Measurement in biology is methodized by theory ⋆

Maël Montévil[a]

[a] *Institut de Recherche et d'Innovation, Centre Pompidou 4 Rue Aubry le Boucher, 75004 Paris, France*

Abstract

We characterize access to empirical objects in biology from a theoretical perspective. Unlike objects in current physical theories, biological objects are the result of a history and their variations continue to generate a history. This property is the starting point of our concept of measurement. We argue that biological measurement is relative to a natural history which is shared by the different objects subjected to the measurement and is more or less constrained by biologists. We call symmetrization the theoretical and often concrete operation which leads to considering biological objects as equivalent in a measurement. Last, we use our notion of measurement to analyze research strategies. Some strategies aim to bring biology closer to the epistemology of physical theories, by studying objects as similar as possible, while others build on biological diversity.

Keywords: Biological measurement, experiments, evolution, systematics, strains, symmetry

Contents

1 Introduction **1**

2 Measurement in physics **2**
2.1 Classical measurement 3
2.2 Quantum measurement 3
2.3 Reference frame 3
2.4 Conclusion 3

3 A theoretical account of biological measurement **3**
3.1 Phylogenetic classification and nomenclature of biological objects 3
3.2 Observed and controlled genealogy 5
3.3 Historical contexts 6
3.4 Synchronic aspects of measurement 7
3.4.1 Current context 7
3.4.2 Choosing or eliminating individuals . . 7
3.4.3 Data acquisition 8
3.5 Irreducibility of biological variation 8
3.6 Recapitulation 9

4 Discussion **9**
4.1 The radical materiality of biological phenomena 9
4.2 Symmetry and symmetrisation 9
4.3 Measurement strategies 11

5 Conclusion **12**

⋆Published as: Montévil, M. (2019) Measurement in biology is methodized by theory. *Biology and Philosophy* 34: 35. `http://dx.doi.org/10.1007/s10539-019-9687-x`.

Email address: `mael.montevil@gmail.com` (Maël Montévil)
URL: `https://montevil.theobio.org` (Maël Montévil)

1. Introduction

Science and more specifically biology and medicine are facing a crisis where systematic attempts to reproduce experiments published in reputable journals fail in the majority of cases (Begley & Ellis, 2012; Baker, 2016). The management and organization of scientific institutions have been investigated, and the pressure to publish has been heavily criticized (Begley & Ioannidis, 2014; Lancet, 2018). In the case of experimental biology, theoretical and philosophical analyses can also play a role to understand and respond to this crisis (Nadin, 2017). There are aspects proper to biological experiments that should be analyzed systematically in light of the current understanding of living beings. This discussion is also particularly relevant now that the scientific focus on (Big) Data analyses bears the risk of forgetting that data are generated in specific empirical conditions (Leonelli, 2014). Data detached from these conditions without proper justification do not carry a genuine scientific meaning.

A scientist cannot assume that her access to reality is one of an omniscient daemon. Understanding what it means to observe natural phenomena is fundamental. This question is multi-faceted. Part of it pertains to the complementary knowledge advocated by Chang (2004), but part of the answer should be principled, in the relevant theoretical framework. We concur with Einstein's epistemological statement: "whether you can observe a thing or not depends on the theory which you use. It is the theory which decides what can be observed" (A. Einstein quoted in Salam, 1990). In physics, measurement is described in theories and is a fundamental part of their formulation. The notion of measurement embedded in theories provides a general link between the output of measurement and the theoretical and mathematical description of the objects of study. For example, measurements in classical mechanics provide approximate results while they change objects in quantum mechanics. There are many other aspects of measurement which are philosophi-

cally important; however, in this article, we aim to ground widely shared practices on theoretical principles.

Biologists often use physical concepts, and measurement is no exception. The notion of measurement of classical mechanics is widely used in biology. Moreover, Wagner (2010) and Houle et al. (2011) advocate the use of measurement theory in biology. This setting leads us to inquire whether biology requires a distinct notion of measurement. In the literature, there is at least one such account: following the informational metaphor, molecular biology often considers measurement as a classical measurement applied to *finite, entirely discrete* features: the sequences of nucleotides. A classical measurement has a limited precision, but knowing finite, discrete structures with a sufficient *finite* precision means knowing them *exactly* (Schrödinger, 1944). The same reasoning applies *mutadis mutandis* to other discrete structures such as the topology of networks (Huneman, 2018). This reasoning only applies to the discrete aspect of the objects, and not the continuous ones such as position in physical space.

This point of view is in contrast with experimental methodologies which are very rich and sometimes subtle (Weber, 2004; Kohler, 1994). In this paper, we argue that general *theoretical principles* of biology leads to a theoretical account of biological measurement which clarifies several aspects of experimental methodologies.

Measurement requires commensurability. For example, measuring the length of an object requires to identify the distance between its edges with the length of another object such as a ruler. It also requires abstract constructs: in this example not only a theory of space (or space-time) but also assumptions on the object measured. These assumptions ensure that the measurement has a meaning (Houle et al., 2011). For example, when measuring a length, is the object solid, or flexible, does it have well-defined boundaries, like a box, or not, like a cloud. As a result, measurement is never only about a single object (token). In biology, the measurement of a part or an aspect of an organism may be performed by the commensurability with a physical object, for example, the length (in meters) of this organism, here and now, measured in physical units. However, this alone is only sufficient to know if we can put it "as is" in a box of a given length. The biological meaning of a length and the procedure to assess it are very different for a tree or a snake. Therefore, we posit that biological measurement is not only about the intended part or aspect, but also has to accommodate the organism measured and its commensurability with other organisms. We will develop mostly the latter idea since it has not been systematically analyzed and raises questions which are proper to biology. To address the specificities of biological measurement and conceptualize the commensurability of organisms, we need theoretical insights on organisms.

We use the principles proposed recently for a theory of organisms (Mossio et al., 2016; Montévil et al., 2016; Soto et al., 2016a). This framework provides a conceptual continuity between the understanding of organisms and evolution. In particular, it emphasizes historical analyses both for phylogenesis and ontogenesis.

In this framework, biological objects are not defined theoretically like objects in physical theories. The theoretical definition of objects is mathematical in physics. Despite quantitative differences, the changes of a well-defined object are assumed to follow an underlying mathematical structure. Invariants and invariant preserving transformations (symmetries) define these mathematical structures (Van Fraassen, 1989; Longo & Montévil, 2014). For example, a falling stone follows the same equation during its fall despite its changes of position and velocity, and a falling log would follow the same equation. As a result, physicists can talk about the generic phenomenon of falling bodies. Physical notions of measurement apply to generic objects, and the reproducibility of physical experiments is guaranteed, at least statistically, once the same generic conditions apply.

By contrast, biological objects are historical in the sense that their organizations stem from an evolutionary and individual history and continue to produce a history. This idea has been developed theoretically and called the principle of variation (Montévil et al., 2016). To an extent, this principle is in line with earlier ideas, in particular, the contingency thesis of Beatty (1995) and the centrality of historicity defended by Gould (2002, chap. 11) in a critical assessment of the work of D'Arcy Thomson. For example, a falling tetrapod is not a purely physical notion since "tetrapod" is a biological concept. In the atmosphere, tetrapods do not just fall, some fly and others are gliders. All these behaviors require different equations, and these changes of equation depend on the underlying evolutionary history. This basic example illustrates the general idea that biological objects should not be conceived as generic and are prone to more profound changes than objects in physics, including the appearance of new possibilities (Montévil, 2018). Moreover, biological objects are contextual in the sense that their organizations depend on their past and current contexts. In other words, describing biological objects does not just involve many quantities, but quantities which are endowed with different biological meaning, and new relevant quantities can appear over time.

In a nutshell, biologists manipulate objects which are understood theoretically as the result of a history and continue to produce a history: diachronic objects. With these ideas stemming from the theory of evolution in mind, experimental reproducibility is not a straightforward notion. Biological objects tend spontaneously to vary whereas perfect reproducibility, even statistically, would require fixed physiology and development, at least at an abstract level.

In section 2, we introduce how several physical theories define measurement and the epistemological and theoretical roles this notion plays. Section 3 discusses the theoretical nature of biological measurement. Biological measurements accommodate natural histories and contexts, not just quantities. Section 4 explores several implications of our framework. In particular, we classify different research strategies to handle biological measurement.

2. Measurement in physics

In order to exemplify our aims in biology, we discuss briefly how the main physical theories conceptualize measurement. We are interested in measurement considered *in principle* in general

theoretical frameworks and not in specific experimental situations. For the theory, what does "obtaining quantities" in experiments or observations means? These accounts are sufficiently general to be valid for *any* practical situation in the corresponding theory, and they have deep practical and theoretical consequences.

2.1. Classical measurement

In classical mechanics, a system has a pointwise state in the space of possible states. The empirical access to this state is approximate: a measurement has a finite precision, ϵ, which can *in principle* be arbitrarily small. Thus, the state of a system is a point, and the result of the measurement is an interval. Classical measurement is a *metrical* notion: it stems from the concept of distance.

Classical dynamics are deterministic, but measurements may or may not allow to predict the subsequent trajectory. Unpredictable dynamics such as chaotic dynamics are called sensitive to initial conditions. The notion of measurement articulates determinism and randomness in the sense of theoretical impredictability (Gillies, 2012; Longo & Montévil, 2017). This example shows that a simple notion of measurement can have far-reaching conceptual consequences.

2.2. Quantum measurement

In quantum mechanics, measurement involves the commensurability of a microscopic object and a macroscopic object. Quantum measurement changes the object and leads to quantum randomness. Informally, a quantum state can be decomposed *for a given measurement* as the superposition (the sum) of different states called eigenstates. Each of them corresponds to a single obtainable result. Performing the measurement means that the state of the system becomes an eigenstate associated with the quantity obtained. The other eigenstates in the initial superposition disappear irreversibly. Quantum measurement has an *algebraic* (and geometric) nature.

There is an internal coherence to this notion. Performing the same measurement twice in a row will lead to the same result because the state of the object is already an eigenstate associated with this result: the result obtained in the first measurement is the only possible outcome in the second.

Different observables do not necessarily lead to the same decomposition. An eigenstate which corresponds to a specific position, for example, does not correspond to a specific velocity and the other way around. Then, measuring the position, measuring the velocity and measuring the position again will not necessarily lead to the same position twice. Lastly, some authors argue that, in an experiment, a measurement is needed to put the system in a known initial state (Mugur-Schächter, 2002). The typical theoretical structure of an experiment is then: measurement, time evolution (Schrödinger equation typically), measurement.

2.3. Reference frame

Experimenters choose space-time reference frames arbitrarily to represent concrete situations and describe features such as positions quantitatively. Relativity (Galilean, special and general) states how the description of a situation in one reference frame can be transformed into the description of the same situation in another reference frame and ensures that these descriptions are coherent. This concept overcomes the arbitrary choices of reference frames, and its mathematical nature is geometric.

2.4. Conclusion

The concepts of physical measurement we described are principles in their respective theory, and they are very different. Their common point is that they all describe the role of the experimenter and its instruments in an abstract and very concise way.

3. A theoretical account of biological measurement

To describe our theoretical notion of biological measurement, we rely mainly on the principle of variation (Montévil et al., 2016). This principle builds on evolutionary biology and states that biological objects can vary in a stronger sense than objects described by physical theories. The latter change, but physicists understand their changes by underlying stable mathematical structures. Instead, biological variations in the strong sense require changing mathematical structures. Biological objects are formed by a cascade of such variations and the notions of historicity and contextuality become fundamental. Figure 1 summarizes this perspective which guides our analysis of biological measurement.

In physics, objects can be highly simplified and remain relevant for physics. For example, it is sound to study a material composed only of iron. In biology, this is not the case. For example, looking at one or several molecules alone pertains to biochemistry, not biology. In biology, the measured features of organisms or cells, such as the concentration of molecules or the shape of tissues, are measured *in* organisms or cells and are generally produced and maintained by them (Mossio et al., 2016; Montévil & Mossio, 2015). Therefore, our discussion of biological measurement is not limited to the parts observed *per se*. Instead, our approach of measurement accommodates both the parts observed and the organisms associated. Both are reported carefully in empirical studies, and we posit that they are elementary aspects of biological measurement. This section may be seen as the theorization of a typical "method section" in any experimental paper in biology.

3.1. Phylogenetic classification and nomenclature of biological objects

Reporting a biological measurement starts with describing the organisms observed and naming them. The theoretical and philosophical underpinnings of these names are an essential aspect of biological measurements. The standard, general way to name organisms is to use systematics. Biologists always use this method, even though other methods can complement it, as discussed in the following section.

We want to emphasize two aspects of this method that impact the concept of measurement. The first is the definition of the names themselves and the second is the phylogenetic classification of living beings (de Queiroz, 1992; Lecointre & Le Guyader, 2006).

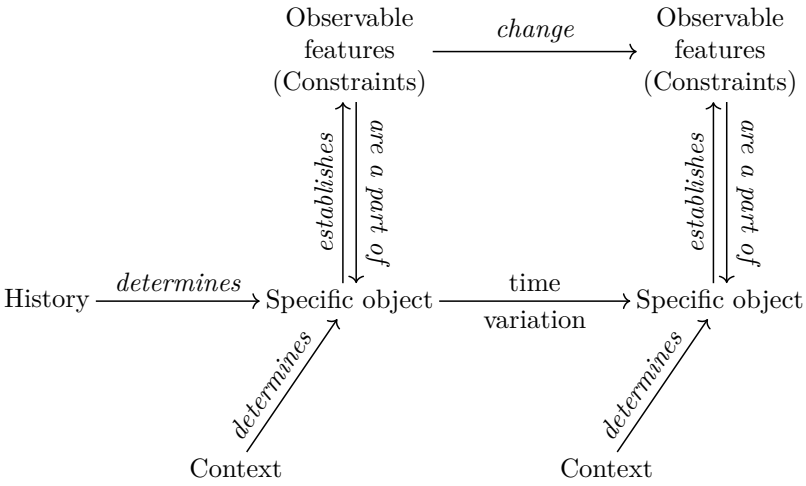

Figure 1: *Theoretical structure of biological objects*, after Montévil et al. (2016). In biology, organisms are not described theoretically by invariants and invariant preserving transformations (symmetries) which would provide a generic meaning to the features observed. Instead, their regularities are constraints that come from an history and collectively maintain each other in a given context. These constraints can change over time as the objects continue to generate a history over physiologic, developmental and evolutive time scales. An account of biological measurement has to accommodate simultaneously the measured aspects (constraints) and the rest of the organism which we describe as a specific object.

In order to provide stability to the meaning of the names used to describe living beings, systematics establish and follow strict rules to describe new species and other clades (e.g., genus, family). Nomenclature codes use the principle of typification. Typification means that defining a name requires a type. For example, the definition of a name at the family level requires a genus-level type, a genus-level name requires a species type, and describing a new species (or subspecies) requires referring to one specimen (holotype) or several specimens (syntype) which are kept in a collection (CZN International, 1999; McNeill et al., 2012, art. 72.3 and 40 resp.). Typification ensures the stability of the definition of names even if the classification changes. Name-bearing types are required to be in a biologically inactive state and thus are fixed reference objects (McNeill et al., 2012, art. 8.4).

Typification implies that the definition of biological names ultimately depends on specific, static, material objects (Grandcolas, 2017). This situation is in contrast with the theoretical definitions in the International System of Unit based on physical theories. For example, a meter is the distance traveled by light in vacuum in 1/299792458 seconds. This definition refers to matter but does not need the conservation of a specific object. Instead, it uses the generic, theoretical object called "light in the vacuum" which has an invariant velocity in both special and general relativity[1].

Names associated with specific material objects (types) are not sufficient for scientific practices. In order to endow names with a more general meaning, systematics uses the phylogenetic classification method (de Queiroz, 1992; Lecointre & Le Guyader, 2006). This method classifies living beings by estimating their genealogy. The genealogy is a theoretical concept that stems from the theory of evolution; however, the genealogy of current organisms spans billions of years, and human observers cannot access it directly. As a result, the phylogenetic classification uses different concepts than a genealogical tree. For example, it is impossible to determine whether a fossil species is an ancestor of a current species, but it is possible to establish that they are closely related genealogically. The phylogenetic method distinguishes a theoretical level and an observable level which is reminiscent of the distinction between a state and what can be observed in physics.

The phylogenetic classification assesses the evolutionary proximity between different organisms. Systematists start with the characters characterizing the different organisms, including DNA sequences. These characters are used by a computational method which provides a nested hierarchy of groups, see figure 2. These methods typically assume that the most likely situation minimizes the number of evolutionary changes, and in particular the appearance of novelties. These analyses lead to classifications where acceptable groups, called monophyletic groups or clades, are the descent of a common theoretical ancestor. The classification can then be used for taxonomic purposes. Of course, evolutionary reasonings guide the choice of the characters and the computational method used, and these choices are commonly debated.

Clades are defined by their estimated historical origin and not by their current ecological status or physiology. Since the definition of clades is based on a historical analysis, it accommodates the diversity and diversification of living beings straightforwardly. For example, the famous goat (*Capra aegagrus hircus*) discussed by West-Eberhard (2003) is a paradigmatic example of developmental plasticity because it is bipedal: a significant change occurred in a single specimen. Despite its peculiarities, this specimen is still part of the subspecies *C. aegagrus hircus* because the subspecies is defined by its historical origin, not by its

[1]Historically, the definition of a meter has first been theoretical, then it used a standard prototype. The current definition is again theoretical.

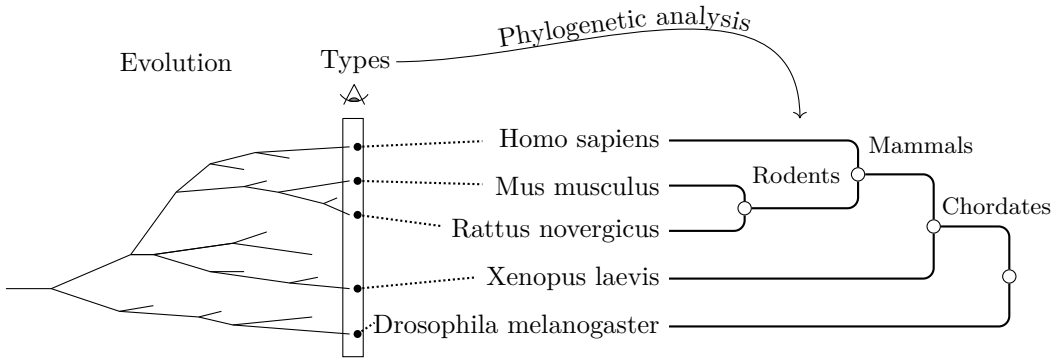

Figure 2: *Principle of the phylogenetic classification.* Left : a schematic representation of the genealogy of a few species over evolutionary timescales. This genealogy is not observable as such. Middle: the consequence of evolution is the presence of diverse life forms, some of which are used by biologists as types. Name-bearing types formally define names. Names are then extended to the specimens of the same species. Right: the characters that the specimens share and do not share are used to assess their evolutionary proximity with a mathematical model of evolution. Acceptable groups are defined as the descent of a theoretical common ancestor and lead to a classification.

properties.

Biological observations typically refer to a specific clade, usually a species or subspecies. By definition of a clade, this only ascertains a given shared theoretical ancestor. This common past involves similarities between the specimens studied, but it does not guarantee that the properties of interest in a given investigation will be similar or even exist.

3.2. Observed and controlled genealogy

The design and description of typical biological experiments use genealogical elements that go beyond what systematics can provide. Genealogical knowledge is provided by the direct observation of the lineages leading to the specimens studied and can be more or less comprehensive. Of course, direct genealogical knowledge is limited to the historical period where biologists follow the appropriate methods, that is to say, about a century at best.

Usually, direct genealogical knowledge goes with more or less control over the genealogy. In the case of organisms reproducing sexually, there are two main strategies to control genealogies: establishing inbred or outbred strains, see figure 3A. Inbred strains stem from several generations of inbreeding. By enforcing this behavior, biologists aim to obtain a genetically homogeneous population. Inbred strains still change over time at least as a consequence of genetic drift. These changes lead to the definition of substrains that have biologically relevant differences and are not interchangeable (Simpson et al., 1997). By contrast, outbred strains aim to maintain heterozygote populations while keeping as much genetic homogeneity as possible. These strains are more genetically labile than inbred strains and are often considered more variable phenotypically (Chia et al. 2005; Festing 2014, however see Jensen et al. 2016).

A specific nomenclature for strains completes the nomenclature deriving from systematics. For example, a widespread strain in biomedical research is the inbred mouse strain C57BL/6 (Black 6) (Festing, 2014). Naming strains to report an experiment includes the breeding institution. For example, C57BL/6NCrl are Black 6 mice from the National Institutes of Health (N) and which are bred by Charles River Laboratory (Crl) (Sacca et al., 2013).

The choice of strain can profoundly impact experimental results. For example, Black 6 mice have singular features such as their nociception (sensation of pain) (Mogil et al., 1999). Isaacs (1986) tested the incidence of tumors in rats exposed to the carcinogen DMBA and found that this incidence is 0%, 15%, 40% and above 90% depending on the strain used. The sensitivity to endocrine disruptors also depends on the strain (Spearow et al., 1999).

In the case of cells, the situation is overall similar to the case of animals. Cell lines and sub-lines are established, named, and exchanged between laboratories. For example, the first laboratory immortal human cell line, the HeLa cell line, originated from a single patient, Henrietta Lacks (who died in 1951) and thus HeLa cells have a common origin. This cell line is widely used, and more than 99000 references in PubMed mention it (08/2018). Cell lines have two specificities (fig. 3B). First, a single cell can originate a clonal population in common cases. Second, the use of frozen samples enables biologist to "stop" biological time. Biologists use these operations to obtain populations of cells that are far closer genealogically to their common ancestor than cells which would be proliferating with variations in culture.

Both animal strains and cell lines can be modified for research purposes, either by artificial selection for a specific trait or by genetic engineering, a subject extensively discussed by Kohler (1994) in the case of *Drosophila melanogaster*. These modifications are not only aiming for a specific new trait; they include ruling out animals with spontaneous, problematic mutations.

It is standard practice to communicate live sample between research laboratories or between breeding institutions (Kohler, 1994, chap. 5). Communicating live samples is required for biologists to ensure that the specimens studied in different laboratories are close genealogically and carry the same spontaneous or artificial changes if any. Commitment to perform these exchanges is required to publish in many journals. Replicating an experi-

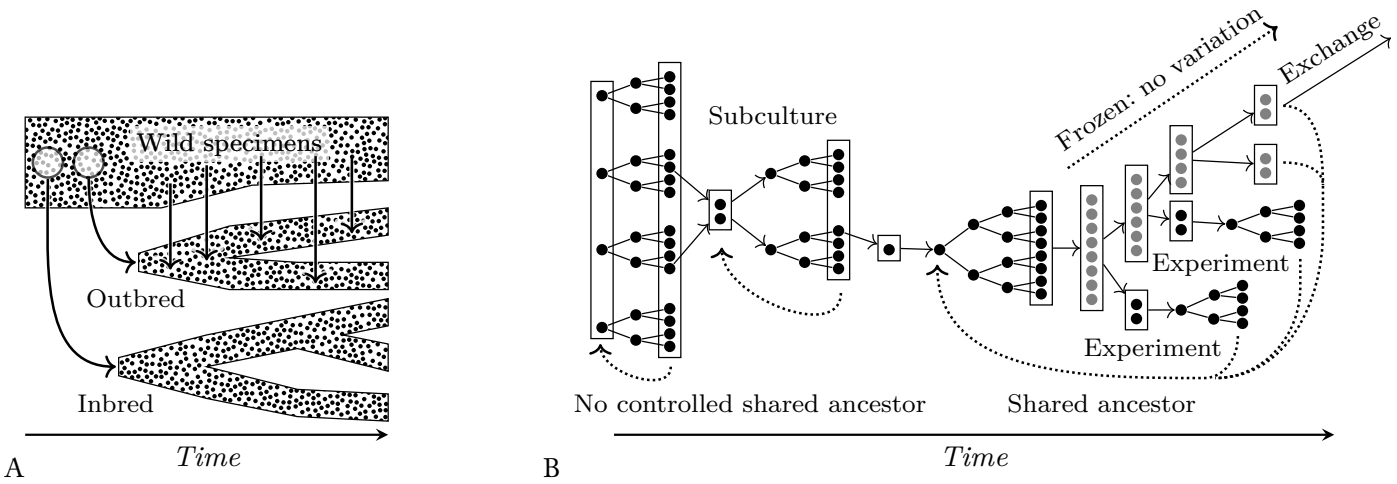

Figure 3: *Observed and controlled genealogies.* A: *A schematic representation of strain breeding.* Biologists use wild or domesticated specimens to start controlled strains. In the case of inbred strains, there is no crossing with specimens external to the strain. In the case of outbred animals, some diversity is regularly introduced. Substrains may be defined, either because they are the result of genetic manipulations, selection in outbred strains, or just as the result of genetic drift. B: *Controlled genealogies in the case of cells.* Doing a standard subculture is not enough to ensure that the individuals of a population share a recent common ancestor. To ensure a recent common ancestor, biologists typically perform a highly diluted subculture which isolates a single cell. This cell and its descent proliferate, and their proliferation leads to a new population. This population can then be frozen in order to stop biological processes, and in particular to stop proliferation and the associated variations (Soto et al., 2016a). Subsets of this frozen population can be used to perform experiments and be shared with other laboratories. The cells obtained using this method share a known, recent common ancestor and are often used to reproduce experiments.

ment using specimens from a controlled genealogy requires an exchange of matter, a point that we discuss in section 4.1.

Genealogies are not limited to cell division and sexual reproduction. Viruses lead to horizontal transfers, biologists use a diversity of manipulations, such as chimera obtained by the fusion of different zygotes. Last, some authors consider that microbiomes should be considered as parts of organisms which implies that several lineages come together to form a holobiont (Gilbert, 2014). These examples are beyond the basic concept of genealogy but fit a broader concept of genealogy *sensu* the historical origins of specimens.

The use of controlled strains and cells lines is not universal in biological experiments. For example, cells may come directly from recent human samples, and animals may come from captures in the wild. However, the practice of using sometimes very tightly controlled genealogies is widespread, in particular in biomedical research. The active control of genealogies, including modifications, leads to a situation where the natural history of the specimens is entangled with the human history of biological sciences (see Kohler, 1994, for a discussion in the case of *D. melanogaster*).

The knowledge and control over part of the recent genealogy of the specimens experimented upon is a supplement to the phylogenetic method of classification. It ensures that the specimen studied have a recent shared past. Even though this control is tighter than with the classifications of systematics alone, the same theoretical and philosophical limitations apply: the description is historical and does not ensure that the specimens have the very same organizations. Nevertheless, several methods provide partial control over biological organizations. For example, inbred strains are (almost) homogeneous genetically, and some aspects of animal phenotypes are controlled regularly in breeding institutions. Thus, these methods provide precise knowledge and control over the historical origin of the specimens studied and limited direct control over their organizations.

3.3. Historical contexts

Knowledge and control of the past of organisms and cells used in an experiment are not limited to their genealogy. Their past contexts are also relevant. By context, we mean the environment, including the possible interactions with other organisms. The control of past contexts can go from the timescale of many generations to the timescale of ontogenesis or even the shorter time scales preceding the experiment.

In the case of cell culture, the control and knowledge of the context stem first from the use of a standardized medium, temperature, and protection from contaminations. Even the choice of supplies such as centrifugal tubes used with the medium can have dramatic consequences on cellular behaviors (Soto et al., 1991). Another critical parameter is the density of cells. When this density is too low, the lack of quorum effect can change cellular behaviors. On the opposite, when the density is too high, the cells constrain each other's proliferation. Moreover, cells typically need time to adjust to a change of conditions such as a change of medium (Longo & Montévil, 2011b). All these factors are important since they determine the status of the cells subjected to the experimentation. In order to perform controlled experiments, experimenters choose an initial status that can be obtained consistently in a cell population (homogeneity) and different replicates (reproducibility). The most straightforward condition that can be obtained and sustained consistently is unconstrained proliferation.

The case of animals is similar to the case of cells. In laboratory conditions, the control of the context includes typically the

temperature, light cycle, the nature and quantity of food, avoiding pathogens, and the number of animals per cage. For example, Heindel et al. (2015, section 2.6) describe the context in which animals are raised before and during a large scale experiment. However, their past context can be considered problematic. This work aims to study the effects of the endocrine-disrupting chemical bisphenol A (BPA). The animal experimented upon are raised in BPA free cages, but they originate from strains which are raised in polycarbonate cages by the animal provider, and polycarbonate leaks BPA. The exposure of pregnant females to BPA have known effects spanning two generations ("grandmother effect", Susiarjo et al., 2007) and there are other known and probably also unknown epigenetic factors.

Understanding the importance of past contexts requires a short theoretical discussion on heredity. Under the assumption that DNA sequences are the only form of heredity, contexts before an experiment are relevant only during development. However, this assumption is not valid in general, and epigenetic inheritance is a widespread phenomenon (Jablonka & Raz, 2009; Jablonka et al., 2014; Danchin et al., 2019). Let us introduce a simple example that does not require recent advances in epigenetics. MMTV is a retrovirus which can be inherited exogenously from the milk of an infected host to another animal, usually its descent (Dudley et al., 2016). If, say, inbred mice are fed milk from contaminated mice of another strain, then these mice will carry MMTV and transmit it to their descent. A contaminated female will lead to a substrain which is genetically identical to the original inbred strain (as long as the retrovirus does not alter mice DNA) but has critical immunological and oncologic differences.

Many strategies such as working with inbred strains or clonal cell populations strive for genetic uniformity. These strategies could be extended formally to known forms of epigenetic heredity. However, the knowledge and control of past contexts over several generations is an *indirect*, partial way to control known and unknown epigenetic heredity, in combination with the control of genealogies. As a conclusion, past contexts over several generations are relevant.

The context at the timescale of one generation is also relevant, as advocated by the concept of ecological developmental biology (Gilbert & Epel, 2009). Even the position of a fetus relative to its male and female siblings in the uterus has a measurable impact (Ryan & Vandenbergh, 2002). The context matters at shorter timescales too. For example, to measure heart rate or blood pressure on a rat, biologists need to take into account the memory and anticipation associated with the procedure (Longo & Montévil, 2011b; Nadin, 2017, for conceptual frameworks). In this particular case, the stress induced by the measurement impacts the heart rate and can be limited by training the animal, that is, changing its anticipations (Gross & Luft, 2003).

The context in which organisms and cells live before the experiment matters from the timescale of several generations to the timescales of development and physiology. The work on past contexts complements the one on genealogies as a method to manage the past of the specimens studied. It follows that the same epistemological limitations apply.

3.4. Synchronic aspects of measurement

The aspects of measurement discussed above are mostly diachronic: they pertain to the past of objects. By contrast, this section analyses aspects relevant during the observation of intended features.

3.4.1. Current context

Overall, the discussion in the previous section applies also to the context during an experiment. The context contributes to the definition of the specimens and quantities observed. This contribution is both practical and theoretical. It is practical because it describes the necessary operations required to perform the same measurement beyond using the same apparatus and reading its results. It is theoretical because the meaning of the results depends on these operations.

To illustrate the importance of the context, let us consider the example of mammal metabolism observed by the oxygen consumption rate. This rate seems to be a simple empirical quantity; however, it depends on the activity of the organism observed and its relevant components. To compare the metabolism of different organisms, biologists define different kinds of physiological activity. The target activities have to be meaningful and achievable for all the organisms considered, which may be difficult when measurement applies to the many different species of a large clade. In all cases, the meaning of the results depends on the nature of the activity chosen (fig. 4). Metabolic rates have several definitions (Longo & Montévil, 2014, chap. 2 for a review):

- The field metabolic rate (FMR) corresponds to the activity of organisms in an ecosystem, without constraints from the observer (fig. 4E1).
- The basal metabolic rate (BMR) considers organisms at rest, that is to say, undisturbed, non-sleeping organisms in a thermoneutral environment and in a post-absorptive state. Evolution leads to a diversity in the activities of organisms and the BMR levels down the impact of this diversity on the metabolism (fig. 4E3). It is not always possible to instantiate this definition; for example, ruminants are never in post-absorptive state (fig. 4E2).
- The maximum metabolic rate (MMR) considers the maximum level of sustainable activity. By focusing on the upper boundary of the metabolism, only the determinants of this boundary are relevant and not the various characters involved in biological activities (fig. 4E4).

By choosing different contexts, biologists co-determine what is observed even when the same measurement apparatus is used to observe the same part. The BMR and MMR show that it is even possible to choose observations that focus on properties shared by different species by leveling down the weight of the organizational diversity stemming from history.

3.4.2. Choosing or eliminating individuals

Filtering of individuals is a method to control strains: breeders disregard animals with deleterious mutations, diseases, or other peculiarities. Sometimes, only minimal control over the

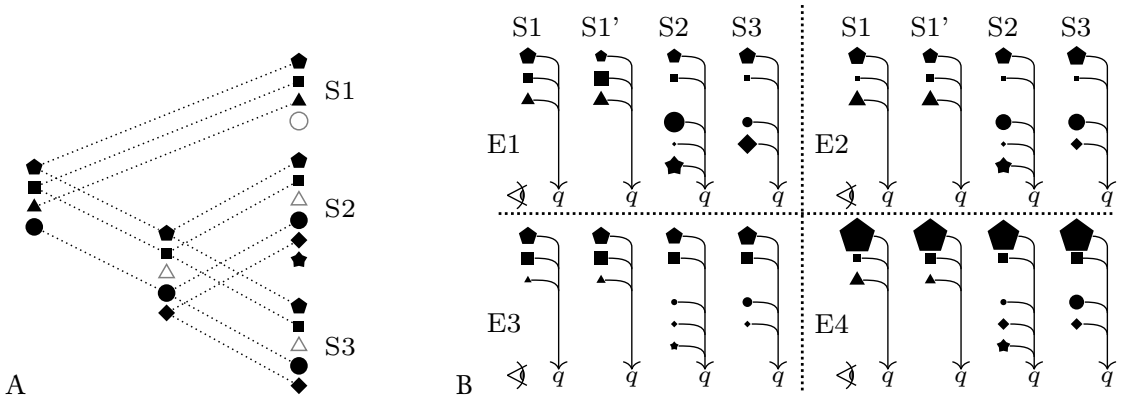

Figure 4: *Different measurements of the same quantity.* A: *A schematic representation of the appearance and disappearance of relevant characters.* Dotted lines represent relations of homology. White shapes are characters which disappeared. B: *Four different ways to measure a quantity q.* S1 and S1' are two similar specimens. All represented characters impact q. The size of a symbol represents the impact of the corresponding character on q in the given context. E1: A measurement performed without specific care for the characters contributing to q, e.g., the field metabolic rate. E2: A measurement performed in a standardized way for S1 but not for the other species. E3: The animal performs no specific activity which reduces the weight of several characters, e.g., the basal metabolic rate. In this case, only homologous characters remain quantitatively relevant. E4: A constraint dominates the determination of the measured quantity despite the diversity of relevant characters, e.g., the maximum metabolic rate.

past context and genealogy is possible. For example, in humans, most methods above would be unethical. Choosing individuals having specific characteristics and eliminating individuals with unwanted characteristics is an alternative method of control on the organisms investigated.

Filtering of individuals is possible during experiments; however, it impacts the meaning of the results. For example, in the case of a toxicological experiment, unexpected variations should be reported since they may be relevant to understand the effect of the chemical studied and may be investigated in other studies. However, if we want to study the "normal" physiology of insulin after long-term exposure to high-sugar diet, then it is necessary to rule out diabetic animals. Last, the quantities of interest cannot be measured at the expected time point in the case of individuals who meet an untimely death, which is an uncontrolled filter.

Filtering of individuals by their properties is a complementary way to control biological objects. Performing this filtering enables biologists to discard specimens which have gone through unwanted variations, or which have not gone through expected variations. Criteria can range from developmental anomalies, mutations, pathologies to animals frightened during measurement.

3.4.3. Data acquisition

Biological measurements typically provide quantities, and this process has an anhistorical dimension that is comparable to physics. The notion of measurement of classical physics is relevant in biology. When measuring a continuous quantity such as the velocity, the measurement is never exact and provides an interval instead of a single quantity (§ 2.1). Other physical notions such as reference frames can also be relevant. Wagner (2010) and Houle et al. (2011) fruitfully import concepts of measurement theory in biology which are relevant for the synchronic aspect of measurement. Since these aspects are not properly biological, we will not develop them further here.

The principle of variation implies that an observed feature can become ill-defined or acquire a different meaning. Here, biology goes beyond standard measurement theory since the changes of biological objects lead to a collapse of the original meaning of the quantities observed. For example, the heart rate is defined by beat-to-beat intervals, but pathological situations such as torsade de pointes escape the standard definition of a heartbeat, and the notion of heart rate becomes ill-defined. Similarly, the properties of the hind legs of the bipedal goat discussed above have a different meaning than in its quadruped counterparts.

Last, most experimental protocols in biology use control groups which are not subjected to the transformations investigated (Johnson & Besselsen, 2002). Control groups enable experimenters to assess the organization of specimens having the same historical origin and exposed to the same context than the organisms subjected to a putative difference maker. Controls enable biologists to estimate whether the results stem from the context, spontaneous variations, or conditions tested. Biological objects are labile, and control groups are the closest reference point possible to the objects tested.

3.5. Irreducibility of biological variation

Despite the use of methodologies providing tight control over biological objects, the principle of variation entails that there are always possible qualitative variations. Variations can impact the observed features directly, making them variable, changing their meaning or even possibly making them ill-defined. Populations which are too similar are evidence of malpractice (Bolland et al., 2016). When observing a given feature among several specimens, biologists report "not applicable" (NA) for a specimen when qualitative variations are too significant. For example, pathological heartbeats that do not follow the same sequence of events that regular heartbeats lead to beat-to-beat intervals that do not have the same meaning. This kind of departures appears for theoretical reasons and not only as a result of experimental errors or

as the result of the improper theoretical definition of the target quantities.

Observable, qualitative variations can be shown experimentally even for clonal cells, for example as a result of asymmetries in cellular division (Cai et al., 2006; Stewart et al., 2005; Lindner et al., 2008; Soto et al., 2016a) or for dynamical reasons (Braun, 2015). Of course, the development of multicellular organisms also leads to a high level of variations. Variations occur even when comparing an individual with itself at another time point, even in the case of close time points. For example, many physiological time series are non-stationary (West, 2006; Longo & Montévil, 2014). Stationary time series follow the same distribution over time which implies that the mean is a stable quantity. By contrast, non-stationarity implies that assessing the average at different times will not necessarily yield the same results. As a consequence, it is not possible to characterize an organism by precise values of physiologic quantities, and precise results are only valid at a specific time point.

3.6. Recapitulation

To sum our theoretical approach up, biological measurement has to accommodate simultaneously the aspect observed and the organism in which it takes place. We propose the following principles :

1. Measurement has a synchronic dimension for the aspect or part of interest (§ 3.4.3). Usually, the concept of measurement from classical physics is relevant, that is to say, measurement as limited precision. Concepts of measurement theory can also be used (Houle et al., 2011).
2. The measurement is relative to/constituted by the history and contexts of the organisms of interest. Historicity, here, means a cascade of context-dependent, qualitative variations. A measurement includes a specific way to manipulate and describe these contexts and natural histories, for example, referring to a theoretical or concrete common ancestor.
   (a) *Genealogy* handles an uncontrolled history that is *shared* by the different organisms studied. Methods include the phylogenetic classification (§ 3.1) and direct genealogical control in the case of strains and cell lines (§ 3.2).
   (b) *Past and current contexts* (environment/interactions) can be (partially) known in the field or controlled in laboratories or breeding institutions. Relevant contexts include past contexts over several generations, during the development or shortly before observations (§ 3.3), and current contexts, during the experiment and observations (§ 3.4.1).
   (c) *Choosing or eliminating individuals* can be used to observe or eliminate specific histories or variations (pathological cases, unwanted behavior, ontogenetic or phylogenetic histories, etc., § 3.4.2).
3. Uncontrolled variations can always impact the measurement, including the very definition of the features observed (§ 3.5).

4. Discussion

4.1. The radical materiality of biological phenomena

The role of matter in experiments is critical to their epistemological analysis (Morgan, 2002). In physics, theories define objects mathematically, by invariants and invariant preserving transformations. This epistemological structure justifies that the same theoretical object can be instantiated independently *de novo*. For example, the speed of light in the vacuum can be assessed on two independent light beams: it is an invariant of the theory. By contrast, biological objects stem from an history. It follows that empirical knowledge in biology cannot be abstracted from concrete material objects (tokens) materializing this history. In this perspective, biological phenomena display a radical materiality (Soto et al., 2016b). Our discussion on biological measurement illustrates this idea. Biological names, in systematics, are not defined by a theoretical construct, they are defined by specific specimens called name-bearing types (§ 3.1). Then, experimenting with individuals of a species associated with this name means experimenting on individuals which descend from an ancestor shared by both the specimens experimented upon and the name-bearing type. These specimens possess a diachronic, material continuity over time: the genealogy. The same reasoning applies to the controlled strains and cell lines; the exchange of living specimens between laboratories is the further materialization of this philosophical idea.

In general, we can distinguish between different categories of theoretical situations in the relationship between matter and theoretical definitions. The methods to reproduce observations characterize them:

1. The description of objects is generic, and the same theoretical object can be instantiated empirically twice without communication of matter, as discussed by Feynman & Gleick (1967, chap. 4). It is the case in current physical theories.
2. The object's behavior is the specific result of a history.
   (a) Scientists use the permanence of the material object studied. They may be fixed artificially. For example, the name-bearing types used in systematics serve as static references for future observations. They may also continue to change over time, for example, in the case of the biosphere.
   (b) Scientists use the fact that the objects reproduce. This property provides an exponential amount of objects sharing a common past. The study of living organisms and cells falls typically in this category (case studies such as types above are an exception).

4.2. Symmetry and symmetrisation

Symmetries play a central role in physics (Feynman & Gleick, 1967; Van Fraassen, 1989; Bailly & Longo, 2011; Longo & Montévil, 2014) and will enable us to provide a more in-depth analysis of biological measurement. Symmetries are transformations which do not change the relevant aspects of a given object. For example, the equation describing free fall does not change for an experiment performed one century ago or today. Time

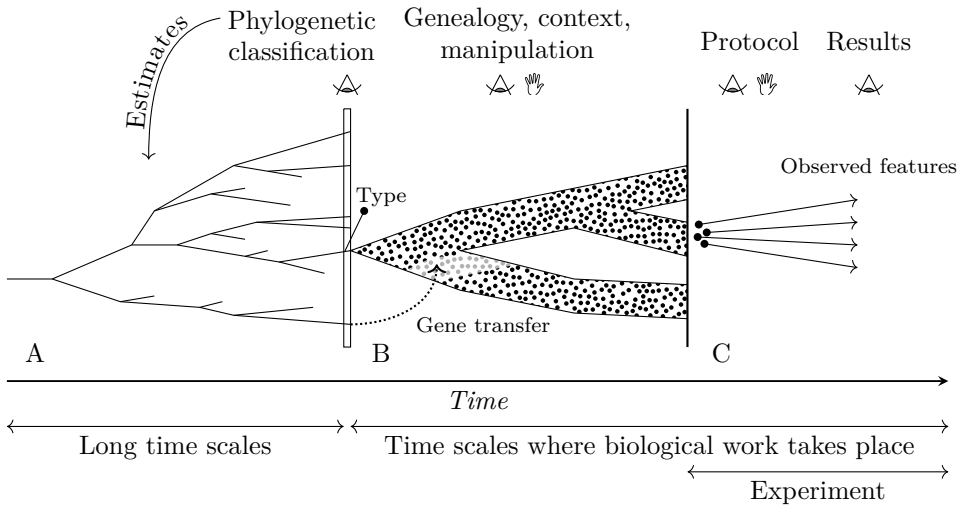

Figure 5: *Recapitulation of the diachronic elements used to define the objects of a typical experiment.* The whole construct illustrated is required to describe the measurement performed. A: The objects are the result of an evolutionary history, which is not directly accessible but can be estimated by the phylogenetic method. B: Specimens of a given species can be used to breed a strain in controlled conditions. C: Elements of this strain are used in an experiment to obtain data.

translation is a transformation that does not change the theoretical description of the object: a symmetry. Moreover, the same equation applies regardless of the nature of the object which is another fundamental symmetry. For example, concerning free fall, if experimenters replace an iron bead with another one, or a copper or wood bead, the phenomenon remains the same: permuting (interchanging) these objects is a symmetry. Symmetries can be either exact, in the sense that they stem from fundamental principles, or approximate.

The concept of experimental reproducibility is a notion of symmetry. The reproducibility of an experiment means that the same set of observations can be performed by different observers, on different material objects, at different times and places.

Moreover, in a given experiment, biologists typically use different specimens exposed to the same conditions in order to perform statistical tests. The tests assume that these specimens follow the same probability distribution, that is, the tests assume that behind the quantitative variations observed there is a single abstract mathematical object (the probability distribution): this is again an assumption of symmetry.

However, biological objects are the result of a history and continue to generate a history. Interpreting this notion in terms of symmetries leads to assert that when time flows, describing the changes of biological objects can require changes of symmetry that do not stem from the description of the initial objects. These changes are the core of the principle of variation (Longo & Montévil, 2011a; Montévil et al., 2016; Montévil, 2018). As hinted to in the introduction, these variations conflict with the aim to perform reproducible experiments. In biology unlike in physics, the symmetries associated with reproducibility are not granted theoretically. Instead, they depend on the measurement as summarized in section 3.6.

Since biological regularities are more labile than physical ones, symmetries are not provided directly by the theory. Instead, they are co-established by the measurement process and the biological objects used. We propose to call this practical and theoretical operation "symmetrization". Biologists typically work on specimens of the same species or more generally specimens with a shared common past. In experiments, they assume a partial equivalence between these specimens and how they are organized. In other words, biologists posit an approximate symmetry between the organization of different organisms and their response to experiments. Control over past contexts is also a symmetrization of the specimens studied and are often designed with this issue in mind. For example, in section 3.3, we have discussed how biologists aim for cells *in vitro* to be in a consistent state over time, that is to say, how biologists symmetrize cells. The different methods described in section 3 should be seen as different symmetrizations.

Different symmetrizations can be performed during an experiment or even during data acquisition. Choosing a symmetrization or another endows the results with entirely different biological meanings. Figure 4 illustrates this idea and shows different ways to make organisms equivalent. In figure 4E1, by being in the field, organisms express their historically (evolutionary) relevant activities and these activities are diverse. In figure 4E3, different organisms are mostly restrained to activities that are common to them: the experimenter performs a stronger symmetrization by limiting the characters involved in the determination of the observed quantity.

We can distinguish two kinds of symmetrization: concrete and epistemic symmetrizations. Concrete symmetrizations involve the action of biologists on objects. For example, establishing inbred strains is a concrete symmetrization of genomes, and the symmetrizations illustrated in figure 4 are also concrete symmetrizations. By contrast, epistemic symmetrizations do not change material objects and are limited to determining what is considered equivalent, that is to say, symmetric by permutation.

For example, we mentioned that the position of a fetus in utero has measurable consequences. Taking this aspect into account or not corresponds to different epistemic symmetrizations. The concept of epistemic symmetrization is particularly relevant for statistical analyses and subsequent biological reasonings.

As a consequence of the principle of variation, the concept of epistemic symmetrization is always relevant: biologists have to symmetrize organisms which are not genuinely symmetric. Concrete symmetrizations occur in most experiments, but not in observations without experiments such as the observation of specimens in systematics. Performing concrete symmetrization constrains the kind of biological objects studied. For example, it is far easier to symmetrize cells in culture by maintaining unconstrained proliferation.

In conclusion, biological measurement as summarized in 3.6 describes both the concrete and epistemic symmetrizations performed to obtain experimental results and endow them with biological meaning.

4.3. Measurement strategies

Baxendale (2018) proposes to map scientific practices on a continuum of strategies defined by their stances concerning reductionism. In this section, we apply a similar approach to measurement strategies. Our concept of biological measurement leads to the notion that measurement depends on symmetrizations, but symmetrizations can be performed more or less tightly and at different levels. To represent these strategies, we propose to organize measurement strategies along three axes as illustrated in figure 6. In this section, we discuss only the different measurement strategies, and we do not imply that they necessarily succeed or that they prevent the joint use of other strategies.

The first axis describes whether the measurement is variable or on the opposite reproducible. Here, reproducibility means that the measurement generates data consistently with different specimens. For example, using inbred strains generally leads to more reproducible results than wild specimens.

The second axis describes whether the measurement targets singular or general objects. Working on the metabolic rate of mammals is more general than working on a single species by measuring wild specimens. Both are more general than outbred strains and *a fortiori* inbred strains, where the genotype is symmetrized. Reciprocally, inbred strains are more singular than outbred strains and so on.

The last axis assesses whether the measurement defines objects coherent with their evolutionary past or instead whether the objects are more or less profoundly altered. For example, inbred strains are homozygotes for all genes which is not the case of mammals in their evolutionary history. Similarly, the basal metabolic rate is far less representative of a species past evolution than the field metabolic rate — but the latter depends on the field.

A qualitative axis emerges in this three-dimensional space, see figure 6. This axis is given by the strategies which increase simultaneously the reproducibility, singularity, and alterations associated with the measurement. In other words, these strategies lose generality and alter the specimens in order to increase the reproducibility of the measurement. At the limit, these strategies aim to generate specimens which are as symmetrized as possible and would have the same status than the objects of physics: these strategies aim the genericization of biological organizations and use many methods to reduce diversity. For example, in the case of cells, samples are frozen to prevent spontaneous variations between experiments. The focus on model organisms at the level of the research community is also a collective strategy of genericization. In situations like clinical trials, on humans, genericization methods are limited for ethical reasons. In contrast to genericization strategies, other strategies on the same axis aim to gain generality and coherence with evolutionary history but at the cost of more variability in the results. It follows that these strategies face more difficulties to obtain statistically significant results.

In order to face the reproducibility crisis and to obtain significant results with fewer animals, it is common to promote strategies genericizing specimens (Festing, 2014; Chia et al., 2005). However, these strategies bear the cost of studying singular organizations: the results obtained may not even be representative of the species studied, and we have seen that strains of the same species have distinct properties. Another example is that the conditions of the laboratory reduce exposure to pathogens in order to symmetrize the life history of animals studied, which is part of the alteration axis. However, this situation leads to immunological functions that differ from wild animals (Abolins et al., 2010).

The genericization of specimens aims, at the limit, to study a single, reproducible organization and is thus highly singular. Results may depend on the specificities of these organizations and their contexts in unknown ways. Therefore, these strategies are vulnerable to minor departures from the genericization performed initially. For example, performing measurement in different laboratories always involve a change of context despite the explicit control of many factors. Genericizations aim reproducibility in the sense of specimens that are very similar, but the reproducibility of experiments is made difficult by the lack of generality of the measurement.

In figure 6, there are only two cases which are far from this axis. The first corresponds to measurements like the basal metabolic rate, see figure 4E3. This measurement is reproducible and nevertheless general. Its downside is that the organisms are put in a specific state to level down the consequences of the diversity of the characters impacting the measured quantity. Its strategy is to symmetrize a shared aspect of organisms when the genericizations discussed above symmetrize complete organizations.

Case studies are the second strategy departing from the main axis. Case studies focus on a single individual and reproducibility is not a goal. For example, Patterson & Linden (1981) study the intelligence of non-human primates. To do so, the authors did not develop standardized conditions and protocols. Instead, they taught sign language to several specimens and focused on a particularly gifted gorilla, Koko, who mastered up to 2000 symbols. Other examples are works on types in systematics, the study of the bipedal goat discussed by West-Eberhard (2003) and the cloned sheep, Dolly, which is one success among 277

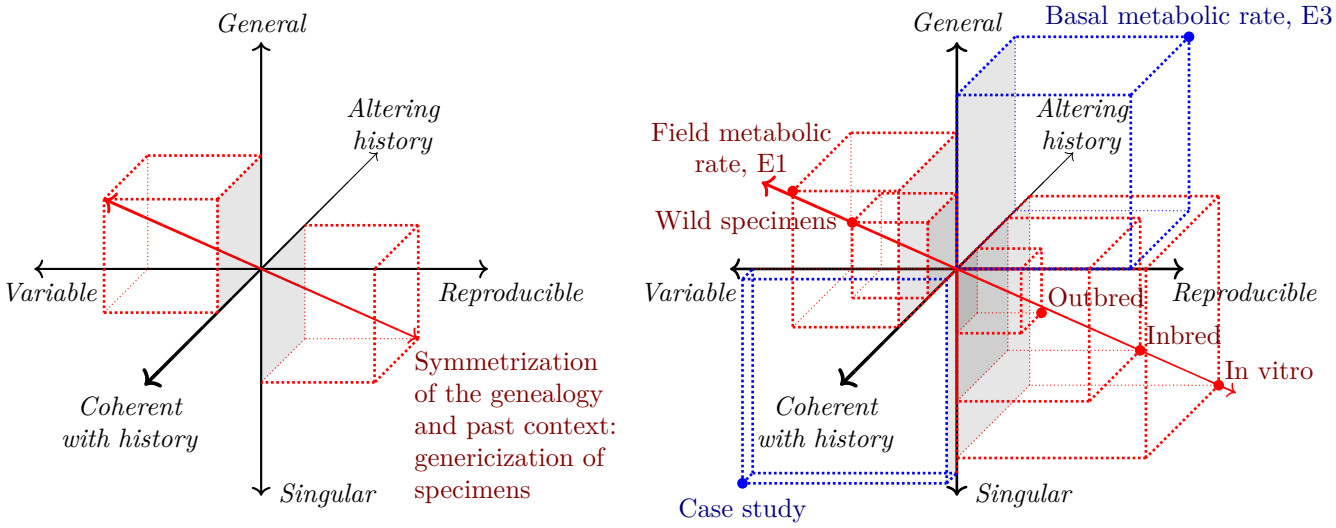

Figure 6: *Different measurement strategies.* The three axes correspond to measurements that are reproducible vs. variable, general vs. singular and coherent with their evolutionary past vs. altered by experimenters. Left: axis where many strategies lie. On one end of the spectrum, these strategies aim to produce specimens that are as close to each other as possible by controlling them tightly. On the other end, experimenters relax this control and aim to make more general measurements. Right: different cases are represented in the space describing measurement strategies. Most measurement strategies are on the axis represented on the left, but departures from this axis are equally interesting since they represent other ways to approach biological phenomena empirically, e.g., case studies.

attempts and remained the only success for a long time (Wilmut et al., 1997). While the study of types does not involve alterations, teaching Koko or cloning a sheep do: case studies are diverse for the third axis.

Case studies are sometimes neglected by experimenters who strive to design reproducible experiments in order to study mechanisms. For example, the success of cloning Dolly without reproducing this feat led to an intense debate, especially when evidence accumulated that Dolly was indeed cloned from an adult cell (Solter, 1998). However, in our conceptual framework, case studies have a specific epistemic role. They are sufficient to prove the existence of a possibility in a theoretical context where biological possibilities are not predefined (Montévil, 2018). The bipedal goat shows the extent of developmental plasticity and studying a type is sufficient to defend the existence of a new species. Case studies can be extensive; for example, the anatomy of the bipedal goat has been described in details. In case studies, the analysis of a single part in several organisms is typically replaced by the analysis of many parts or aspects in a single organism. Last, the study of types in systematics plays a pivotal role in the general architecture of biological knowledge to name biological objects.

Representing different measurement strategies by the symmetrizations performed is fruitful. These strategies are different responses to the difficulties raised by the historical and varying nature of biological objects.

5. Conclusion

Our theoretical notion of measurement accommodates how biologists manipulate immensely complex objects, organisms and cells typically, which are the result of a history and continue to produce a history by generating qualitative variations. The concept of biological measurement which we propose accommodates simultaneously the organisms or cells and their part or aspect of interest which may be quantified. In our framework, a measurement is relative to a history and context. To develop reproducible experiments, biologists observe specimens with a shared past. This shared past is ascertained by systematics and by direct knowledge and control of both their genealogy and past contexts. In the study of objects defined by their history, the objects which can be considered equivalent are objects having a shared past. In this context, we call symmetrization the concrete and theoretical operations which establish and posit the equivalence of different objects with more or less tightly controlled shared pasts and contexts. Symmetrization also includes the operations performed during the observation which can constrain and structure variability.

The notion of biological measurement is compatible with different research strategies and leads to a framework to map them. In this framework, we find two polar opposites. In one end, strategies strive to genericize biological organizations at the cost of studying singular organizations and altering them. To implement these strategies, biologists developed a plethora of methods. They expose objects to similar contexts and ensure that they have recent, controlled common ancestors. In some cases, biologists freeze samples to prevent them from undergoing variations between experiments. On the other end of the spectrum, the objects studied are more general (e.g., diverse genetically) and coherent with their evolutionary history, but they are also more variable. There are strategies which escape this opposition, for example, case studies or methods to level down the diversity relevant for the part studied while the rest of the organizations remain diverse.

Having a clear notion of what it means to access biological

objects empirically is critical for biological knowledge. In this paper, we provide only an outline of biological measurement, and this notion deserves further discussions, focusing on both general and specific situations. Nevertheless, since our notion builds on solid ground, namely the theory of evolution and extensions for organisms, we hope that our work will be of use for further research. We have shown that biological measurement has significant differences with the notions of measurement in physics. Depending on the perspective, biological measurement may be seen as an extension of classical measurement in order to accommodate the historicity and variability of biological objects, or as a different concept altogether because the objects are not described theoretically by underlying equations. In all cases, acknowledging the specificities of biological measurement should provide new systematic ways to approach biological observations critically and ultimately to promote experimental reproducibility.

Acknowledgments

I am grateful to Ana Soto, Giuseppe Longo, Carlos Sonnenschein, Guillaume Lecointre, Matteo Mossio, Arnaud Pocheville and Véronique Thomas-Vaslin for their comments on previous versions of this article and helpful discussions. I also thank the two anonymous reviewers and the editor for their candid comments.

References

Abolins, S. R., Pocok, M., Hafalla, J., Riley, E., & Viney, M. (2010). Measures of immune function of wild mice, *Mus musculus*. *Molecular Ecology*, *20*, 881–892. doi:10.1111/j.1365-294X.2010.04910.x.

Bailly, F., & Longo, G. (2011). *Mathematics and the natural sciences; The Physical Singularity of Life*. London: Imperial College Press.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*, 452–454. doi:10.1038/533452a.

Baxendale, M. (2018). Mapping the continuum of research strategies. *Synthese*, . doi:10.1007/s11229-018-1683-1.

Beatty, J. (1995). The evolutionary contingency thesis. In G. Wolters, & J. Lennox (Eds.), *Conceptual Issues in Evolutionary Biology*. (pp. 45–81). Pittsburgh: University of Pittsburgh Press.

Begley, C., & Ioannidis, J. (2014). Reproducibility in science. *Circulation Research*, *116*, 116–126. doi:10.1161/CIRCRESAHA.114.303819.

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, *483*, 531–533. doi:10.1038/483531a.

Bolland, M. J., Avenell, A., Gamble, G. D., & Grey, A. (2016). Systematic review and statistical analysis of the integrity of 33 randomized controlled trials. *Neurology*, . doi:10.1212/WNL.0000000000003387.

Braun, E. (2015). The unforeseen challenge: from genotype-to-phenotype in cell populations. *Reports on Progress in Physics*, *78*, 036602. doi:10.1088/0034-4885/78/3/036602.

Cai, L., Friedman, N., & Xie, X. (2006). Stochastic protein expression in individual cells at the single molecule level. *Nature*, *440*, 358 – 362. doi:10.1038/nature04599.

Chang, H. (2004). *Inventing temperature: Measurement and scientific progress*. Oxford University Press.

Chia, R., Achilli, F., Festing, M., & Fisher, E. (2005). The origins and uses of mouse outbred stocks. *Nature genetics*, *37*, 1181. doi:10.1038/ng1665.

CZN International (1999). *International code of zoological nomenclature*. London: International Trust for Zoological Nomenclature.

Danchin, E., Pocheville, A., Rey, O., Pujol, B., & Blanchet, S. (2019). Epigenetically facilitated mutational assimilation: epigenetics as a hub within the inclusive evolutionary synthesis. *Biological Reviews*, *94*, 259–282. doi:10.1111/brv.12453.

Dudley, J., Golovkina, T., & Ross, S. (2016). Lessons learned from mouse mammary tumor virus in animal models. *ILAR Journal*, *57*, 12–23. doi:10.1093/ilar/ilv044.

Festing, M. (2014). Evidence should trump intuition by preferring inbred strains to outbred stocks in preclinical research. *ILAR Journal*, *55*, 399–404. doi:10.1093/ilar/ilu036.

Feynman, R., & Gleick, J. (1967). *The character of physical law*. MIT press.

Gilbert, S. F. (2014). A holobiont birth narrative: The epigenetic transmission of the human microbiome. *Frontiers in Genetics*, *5*. doi:10.3389/fgene.2014.00282.

Gilbert, S. F., & Epel, D. (2009). *Ecological developmental biology: integrating epigenetics, medicine, and evolution*. Sinauer Associates Sunderland.

Gillies, D. (2012). *Philosophical theories of probability*. Routledge.

Gould, S. (2002). *The structure of evolutionary theory*. Harvard University Press.

Grandcolas, P. (2017). Loosing the connection between the observation and the specimen: a by-product of the digital era or a trend inherited from general biology? *Bionomina*, *12*, 57–62. doi:10.11646/bionomina.12.1.7.

Gross, V., & Luft, F. (2003). Exercising restraint in measuring blood pressure in conscious mice. *Hypertension*, *41*, 879–881. doi:10.1161/01.HYP.0000060866.69947.D1.

Heindel, J., Newbold, R., Bucher, J., Camacho, L., Delclos, K. B., Lewis, S. M., Vanlandingham, M., Churchwell, M. I., Twaddle, N. C., McLellen, M., Chidambaram, M., Bryant, M., Woodling, K., da Costa, G. G., Ferguson, S. A., Flaws, J., Howard, P. C., Walker, N. J., Zoeller, R. T., Fostel, J., Favaro, C., & Schug, T. T. (2015). NIEHS/FDA CLARITY-BPA research program update. *Reproductive Toxicology*, *58*, 33 – 44. doi:10.1016/j.reprotox.2015.07.075.

Houle, D., Pélabon, C., Wagner, G. P., & Hansen, T. F. (2011). Measurement and meaning in biology. *The Quarterly Review of Biology*, *86*, 3–34. doi:10.1086/658408.

Huneman, P. (2018). Diversifying the picture of explanations in biological sciences: ways of combining topology with mechanisms. *Synthese*, *195*, 115–146. doi:10.1007/s11229-015-0808-z.

Isaacs, J. T. (1986). Genetic control of resistance to chemically induced mammary adenocarcinogenesis in the rat. *Cancer Research*, *46*, 3958–3963.

Jablonka, E., Lamb, M. J., & Zeligowski, A. (2014). *Evolution in Four Dimensions, revised edition: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. London: MIT press.

Jablonka, E., & Raz, G. (2009). Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *The Quarterly review of biology*, *84*, 131–176. doi:10.1086/598822.

Jensen, V. S., Porsgaard, T., Lykkesfeldt, J., & Hvid, H. (2016). Rodent model choice has major impact on variability of standard preclinical readouts associated with diabetes and obesity research. *American journal of translational research*, *8*, 3574.

Johnson, P. D., & Besselsen, D. G. (2002). Practical aspects of experimental design in animal research. *ILAR journal*, *43*, 202–206. doi:10.1093/ilar.43.4.202.

Kohler, R. E. (1994). *Lords of the fly: Drosophila genetics and the experimental life*. University of Chicago Press.

Lancet (2018). Uk life science research: time to burst the biomedical bubble. *The Lancet*, *392*, 187. doi:10.1016/S0140-6736(18)31609-X.

Lecointre, G., & Le Guyader, H. (2006). *The tree of life: a phylogenetic classification* volume 20. Harvard University Press.

Leonelli, S. (2014). What difference does quantity make? on the epistemology of big data in biology. *Big Data & Society*, *1*, 2053951714534395. doi:10.1177/2053951714534395.

Lindner, A., Madden, R., Demarez, A., Stewart, E., & Taddei, F. (2008). Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proceedings of the National Academy of Sciences*, *105*, 3076 – 3081. doi:10.1073/pnas.0708931105.

Longo, G., & Montévil, M. (2011a). From physics to biology by extending criticality and symmetry breakings. *Progress in Biophysics and Molecular Biology*, *106*, 340 – 347. doi:10.1016/j.pbiomolbio.2011.03.005.

Longo, G., & Montévil, M. (2011b). Protention and retention in biological systems. *Theory in Biosciences*, *130*, 107–117. doi:10.1007/s12064-010-0116-6.

Longo, G., & Montévil, M. (2014). *Perspectives on Organisms: Biological time, symmetries and singularities*. Lecture Notes in Morphogenesis. Dordrecht: Springer. doi:10.1007/978-3-642-35938-5.

Longo, G., & Montévil, M. (2017). Comparing symmetries in models and simu-

lations. In M. Dorato, L. Magnani, & T. Bertolotti (Eds.), *Springer Handbook of Model-Based Science*. Springer. doi:10.1007/978-3-319-30526-4.

McNeill, J., Barrie, F., Buck, W., Demoulin, V., Greuter, W., Hawksworth, D., Herendeen, P., Knapp, S., Marhold, K., Prado, J. et al. (2012). *International Code of Nomenclature for algae, fungi and plants*. Konigstein, Germany: Koeltz Scientific Books.

Mogil, J. S., Wilson, S. G., Bon, K., Lee, S. E., Chung, K., Raber, P., Pieper, J. O., Hain, H. S., Belknap, J. K., Hubert, L., Elmer, G. I., Chung, J. M., & Devor, M. (1999). Heritability of nociception i: Responses of 11 inbred mouse strains on 12 measures of nociception. *Pain*, *80*, 67 – 82. doi:10.1016/S0304-3959(98)00197-3.

Montévil, M. (2018). Possibility spaces and the notion of novelty: from music to biology. *Synthese*, . doi:10.1007/s11229-017-1668-5.

Montévil, M., & Mossio, M. (2015). Biological organisation as closure of constraints. *Journal of Theoretical Biology*, *372*, 179 – 191. doi:10.1016/j.jtbi.2015.02.029.

Montévil, M., Mossio, M., Pocheville, A., & Longo, G. (2016). Theoretical principles for biology: Variation. *Progress in Biophysics and Molecular Biology*, *122*, 36 – 50. doi:10.1016/j.pbiomolbio.2016.08.005.

Morgan, M. (2002). Experiments without material intervention: Model experiments, virtual experiments and virtually experiments. In H. Radder (Ed.), *Scientific Experimentation and Its Philosophical Significance*. Pittsburgh University Press.

Mossio, M., Montévil, M., & Longo, G. (2016). Theoretical principles for biology: Organization. *Progress in Biophysics and Molecular Biology*, *122*, 24 – 35. doi:10.1016/j.pbiomolbio.2016.07.005.

Mugur-Schächter, M. (2002). Objectivity and descriptional relativities. *Foundations of Science*, *7*, 73 – 180. doi:10.1023/A:1016095424229.

Nadin, M. (2017). Rethinking the experiment: necessary (r) evolution. *AI & SOCIETY*, (pp. 1–19). doi:10.1007/s00146-017-0705-8.

Patterson, F., & Linden, E. (1981). *The education of Koko*. Holt, Rinehart, and Winston.

de Queiroz, K. (1992). Phylogenetic definitions and taxonomic philosophy. *Biology and Philosophy*, *7*, 295–313. doi:10.1007/BF00129972.

Ryan, B. C., & Vandenbergh, J. G. (2002). Intrauterine position effects. *Neuroscience & Biobehavioral Reviews*, *26*, 665 – 678. doi:10.1016/S0149-7634(02)00038-6.

Sacca, R., Elder, B., & Wasson, K. (2013). *The C57BL/6 Mouse: The Role of the C57BL/6N Mouse in the Creation of Future Genetically Engineered Models*. Charles River Laboratories International, Inc.

Salam, A. (1990). *Unification of Fundamental Forces*. Cambridge University Press.

Schrödinger, E. (1944). *What Is Life?*. Cambridge U.P.

Simpson, E. M., Linder, C. C., Sargent, E. E., Davisson, M. T., Mobraaten, L. E., & Sharp, J. J. (1997). Genetic variation among 129 substrains and its importance for targeted mutagenesis in mice. *Nature genetics*, *16*, 19. doi:10.1038/ng0597-19.

Solter, D. (1998). Dolly is a clone—and no longer alone. *Nature*, *394*, 315.

Soto, A. M., Justicia, H., Wray, J. W., & Sonnenschein, C. (1991). p-nonyl-phenol: An estrogenic xenobiotic released from "modified" polystyrene. *Environmental Health Perspectives*, *92*, 167–173. doi:10.1289/ehp.9192167.

Soto, A. M., Longo, G., Montévil, M., & Sonnenschein, C. (2016a). The biological default state of cell proliferation with variation and motility, a fundamental principle for a theory of organisms. *Progress in Biophysics and Molecular Biology*, *122*, 16 – 23. doi:10.1016/j.pbiomolbio.2016.06.006.

Soto, A. M., Longo, G., Noble, D., Perret, N., Montévil, M., Sonnenschein, C., Mossio, M., Pocheville, A., Miquel, P.-A., Hwang, S.-Y. et al. (2016b). From the century of the genome to the century of the organism: New theoretical approaches. *Progress in Biophysics and Molecular Biology, Special issue*, (pp. 1–82).

Spearow, J. L., Doemeny, P., Sera, R., Leffler, R., & Barkley, M. (1999). Genetic variation in susceptibility to endocrine disruption by estrogen in mice. *Science*, *285*, 1259–1261.

Stewart, E., Madden, R., Paul, G., & Taddei, F. (2005). Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biology*, *3*, e45. doi:10.1371/journal.pbio.0030045.

Susiarjo, M., Hassold, T. J., Freeman, E., & Hunt, P. A. (2007). Bisphenol a exposure in utero disrupts early oogenesis in the mouse. *PLoS genetics*, *3*, e5. doi:10.1371/journal.pgen.0030005.

Van Fraassen, B. (1989). *Laws and symmetry*. Oxford University Press, USA.

Wagner, G. P. (2010). The measurement theory of fitness. *Evolution*, *64*, 1358–1376. doi:10.1111/j.1558-5646.2009.00909.x.

Weber, M. (2004). *Philosophy of experimental biology*. Cambridge university press.

West, B. (2006). *Where medicine went wrong: Rediscovering the path to complexity*. Teaneck, NJ: World Scientific.

West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford University Press.

Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J., & Campbell, K. H. (1997). Viable offspring derived from fetal and adult mammalian cells. *Nature*, *385*, 810–813. doi:10.1038/385810a0.