Apes in Lab Coats: Scientific Exploratory Behaviour in *Homo sapiens*

Dennis Lendrem (dennis.lendrem@ncl.ac.uk)
Newcastle University https://orcid.org/0000-0001-6268-5509

Article

Keywords:

Posted Date: May 25th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1650096/v1

Abstract

This study reports observations of scientists attempting to solve a standard experimental problem - DNA amplification through the Polymerase Chain Reaction (PCR) - in a 'virtual' laboratory. One striking observation is that scientists employ search strategies not dissimilar to the foraging strategies of apes and other animals. While such search strategies may be appropriate when exploring two- or three-dimensional environments, they are highly unsuited to exploring the complex multi-dimensional spaces representing many scientific and technical challenges. The evolution of exploratory behaviour in *Homo sapiens* may represent a major obstacle to the uptake of better tools. The wider adoption of mathematical and statistical tools designed for the exploration and mapping of such spaces may be constrained by our evolutionary biology.

Introduction

Many scientists believe that small experiments, guided by scientific intuition alone, are more efficient than Design of Experiments (DOE) tools. This belief is strong and persists, even in the face of data demonstrating that it is clearly wrong.[1,2,3] This situation is perpetuated by scientific teaching in support of One-Factor-At-A-Time (OFAT) approaches to scientific experimentation. Though widely discredited, OFAT persists partly because it is taught in schools as part of the "scientific method," partly because of the illusion that it is an efficient methodology, and partly because OFAT methods have the beguiling property of generating data confirming initial scientific beliefs[3,4].

In stark contrast, designed experiments permit the efficient mapping of highly complex, multidimensional spaces[1,2,3]. Fellermann *et al* [5] developed a web-based application introducing multifactorial experimental design concepts and supporting teaching of the polymerase chain reaction—the Virtual PCR simulator[5]. Learners select experimental settings and receive results of their simulated reactions quickly, allowing rapid iteration between data generation and analysis. This enables scientists to perform experiments within the constraints of a short teaching session and in the relative safety of a virtual environment.[5] The Virtual PCR Simulator not only allows scientists to conduct experiments, it records their experimental decisions giving a unique insight into scientific behaviour. We used this tool to observe the multidimensional problem-solving strategies employed by scientists in a simulated laboratory environment.

Results

Illustrative patterns of scientific exploratory behaviour are shown in **Fig. 1**. The points in each plot show the settings selected for two of the key variables – volume of polymerase and volume of dNTP – as the search progressed. There was a range of scientific exploratory strategies.

Some scientists chose to fix values of both primers and dNTP and perform experiments in the remaining space (eg. scientist 35, scientist 61 and scientist 63 ran all their experiments at a fixed volume of primer

and dNTP).

Others chose to fix one or the other - either the volume of primers was fixed and the volume of dNTP varied, or the volume of dNTP was fixed and the volume of primers explored (eg. 1, 4, 5, 8, 13, 14, 23, 49, 57). Some scientists chose to fix one or other variable before exploring the other (eg. scientists 27, 40, 45). Yet others chose to vary both dNTP and primers **at the same time** making it impossible to separate out the impact of each of the variables (eg. scientists 11, 26, 33 & 42). The end-result is that the relative importance of the variables can be impossible to determine and very little of the design space is properly explored. While some scientists restrict themselves to relatively small changes from an initial value (eg. scientists 17, 21, 29, 41, 44, 58, 60, 62, 65 & 66). Others make more sweeping changes across a wider range of volumes (eg. scientists 6, 20, 22, 30, 38, 50 & 59).

To visualize the movements of scientists in multi-dimensional space, we used PCA to reduce the twelve-factor design space to just two dimensions and mapped scientist movements in the resulting canonical space – see Fig. 2. Mapping movements in canonical space highlights that there are large regions of the problem space for which we have no information.

Upon closer inspection, larger movements within canonical space, seem to be driven by the success of initial results. In **Fig. 3** the search paths of four of the scientists are shown in more detail with elapsed time (in minutes) as indicated. A striking feature of these search paths is their similarity to foraging search strategies in other animals. For example, after one or two initial experiments yielding unpromising results, scientists might make a Lévy flight[8] to some other region. They then wander around that region of the design space to fine-tune their results – see **Scientists A**, **C** and **D**. Alternatively, if the initial results are promising some scientists might make an excursion to a choice of more radical settings – see **Scientist B**. Fingers burnt, they then retreat to their original settings and conduct smaller excursions from that point.

Discussion

Many accounts of the 'scientific method' suppose a normative model with scientists rigorously and systematically exploring the design space generating data to support or reject a hypothesis[9,10]. The first striking feature of our data is that many scientists perform rather less rigorous, less than systematic searches of the design space preferring to make an initial guess and then follow scientific hunches[11]. This leaves much of the design space unexplored. Indeed many variables may remain completely unexplored. The impact of key variables may remain unknown raising questions about the robustness to variation in such variables[12]. Worse still, these variables may go unreported, contributing to the difficulty involved in replicating scientific experiments from different laboratories[13].

A second, striking feature of the experimental search strategies employed by scientists is their similarity to foraging strategies[14,15] resembling the search behaviour of apes and other animals [16,17,18,19]. Such foraging strategies evolved to exploit patch resources and may be characterized as area-restricted search

patterns in regions of high reward punctuated by longer Lévy flights as foraging returns decline. While these strategies may have served us well in our evolutionary past, such search strategies may constrain our scientific search strategies. While some authors suggest that higher primates - such as chimpanzees - may exhibit more complex foraging behaviours[20] the data to support more complex strategies may be weak[21]. These more complex behaviours may be captured using simple algorithms and heuristic rules[22]. Such foraging behaviours may not be dissimilar to those of isolated tribes of hunter-gatherers and early hominids [15,16,17,18,19]. If scientific search strategies have evolved from foraging strategies better suited to exploring two- or three-dimensional environments, this may place serious constraints on the ability of *Homo sapiens* to navigate multidimensional spaces.

Of course, the scientific behaviours we report here are from a series of simulated laboratory experiments. While there may be good evidence to support the use of such simulation tools[23,24,25,26] we might question the ecological validity of these data. Do they simply arise as a result of some kind of ludic artefact?[27] We think not. Participants report that the problem is engaging, realistic, and that the simulation mirrors real life decision making accurately. We believe these workshops are a realistic approximation to real-world scientific behaviour in the laboratory.

Many authors have lamented the failure of scientists to adopt DOE tools and note that good sound, rational argument alone has not always been successful in persuading scientists to adopt those tools[1,2]. There may be a number of reasons for this. However, this study suggests that the evolution of exploratory behaviour in *Homo sapiens* may present a major obstacle to the uptake of DOE tools. While these tools work best in conjunction with prior scientific knowledge, DOE tools are often counter-intuitive demanding that scientists focus less on immediate returns in exchange for better information about unexplored regions of the design space.

The scientific problems we now wrestle with are often highly dimensional[3]. DOE tools require that scientists think and work in multiple dimensions. *Homo sapiens* did not evolve to function in such spaces. Our evolution left us poorly equipped to deal with such problems. One of the difficulties we face when exploring highly multidimensional spaces is that they are often not amenable to the low dimensional tools that have served us well until now. Resistance to the uptake of DOE tools, and indeed many other mathematical and statistical tools designed for multidimensional problem solving, may be driven by our evolutionary biology.

Methods

Participants

Scientists were recruited through the Summer School in Experimental Design at Newcastle University. Scientists ranged in age from 22 through to 50 years of age. All held scientific degrees in biotechnology, bioengineering or medical sciences to PhD level with 42% recruited from industry and 58% from academia. In total, observations were recorded for 69 scientists performing a minimum of 10 experiments

over a 30-minute scientific exploratory play workshop in which they were invited to use a Virtual PCR simulator to identify PCR settings producing the maximum yield of DNA in a DNA amplification experiment. Scientific movements - including the time and choice of settings for each experimental run - within the design space were tracked using anonymized data. Individuals were identified with a random eight-character digital identifier.

The Simulator

Described in full elsewhere[5] the Virtual PCR simulator is available on the Newcastle University server at http://virtual-pcr.ico2s.org/ where it records design decisions, experiments, and results for all experiments. DNA amplification through PCR is a complex multi-dimensional problem that involves cycling through three key steps – denaturation, annealing, and extension. Scientists were able to adjust settings for 12 experimental factors including the temperature and duration of each these three key steps during each cycle. In addition, there are four reagent volumes and one categorical choice of polymerase. Opportunities were given for questions about PCR. Working individually, scientists were allotted 30 minutes to identify settings maximizing the yield of the process - though most completed the task to their own level of satisfaction in less than this time. Data were exported as csv files and imported into SAS JMP Pro[7] for statistical analysis and graphical visualization. In addition to plotting the movements of individual scientists in two- and three-dimensional space, we used principal components analysis (PCA) to reduce the dimensionality of the design space to the first two key components, before plotting movements in multi-dimensional canonical space.

Declarations

Conflicts of Interest

No conflicts of interest to report. Newcastle University Ethics Approval Reference 22393/2022.

Acknowledgements

The author thanks the following for valuable feedback on earlier versions of the manuscript: Lendrem BC, Howard TP, Fellermann H, Barnes M, Moore M, Ydenberg R, Linsley M, Ridley MW, Maldonado R, Shirt-Ediss B, Kozyra V, Lendrem BAD, Pratt AG, Isaacs JD. Data downloads and pre-processing of data from the Virtual Simulator: Fellermann H and Lendrem BAD. The BBSRC STARS award (BB/P004342/1) provided financial support to the workshop, and the author was supported in part by the National Institute of Health Research Biomedical Research Centre at Newcastle University and the Newcastle-upon-Tyne Hospitals National Health Service Trust.

Contributions

HF, BSE and VK wrote the software and mathematical model. TPH advised on principles and applications of PCR. HF, ML, DL and TPH tested and iterated the simulator in a teaching environment. HF and BADL performed data downloads and pre-processing of the data from the Virtual Simulator. MM advised on SAS JMP visualization tools. RY and MWR commented on foraging theory, and RM, AGP and JDI shared scientific and biomedical insights.

References

1. Box, G.E.P., Hunter, J.S. and Hunter, W.G. (2005) *Statistics for experimenters*. Wiley New York.
2. Montgomery, D.C. (2012) *Design and analysis of experiments*. 8th edn. John Wiley & Sons.
3. Lendrem, D.W., Lendrem, B.C., Woods, D., Rowland-Jones, R., Burke, M., Chatfield, M., Issacs, J.D. and Owen, M.R. (2015) 'Lost in Space: Design of Experiments and Scientific Exploration in a Hogarth Universe', Drug Discov Today, 20(11), 1365–1371.
4. Lendrem, D.W., Lendrem, B.C., Rowland-Jones, R., D'Agostino, F., Linsley, M., Owen, M.R. and Isaacs, J.D. (2016) 'Teaching examples for the design of experiments: geographical sensitivity and the self-fulfilling prophecy', Pharmaceutical Statistics, 15(1), 90–92.
5. Fellermann, H., Shirt-Ediss, B., Kozyra, J., Linsley, M., Lendrem, D., Isaacs, J. and Howard, T. (2019) 'Design of experiments and the virtual PCR simulator: An online game for pharmaceutical scientists and biotechnologists', Pharm Stat. 18(4), 402–406.
6. Fellermann, H.S.-E., B;Kozyra, J;Linsley, M;Lendrem, D.W.;Howard, T; (2018) *The PCR Simulator: an on-line application for teaching Design of Experiments and the polymerase chain reaction.*
7. *SAS JMP® Pro* (2018) (Version 14) [Computer program]. SAS Institute.
8. Meyers, R.A. (ed.) (2009) *Encyclopedia of Complexity and Systems Science*. New York: Springer.
9. Medawar, P.B. (2013) *Induction and Intuition in Scientific Thought*. Taylor & Francis.
10. Rosenberg, A. (2011) *Philosophy of science: a contemporary introduction*. Routledge.
11. Dunbar, K. (2000) 'How scientists think in the real world: Implications for science education', Journal of Applied Developmental Psychology, 21(1), 49–58.
12. Box, G.E.P., Tiao, G.C. and Bisgaard, S. (2000) *Box on quality and discovery: with design, control, and robustness*. John Wiley & Sons.
13. Yong, E. (2012) 'Replication studies: Bad copy', Nature, 485(7398), 298–300.
14. Stephens, D. and Charnov, E.L. (1982) ' Optimal foraging: some simple stochastic models.', Behavioral Ecology and Sociobiology, 10(4).
15. Wosniack, M.E., Santos, M.C., Raposo, E.P., Viswanathan, G.M. and da Luz, M.G.E. (2017) 'The evolutionary origins of Levy walk foraging', PLoS Comput Biol, 13(10).
16. Barrett, B.J., McElreath, R.L. and Perry, S.E. (2017) 'Pay-off-biased social learning underlies the diffusion of novel extractive foraging traditions in a wild primate', *Proceedings of the Royal Society B-Biological Sciences*, 284(1856).

17. Hills, T.T., Kalff, C. and Wiener, J.M. (2013) 'Adaptive Levy Processes and Area-Restricted Search in Human Foraging', PLoS One, 8(4).
18. Raichlen, D.A., Wood, B.M., Gordon, A.D., Mabulla, A.Z.P., Marlowe, F.W. and Pontzer, H. (2014) 'Evidence of Levy walk foraging patterns in human hunter-gatherers', Proc Natl Acad Sci U S A, 111(2), 728–733.
19. Reynolds, A., Ceccon, E., Baldauf, C., Medeiros, T.K. and Miramontes, O. (2018) 'Levy foraging patterns of rural humans', PLoS One, 13(6).
20. Sayers, K. and Menzel, C.R. (2012) 'Memory and foraging theory: chimpanzee utilization of optimality heuristics in the rank-order recovery of hidden foods', Animal Behaviour, 84(4), 795–803.
21. Trapanese, C., Meunier, H. and Masi, S. (2018) 'What, where and when: spatial foraging decisions in primates.', Biological Reviews.
22. Janson, C. (2014) 'Death of the (Traveling) Salesman: Primates Do Not Show Clear Evidence of Multi-Step Route Planning', American Journal of Primatology, 76(5), 410–420.
23. Waldrop, M.M. (2013) 'The virtual lab', Nature, 499(7458), 268–270.
24. Koretsky, M.A., D;Barnes, C;Kimura, S; (2008) 'Enhancement of student learning in experimental design using a virtual laboratory', IEEE Trans Education, 51(1), 76–85.
25. Vogel, J.J., Vogel, D.S., Cannon-Bowers, J., Bowers, C.A., Muse, K. and Wright, M. (2016) 'Computer gaming and interactive simulations for learning: a meta-analysis.', *J Educ Comput Res.*, 34(3), 229-243.
26. Darius, P.P., KM;Schrevens E; (2007) 'Virtual experiments and their use in teaching experimental design.', International Statistical Review, 75(3), 281–294.
27. Taleb, N.N. (2010) *The black swan: the impact of the highly improbable*. 2nd edn. New York: Random House Trade Paperbacks.

Figures

Figure 1

Typical search patterns in two-dimensional space for scientists exploring PCR DNA amplification. Settings for two critical variables – the volume of primers and the volume of dNTP – are plotted for sequences of experiments. The exploratory behaviour of scientists varies markedly. See text for details.

Figure 2

Individual patterns of exploration for scientists mapped into multidimensional space. Movements in canonical space are plotted for scientists during an initial exploratory play workshop. Each point represents an experiment in multidimensional space.

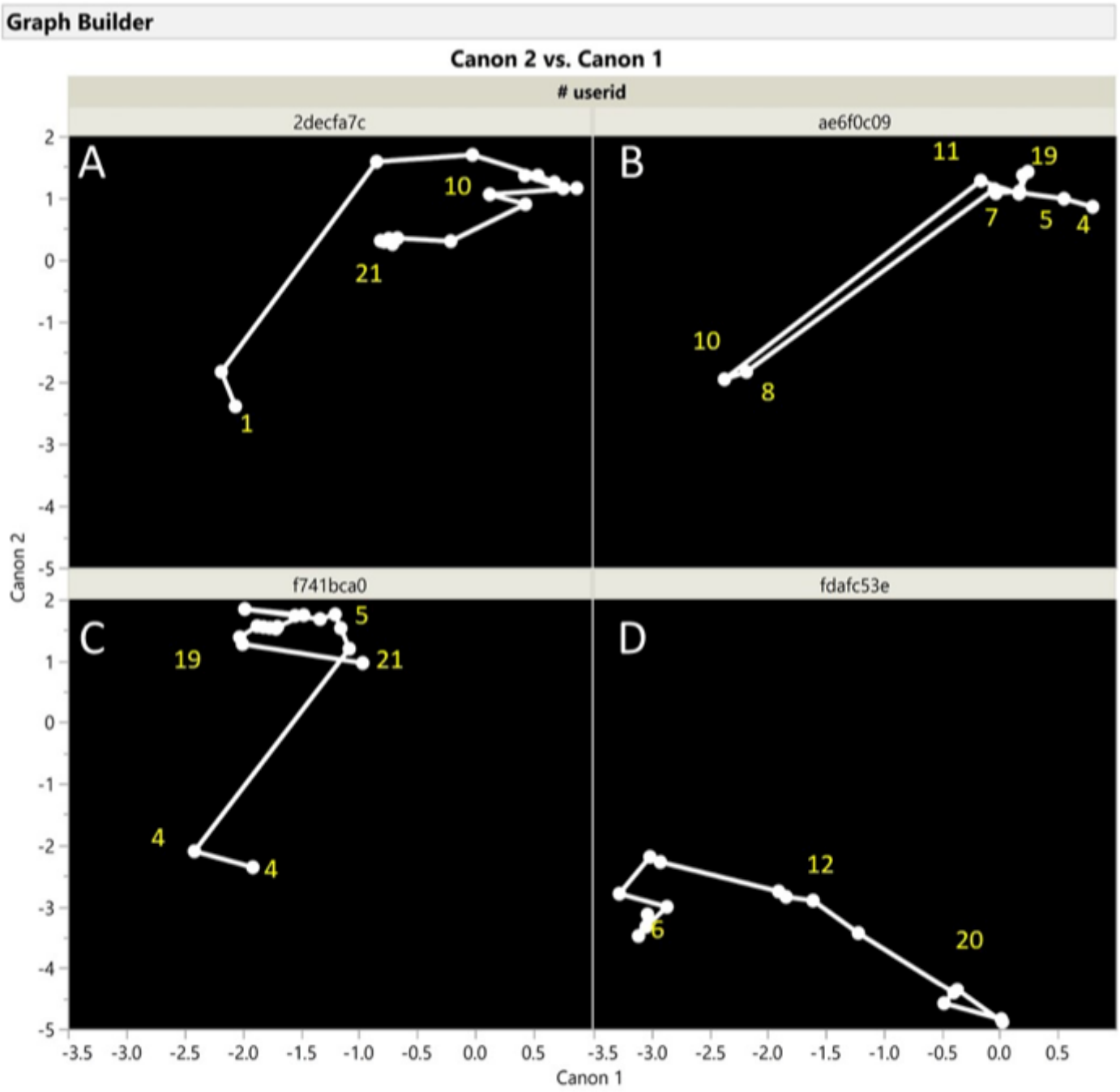

Figure 3

Individual search paths for scientists mapped into multidimensional canonical space. Movements in canonical space are plotted for four of the scientists during an initial exploratory play workshop. Each point represents an experiment in multidimensional space. The numbers in yellow are the elapsed time in minutes since the start of the task. **Scientist A** begins with two experiments in the lower-left quadrant of

canonical space. When the results prove unpromising, experimentation shifts quickly to the upper-right quadrant and the rest of the experiments are conducted in this region of multidimensional space. In contrast, **Scientist B** performs several initial experiments in the upper-right quadrant before moving to the lower-left quadrant. When the results are unpromising, we observe a rapid retreat to the original settings followed by short exploratory flights from this region. **Scientist C** performs initial experiments in the lower-left quadrant before moving to a more promising region of the multidimensional design space upper-right quadrant. Finally, **Scientist D** takes rather more time to begin experiments in the lower-left quadrant before moving eventually to a more promising region in the lower-right hand quadrant. Such search patterns are characteristic of many foraging animals including apes and other animals. See text for details.