PROCEEDINGS B

royalsocietypublishing.org/rspb

Biological science practices

Check for updates

Cite this article: Schickore J, Magalhães S, Gómez-Schiavon M. 2026 Harnessing complexity may strangle it: what is a well-controlled experiment? *Proc. R. Soc. B* **293**: 20253130. https://doi.org/10.1098/rspb.2025.3130

Received: 4 December 2025
Accepted: 9 April 2026

Subject Category:
Ecology

Subject Areas:
computational biology, ecology, systems biology

Keywords:
experimental design, control, emergent properties, causality

Author for correspondence:
Mariana Gómez-Schiavon
e-mails: mgschiavon@liigh.unam.mx; mgschiavon@gmail.com

Harnessing complexity may strangle it: what is a well-controlled experiment?

Jutta Schickore[1,2], Sara Magalhães[2,3] and Mariana Gómez-Schiavon[2,4,5]

[1]Department of History and Philosophy of Science and Medicine, Indiana University, Bloomington, IN, USA
[2]Wissenschaftskolleg zu Berlin, Institute for Advanced Study, Berlin, Germany
[3]cE3c—Center for Ecology, Evolution and Environmental Changes & CHANGE—Global Change and Sustainability Institute, Departamento de Biologia, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal
[4]Laboratorio Internacional de Investigación sobre el Genoma Humano (LIIGH), Universidad Nacional Autónoma de México, Querétaro, Mexico
[5]ANID—Millennium Science InitiativeProgram—Millennium Institute for Integrative Biology (iBio), Santiago, Chile

MG-S, 0000-0002-0955-7257

This article discusses a core ideal of biological experimentation, the goal of controlling as strictly as possible. We present the epistemological argument underlying the ideal of a perfectly controlled experiment. While most scientists agree that the ideal is unattainable in practice, they still uphold the pragmatic goal that one should strive to approximate the ideal as much as possible in all cases. We argue that this desire for strict control is problematic and outline an alternative understanding of control in experimentation. First, we distinguish four different ways in which real-world biological experiments can fall short of the ideal of perfect control. Then, we argue that strict control does not always yield better experimental results. Certain core features of biological systems actually suggest that strict control can be counterproductive for learning from experiments. We thus argue for 'controlling well', for striking a balance between holding variables fixed while allowing some degrees of freedom. What it means to control well cannot be specified in general terms; it is context- and purpose-dependent and needs to be determined relative to a particular experimental project.

1. Introduction: the ideal of perfect control

Students learn in science classes that rigorous control is key to extracting information from an experimental procedure. There is an epistemological warrant underlying this idea: it is that to determine whether an intervention on an independent variable is the real cause of change in the dependent variable, all other variables but the independent variable have to be held constant. If this is the case and the independent variable is completely isolated, then all possible confounding variables are excluded. The experimenter is warranted to conclude that any change occurring in the system after the experimental intervention is caused by the change in the independent variable alone [1]. From this epistemological perspective, 'perfect' control (in the sense just described) is therefore a core ideal of experimentation.

In practice, however, no experiment will ever be perfectly controlled, as virtually every scientist will agree. Similarly, as textbooks on experimental design discuss in detail the many complex strategies to ensure control over independent variables, experimental conditions, methods, reagents and experimenters [2, pp. 137–219], they imply that the more these features can be controlled, the better the interpretation of the experiment [2, p. 136]. Most scientists still uphold the pragmatic aim of *strict* control as something that they should strive to approximate as best as they can in all cases [3]. We outline an alternative understanding of control in experimentation.

THE ROYAL SOCIETY PUBLISHING

In this article, we have two goals. We first clarify four different ways in which real-world biological experiments fall short of the ideal of *perfect* control. We do this to ensure a shared understanding of what this ideal entails and why it is practically unattainable.

Our second aim is to develop the argument that *strict* control of variables *should not* always be the guide for experimental design. Building on Wallis [4] and Desjardins *et al*. [5], we argue that *stricter* control does not always yield better experimental results. Certain core features of biological systems—such as their intrinsic complexity and entanglement with environments, their stochastic nature (i.e. the inherent randomness or probabilistic behaviour of biological processes) and their emergent properties (i.e. collective behaviours that arise from interactions among components, not predictable from the parts alone)—suggest that strict control can be counterproductive for learning from experiments. In these cases, relaxing strict control may yield deeper insights into the target system and its relation to the environment by allowing key aspects of that system to manifest.

Of course, we do not suggest foregoing control altogether. Rather, we argue for 'controlling well', by striking a balance between holding variables fixed while allowing some degrees of freedom. That is, contrary to the widespread belief that an appropriately controlled biological experiment is one where variables are strictly—as much as possible—controlled, we argue that in some situations, controlling well implies deliberately *not* controlling strictly. What it means to control well cannot be specified in general terms; it is context- and purpose-dependent and needs to be determined relative to particular experimental projects. Our article offers examples of how this purpose-dependence can play out in biological experimentation.

We thus advocate for an experimental mindset that aligns with the complexity of the systems under study by not always striving to minimize variability or isolate single variables. Adopting this mindset goes beyond simply acknowledging that the ideal of perfect control cannot be achieved in practice; it encourages a fundamental shift in how we conceive control and rigour in experimentation—from pursuing the aim of strict control to the purpose-dependent practice of well-controlled experiments.

We proceed in the following steps:

(i) We discuss the commonplace that perfect control is practically unattainable because it is not possible to control everything. We show that 'not controlling everything' can mean different things: first, some variables do not need to be controlled for the experiment to work. Second, we may not have the technical tools to control a variable even though we know it should be controlled. Third, we may fail to control for something we should control. Fourth, controlling necessarily involves decisions that have implications for what can and cannot be concluded from the experiment.

(ii) We discuss important features of biological phenomena, as they are currently understood, including the complexity of causal chains, object–environment entanglement, biological noise and emergent properties. We argue that, in light of these features, the pragmatic aim of *strict* control of variables warrants reconsideration. In some contexts, relaxing the control of certain variables may in fact offer a *more* scientifically rigorous approach, as it allows key aspects of biological systems to be properly observed and interpreted.

(iii) We introduce the concept of a well-controlled experiment and argue that the aim of 'controlling well' is more suitable to the practice of biological experimentation than the general aim of strict control. We discuss the implications of this shift for scientific communication practices and for our general understanding of biological experimentation.

2. It is not possible to control everything

A perfectly controlled experiment is one in which a representative target system is completely isolated from its environment and manipulated by making changes to its properties while holding everything else fixed [3]. It is commonplace that not everything is controlled in an actual biological experiment. However, what exactly this means is rarely discussed. We distinguish different ways in which an experiment can fall short of the ideal of perfect control.

(a) Some variables do not need to be controlled

In the most literal and trivial sense, not *every single variable* other than the target system is held fixed in a real-world experiment. Controls typically do not extend to the environment outside the laboratory or setting for the experiment (except in rare cases where extremely sensitive measuring instruments need to be shielded from external noise and vibrations). But even within the laboratory, many background factors can be ignored because they are not expected to have an effect on the outcome of the experiment. For example, carrying out an experiment at different hours of the day or at different times of the year might have an important impact on its outcome because it interferes with the circadian, lunar or seasonal rhythms of organisms [6–8]. But it typically does not matter whether the experiment is done on a Monday or Wednesday.

(b) We may not have the appropriate technology to control everything

Some variables are just not fully controllable because the appropriate technology is not (yet) available. Numerous historical cases illustrate how challenging it can be to develop the technologies for controlling variables that should be controlled. In his studies of psychic phenomena, the physicist Michael Faraday, trying to prevent participants in table turning experiments from making involuntary movements, designed a contraption consisting of a stack of cardboard sheets, arranged like a voltaic pile with pellets of wax in between, which would be placed between the hands of the séance participants and the tabletop. The sheets were arranged and marked in such a way that their displacement would give evidence of hand movements prior to the movement of the table [9]. The experimental psychologist Carl Stumpf designed and set up a room-sized arrangement of tubes to be able to compare how his experimental subjects perceived natural and machine-generated vowels [10].

This difficulty is not just a thing of the past but pertinent nowadays. For instance, most biological phenomena are highly dependent upon abiotic conditions such as temperature, humidity, daylight cycle or air pressure. Laboratory experiments are typically conducted in controlled environments, such as climate chambers or incubators, which regulate these features. However, these technologies do not always work flawlessly, and abiotic conditions often fluctuate even within these enclosures.

In fact, this is the first point where one may begin to wonder whether such strict control is even desirable. If the conclusions that are drawn from an experiment are contingent upon the very specific set of conditions found in a particular shelf of a particular incubator in a particular laboratory, then this lack of generality compromises the general interest of the finding at stake. That is, the external validity of those results may be very reduced, a phenomenon coined the 'standardization fallacy' [11].

(c) We may fail to control for confounding factors

Given that we cannot control everything, we must select which variables to control. This inevitably carries the risk of overlooking confounder effects that may influence the outcome.

An example concerns photosynthesis research, in which the plant tissues and cells under study are grown on specific culture media. The growth of cultured plant tissues is determined by the initial levels of growth substances and nutrients in the culture medium as well as by the relative rates of their disappearance as the plant grows. For a long time, it was assumed that blue and near-UV light, well-known inhibitors of elongation growth in seedlings, acted on light-sensitive systems in the plant tissues. However, an experimental study on the model organism *Arabidopsis* revealed that the growth inhibition was primarily due to photochemical changes in the culture medium rather than to photosensory functions of the plant tissue itself [12,13]. Indeed, when culture media are exposed to plant material, their composition changes as cells metabolize components. Abiotic factors, such as light and heat, can change the chemical composition of culture media as well. Therefore, studies of plant metabolism can only lead to sound conclusions if these factors are appropriately monitored.

In contrast to the examples described in §2b, where experimenters are aware that their technical abilities to control certain factors are limited and can try to develop workarounds, the case of photosynthesis research is an example where only hindsight revealed that the experimenters failed to control an important variable. In this case, the realization that an unknown confounder was at work leads to discovery [14]. Our impulse to control as much and as strictly as possible, may reflect a deeper fear of missing such important confounders. But all we can do is think hard about variables that may have an impact on the experiment and pre-emptively try to exclude them. There is no foolproof way of avoiding confounders.

Researchers make efforts pre-emptively to control for potential confounders in the experimental set-up. For example, because male spider mites from different species are indistinguishable to the naked eye, experimenters place a tiny amount of dust on their backs in mate choice experiments. Although spider mites have very poor eyesight, and results show that the dust colour does not affect choice, they are still randomized between treatments across replicates [15].

In some cases, such pre-emptive controlling for potential confounders has, in fact, turned out to be instrumental. For example, fruit flies use a combination of visual and odour cues to reach an odour source [16]. Because this is known, olfactometer studies generally randomize the side of the cues offered in olfactometers among trials, while also excluding all possible visual cues from the set-up [17]. If a confounder is found, the latter can still be dealt with statistically, but at the cost of a loss in statistical power [18].

(d) Deciding what to control and what not to control has downstream effects on experimentation

As noted, an implication of the impossibility to control everything is that scientists must select what to control in an experiment. This selection has consequences for what can be learned from that experiment [19,20]. This is not just a question of increasing the generality of the conclusions that can be drawn from an experiment (i.e. a question of external validity). The decision on which variables will be controlled for in an experiment gears the experiment towards a particular research question, while limiting other types of endeavours, much in the same way as a train, once on a particular track in a particular direction, will not be able to reach destinations that lie elsewhere.

A prominent example of how control decisions shape research outcomes is the widespread use of a few model organisms in biology. These systems are attractive for many reasons, including ease of supply and community support. One of those features is that they are highly tractable, that is, they enable a high degree of experimental control—genetic tractability, standardized conditions and reproducibility [21]. However, this tight control also limits what can be observed or inferred. For instance, *Drosophila melanogaster* has shaped research agendas in sexual selection and immune response, yet its particular mating system (last male sperm precedence [22]) and immune architecture (reliance on Toll and IMD pathways [23]) constrain the kinds of questions that can be addressed. Traits absent in this model, such as alternative reproductive strategies [24–26] or different types of defences against pathogens [27–29], may remain neglected for long time periods—not necessarily because they are unimportant, but because they are not present in the few model organisms studied by most laboratories.

The deliberate exclusion of some variables in the experimental design, while warranted within a specific project, may limit our ability to draw conclusions from the experiment. For example, the design of a maze used to study rat behaviour shapes which aspect of that behaviour can be observed. If a maze is stripped of environmental cues—such as visual, auditory or olfactory signals—the experimenter can observe how a rat behaves and gradually improves its orientation in a barren, simplified setting. This kind of set-up allows for tight experimental control and allows reintroducing individual cues one at a time to study their specific effects on behaviour. However, when maze design incorporates multiple, variable cues and takes into account ecological

assumptions about the animal's foraging behaviour, it opens the door to a different kind of insight. By simulating more realistic and complex environments, such experiments increase their ecological validity [30] and allow researchers to investigate how rodents navigate uncertainty, integrate diverse sources of information and flexibly adapt to changing conditions. As a result, the understanding of rodent cognition and behaviour that emerges from such designs is often richer and more nuanced [31].

The point here is not that one set of experiments is 'better controlled' than the other, or that one of the approaches is correct and the other is wrong [20]. Rather, each set of experiments comes with a specific set of controls, and this imposes constraints on what we can extract from the data. Each experimental system keeps certain parts of the setting 'fixed', but what is kept fixed is different, and therefore different causal relations are tested and different insights are gained from each experimental design.

3. It is not desirable to control everything

In §2, we elaborated on the commonplace that it is not possible to control 'everything' in an experiment, and distinguished different ways in which an experiment can fall short of the ideal of perfect control. Some forms of incomplete control are trivial and unproblematic (see §2a), but others are not, because they are consequential for the knowledge that can be gained from the experiment (§2d). In this section, we turn to the question of whether an experiment that is more strictly controlled is always more informative than an experiment that is less strictly controlled. We argued that this is not the case and show why it may be desirable to relax control in certain cases, rather than strictly control the parameters.

(a) The problem of overfitting

The desire to control strictly can be problematic because it may lead to overfitting of the model—statistical or experimental—to the data. Notably, an experiment is itself a model of a target system, implemented materially rather than mathematically. Overfitting reduces the generality of a model and the conclusions that can be drawn from it, and increases the risk of collinearity among explanatory variables, thus complicating interpretation [32]. In such cases, models may capture idiosyncrasies of the dataset as if they were meaningful patterns, leading to false positives—factors that appear significant but do not in fact affect the outcome of the experiment.

While a statistical model overfits when it captures idiosyncrasies of the training data, an experimental model overfits when it captures idiosyncrasies of the experimental set-up rather than stable features of the target phenomenon. In modern biology, where experiments and modelling typically work together, overfitting not only can originate at the level of the statistical model but also can be induced upstream by experimental design choices. Certain experimental designs may create conditions under which downstream models, even when statistically sound, end up fitting context-specific structure. For instance, controlling for one variable may inadvertently introduce bias by overcorrecting or masking the effects of another, particularly when the causal structure of the system is not fully understood. Although model-selection criteria such as the Akaike information criterion may help assess parsimony at the level of statistical inference, they do not address the risk of false positives, biases introduced by experimental design or dependencies among variables.

A common instance of overfitting arises in the form of batch effects in omics data, where researchers aim to isolate the biological signal from the many sources of technical variation inherent in high-throughput methods. In multi-sample transcriptomic experiments, differences in sample preparation, sequencing runs, reagent batches and even laboratory environments can introduce variation unrelated to the biological variables under study. If not properly addressed, such batch effects can become confounded with experimental factors, leading models to mistake technical variability for biological signal. A recent comparative study of batch-correction methods across transcriptomics, proteomics and metabolomics found that commonly used tools often failed when batch effects aligned with biological structure, resulting in distorted conclusions [33]. Other work has shown that models trained on such data may perform well in-sample but fail to generalize due to reliance on laboratory-specific variation [34]. These examples underscore how fitting to technical variation can limit both inference and model generality.

(b) The problem of biological complexity

The desire to isolate a system as much as possible and control parameters strictly may be counterproductive because it does not appropriately acknowledge certain unavoidable features of biological systems—arising both from their internal organization and from their interactions with other systems and environments—namely, complex causality structure within a system or between a system and its environment, stochasticity and emergent properties. After defining these concepts, we argue that acknowledging their role calls for a shift away from the pursuit of strict control and towards experimental strategies more appropriately aligned with biological complexity—suggesting that, in many cases, embracing this complexity can yield more informative and insightful outcomes (figure 1).

(i) The chain of causality within a system is complex

So far, we have dealt with limitations that still consider that different independent variables exert separate, additive effects on a particular dependent variable. However, this paradigm does not apply to more complex phenomena, as discussed within the field of ecology [35]. The acknowledgement of complex data structures has led to significant advancements in the field of statistics and its use in biology. Indeed, even standard statistical models allow coping with interactions among independent variables. In

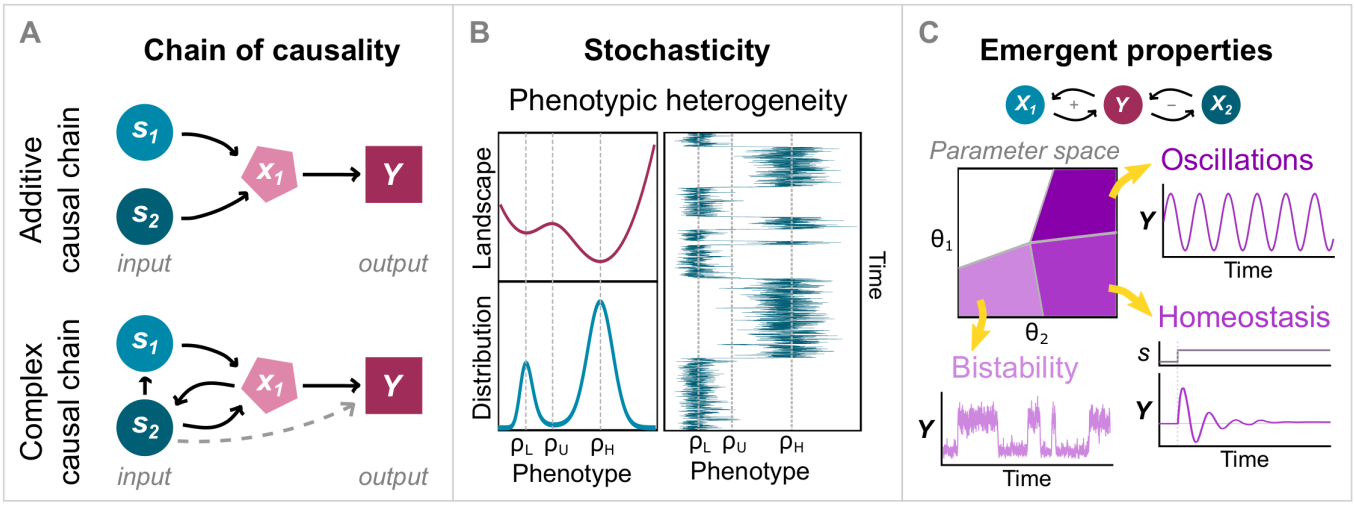

Figure 1. Conceptual illustration of the problem of biological complexity. (A) The chain of causality in biological systems is complex. In simplified models, inputs are assumed to affect an output additively, independently, and in isolation. In more realistic scenarios, variables may interact, share upstream causes or participate in feedback loops—and some interactions may remain unknown. These features challenge the assumption of independent, linear effects and call for statistical or conceptual frameworks capable of capturing complex causality. (B) Stochasticity enables phenotypic heterogeneity. In this conceptual example, intrinsic noise drives transitions between two stable phenotypes within an otherwise genetically homogeneous population. The resulting population-level distribution reflects both the architecture of the underlying regulatory system and the role of stochasticity—illustrating how noise can generate adaptive behaviour in complex biological systems. (C) Emergent properties arise from the interplay of system architecture and parameters. This conceptual example illustrates a regulatory system with both positive and negative feedback. Depending on feedback strengths, the same components can give rise to distinct emergent behaviours—including bistability, oscillations and homeostasis. These dynamics are not hardcoded in the components but arise from their interactions, emphasizing that emergence is a property of the system as a whole. Capturing such behaviours requires experimental designs that preserve—not constrain—this interactive complexity.

the last decades, the use of model selection [36], structural equation models [37] and machine learning [38] has allowed tailoring statistical models to complex data structures and refraining from assuming a particular structure in the first place. However, these advanced statistical techniques are still mostly used in datasets in which most factors cannot be controlled (e.g. in field experiments), allowing to mitigate potential confounding patterns.

We argue that the systematic attempt to harness complexity into a few independent and dependent variables may prevent us from detecting highly relevant phenomena. Indeed, even the laboratory environment itself is sometimes not a neutral bystander for the organisms that are being tested: rather, it may represent a selective force to which organisms may adapt. This has been repeatedly shown, in particular in a body of work on laboratory adaptation in *Drosophila subobscura* [39–41]. Such adaptation may thus be conflated with adaptation to other selection pressures that are being explicitly studied. The existence of control populations evolving in a control environment does not eliminate the problem, as laboratory adaptation can interact with adaptation to different environments differently [42,43].

(ii) System–environment entanglement

Target systems in biological experimentation are often entangled with their immediate environment. For instance, in experimental studies of songbird development, adopting a study design whereby birds were housed together and in a variety of housing options led to deeper insights into song learning than one in which juvenile birds were isolated in strictly controlled laboratory cages and exposed to isolated songs [5]. Also, an experimental evolution study showed that spider mites evolving on plants with cadmium had higher growth rates on those plants than control populations unexposed to that environment [44]. However, such populations also increased their sensitivity to intraspecific competition. This change led to a shift in the outcome of the interaction with a competitor, from exclusion to coexistence, which would not have occurred if only changes in growth rate had been considered. Increasing the growth rate leads to a higher density, which in turn affects intraspecific competition. One could have fallen into the temptation of actually controlling for this side effect of a higher growth rate during experimental evolution, but this would be biologically artificial, as higher density is a necessary by-product of higher growth rate in real systems (for a given amount of resources). Therefore, in this case, leaving some degrees of freedom to the evolving populations improved our understanding of this and other systems.

(iii) Stochasticity

The ubiquity of biological noise—arising from both internal processes and interactions with the environment and other systems—is now widely recognized [45,46]. It arises from the discrete nature of biological units (e.g. molecules, cells, organisms) and the stochasticity of their interactions (e.g. biochemical reactions, cell divisions, ecological encounters). Theoretical work has shown that such noise cannot be entirely eliminated [47], and both modelling and experimental studies have demonstrated that it is

not merely a nuisance—it plays a functional role in shaping biomolecular processes, phenotypic variability, population-level dynamics and even evolutionary trajectories [48]. Consequently, designing a truly deterministic experimental set-up is impossible, fundamentally undermining the feasibility of perfect control in biological research.

This stochasticity is not just unavoidable—it is integral. Relying on large sample sizes or simply reporting variance is not equivalent—those practices still assume a fundamentally deterministic system with noise 'added on top'. Instead, biological systems are often fundamentally stochastic, with deterministic behaviour better viewed as a limiting approximation. In this light, experimental design should not aim to suppress noise but to engage with it. There are two main reasons for this. First, biochemical noise can give rise to non-intuitive emergent behaviours, as discussed below. Second, the shape and structure of the noise—its distribution, correlations and temporal patterns—can be highly informative about the underlying mechanisms and interactions.

A compelling illustration comes from the theoretical work on epigenetic switching in fluctuating environments. Using *in silico* evolution of a self-activating genetic circuit, Gómez-Schiavon & Buchler [49] showed that stochastic gene expression—a direct consequence of biological noise—can enable heritable phenotypic switching without genetic mutation as an adaptation strategy. Crucially, this noise-driven mechanism outperformed genetic adaptation in fast-changing environments, not only because it enabled faster responses after environmental shifts but also because it attenuated the detrimental effects of noise on population fitness. These results demonstrate that noise is not merely a source of error but a functional element that shapes evolutionary trajectories [50]. Ignoring it risks obscuring both the mechanisms of adaptation and the evolutionary trade-offs involved.

(iv) Emergent properties

Emergent properties refer to system-level behaviours that arise not from individual components in isolation but through their interactions within the system. Classic examples include bistability (e.g. cell fate decisions; seed germination timing), oscillations (e.g. circadian rhythms; predator–prey cycles), negative frequency dependence (competition, coevolution), robustness (e.g. homeostatic tissue growth; compartmentalized structure in biofilms) and spatial pattern formation (e.g. phase separation; vegetation patterns in arid ecosystems). However, the landscape of emergent behaviours likely extends far beyond these textbook cases. Crucially, some of these phenomena—such as stochastic switching or noise-induced oscillations—emerge specifically from the interplay between system architecture and biological noise. Identifying such properties is often non-trivial, requiring not only a systems-level perspective but also an appreciation for the generative role of biological variability. An overly rigid experimental design that seeks to strictly control all variables may oversimplify the system and its interactions to the point where emergent behaviours are no longer sustained. This introduces a critical, often overlooked trade-off: the more tightly we constrain an experiment in the name of control, the more we may exclude the very complexity that defines biological systems.

A striking example of emergent behaviour comes from a synthetic gene circuit designed to activate its own expression via a non-cooperative positive feedback loop [51]. Based on the circuit's architecture, no bistability was expected. However, when implemented in *E. coli*, the system exhibited robust bistable gene expression. This counterintuitive outcome emerged from an unexpected interaction between the circuit and host physiology: circuit activation slowed down cell growth, which in turn reduced protein dilution, reinforcing the accumulation of the activator. In other words, bistability arose not from the circuit alone, but from the feedback between circuit dynamics and cellular processes—an interaction that would be invisible in a model or experiment treating the host as a fixed background. This example underscores that overly simplified or tightly constrained experimental set-ups may preclude the emergence of system-level behaviours, not because such behaviours are absent, but because the experimental context has been stripped of the interactions that give rise to them.

4. Controlling well

Given all of the above, designing controlled biological experiments often requires grappling not only with lack of knowledge about the system under study (i.e. grappling with the issues discussed in §2) but also with key biological features such as causal complexity and object/environment entanglement, the system's intrinsic stochasticity and emergent behaviours. The variety of issues precludes a one-size-fits-all concept of experimental control, and proposing one would run counter to the very spirit of the proposal to reevaluate the aim of strict control. Instead, we advocate the notion of 'controlling well', which is intrinsically context-dependent. The examples presented are intended to provoke reflection and foster a broader discussion about alternative approaches to experimental design—approaches that move away from the desire for ever stricter control and towards more realistic, context-sensitive and purpose-dependent forms of scientific rigour.

Some strategies that have successfully addressed—and even leveraged—such challenges are available. We argue that they are best framed as intentional relaxations of experimental control. For instance, instead of circumscribing an experiment to a given set of conditions, researchers have adopted the strategy of randomizing replicates across different conditions, accounting for the environmental variation in the block effect of their statistical model [52]. Still, there may be idiosyncrasies associated with the laboratory itself, such that the same experiment performed in a different environment would yield different results [53]. One should perform the same experiment in a variety of *different* laboratories and account for a laboratory effect in the statistical model [53,54].

Bayesian frameworks exemplify how probabilistic reasoning can directly engage with biological variability, treating it not as a lack of control, but as a valuable source of information. These approaches go beyond point estimates, offering full probability distributions that quantify uncertainty and reveal features of the system that deterministic models might obscure. For instance, Gómez-Schiavon *et al.* [55] used single-molecule data (smFISH) and Bayesian inference to analyse noisy single-cell gene expression data. They demonstrated that cell-to-cell variability in transcript abundance—traditionally dismissed as noise—can reveal the kinetics of promoter switching, including stimulus-dependent rates. Rather than averaging out variability (a measure of indirect

control), they used its full distribution to extract mechanistic insight—underscoring how strict control strategies that suppress variance can be counterproductive to understanding biological systems.

Another example concerns pre-clinical tests in animal models for human diseases. For instance, if the purpose is to translate laboratory results to clinical application, it may be better to experiment with 'dirty mice' (those with microbial exposures) than with standardized mouse models [4]. If the aim is to obtain reproducible laboratory results, studies should use data from several laboratories, because it has been found that studies based upon single laboratory trials are less reproducible [56]. Furthermore, Usui *et al.* [57] challenged the assumption that strict standardization enhances replicability. Through a meta-analysis of ischaemic stroke models in animals, they showed that reducing within-laboratory variability can produce results that fail to generalize. By introducing structured biological variability, they were able to identify interventions that are broadly effective across individuals, as well as those whose effects vary substantially and may require tailored application—thus providing a framework for refining experimental design. By deliberately including sample heterogeneity—a strategy they call heterogenization—they improved both the reproducibility and the ecological validity of experimental outcomes. Instead of minimizing inter-individual differences, their approach embraces these differences to uncover biologically meaningful patterns.

Ultimately, replacing the notion of 'strictly controlled' experiment with that of a 'well-controlled' experiment means shifting the focus from suppressing or averaging away variability, and thereby overlooking complexity, stochasticity and emergent properties, to understanding their roles. In complex systems, control is not about simplification, but about navigating uncertainty with purpose and sensitivity to the context of an investigation.

5. Discussion

In §2, we distinguished four ways in which the ideal of perfect control is not met in actual biological practice: the case where we disregard certain parameters because they do not have an impact on the experiment; the case where we are limited in our ability to control because of technical limitations; the case where we fail to control for a confounder (where parameters we thought unproblematic turn out to be problematic); and the conscious decision for a particular experimental design and controls appropriate to it, which excludes others.

It may seem that the problems described in §2 are just temporary technical problems that will be solved as we keep experimenting and refining our experimental designs. However, we think that there is a deeper issue underlying the fact that perfect control is not possible.

As we discussed, each set of experiments comes with a specific set of controls, and this imposes constraints on what we can extract from the data. Each experimental system keeps certain parts of the setting 'fixed', but what is kept fixed is specific to it, and in this sense, different causal relations are tested, which means that different insights are gained from different experimental designs.

Assessing which variables should be controlled is a key task of the experimental design, which often advances by trial and error but also relies on the experimenters' background knowledge about the target system. Once scientists have settled for a particular approach, they will likely continue in this line of research, not the least because time and money have already been invested. This entails a contingency in the direction of scientific research in the sense that certain other potentially promising routes are not followed. One way to mitigate the narrowing of research paths that these decisions introduce is to clarify the assumptions underlying a particular experimental design. Indeed, it would be useful if researchers were explicit about *what they choose not to control and why*, both when they design the experiment and when they write it up in a publication.

In §3, we provided several reasons why aiming for strict control may be undesirable. We first discussed how the pursuit of strict control can lead to overfitting, reducing the generality of models and obscuring meaningful biological patterns. We then argued that such efforts often fail to account for essential features of biological systems—namely, complex causal structures (including between systems and their environments), intrinsic stochasticity and emergent properties. These are not peripheral complications but defining characteristics of living systems.

One of the key implications of the discussion in §3 is that we should design experiments with—not against—those essential biological features. Too strict control of the components of target systems in biological experimentation may preclude the study of emergent properties, i.e. properties of the system that cannot be reduced to properties of the system's components. It may also lessen generalizability and obscure insights into the object/environment entanglement of systems. We thus propose the concept of 'controlling well', the context- and purpose-dependent application of control measures. What 'controlling well' means needs to be spelled out with regard to specific contexts of experimentation.

We advocate for an approach to biological experimentation that does not always demand specific, precise hypotheses. This may involve exploratory research, where the goal is not to test a specific prediction but to observe unexpected phenomena that may only become visible when experimental constraints are loosened. This type of research may lead to observations that will then be at the base of hypotheses that can be tested at later stages of inquiry. In addition, we also want to emphasize a second, less commonly recognized possibility: hypothesis-driven research that nonetheless embraces a degree of uncertainty and complexity. In this case, the hypothesis is clear and explicitly stated, but the experimental system is not artificially simplified to ensure clean, isolated measurements. This approach allows for the possibility that emergent properties—in both the everyday sense of unexpected results and the philosophical sense of irreducible system-level behaviours—may arise and be meaningfully studied.

Notably, both points run contrary to current publication practices. Papers in scientific journals typically present formalized, logical arguments—not the behind-the-scenes narrative of how the work actually unfolded [58]. Disciplinary traditions of scientific writing prescribe the order of the sections that a scientific paper should contain (traditionally Introduction, Methods, Results, Discussion) and the format and style in which the content is presented. In most cases, methods sections describe the materials,

instruments, target systems and overall design used in a study in a highly formalized manner. They rarely include details about the actual process of experimental design—such as the reasoning behind key decisions, including those consequential decisions about what to control and what not to control, the modifications made along the way, the challenges encountered or the alternative approaches that were considered and consciously excluded. In principle, electronic publications would easily allow for the inclusion of such details in the electronic supplementary material. However, it is not clear to what extent this is currently done and to what extent the supplementary sections are read. To make room for more meaningful, broader discussions of experimental control and to ensure that they are read, it would make sense to expand the methods sections in the main body of the publication. Just as importantly, reviewers would need to tolerate and even encourage more narrative styles in experimental reporting.

Strategies of relaxing control sit uneasily within the current funding and publishing landscape, which still tends to reward strictly controlled experiments aligned with the old Popperian ideal of falsification. Such studies lend themselves to familiar statistical metrics like *p*-values, effect sizes and replication. They often leave little room for alternative forms of scientific rigour—such as rich, contextual observation or sensitivity to unanticipated interactions. As a result, proposals that are methodologically conservative and narrowly focused are often privileged, even when more open-ended approaches might reveal novel or complex features of biological systems [59]. It would be useful to leave some space for (and fund!) observation within hypothesis-driven designs, for example, by routinely measuring more variables than those explicit in the hypothesis at stake. A classic example is Barbara McClintock's discovery of transposable elements—work grounded in attentive observation and anomaly-following rather than rigid hypothesis testing. As Keirns [60] argues, this kind of research might struggle to find support in today's system, despite its foundational impact.

6. Conclusions: towards 'well-controlled' experiments

In this article, we have distinguished different ways in which the ideal of perfect control is not possible to achieve in biological practice and, accordingly, different ways in which an experiment can fall short of the ideal. We have argued that scientists should aim for *well-controlled* experiments to gain a better understanding of the biological world. We do not advocate abandoning control altogether. It depends on the specific situation and how one should strike a balance between controlling and not controlling. Precisely because what it means to conduct a 'well-controlled' experiment is context- and purpose-dependent, it is important that information about design choices is provided in methods sections. With our concept of 'controlling well', we thus advocate for a mindset that acknowledges and embraces the complexity and variability of biological phenomena, the diversity of biological practice, as well as the value of explicit discussion of control measures in scientific papers.

Although we have grounded our discussion in biology—being our area of expertise—we believe that underlying issues resonate across scientific disciplines—and even beyond. In many domains, a competitive environment focused on productivity has led to the fallacy that harnessing the complexity of systems at the stage of the experimental design would produce guaranteed results. Here, we argue that this is not necessarily the case, as this approach may produce non-representative results and/or fail to identify important features of the system. Instead, we advocate for a scientific landscape that incorporates other forms of acquiring knowledge. In this light, relaxing control is not merely a methodological choice—it is also a call to preserve space for surprise, creativity and truly innovative thinking.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. No datasets were generated or analysed during the current study.

Declaration of AI use. AI-assisted technologies (ChatGPT, OpenAI) were used only for grammar checking and minor language edits. All scientific content was produced and verified by the authors, and all AI-generated suggestions were reviewed and edited for accuracy.

Authors' contributions. J.S.: conceptualization, writing—original draft, writing—review and editing; S.M.: conceptualization, writing—original draft, writing—review and editing; M.G.-S.: conceptualization, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. This work was supported by Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT UNAM; IA203524 to M.G.-S) and by ANID—Millennium Science Initiative Program—Millennium Institute for Integrative Biology (iBio; ICN17_022 to M.G.-S).

Acknowledgements. We wish to thank the History and Philosophy of Sciences group at the Wissenschaftskolleg zu Berlin for fruitful discussions and Cheryl Misak, Inês Fragata, João Picão Osório and Racha Kirakosian as well as two anonymous reviewers for helpful comments on the manuscript. Additionally, the authors want to thank Luis A. Aguilar from the Laboratorio Nacional de Visualización Científica Avanzada and Jair Santiago García Sotelo, Iliana Martínez, María Arteaga and Eglee Lomelín from the Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional Autónoma de México, Santiago de Querétaro, Mexico, for their support.

References

1. Potochnik AM, Colombo M, Wright C. 2019 *Recipes for science: an introduction to scientific methods and reasoning*. New York, NY: Routledge. (doi:10.4324/9781315686875)
2. Glass DJ. 2014 *Experimental design for biologists*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
3. Currie A, Levy A. 2019 Why experiments matter. *Inquiry* **62**, 1066–1090. (doi:10.1080/0020174X.2018.1533883)
4. Wallis JM. 2025 Relinquishing experimental control to improve translation. *Synthese* **206**, 160. (doi:10.1007/s11229-025-05219-9)
5. Desjardins E, Oswick D, Fox CW. 2023 On the ambivalence of control in experimental investigation of historically contingent processes. *J. Philos. Hist*. **17**, 130–153. (doi:10.1163/18722636-12341492)
6. Stoleru D, Peng Y, Agosto J, Rosbash M. 2004 Coupled oscillators control morning and evening locomotor behaviour of *Drosophila*. *Nature* **431**, 862–868. (doi:10.1038/nature02926)

7. Reed WL, Clark ME, Parker PG, Raouf SA, Arguedas N, Monk DS, Snajdr E, Nolan V, Ketterson ED. 2006 Physiological effects on demography: a long-term experimental study of testosterone's effects on fitness. *Am. Nat*. **167**, 667–683. (doi:10.1086/503054)
8. Andreatta G, Tessmar-Raible K. 2020 The still dark side of the moon: molecular mechanisms of lunar-controlled rhythms and clocks. *J. Mol. Biol*. **432**, 3525–3546. (doi:10.1016/j.jmb.2020.03.009)
9. Cristalli C. 2024 A 'careful examination of all kind of phenomena': methodology and psychical research at the end of the nineteenth century. In *Elusive phenomena, unwieldy things* (eds J Schickore, WR Newman), pp. 149–177. Cham, Switzerland: Springer. (doi:10.1007/978-3-031-52954-2_6)
10. Kursell J. 2024 Carl Stumpf and control groups. In *Elusive phenomena, unwieldy things* (eds J Schickore, WR Newman), pp. 125–148. Cham, Switzerland: Springer. (doi:10.1007/978-3-031-52954-2_5)
11. Würbel H. 2000 Behaviour and the standardization fallacy. *Nat. Genet*. **26**, 263–263. (doi:10.1038/8154110.1038/81541)
12. Stasinopoulos TC, Hangarter RP. 1990 Preventing photochemistry in culture media by long-pass light filters alters growth of cultured tissues. *Plant Physiol*. **93**, 1365–1369. (doi:10.1104/pp.93.4.1365)
13. Hangarter RP, Stasinopoulos TC. 1991 Repression of plant tissue culture growth by light is caused by photochemical change in the culture medium. *Plant Sci*. **79**, 253–257. (doi:10.1016/0168-9452(91)90114-n)
14. Redish AD, Kummerfeld E, Morris RL, Love AC. 2018 Opinion: reproducibility failures are essential to scientific inquiry. *Proc. Natl Acad. Sci. USA* **115**, 5042–5046. (doi:10.1073/pnas.1806370115)
15. Clemente SH, Rodrigues LR, Ponce R, Varela SAM, Magalhães S. 2016 Incomplete species recognition entails few costs in spider mites, despite first-male precedence. *Behav. Ecol. Sociobiol*. **70**, 1161–1170. (doi:10.1007/s00265-016-2124-0)
16. Saxena N, Natesan D, Sane SP. 2018 Odor source localization in complex visual environments by fruit flies. *J. Exp. Biol*. **221**, 172023. (doi:10.1242/jeb.172023)
17. Roberts JM, Clunie BJ, Leather SR, Harris WE, Pope TW. 2023 Scents and sensibility: best practice in insect olfactometer bioassays. *Entomol. Exp. Appl*. **171**, 808–820. (doi:10.1111/eea.13351)
18. Arnqvist G. 2020 Mixed models offer no freedom from degrees of freedom. *Trends Ecol. Evol*. **35**, 329–335. (doi:10.1016/j.tree.2019.12.004)
19. Schickore J. 2019 The structure and function of experimental control in the life sciences. *Philos. Sci*. **86**, 203–218. (doi:10.1086/701952)
20. Schickore J. 2025 *Controlled experiments*. Cambridge, UK: Cambridge University Press. (doi:10.1017/9781009348966)
21. Dietrich MR, Ankeny RA, Crowe N, Green S, Leonelli S. 2020 How to choose your research organism. *Stud. Hist. Philos. Biol. Biomed. Sci*. **80**, 101227. (doi:10.1016/j.shpsc.2019.101227)
22. Price CSC, Dyer KA, Coyne JA. 1999 Sperm competition between *Drosophila* males involves both displacement and incapacitation. *Nature* **400**, 449–452. (doi:10.1038/22755)
23. Hoffmann JA, Reichhart JM. 2002 *Drosophila* innate immunity: an evolutionary perspective. *Nat. Immunol*. **3**, 121–126. (doi:10.1038/ni0202-121)
24. Zuk M, Garcia-Gonzalez F, Herberstein ME, Simmons LW. 2014 Model systems, taxonomic bias, and sexual selection: beyond *Drosophila*. *Annu. Rev. Entomol*. **59**, 321–338. (doi:10.1146/annurev-ento-011613-162014)
25. Watson PJ. 1991 Multiple paternity and first mate sperm precedence in the sierra dome spider, *Linyphia litigiosa* Keyserling (Linyphiidae). *Anim. Behav*. **41**, 135–148. (doi:10.1016/S0003-3472(05)80509-3)
26. Rodrigues LR, Figueiredo ART, Van Leeuwen T, Olivieri I, Magalhães S. 2020 Costs and benefits of multiple mating in a species with first-male sperm precedence. *J. Anim. Ecol*. **89**, 1045–1054. (doi:10.1111/1365-2656.13171)
27. Iwama RE, Moran Y. 2023 Origins and diversification of animal innate immune responses against viral infections. *Nat. Ecol. Evol*. **7**, 182–193. (doi:10.1038/s41559-022-01951-4)
28. Santos-Matos G *et al*. 2017 *Tetranychus urticae* mites do not mount an induced immune response against bacteria. *Proc. R. Soc. B* **284**, 20170401. (doi:10.1098/rspb.2017.0401)
29. Zélé F, Santos-Matos G, Figueiredo ART, Eira C, Pinto C, Laurentino TG, Sucena É, Magalhães S. 2019 Spider mites escape bacterial infection by avoiding contaminated food. *Oecologia* **189**, 111–122. (doi:10.1007/s00442-018-4316-y)
30. Schmuckler MA. 2001 What is ecological validity? A dimensional analysis. *Infancy* **2**, 419–436. (doi:10.1207/s15327078in0204_02)
31. Arnet E. 2024 Controlling away the phenomenon: maze research and the nature of learning. In *Elusive phenomena, unwieldy things* (eds J Schickore, WR Newman), pp. 269–290. Cham, Switzerland: Springer. (doi:10.1007/978-3-031-52954-2_10)
32. Hawkins DM. 2004 The problem of overfitting. *J. Chem. Inf. Comput. Sci*. **44**, 1–12. (doi:10.1021/ci0342472)
33. Yu Y *et al*. 2023 Correcting batch effects in large-scale multiomics studies using a reference-material-based ratio method. *Genome Biol*. **24**, 201. (doi:10.1186/s13059-023-03047-z)
34. Yu Y, Mai Y, Zheng Y, Shi L. 2024 Assessing and mitigating batch effects in large-scale omics studies. *Genome Biol*. **25**, 254. (doi:10.1186/s13059-024-03401-9)
35. Quinn JF, Dunham AE. 1983 On hypothesis testing in ecology and evolution. *Am. Nat*. **122**, 602–617. (doi:10.1086/284161)
36. Johnson JB, Omland KS. 2004 Model selection in ecology and evolution. *Trends Ecol. Evol*. **19**, 101–108. (doi:10.1016/j.tree.2003.10.013)
37. Fan Y, Chen J, Shirkey G, John R, Wu SR, Park H, Shao C. 2016 Applications of structural equation modeling (SEM) in ecological studies: an updated review. *Ecol. Process.* **5**, 19. (doi:10.1186/s13717-016-0063-3)
38. Valletta JJ, Torney C, Kings M, Thornton A, Madden J. 2017 Applications of machine learning in animal behaviour studies. *Anim. Behav*. **124**, 203–220. (doi:10.1016/j.anbehav.2016.12.005)
39. Matos R, Pité MTR, Rego A. 1999 Adaptation to the laboratory environment in *Drosophila subobscura*. *J. Evol. Biol.* **13**, 9–19. (doi:10.1046/j.1420-9101.2000.00116.x)
40. Santos J *et al*. 2012 From nature to the laboratory: the impact of founder effects on adaptation. *J. Evol. Biol*. **25**, 2607–2622. (doi:10.1111/jeb.12008)
41. Fragata I *et al*. 2014 Laboratory selection quickly erases historical differentiation. *PLoS One* **9**, e96227. (doi:10.1371/journal.pone.0096227)
42. Matos M, Rego C, Levy A, Teotónio H, Rose MR. 2000 An evolutionary no man's land. *Trends Ecol. Evol*. **15**, 206. (doi:10.1016/S0169-5347(00)01844-9)
43. Magalhães S, Matos M. 2012 Strengths and weaknesses of experimental evolution. *Trends Ecol. Evol*. **27**, 649–650; (doi:10.1016/j.tree.2012.08.004)
44. Fragata I, Godinho DP, Rodrigues LR, Cruz MA, Zélé F, Godoy O, Magalhães S. 2025 Evolution in response to an abiotic stress shapes species coexistence. *Ecol. Lett*. **28**, e70286. (doi:10.1111/ele.70286)
45. Raj A, van Oudenaarden A. 2008 Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–226. (doi:10.1016/j.cell.2008.09.050)
46. Shoemaker LG *et al*. 2020 Integrating the underlying structure of stochasticity into community ecology. *Ecology* **101**, 2922. (doi:10.1002/ecy.2922)
47. Lestas I, Vinnicombe G, Paulsson J. 2010 Fundamental limits on the suppression of molecular fluctuations. *Nature* **467**, 174–178. (doi:10.1038/nature09333)
48. Ackermann M. 2015 A functional perspective on phenotypic heterogeneity in microorganisms. *Nat. Rev. Microbiol*. **13**, 497–508. (doi:10.1038/nrmicro3491)
49. Gómez-Schiavon M, Buchler NE. 2019 Epigenetic switching as a strategy for quick adaptation while attenuating biochemical noise. *PLoS Comput. Biol*. **15**, e1007364. (doi:10.1371/journal.pcbi.1007364)
50. Weinreich DM, Sgouros T, Raynes Y, Burtsev H, Chang E, Rajakumar S, Bravo IG, Peták C. 2025 The population genetics of biological noise. *bioRxiv*. (doi:10.1101/2025.01.11.632402)
51. Tan C, Marguet P, You L. 2009 Emergent bistability by a growth-modulating positive feedback circuit. *Nat. Chem. Biol*. **5**, 842–848. (doi:10.1038/nchembio.218)

52. Kreutz C, Timmer J. 2009 Systems biology: experimental design. *FEBS J*. **276**, 923–942. (doi:10.1111/j.1742-4658.2008.06843.x)
53. Jaljuli I, Kafkafi N, Giladi E, Golani I, Gozes I, Chesler EJ, Bogue MA, Benjamini Y. 2023 A multi-lab experimental assessment reveals that replicability can be improved by using empirical estimates of genotype-by-lab interaction. *PLoS Biol*. **21**, e3002082. (doi:10.1371/journal.pbio.3002082)
54. Lithgow GJ, Driscoll M, Phillips P. 2017 A long journey to reproducible results. *Nature* **548**, 387–388. (doi:10.1038/548387a)
55. Gómez-Schiavon M, Chen LF, West AE, Buchler NE. 2017 BayFish: Bayesian inference of transcription dynamics from population snapshots of single-molecule RNA FISH in single cells. *Genome Biol.* **18**, 164. (doi:10.1186/s13059-017-1297-9)
56. Voelkl B, Vogt L, Sena ES, Würbel H. 2018 Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLoS Biol*. **16**, e2003693. (doi:10.1371/journal.pbio.2003693)
57. Usui T, Macleod MR, McCann SK, Senior AM, Nakagawa S. 2021 Meta-analysis of variation suggests that embracing variability improves both replicability and generalizability in preclinical research. *PLoS Biol*. **19**, e3001009. (doi:10.1371/journal.pbio.3001009)
58. Schickore J. 2008 Doing science, writing science. *Philos. Sci*. **75**, 323–343. (doi:10.1086/592951)
59. O'Malley MA, Elliott KC, Haufe C, Burian RM. 2009 Philosophies of funding. *Cell* **138**, 611–615. (doi:10.1016/j.cell.2009.08.008)
60. Keirns C. 1999 Seeing patterns: models, visual evidence and pictorial communication in the work of Barbara McClintock. *J. Hist. Biol*. **32**, 163–195. (doi:10.1023/a:1004420726771)