



A common raven (*Corvus corax*) engages with an experimental task that is mounted above the ground.

How STRANGE are your study animals?

Michael M. Webster & Christian Rutz

A new framework for animal-behaviour research will help to avoid sampling bias – ten years on from the call to widen the pool of human participants in psychology studies beyond the WEIRD.

Ten years ago this week, researchers pointed out that many important findings in human experimental psychology cannot be generalized because study participants are predominantly drawn from a small, unrepresentative subset of the world's population: societies that are Western, Educated, Industrialized,

Rich and Democratic (WEIRD)¹. Mounting evidence suggests that there could be similar sampling problems in research on animals. Behavioural studies of a wide range of species – from insects to primates – could be affected, with researchers testing individuals that are not fully representative of the wider populations they seek to understand. For example, certain sampling protocols are likely to trap the boldest animals, potentially skewing experimental results².

It is high time for scientists who work on animal behaviour to identify, and mitigate, potential sampling biases. Simply gathering more data is not a solution, because researchers should always strive to minimize the number of experimental animals used. Instead, we propose a framework with a fitting acronym – STRANGE – that researchers can use to interrogate how unusual their study subjects are. This will aid the design of new studies,

and enable potential biases to be declared and discussed when publishing completed work.

WEIRD humans

In June 2010, human-evolutionary biologist Joe Henrich and his co-authors published a landmark paper challenging the widely held assumption that human behaviour varies little across populations¹. They highlighted that the vast majority of research on human behaviour, and its cognitive basis, was conducted using participants from societies who are often outliers in broader comparisons. For example, WEIRD people are unusual in how they find their way around, what they consider to be fair, and their willingness to punish others¹.

Several research fields chart the behaviour of non-human animals, including comparative psychology, ethology, behavioural ecology, evolutionary biology and

Comment

conservation science. They attempt to uncover the evolutionary origins, developmental pathways, adaptive function, underlying mechanisms and conservation relevance of a wide variety of behaviours. Their common goal is to identify general biological principles that apply within and across taxa. Many animal studies are susceptible to sampling biases.

STRANGE framework

One well-known source of bias is the excessive reliance in some research fields on a few model organisms³, such as fruit flies (*Drosophila* spp.) and laboratory mice (*Mus musculus*), which inevitably limits the generality of findings. Less obvious, but perhaps much more problematic, is the fact that many species exhibit substantial variation in behaviour, and only certain subsets

of that diversity might (unintentionally) turn up in test samples. In primates, for example, individuals of particular ‘personalities’ can be more likely to take part in trials voluntarily⁴ (see ‘Beware of STRANGE’). Life-long participation in experiments can also significantly alter subjects’ natural behaviour.

When developing the framework, our starting observation was that an animal’s behaviour is shaped⁵ by its genetic make-up, experience and social background. Ensuring that subjects are representative with regards to these three factors is paramount. Adding a few specific effects, which are well-documented yet often overlooked, provides the acronym STRANGE. It stands for: Social background; Trappability and self-selection; Rearing history; Acclimation and habituation; Natural

changes in responsiveness; Genetic make-up; and Experience. STRANGE-related biases can affect both laboratory and field studies. They can influence which animals are sampled for testing, the extent to which they participate in experiments and, importantly, the behaviours that they exhibit during trials. This, in turn, can complicate comparisons between studies, hampering both the generalizability and the reproducibility of findings.

Our approach uses a simple test to evaluate the robustness of a completed or planned study. Researchers should ask: are my animal subjects strange when compared to the wider population for which I wish to make inferences, in any of the seven categories of the framework? Here, using selected examples, we showcase how sampling biases could affect the behaviours observed in animal studies (for further examples, see ‘Beware of STRANGE’ and Supplementary Information, Table S1).

Social background. This includes an animal’s social status, the nature and frequency of its interactions with others, and its past opportunities to learn socially from other individuals or their products. For example, when pheasants (*Phasianus colchicus*) were given a spatial-discrimination task to learn which of two holes contained bait, they performed better when they had been housed in groups of five rather than in groups of three⁶.

Trappability and self-selection. These are closely related processes. They mean that individuals with particular traits² are most likely to be caught or to participate voluntarily in experiments (see ‘Beware of STRANGE’). In a classic study on pumpkinseed sunfish (*Lepomis gibbosus*), individuals collected using funnel traps (into which fish have to actively swim) were faster to start eating in the laboratory compared with those trapped using more indiscriminate nets⁷.

Trappability effects are expected to be prominent in bio-logging studies in which animals are fitted with electronic tags for remote observation. For example, when researchers tag surfacing whales or seabirds in breeding colonies that are difficult to access, they might inadvertently obtain non-random samples. Self-selection biases are a well-known – but usually neglected – problem in laboratory and field studies of animal cognition⁴.

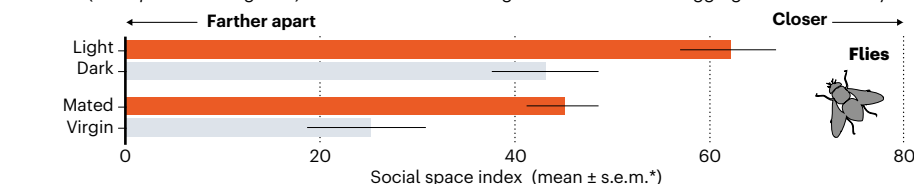
Rearing history. This describes an animal’s developmental experiences, including the extent to which it has been exposed to a stimulating physical environment, to other animals and to humans. One study showed that captive-reared jumping spiders (*Phidippus audax*) were less active, less exploratory and had reduced interest in prey than were those collected from the wild⁸. Exposure to enrichment, social stimulation and exercise during development can affect brain development and, in turn, cognitive and motor performance⁹. This relationship is well established

BEWARE OF STRANGE

Many factors can affect animal behaviour, as these examples show, with their STRANGE category acronyms in brackets (the studies shown explicitly set out to examine these effects). Problems arise when test samples are biased with regards to these categories, and when researchers do not account for this.

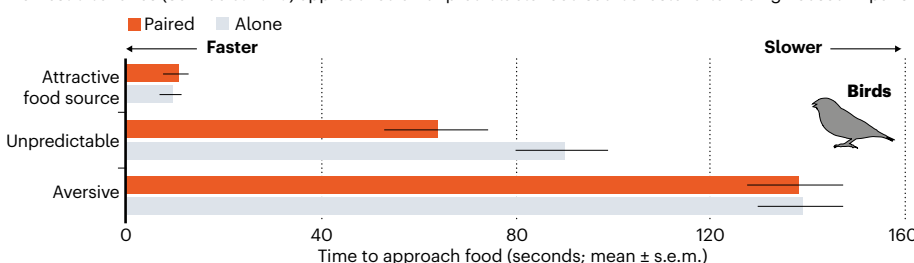
Experience and timing influence sociality (E, N)

Fruit flies (*Drosophila melanogaster*) that were tested in the light or that had mated aggregated more closely¹⁸.



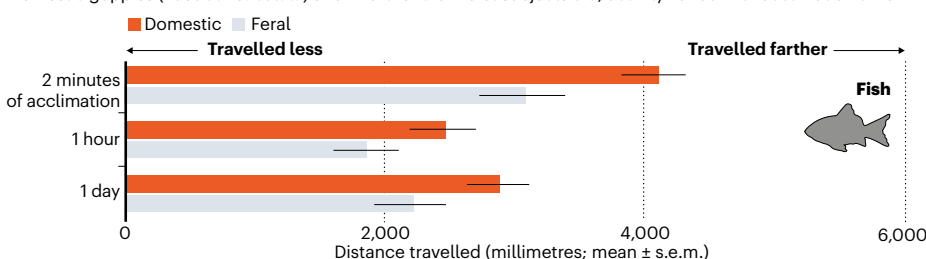
Social contact induces ‘optimism’ (S, R)

Domestic canaries (*Serinus canaria*) approached an unpredictable food source faster after being housed in pairs¹⁹.



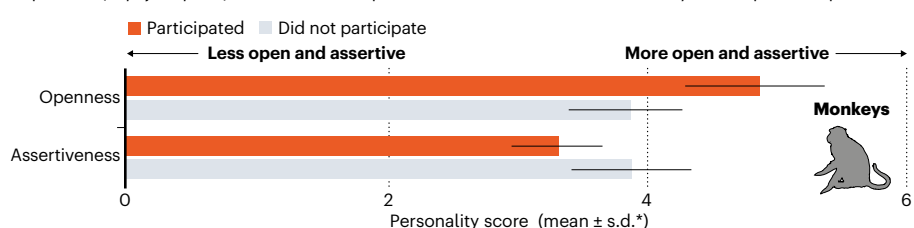
Origin and acclimation affect activity (G, A)

Domestic guppies (*Poecilia reticulata*) swam farther than feral subjects did; activity varied with acclimation time²⁰.



Self-selection varies with personality (T)

Capuchins (*Sapajus apella*) that were more open and less assertive were more likely to take part in experiments⁴.



*s.d., standard deviation; s.e.m., standard error of the mean.



A hawksbill sea turtle (*Eretmochelys imbricata*) tagged with a satellite transmitter.

for rats and mice¹⁰.

Acclimation and habituation. These refer to behavioural changes over time following handling, tagging or exposure to new testing situations. For instance, green turtles (*Chelonia mydas*) spent less time swimming and more time feeding the day after they were tagged with video cameras than they did on the day of tagging. This probably reflects the turtles' recovery from the tagging procedure and their habituation to the cameras¹¹.

Natural changes in responsiveness. These sometimes follow daily, reproductive or seasonal cycles, or the transition from one life stage to another. This means that the timing of experiments is often crucial, as highlighted by a study that found that honeybees (*Apis mellifera*) learn more effectively in the morning than at other times of the day¹².

Genetic make-up. This can have profound effects on behaviour. For example, the experience of losing territorial fights early in life has different effects in wild-type brown rats (*Rattus norvegicus*) than in individuals from a common genetic strain. The laboratory rats end up spending less time investigating intruders as adults¹³. There can also be marked differences in behaviour between wild populations, and between males and females, as has been shown in an experimental study on anti-predator behaviour in Trinidadian guppies (*Poecilia reticulata*), in which females spent more time shoaling than did males, for instance¹⁴.

Experience. This encompasses

opportunities for individual learning, such as participation in earlier experiments. (This means there can be overlap with some of the other framework categories.) After male chaffinches (*Fringilla coelebs*) had been lured by a playback of a rival's song, captured, handled and released again, they sang fewer territorial songs in response to simulated territory

“Concerns arise whenever samples of study subjects are unwittingly biased.”

intrusions than did birds that had not been captured¹⁵. Long-lived animals can accumulate complex experimental histories in research laboratories, which in our view must be better documented and accounted for.

Crucial considerations

Some general points about the framework are worth highlighting. First, there are important conceptual differences between WEIRD and STRANGE. The former identifies attributes of a particular demographic group; the latter refers to a suite of factors that can affect behaviour. The seven categories of STRANGE are not problematic by themselves. In fact, they are often the focus of well-designed research projects, such as those we mention here, or are confounding factors that have been explicitly controlled for. Concerns

arise whenever samples of study subjects are unwittingly biased with regards to any of these categories, and when researchers overlook that fact.

Second, there is overlap and strong interdependence between some STRANGE categories. For example, the origin of an animal – whether it was wild-caught or captive-bred – will often simultaneously affect its genetic make-up, social background, experience and rearing history. Third, we designed the categories to be broad enough to accommodate future extensions. Finally, although STRANGE refers to samples of animal subjects, effects can be moderated by study protocols. For instance, depending on the species, testing with or without others present can significantly affect an animal's willingness to participate, as often observed in fish, birds and primates.

What will our critics say? Some might note that several of the effects we discuss – such as self-selection and experience biases – have been highlighted as problematic in the past. We feel that, because research practice seems to have improved little in response to specific warnings, it is time to introduce a memorable framework that integrates all factors that could affect the generalizability of animal-behaviour studies.

Others might point out that most of the examples we mention simply illustrate drivers of behavioural variation, and not sampling bias. As we will explain, systematic studies are

Comment

urgently required to assess our contention that the potential for bias is both widespread and routinely ignored.

3D reproducibility

It is possible to identify and mitigate STRANGE-related biases at little or no extra cost, with a simple '3D' approach: design, declare, discuss (see also Supplementary information, Box S1).

Design. There are many opportunities for researchers to make their test samples more representative (see Box S1, step 1). For example, we recommend that projects that rely on trapping wild subjects consider using a variety of trap types or bait preparations. They can also sample across multiple populations to reduce systematic biases. Similarly, in studies in which animals effectively select themselves for participation, we encourage researchers to think about ways of altering the testing environment or the task itself, to encourage more inclusive participation. Anecdotal evidence reveals that some crows, for instance, are hesitant to approach experimental tasks that are placed on the floor, but will readily engage with them when they are mounted just one metre above the ground.

Declare. When submitting a manuscript, researchers should supply – and journals should ask for – an objective evaluation of a system's 'STRANGEness' so that editors and reviewers can gauge the scope for bias. Although reporting standards have significantly improved over the past few years, a surprising number of journals – including many specializing in animal behaviour and cognition – still lack robust reporting policies. We urge journals to insist that all behavioural studies report detailed subject-attribute data as set out in the ARRIVE guidelines (Animal Research: Reporting of *In Vivo* Experiments)¹⁶,

with some additions (such as a full declaration of the attributes of non-participating subjects; see Box S1, step 2a).

Our STRANGE framework can then be used together with this information to evaluate the scope for sampling biases (Box S1, step 2b),

"We encourage animal-behaviour researchers to routinely ask how unusual their samples are."

and to describe which precautionary steps were taken to avoid bias, if any (Box S1, step 2c). These declarations enhance transparency, provide a valuable resource for systematic reviews and formal meta-analyses, and will hopefully encourage better planning of future studies.

Discuss. It is essential to detail any potential issues prominently in the main body of research papers (Box S1, step 3). This would force authors to explicitly link their findings to the studied sample, rather than to the population or species as a whole¹⁷.

As well as supporting all steps of this 3D approach, STRANGE provides a convenient memory aid: we encourage animal-behaviour researchers to routinely ask how unusual their samples are. (For further questions about specific categories, see Table S1.)

Next steps

There is a large body of literature demonstrating how the categories in our STRANGE framework can affect the behaviours researchers observe. On the basis of this evidence, our personal research experience and our discussions with many colleagues, we suspect that STRANGE-related problems are widespread.

We now urgently need retrospective analyses of published work that quantify how often test samples are biased with regards to these categories, when researchers failed to account for – or declare – these confounding factors.

Human experimental psychology has made great strides towards addressing sampling biases by improving reporting standards. For example, the Association for Psychological Science recommends that authors identify the participant population, explain their selection and consider how generalizable their findings are. However, despite these efforts, problems are surprisingly persistent, with many published studies still providing insufficient detail about their participants¹⁷.

Animal-behaviour scientists have a lot to learn from the WEIRD debate. We hope that our STRANGE framework will help to improve how animal-behaviour research is conducted, reported and interpreted.

The authors

Michael M. Webster is a lecturer in the School of Biology, University of St Andrews, UK.

Christian Rutz is a professor in the School of Biology, University of St Andrews, UK, and worked on this article while he was the 2019–2020 Grass Fellow at the Radcliffe Institute for Advanced Study, Harvard University, Cambridge, Massachusetts, USA.
e-mail: christian.rutz@st-andrews.ac.uk

1. Henrich, J., Heine, S. J. & Norenzayan, A. *Behav. Brain Sci.* **33**, 61–83 (2010).
2. Biro, P. A. & Dingemanse, N. J. *Trends Ecol. Evol.* **24**, 66–67 (2009).
3. Alfred, J. & Baldwin, I. T. *eLife* **4**, e06956 (2015).
4. Morton, F. B., Lee, P. C. & Buchanan-Smith, H. M. *Anim. Cogn.* **16**, 677–684 (2013).
5. Bateson, P. & Laland, K. N. *Trends Ecol. Evol.* **28**, 712–718 (2013).
6. Langley, E. J., van Horik, J. O., Whiteside, M. A. & Madden, J. R. *Anim. Behav.* **142**, 87–93 (2018).
7. Wilson, D. S., Coleman, K., Clark, A. B. & Biederman, L. *J. Comp. Psychol.* **107**, 250–260 (1993).
8. Carducci, J. P. & Jakob, E. M. *Anim. Behav.* **59**, 39–46 (2000).
9. Thornton, A. & Lukas, D. *Phil. Trans. R. Soc. B* **367**, 2773–2783 (2012).
10. van Praag, H., Kempermann, G. & Gage, F. H. *Nature Rev. Neurosci.* **1**, 191–198 (2000).
11. Thomson, J. A. & Heithaus, M. R. *J. Exp. Mar. Biol. Ecol.* **450**, 15–20 (2014).
12. Lehmann, M., Gustav, D. & Galizia, C. G. *Behav. Ecol. Sociobiol.* **65**, 205–215 (2011).
13. Vidal, J., Buwalda, B. & Koolhaas, J. M. *Behav. Processes* **88**, 76–80 (2011).
14. Magurran, A. E. *Phil. Trans. R. Soc. B* **353**, 275–286 (1998).
15. Budka, M., Matyjasiak, P., Typiak, J., Okotowski, M. & Zagalska-Neubauer, M. *J. Ornithol.* **160**, 673–684 (2019).
16. Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. *PLoS Biol.* **8**, e1000412 (2010).
17. Rad, M. S., Martingano, A. J. & Ginges, J. *Proc. Natl Acad. Sci. USA* **115**, 11401–11405 (2018).
18. Simon, A. F. et al. *Genes Brain Behav.* **11**, 243–252 (2012).
19. Lalot, M., Ung, D., Péron, F., d'Ettorre, P. & Bovet, D. *Behav. Processes* **134**, 70–77 (2017).
20. O'Neill, S. J., Williamson, J. E., Tosetto, L. & Brown, C. *Behav. Ecol. Sociobiol.* **72**, 166 (2018).

Supplementary information accompanies this article; see [go.nature.com/2uefq6g](https://doi.org/10.1038/s41586-020-2000-2).



Sampling biases can affect research on model organisms, such as fruit flies.