance_z-score mediates the effect. A mediator is rather hard to distinguish from a collider, from Social_Dominance_z-score being the cause of both instrument choice and expected death toll, which is the causal relationship I intended to simulate.

Importance = 3: The reported analysis cannot possibly test the three causal hypotheses mentioned in the introduction. It can only detect a statistical relationship. A causal relationship might have been detected by controlling what instruments were available to whom.

Importance = 2: Figure 2 shows that one instruments (the organ) is quite an outlier when it comes to sound amplitude. How much do the results depend on just the eight people who played the organ? It would be good to have an analysis without them, to see how robust the results are.

Discussion

Importance = 1: The various causal hypotheses are not clearly distinguished. A numbered list of hypotheses and referring to the numbers would have made things clearer.

Karakterskala som er benyttet

Bokstavkarakter: https://innsida.ntnu.no/wiki/-/wiki/Norsk/Karakterskalaen