CrossMark

ORIGINAL PAPER IN PHILOSOPHY OF BIOLOGY

Integrative taxonomy and the operationalization of evolutionary independence

Stijn Conix[1]

Abstract There is growing agreement among taxonomists that species are independently evolving lineages. The central notion of this conception, evolutionary independence, is commonly operationalized by taxonomists in multiple, diverging ways. This leads to a problem of operationalization-dependency in species classification, as species delimitation is not only dependent on the properties of the investigated groups, but also on how taxonomists choose to operationalize evolutionary independence. The question then is how the operationalization-dependency of species delimitation is compatible with its objectivity and reliability. In response to this problem, various taxonomists have proposed to integrate multiple operationalizations of evolutionary independence for delimiting species. This paper first distinguishes between a standard and a sophisticated integrative approach to taxonomy, and argues that it is unclear how either of these can support the reliability and objectivity of species delimitation. It then draws a parallel between the measurement of physical quantities and species delimitation to argue that species delimitation can be considered objective and reliable if we understand the sophisticated integrative approach as assessing the coherence between the idealized models of multiple operationalizations of evolutionary independence.

Keywords Integrative taxonomy · Species delimitation · Species classification · Evolutionary independence · Measurement · Operationalization

1 Introduction

Despite numerous unresolved issues in philosophical debates about the nature of species, there is growing consensus among taxonomists that the aim of species

✉ Stijn Conix
   sc881@cam.ac.uk

1   Department of History and Philosophy of Science, University of Cambridge, Free School Lane, CB2 3RH Cambridge, UK

Springer

delimitation is to identify independently evolving lineages (e.g. Fujita et al. 2012; Padial and De La Riva 2010).[1] This paper takes this view on species as a given, and aims to resolve an epistemological problem faced by taxonomists who adopt it for species delimitation. The problem, which I will call the problem of operationalization-dependency (POD henceforth), is that evolutionary independence can be, and in practice is, operationalized in innumerable ways, many of which lead to diverging outcomes. Hence, the results of species delimitation seem to depend as much on taxonomists' choice of operationalization as on the properties of the groups they investigate. This poses a problem, because it seems incompatible with two major desiderata of taxonomy, namely, the reliability and objectivity of classifications. The aim of this paper is to show that the operationalization-dependency of species delimitation does not necessarily obstruct the fulfilment of these desiderata.

To do this, section 2 spells out the POD and presents integrative taxonomy, a commonly adopted solution to this problem. I distinguish between two interpretations of integrative taxonomy, and argue that it is not clear how either of those can support the objectivity and reliability of species classification. Section 3 then turns to the philosophy of measurement, and shows how a model-based account of measurement solves an analogous problem in the measurement of physical quantities. Section 4 discusses a case-study to show that the sophisticated integrative approach to taxonomy solves the POD in a similar way. The paper finishes in section 5 with concluding remarks.

Before I set out to do this, it is worth emphasizing that this paper deals only with the epistemology of species delimitation, and does not make any commitments about the metaphysics of species.[2] Most importantly, my points are compatible both with conventionalist and realist views about species. This is because even if one adopts realism and holds that there is a matter of fact about the evolutionary independence of a lineage, it remains the case that taxonomists can only access this through various diverging operationalizations. Thus, the POD affects species delimitation regardless of the metaphysical views we adopt. This is not to say that the POD is entirely independent from the metaphysics of species. If, as many have argued, pluralism about lineages is true and evolutionary independence comes in degrees, it should be expected that different operationalizations lead to different outcomes (Ereshefsky 1992; Degnan and Rosenberg 2006; Haber 2012; Maddison 1997). Indeed, as will become clear in the last section of the paper, understanding this may be epistemically valuable, for example when knowledge on the differences between various operationalizations can be implemented in species delimitation (Degnan and Rosenberg 2009; Haber 2016). However, even if this metaphysical picture holds, it remains the case that taxonomists have to get on with the task of recognising certain groups of organisms as species. Thus,

---

[1] This can be interpreted either in a backward-looking way, in which species are the lineages that result from independent evolution, or in a forward-looking way, in which species are the entities that evolve (Reydon 2005). As the difference between these is not relevant here, I will not distinguish between them.

[2] That is, aside from the assumption that species delimitation aims to pick out independently evolving lineages. However, an argument analogous to the one in this paper can be made about any other general conception of species that can be operationalized in many ways. This definitely seems to be the case for the view that species are taxa (see Baum 2009), which in the current literature is the main contender for a general view on species.

metaphysically problematic as it may be, they must face the multiplicity of operationalizations and solve the POD.

By focusing on the epistemology of species delimitation rather than the metaphysical nature of species, this paper deviates from the largest part of the philosophical literature on species. Instead, it picks up on the research cited above which highlights the epistemological implications of our growing understanding of the nature of species (Degnan and Rosenberg 2006, 2009; Haber 2012, 2016; Maddison 1997; Sterner 2017). This paper advances such research by developing one particularly important epistemological problem in more detail (i.e., the POD) and proposing a potential solution. This fresh, epistemological perspective demonstrates an interesting and underexplored role for philosophy in questions about species.[3] Moreover, the focus on epistemology is in line with an increasing theoretical interest in the taxonomic literature in methods of species delimitation (Sites and Marshall 2004; Camargo and Sites 2013; Leavitt et al. 2015).

2 Integrative taxonomy and the problem of operationalization-dependency

In order to clarify the POD, it is helpful to consider what taxonomists mean by 'independently evolving lineages'. By 'lineages' they mean ancestor-descendant sequences of populations with a unique evolutionary origin; but what is meant by 'independently evolving' is notoriously hard to elucidate. Most commonly, it is claimed that a lineage evolves independently if the organisms that make up its populations generally face evolutionary pressures as a unit, distinct from other such units (Wiley 1978). In other words, independently evolving lineages are groups of organisms that are cohesive in an evolutionary sense. The result of this evolutionary cohesion is that the organisms that make up these lineages have a shared unique path through evolutionary space. This means that we can characterise the evolutionary independence of a lineage as the evolutionary fate sharing of the organisms that make up the lineage (Barker 2007; Barker and Wilson 2010).

It should be noted that the fate sharing of organisms within a lineage need not be absolute. Indeed, some differences are expected, as intraspecific variation is a crucial requirement for evolution to occur. The distinctness of the fate of organisms between species-lineages is not absolute either. For example, there is a significant sense in which humans and chimpanzees share an evolutionary fate not shared by tigers, but they are clearly part of different species-lineages. In other words, evolutionary fate sharing is a matter of degree, and consequently the evolutionary independence of species-lineages is a matter of degree too. Species, then, are generally characterised as groups of organisms with relatively high degrees of fate sharing. This raises two questions concerning evolutionary independence that are crucial for species delimitation. First, the question of how high this degree of independence should be for a lineage to qualify as a species. I will not go into this issue. Instead, I focus on a more fundamental

---

[3] In this sense, it takes the cue from David Hull (1997, 372–373), who wrote in relation to species that '[u]nfortunately, philosophers have not addressed this question [about operationalization] in any detail […]. For any student of science who is getting a bit bored with the Covering Law Model of Scientific Explanation or incommensurability, the topic of operationalizing concepts is wide open.'

question that needs answering before a particular degree of independence can be chosen. This is the question of how one can determine the degree of evolutionary independence in the first place.[4]

This question is particularly pressing because the characterization of evolutionary independence in terms of evolutionary fate sharing is highly theoretical and, some have argued, vague (Ereshefsky 1991; Wheeler and Meier 2000). Despite the lack of a precise definition, taxonomists engaged in species delimitation have to assess the degree of evolutionary independence. They do this either through the processes that cause evolutionary fate sharing, or through the patterns resulting from these processes. The causes of fate sharing that taxonomists commonly rely on for species delimitation include gene-flow, reproductive interactions and shared selection pressures. The patterns resulting from these processes include phenotypic and genotypic similarity, monophyly, and genealogical relations between organisms.[5] Together, these processes and patterns provide multiple ways of operationalizing the notion of evolutionary independence for species delimitation.

Various authors have argued that this diversity of operationalizations of the single abstract conceptualization of species is important to capture the broad variety of species lineages in the organic world (Mayden 1999, 97; De Queiroz 2007, 879). However, this advantage comes with an important worry on the other side of the coin. It is well known that the various operationalizations associated with evolutionary independence do not always lead to the same results. For example, many groups that are morphologically diagnosable are not reproductively isolated, while there are also many morphologically indistinguishable groups that are reproductively isolated. More generally, various authors show that disagreement between these operationalizations is more common than agreement (Padial et al. 2009; Schlick-Steiner et al. 2010; Willis 2017). In practice, this means that different operationalizations of evolutionary independence tend to pick out different groups as species. It follows then that depending on how taxonomists decide to operationalize the notion of evolutionary independence, different species delimitations result. This is what I call the operationalization-dependency of species delimitation.[6]

Various authors point out that the operationalization-dependency of taxonomy forms a threat to the epistemic value of species delimitation (e.g. Schlick-Steiner et al. 2010; Willis 2017). Fujita et al. (2012, 481), for example, comment on the 'inherent subjectivity' of choosing proxies (i.e. operationalizations) of independent evolution and how this negatively affects the repeatability of species delimitation. More precisely, operationalization-dependency seems to constitute a lack of objectivity of species delimitation. Defining objectivity in detail is of course beyond the scope of this paper; I will use it here to refer to anything that is not dependent on a particular point of view

---

[4] Another way of putting this is to say that the question at hand is about the *measurement* of evolutionary independence. I develop this line of thought further in section 3.

[5] Note that there is no clear distinction between the causes and effects of evolutionary independence. For example, gene-flow and shared selection pressures lead to increased phenotypic and genotypic similarity, which in turn increase reproductive isolation and the similarity in the way organisms respond to selection pressures.

[6] Taxonomists often use the term 'researcher bias' for this. While this term better conveys the fact that researchers' choices indirectly play an important role in species delimitation, I use the term 'operationalization-dependency' instead to avoid confusion with systematic error and to emphasize the importance of operationalization.

or perspective. By this meaning, operationalization-dependency constitutes a lack of objectivity because the outcome of species delimitation is directly dependent on the particular operationalization (i.e. perspective) that is adopted. This lack of objectivity in turn raises worries for the reliability of these operationalizations. Assuming that the diverging outcomes of different operationalizations cannot all be accepted, it must be that they sometimes fail to track evolutionary independence.[7] Given that we can only access evolutionary independence through these operationalizations, there is no independent criterion to decide which ones to trust. It is unclear then why we should trust these operationalizations as reliably picking out independently evolving lineages.

These problems, which I jointly refer to as the POD, are particularly worrying because they form a threat to two further desiderata of biological taxonomy, namely, the stability of species classifications and the comparability of the species they consist of. If species delimitation is subjective and unreliable in the sense described here, classifications are likely to change frequently when different operationalizations are used to investigate the same groups. This instability poses a problem for users of taxonomy, such as conservation legislation and policies, which have to adapt to these changes. Similarly, the dependence of species delimitation on the adopted operationalization puts into question the extent to which various groups recognised as species are comparable. This is important because research in other fields, such as macro-evolutionary research and conservation biology, implicitly depends on species being similar in certain respects (Isaac et al. 2004).

In response to these epistemological problems, particularly the lack of comparability, Mishler (1999) argues that we should 'get rid' of the species rank, thus solving the POD by dissolution.[8] Rather than adopting this radical solution, various taxonomists propose to solve the POD by adopting a novel approach to species delimitation called 'integrative taxonomy', which uses multiple operationalizations of evolutionary independence for species delimitation (Dayrat 2005; Padial et al. 2010; Will et al. 2005). In practice, integrative taxonomy is most commonly applied in the form of the requirement that lineages should be recognized only when all or most operationalizations converge on the same outcome. I will call this 'the standard approach' to integrative taxonomy.[9] The idea behind this standard interpretation is that it is very unlikely that multiple operationalizations that are all wrong would yield the same result. Thus, the robustness of a particular species delimitation over multiple operationalizations is taken as a sign of its reliability on the basis of a no-coincidence argument. This robustness also provides an easy response to the worry that species delimitation is not objective. Given that the standard approach integrates various points of view, species delimitation is no longer dependent on a single perspective. Instead, it is objective in the sense of inter-subjectivity.[10]

---

[7] As noted in the introduction, a metaphysical pluralist could point out here that it might just be that they are all correct. However, given that accepting innumerably many species is not a feasible option in taxonomy, this does not defuse the epistemological problem faced by taxonomists who have to accept some operationalizations as reliable and reject others.

[8] I thank an anonymous reviewer for pointing this out to me.

[9] For examples of this standard approach, see Alström et al. (2008) and Vieites et al. (2009), who claim that groups should only be recognised as species if there is agreement between most or all criteria for species status.

[10] I say more about this sense of objectivity at the end of sections 3 and 4.

However, the standard approach faces an important problem. As discussed above, many currently recognised species, including well-established ones, would not be picked out by all commonly used operationalizations of evolutionary independence. One particularly clear illustration of this comes from the work of Degnan and Rosenberg (2006), which shows that incongruence between the majority gene-tree and population tree of a group is not only possible, but highly likely under certain common conditions. The prevalence of such diverging operationalizations implies that the standard approach risks not recognising many groups that we would want to accept as species, and when applied very strictly, might even make species delimitation impossible. In response to this problem of false negatives, various authors propose a more sophisticated integrative approach, which uses multiple operationalizations but does not require absolute convergence (Padial et al. 2009; Padial et al. 2010; Schlick-Steiner et al. 2010).[11] This sophisticated approach requires taxonomists to decide which operationalizations reliably track evolutionary independence, and which are erroneous or in need of correction. However, this was precisely what led taxonomists to adopt an integrative approach in the first place. Given that this decision is underdetermined by the operationalizations, this means that taxonomists must also rely on background theory for this. Thus, theoretical assumptions play a crucial role in integrating multiple operationalizations on the sophisticated integrative approach. This does not mean, of course, that the standard approach does not rely on background theory at all. Rather, the point is that, unlike the sophisticated approach, the standard approach does not use this background theory to correct and integrate diverging operationalizations.

While the sophisticated integrative approach avoids the problem of false negatives, it faces different but equally troubling problems. Most importantly, it is not clear how the integration of multiple fallible operationalizations, some of which are explicitly assumed to be erroneous, can increase the objectivity and reliability of species delimitation. Unlike the standard approach, the sophisticated approach cannot rely on a no-coincidence argument. Given that the sophisticated account relies on background assumptions to retain, correct and dismiss various diverging operationalizations, any resulting convergence is the result of taxonomists moulding the evidence to fit together. This moulded convergence of operationalizations could hardly serve as the basis for a no-coincidence argument.

The natural response of the integrative taxonomists here is to claim that there is a commonly accepted set of background assumptions that uniquely determine integration. However, as discussed above, evolutionary independence is a vague theoretical notion. This is precisely why taxonomists have to rely on multiple operationalizations for species delimitation in the first place. In other words, operationalizations are required to refine the theory, while the same theory is required to integrate the operationalizations.[12] Consequently, it is not always clear which background assumptions to adopt. Yeates et al. (2011, 210) point out that in the context of tracking

---

[11] Note that by using the term 'false negative' (and 'false positive' in section 4) I am not making the metaphysical claim that there is a particular fixed number of good species, and that the standard approach fails to recognise some of them. Rather, I remain agnostic about this, and use the term with reference to the species that are currently well-established. The point, then, is that the standard approach would fail to pick out many currently well-established species. I thank John Dupré for pointing this out.

[12] This circularity is related to Chang's (1995) 'problem of nomic measurement' and Tal's (2011) 'multiple realizability of unit definitions'. I connect the POD with these measurement-problems in the next section.

evolutionary independence, 'the ways in which any data source [operationalization] may fail are almost limitless'. Two operationalizations can diverge because of limited data, the choice of statistical methods and model of evolution, or simply because of the differential operation of one of the many processes that shape evolutionary change. Given the complexity of the organic world and the practically limitless set of operationalizations, this integration can often be done in many different ways, leading to different results. The upshot of this is that operationalization-dependency enters species delimitation *especially* on the sophisticated integrative approach through the choice between various ways of correcting diverging evidence.[13] In short then, instead of eliminating operationalization-dependency and guaranteeing the objectivity and reliability of species classifications, sophisticated integrative taxonomy seems to add even more fuel to the POD.

3 The model-based account of measurement

The previous section argued that the standard approach to integrative taxonomy is not helpful because it risks overlooking many good species, while it is not clear how the sophisticated approach can guarantee the reliability and objectivity of species delimitation. Given that the problem for the sophisticated integrative approach concerns the reliable determination of the *degree* of evolutionary independence of a lineage, we can think of this as a problem of measurement reliability. In line with this, I now turn to the philosophy of measurement for a solution. This section discusses a problem in the measurement of physical quantities that is highly similar to the POD in species delimitation. I then point to the way the model-based account of measurement makes sense of the objectivity and reliability of measurement despite this problem. This will allow me in section 4 to apply the model-based account of measurement to species delimitation to show that sophisticated integrative species delimitation, like measurement, is objective and reliable in a substantial sense.

The POD in the measurement of physical quantities is easiest to explain using the example of measuring temperature (see Chang 2004; Tal 2017). Temperature is often measured by the rate of expansion of fluids when heated. Different thermometers use different fluids, such as mercury or alcohol, and different containers. This poses a problem, because these fluids and containers do not all expand at the same rate. This means that if one assumes that the readings of each thermometer are linearly correlated with the temperature, they give different results for the same quantities. In order to avoid the conclusion that different thermometers measure different quantities, we have to adopt different, non-linear conversion schemes between the thermometer-readings and the measured temperature. As the choice between these conversion schemes is not determined by the thermometer readings themselves, we must rely on our theoretical understanding of the measurement system and further background assumptions. The same thermometer-readings then can be interpreted as indicating different temperatures depending on which background assumptions are adopted. This seems to lead to a POD in the measurement of temperature, as the outcome of measurement is dependent not

---

[13] See Schlick-Steiner et al. (2010, 428) for an idealized overview of the sophisticated integrative research process which identifies all sources of operationalization-dependency (referred to as 'researcher bias').

only on the measured quantity, but also on the particular set of theoretical assumptions we choose.

This POD is in various ways similar to the POD in species delimitation. First, both evolutionary independence and temperature are abstract theoretical notions that can only be accessed through the various ways in which they are operationalized. Second, these operationalizations sometimes lead to diverging results. This is a problem because there is no standard external to the operationalizations we can turn to in order to decide which operationalizations reliably track the research object, and which are irrelevant or faulty. Thus, the scientists have to rely on background assumptions in order to integrate various operationalizations. Third, this leads to a POD because there may be various plausible ways of integrating different operationalizations, and no straightforward way of choosing between these. That means that the outcomes of research might reflect the particular operationalization rather than the temperature or evolutionary independence they set out to determine. In both cases, this poses the question of how the outcomes of research can be thought to be reliable and objective.

At this point, I would like to refer to the way Eran Tal's model-based view on measurement proposes to answer this question.[14] I focus on this philosophical view on measurement because it was developed specifically to make sense of both the epistemological problems scientists encounter in the practice of measurement, and their success (and sometimes failure) in dealing with these problems. Most broadly, the model-based account views measurement as assigning values to parameters in an idealized model that comprises the measurement process, the measurement environment and the system under measurement. Measurement is considered successful if it can be shown that different operationalizations converge within the limits of divergences that can be theoretically explained and predicted by their models. I lack the space here to provide a full overview of this account. Instead, I will focus on three central aspects that are particularly relevant for the POD:

Indications vs. outcomes Crucial to the model-based view is the distinction between measurement indications, which are the meter-readings or other end-states of the measuring instrument (e.g. the dial is at the mark of 40), and measurement outcomes, which are knowledge claims about the measured quantity inferred from the indications (e.g. the room temperature is 40 ± 1 °C). It is important to see that measurement indications *per se* do not say anything about the measured quantity yet. It is only after they have been interpreted using the theoretical assumptions and conversion schemes discussed above that they become knowledge claims. For example, it may be part of our theoretical understanding of the thermometer that it is only accurate within 1 °C, so the measurement outcome must include an error range.

Idealization The distinction between outcomes and indications brings to light the importance of idealization in measurement. Indications are the result of the interaction between the researcher, instrument, measured quantity, and environment. This means that variations in indications reflect not only variations in the measured quantity, but also various sources of background noise. The aim of moving from indications to outcomes is to pick out the relevant variation, and neglect variation due to background

---

[14] The discussion in this section and the next draws heavily from Tal (2017), see also Tal (2011, 2013)

noise. Because this inference is underdetermined by the operationalizations, scientists also have to rely on background knowledge for this. This is done by representing the instrument in a manner that distinguishes between the idealized measurement procedure and the various ways in which the actual procedure diverges from this ideal. For example, different glass containers expand differentially when heated, and thus influence the thermometer indications. In the model of the thermometer, this influence is accounted for and idealized away as reflecting the particular choice of container rather than variations in the measured quantity. The idealized model thus makes possible the inference of measurement outcomes from measurement indications (e.g. after correction, the dial that is at the 40 mark is interpreted as 39 °C).

Coherence Given that measurement indications are influenced by a wide range of factors, there are usually multiple ways of correcting operationalizations, and hence, of constructing such models. This raises the question of how scientists are to choose between these, and more generally, how they should decide whether a particular measurement outcome successfully measures the quantity. Tal argues that the use of multiple operationalizations together with considerations of coherence play a crucial role in this. He holds that various operationalizations can be considered to track a quantity successfully 'once their idiosyncrasies are idealized away in a mutually coherent fashion' (Tal 2017, 248). For example, the thermometer indicating 40 °C can be considered coherent with one indicating 42 °C if their theoretical models predict that both are only accurate within 1 °C (e.g. because of the impact of the material the containers are made of). Note that there are two levels of models at work here. First, there is a model of each operationalization and the various ways it deviates from the ideal procedure. This model allows us to infer outcomes from the indications of that operationalization. Then there is a higher-level model that incorporates these various operationalizations and allows us to assess their mutual coherence.

In short then, the model-based account defines successful measurement as the convergence of outcomes under the models of multiple operationalizations. By tying the success of measurement to the coherence between various operationalizations of a quantity, this view explains how measurement can be reliable despite its being based on integrating diverging and fallible operationalizations. By locating this coherence in the idealized model of various operationalizations, the view explains how coherence is possible despite divergence between indications.

This model-based view explains how the POD is compatible with the reliability and objectivity of measurement. As Basso (2017) points out, the reliability of measurement on the model-based account depends on the robustness of measurement outcomes over multiple operationalizations. Importantly, this robustness does not lie in the simple convergence of the indications of various operationalizations; rather, it lies in the convergence of the idealized representations of these operationalizations. What robustness indicates then is the fit between the actual convergence of operationalizations and the convergence expected by the idealized model of these operationalizations. This fit can then be tested by checking whether the predictions made on the basis of the model are borne out. A bad fit indicates a lack of coherence due to error that has not been accounted for. A good fit, on the other hand, implies that the indications are not erroneous beyond what we can correct for in the model, and hence, that the outcomes reflect the quantity we set out to measure.

At first sight, one might think that this model-based account of the reliability of measurement is incompatible with the objectivity of measurement. Given that successful measurement is defined with reference to idealized models, it is clear that measurement outcomes are dependent on the choice of model and theoretical assumptions. Thus, they are not objective in the sense of independence from any perspective. However, Tal argues that there is another substantial sense in which measurement on the model-based account is objective. He argues that because the success of measurement depends on the convergence between multiple idealized operationalizations, it is objective in the sense of 'perspective invariance' (Tal 2017, 248). That is, even if good measurement outcomes are not perspective-independent, they are also not dependent on one particular perspective. Instead, convergence of different operationalizations guarantees that the outcome is robust across a range of perspectives. This 'robust perspectivism', as Tal (2017, 248-249) calls it, shows how the POD and objectivity of research outcomes are compatible. First, the dependency of measurement on idealized models poses no threat to the perspective-invariance of measurement outcomes. Second, these models even form a necessary requirement for outcomes to be considered successful in the first place: it is only in the idealized model of the measurement process that various diverging operationalizations can be seen as coherent.

4 Applying the model-based account to species delimitation

In section 2, I argued that it is unclear how sophisticated integrative taxonomy provides a solution to the POD given that the threat of operationalization-dependency still persists through the choice of background assumptions. This section argues that an interpretation of the sophisticated approach to integrative taxonomy along the lines of the model-based approach to measurement shows how species delimitation can be reliable and objective despite the POD. In order to do this, I discuss a recent integrative taxonomic study and show how the model-based account of measurement can be applied to it.

The example concerns a taxonomic study on *Opiliones* (colloquially known as daddy longlegs) of the genus *Megabunus* by Wachter et al. (2015), and more precisely the part of their study that focuses on the classification of the organisms recognised as *Megabunus rhinoceros* (*M. rhinoceros* from hereon) before the study. The researchers, who explicitly adopt the view of species as independently evolving lineages, integrate multiple operationalizations to track independently evolving lineages in this group. They sample six nuclear markers (nuDNA) and one fragment of a mitochondrial DNA (mtDNA), which are analysed using two and three different methods respectively.[15] In addition, they sample chemical characters, traditional morphometric characters, and use geometric morphometrics to track changes in the shape of the head and thorax. These operationalizations are chosen because they have previously been used successfully for species delimitation in related taxa.

Using a discovery approach (i.e. using these operationalizations to find lineages without starting from a particular classification hypothesis), these operationalizations

---

[15] They use the GMYC model, the bPTP model and a statistical parsimony network for the mtDNA, and STRUCTURE and mr.Bayes for the nuDNA. For more on these and other common methods, see Camargo and Sites (2013) and Leavitt et al. (2015).

lead to widely different classifications of the organisms traditionally assigned to *M. rhinoceros*. The nuDNA data pick out one or three lineages, depending on the method of analysis. The mtDNA fragment suggests recognising six, four or three lineages, depending on the method of analysis. No lineage was consistently recovered over the different DNA-based operationalizations. Finally, the chemical evidence and both sets of morphological evidence pick out a single lineage. Thus, the integration of these operationalizations requires the authors to resolve the divergence between at least five different classifications.

They do this by relying on a combination of converging operationalizations and evolutionary explanations for incongruent evidence. They propose to recognise the three lineages picked out by the nuclear markers as independently evolving lineages. The divergence between the nuDNA and mtDNA operationalizations is explained by referring to incomplete lineage sorting and introgression of mtDNA.[16] Divergence between multiple analysis methods of mtDNA is explained by particular shortcomings of the methods. Divergence between the nuDNA and the morphometric evidence is explained by referring to morphological stasis. Finally, they claim that the failure of the chemical evidence to pick out the lineages is not surprising given the relatively high failure rate of this method in related groups. According to the authors, these explanations account for the fact that the three lineages are evolving independently despite the failure of various operationalizations to pick them out.

These results are then taken as the input for hypothesis-driven research which evaluates the proposed classification using the same operationalizations with different data or methods of analysis. These hypothesis-based tests clearly confirm the classification. For example, even though geometric morphometrics failed to pick out the three lineages, significant phylogenetic signal was found in the data, which suggests that morphological and molecular characters evolved together.[17] Similarly, the researchers found new morphological differences in male genitalia, forelegs and mouthparts that reliably pick out the three lineages. The authors conclude their case by formally recognising three independently evolving lineages.

Wachter et al. (2015) explicitly claim that their integrative approach increases the rigor (p. 877), reliability (p. 863) and objectivity (p. 864) of their results. This echoes claims in many other studies using a sophisticated integrative approach. To show how this is possible despite the POD, the remainder of this section will show how the model-based approach of measurement applies to Wachter et al.'s study of *M. rhinoceros*.

First, there is a clear distinction in the case-study between the various signs of evolutionary independence, such as chemical or morphometrical distinctness, and the way these signs are interpreted in the final knowledge claims about the status of the lineages. This distinction, which is parallel to that between measurement indications and outcomes, is illustrated by the difference in meaning between the results of various

---

[16] Introgression and incomplete lineage sorting are two phenomena that commonly cause divergence between the genealogical histories of various genes within one organism or species. Introgression occurs when genes of one species invade another species through backcrossing of hybrids with one of the parental species. Incomplete lineage sorting occurs through the differential sorting of ancestral polymorphisms in different lineages.

[17] Phylogenetic signal is a measure for the dependence between species' characters that is due to phylogenetic relations. In the context of morphometric data, it can be defined as the tendency of related species to be more similar to each other than to species drawn at random from the same phylogenetic tree.

operationalizations at the start and end of the study. As noted above, the various mtDNA-based methods initially support six, four or three species. However, this is not the form in which the mtDNA evidence figures in the knowledge claims about evolutionary independence that the authors posit at the end of the research. While it still plays a role there, it has substantially changed and is taken to support the three-species classification suggested by the nuDNA. Similar examples are provided by the morphological evidence. The point is that these various operationalizations *per se* are not interpreted as knowledge about the independence of a lineage. It is only after being interpreted in line with other evidence and background assumptions that they are used to make claims about independently evolving lineages.

Second, idealization plays a crucial role in inferring outcomes of species delimitation from indications of evolutionary independence. Just like physicists idealize the way the thermometer tracks temperature, Wachter et al. idealize the way various operationalizations track evolutionary independence. This is shown by the fact that they start by assuming a direct relation between the various operationalizations and evolutionary independence: if there are two lineages, then we should expect two morphologically distinct groups, two monophyletic groups, etc. As it is generally known that morphology, chemical characters and monophyly of nuDNA and mtDNA very often fail to track evolutionary independence, this is a highly simplified and idealized model of these operationalizations and their relation to evolutionary independence. Divergence from these idealized representations of the operationalizations is then accounted for by evolutionary explanations, just like error in temperature measurement was accounted for by referring to the impact of the glass container. Thus, these explanations connect the actual operationalizations to their idealized models, and allow the taxonomists to infer information about evolutionary independence from their indications.

For example, the authors explicitly adopt monophyly for mtDNA as a criterion for evolutionary independence. This is an idealization, as the effects of incomplete lineage sorting and introgression make it unlikely that any actual lineage is monophyletic for all its genes. Understanding the criterion as an idealization helps explain why the lack of monophyly of the three nuDNA lineages for the mtDNA fragment is not interpreted as evidence against their evolutionary independence. Instead, the results of the various mtDNA methods are corrected on the basis of background knowledge about these methods and evolutionary processes. More precisely, as it is known that the GMYC-model tends to track population structure rather than species divergence in groups with life history traits like *M. rhinoceros*, this method is taken to overestimate the number of lineages. Similarly, because of the slower rate of evolution and larger effective population size of the nuDNA compared to the mtDNA, it is expected that there are ancestral polymorphisms retained in the nuDNA and not in the mtDNA. Finally, introgression of mtDNA genes between two of the nuclear lineages can be expected given the close proximity of their sampling locations and likely secondary contact between them after expansion between glacial periods.[18] These three points explain the extent to which the mtDNA operationalizations failed to track evolutionary independence. This allows the

---

[18] The role of glacial refugia and expansion between glacial periods and their role in diversification in these taxa is investigated in Wachter et al. (2016). The findings in this study agree well with the conclusions of the 2015 paper, thus further tightening the coherence.

authors to interpret these operationalizations as supporting the hypothesis, even though they do not meet the criterion of evolutionary independence (i.e. being monophyletic for the tested genes) that the authors adopted at the start.

Third, the idealization of the operationalizations is necessary for Wachter et al. to interpret them in a mutually coherent manner. This coherence then plays a crucial role in how they determine the success of species delimitation. This is noted explicitly by Wachter et al. (2015, p. 2; see also their reference to Schlick-Steiner et al. 2010) when they state that species delimitation under their integrative approach is successful only if all evidence is congruent, or if there are plausible evolutionary explanations for any incongruent evidence. The reference to plausible evolutionary explanations indicates that the required coherence is not that between multiple indications of evolutionary independence, but rather that between the outcomes interpreted in light of other operationalizations and background knowledge. Thus, the convergence of operationalizations is taken to be convergence under the idealized models of the operationalizations. The robustness of the result over the idealized outcomes of multiple operationalizations is then taken as an indication of its reliability.

It is instructive here to compare the role of coherence and robustness in this sophisticated interpretation with that in the standard interpretation. As discussed in section 2, the standard interpretation infers the reliability of species delimitation from the robustness of the result over multiple operationalizations. For example, if mtDNA and nuDNA operationalizations independently pick out the same groups as species, the standard approach takes this to indicate the reliability of species delimitation. This implicitly relies on a no-coincidence argument: it is unlikely that these operationalizations would indicate the same result unless they are both tracking the world reliably. Given that there is no actual convergence in the case study, and the authors explicitly *mould* the data to be convergent, this no-coincidence argument is not doing the work here. Of course, the convergence of multiple operationalizations still plays a role. For example, Wachter et al. consider their hypothesis successful because the mtDNA and nuDNA operationalizations converge within the error ranges and corrections explained by incomplete lineage sorting and introgression. However, this convergence is not that between the outcomes of multiple operationalizations interpreted independently, as is the case on the standard approach. Instead, the convergence here is that between multiple operationalizations that were originally incongruent, and have then been interpreted and adapted in light of each other and background knowledge. Thus, the reliability of species delimitation in this case does not derive from the no-coincidence argument, but from the fit between the convergence (and divergence) that is predicted and explained by the models, and the actual convergence (and divergence) found by the operationalizations. In this way, the sophisticated approach relies on convergence between multiple operationalizations for reliable species delimitation while avoiding the standard approach's problem of false negatives.

At this point, there is an important worry that needs to be addressed. One could argue that given the theoretical vagueness of evolutionary independence and the complexity of the organic world, evolutionary explanations for diverging evidence are often easy to come by. For example, incomplete lineage sorting and introgression, both cited by Wachter et al., occur so frequently in the organic world that one might think they can be easily invoked to explain the divergence between various DNA-based operationalizations. If this is true, then the coherence of idealized operationalizations, which the model-based view takes to indicate reliability, is also easy to come by. The

risk then is that operationalizations are mistakenly taken to be reliable, and groups are mistakenly accepted as independently evolving lineages, due to taxonomists trying to avoid inconsistency by 'cleaning up' the data. Thus, it seems that the sophisticated approach to integrative taxonomy only avoids the standard approach's risk of false negatives by risking equally undesirable false positives.[19]

While it is true that the sophisticated approach runs a higher risk of false positives than the standard approach, this does not imply that it fails to increase the reliability of species delimitation. To see this, it is crucial to note that taxonomists can test the fit between the actual and modelled convergence. If further tests reveal outcomes that fall outside the error ranges of explanations predicted by the model, this indicates that at least one of the operationalizations is unreliable in a way that is not accounted for in the model, or that one or more of the explanations are incorrect. In other words, the problem noted above can be mitigated by conducting further tests of the coherence and fit that are taken as indicators of reliability. For example, going back to incomplete lineage sorting, Degnan and Rosenberg (2009) show that the pattern of incomplete lineage sorting is highly predictable. Thus, if researchers use incomplete lineage sorting as an explanation, they can test the robustness of the model by investigating whether the deviation from the idealized model (i.e. lack of monophyly for mtDNA) has the expected pattern. If the pattern they find deviates from the expected pattern of incomplete lineage sorting, there is no coherence between the various idealized models.

The hypothesis-driven reanalysis of the morphological evidence in Wachter et al.'s study also provides a good illustration of such a test. The authors explain the divergence between the morphological evidence (one lineage) and the species hypothesis (three lineages) by referring to morphological stasis. This explanation is plausible because well-established, morphologically indistinguishable species have been recognised in related taxa that, just like *M. rhinoceros*, have strongly isolated populations and low dispersal abilities (Keith and Hedin 2012). Importantly, despite the morphological stasis, these related species showed morphological divergence of male genitalia. According to the model then, *M. rhinoceros* groups are expected to show morphological stasis, like their relatives, except with respect to male genitalia. The authors test for this, collect new data, and, in line with the expectation, find significant morphological differences. Thus, this reanalysis acts as a test of the fit between the actual divergence and expected divergence, and suggests that the coherence is not merely the result of taxonomists cleaning up the data. A similar example comes from the test for phylogenetic signal in the change of the head and thorax shape. While geometric morphometrics did not pick out the three lineages, the model predicts that there should be phylogenetic signal in any shape changes. This prediction of the model is confirmed by the follow-up test.

In short then, the sophisticated approach holds that the reliability of species delimitation is indicated by the fit between the actual convergence and the convergence explained and predicted by the model. Crucially, this fit can then be tested by follow-up research like Wachter et al.'s hypothesis-based tests. In this way, the sophisticated integrative approach avoids the POD and increases the reliability of species delimitation despite relying on the extensive use of background assumptions.

Having shown how the model-based interpretation of the sophisticated approach increases reliability, we can now turn to the equivalent question of objectivity.

---

[19] I thank Eran Tal and Hasok Chang for pointing out this worry to me.

It is clear that species delimitation is not objective in the sense of perspective-independence, because the attribution of evolutionary independence relies on a particular idealized representation of the operationalizations. However, this does not mean that the outcome of species delimitation simply reflects the particular operationalization of choice. The convergence of various operationalizations (and further tests of coherence) guarantee that the attribution of evolutionary independence is not merely the artefact of methodological choices, but is stable across a range of operationalizations and viewpoints. Thus, species delimitation on the sophisticated integrative approach is objective in the sense of a robust perspective-invariance. This perspective-invariance suggests that integrative species delimitation is not likely to turn out different for each taxonomist and her particular methodological choices, and thus can lead to stable and repeatable species classifications. Note that this robust perspective-invariance applies only to the idealized representation of the operationalizations; this illustrates the crucial role that the model plays in the way sophisticated integrative taxonomy increases the objectivity of species delimitation.

5 Conclusion

This paper has connected species delimitation and the measurement of physical quantities in order to show that the former, like the latter, can be considered reliable and objective despite the POD. More specifically, it has argued that the sophisticated approach to integrative taxonomy can be seen as assessing the convergence between the outcomes of multiple idealized operationalizations of evolutionary independence. This convergence of outcomes under idealized models indicates the reliability of the resulting species classifications. Moreover, the robustness and perspective-invariance of the resulting classifications provide good reasons to believe that they are objective in a substantial sense.

Although further development of these ideas is needed, I think they establish at least three significant points. First, they form an argument against the standard approach to integrative taxonomy, which holds that species hypotheses require the congruence of multiple operationalizations of evolutionary independence, and should not be accepted if various lines of evidence disagree (Meiri and Mace 2007; Alström et al. 2008; Vieites et al. 2009). In addition to this, the arguments support the sophisticated approach and are in line with recent papers on integrative taxonomy by Schlick-Steiner et al. (2010) and Padial and De La Riva (2010). Second, the POD as presented in this paper is an important problem that has remained mostly under the radar in the philosophical debate on species. I believe that raising this problem is valuable independently of whether the solution offered here is considered successful. Related to this, the paper shows how an investigation of the operationalization of species concepts and related epistemological questions can proceed without getting tangled up in the metaphysical debates of the species-problem. This is important, as the past decade has seen an explosion of new and varied methods of species delimitation, most significantly multi-species coalescent-based methods, and a philosophical investigation of these methods and their assumptions is highly due (Camargo and Sites 2013). Finally, and more speculatively, this paper proposes the idea of thinking about species delimitation as a process of *measuring* evolutionary independence. In doing so, the paper hints at a way in which the

notoriously vague notion of evolutionary independence could be clarified. Just like the use of multiple operationalizations played a crucial role in establishing a precise and well-defined scale of physical quantities (see Chang 2004 for the example of temperature), the practice of integrative taxonomy could lead to a more precise notion of evolutionary independence.

References

Alström, P., Rasmussen, P. C., Olsson, U., & Sundberg, P. (2008). Species delimitation based on multiple criteria: the Spotted Bush Warbler Bradypterus thoracicus complex (Aves: Megaluridae). *Zoological Journal of the Linnean Society, 154*(2), 291–307. https://doi.org/10.1111/j.1096-3642.2008.00418.x.

Barker, M. J. (2007). The empirical inadequacy of species cohesion by gene flow. *Philosophy of Science, 74*(5), 654–665. https://doi.org/10.1086/525611.

Barker, M. J., & Wilson, R. A. (2010). Cohesion, gene flow, and the nature of species. *The Journal of Philosophy, 107*(2), 61–79. https://doi.org/10.5840/jphil201010721.

Basso, A. (2017). The appeal to robustness in measurement practice. *Studies in History and Philosophy of Science Part A.* https://doi.org/10.1016/j.shpsa.2017.02.001.

Baum, D. A. (2009). Species as ranked taxa. *Systematic Biology*, 74–86. https://doi.org/10.1093/sysbio/syp011.

Camargo, A., & Sites, J. (2013). Species delimitation: A decade after the renaissance. In I. Pavlinov (Ed.), *The species problem - ongoing issues* (pp. 225–247). Rijeka: InTech.

Chang, H. (1995). Circularity and reliability in measurement. *Perspectives on Science, 3*, 153–172.

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

Dayrat, B. (2005). Towards integrative taxonomy. *Biological Journal of the Linnean Society, 85*(3), 407–415. https://doi.org/10.1111/j.1095-8312.2005.00503.x.

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology, 56*(6), 879–886. https://doi.org/10.1080/10635150701701083.

Degnan, J. H., & Rosenberg, N. A. (2006). Discordance of species trees with their most likely gene trees. *PLoS Genetics, 2*(5), e68.

Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution, 24*(6), 332–340.

Ereshefsky, M. (1991). Species, higher taxa, and the units of evolution. *Philosophy of Science, 58*(1), 84–101.

Ereshefsky, M. (1992). Eliminative pluralism. *Philosophy of Science, 59*(4), 671–690.

Fujita, M. K., Leaché, A. D., Burbrink, F. T., McGuire, J. A., & Moritz, C. (2012). Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution, 27*(9), 480–488. https://doi.org/10.1016/j.tree.2012.04.012.

Haber, M. H. (2012). Multilevel lineages and multidimensional trees: the levels of lineage and phylogeny reconstruction. *Philosophy of Science, 79*(5), 609–623.

Haber, M. H. (2016). The individuality thesis (3 ways). *Biology & Philosophy, 31*(6), 913–930.

Hull, D. L. (1997). The ideal species concept-and why we can't get it. In M. Claridge, H. Dawah, & M. R. Wilson (Eds.), *Species: the units of biodiversity* (pp. 357–380). London: Chapman & Hall.

Isaac, N., Mallet, J., & Mace, G. M. (2004). Taxonomic inflation: its influence on macroecology and conservation. *Trends in Ecology & Evolution, 19*(9), 464–469. https://doi.org/10.1016/j.tree.2004.06.004.

Keith, R., & Hedin, M. (2012). Extreme mitochondrial population subdivision in southern Appalachian paleoendemic spiders (Araneae: Hypochilidae: Hypochilus), with implications for species delimitation. *Journal of Arachnology, 40*(2), 167–181. https://doi.org/10.1636/A11-49.1.

Leavitt, S. D., Moreau, C. S., & Lumbsch, H. T. (2015). The dynamic discipline of species delimitation: Progress toward effectively recognizing species boundaries in natural populations. In *Recent Advances in Lichenology* (pp. 11–44). New Delhi: Springer.

Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology, 46*(3), 523–536.

Mayden, R. L. (1999). Consilience and a hierarchy of Species concepts: advances toward closure on the species puzzle. *Journal of Nematology, 31*(2), 95.

Meiri, S., & Mace, G. M. (2007). New taxonomy and the origin of species. *PLoS Biology, 5*(7), e194. https://doi.org/10.1371/journal.pbio.0050194.

Mishler, B. D. (1999). Getting rid of species. In R. A. Wilson (Ed.), *Species: New Interdisciplinary Essays* (pp. 141–185). Cambridge: MIT Press.

Padial, J. M., & De La Riva, I. (2010). A response to recent proposals for integrative taxonomy. *Biological Journal of the Linnean Society, 101*(3), 747–756 https://doi.org/10.1111/j.1095-8312.2010.01528.x.

Padial, J. M., Castroviejo-Fisher, S., Köhler, J., Vilà, C., Chaparro, J. C., & De la Riva, I. (2009). Deciphering the products of evolution at the species level: the need for an integrative taxonomy. *Zoologica Scripta, 38*(4), 431–447. https://doi.org/10.1111/j.1463-6409.2008.00381.x.

Padial, J. M., Miralles, A., De la Riva, I., & Vences, M. (2010). The integrative future of taxonomy. *Frontiers in Zoology, 7*(16), 1–14.

Reydon, T. A. C. (2005). On the nature of the species problem and the four meanings of "species". *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 36*(1), 135–158. https://doi.org/10.1016/j.shpsc.2004.12.004.

Schlick-Steiner, B. C., Steiner, F. M., Seifert, B., Stauffer, C., Christian, E., & Crozier, R. H. (2010). Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual Review of Entomology, 55*(1), 421–438. https://doi.org/10.1146/annurev-ento-112408-085432.

Sites, J. W., & Marshall, J. C. (2004). Operational criteria for delimiting species. *Annual Review of Ecology, Evolution, and Systematics, 35*, 199–227.

Sterner, B. (2017). Individuating population lineages: a new genealogical criterion. *Biology and Philosophy, 32*(5), 683–703. https://doi.org/10.1007/s10539-017-9580-4.

Tal, E. (2011). How accurate is the standard second? *Philosophy of Science, 78*(5), 1082–1096. https://doi.org/10.1086/662268.

Tal, E. (2013). Old and new problems in philosophy of measurement. *Philosophy Compass, 8*(12), 1159–1173. https://doi.org/10.1111/phc3.12089.

Tal, E. (2017). A model-based epistemology of measurement. In A. Nordmann & N. Mößner (Eds.), *Reasoning in measurement* (pp. 233–253). London and New York: Routledge.

Vieites, D. R., Wollenberg, K. C., Andreone, F., Köhler, J., Glaw, F., & Vences, M. (2009). Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. *Proceedings of the National Academy of Sciences, 106*(20), 8267–8272. https://doi.org/10.1073/pnas.0810821106.

Wachter, G. A., Muster, C., Arthofer, W., Raspotnig, G., Föttinger, P., Komposch, C., … Schlick-Steiner, B. C. (2015). Taking the discovery approach in integrative taxonomy: decrypting a complex of narrow-endemic Alpine harvestmen (Opiliones: Phalangiidae: Megabunus). *Molecular Ecology, 24*(4), 863–889. https://doi.org/10.1111/mec.13077

Wachter, G. A., Papadopoulou, A., Muster, C., Arthofer, W., Knowles, L. L., Steiner, F. M., & Schlick-Steiner, B. C. (2016). Glacial refugia, recolonization patterns and diversification forces in Alpine-endemic Megabunus harvestmen. *Molecular Ecology, 25*(12), 2904–2919. https://doi.org/10.1111/mec.13634.

Wheeler, Q., & Meier, R. (2000). *Species concepts and phylogenetic theory: a debate*. New York: Columbia University Press.

Wiley, E. O. (1978). The evolutionary species concept reconsidered. *Systematic Biology, 27*(1), 17–26. https://doi.org/10.2307/2412809.

Will, K., Mishler, B., & Wheeler, Q. (2005). The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology, 54*(5), 844–851. https://doi.org/10.1080/10635150500354878.

Willis, S. C. (2017). One species or four? Yes!...and, no. Or, arbitrary assignment of lineages to species obscures the diversification processes of neotropical fishes. *PLoS One, 12*(2), e0172349. https://doi.org/10.1371/journal.pone.0172349.

Yeates, D. K., Seago, A., Nelson, L., Cameron, S. L., Joseph, L., & Trueman, J. W. H. (2011). Integrative taxonomy, or iterative taxonomy? *Systematic Entomology, 36*(2), 209–217. https://doi.org/10.1111/j.1365-3113.2010.00558.x.