VII-2008

Hello Carl!

Ok, one awkward moment down. Let me start by saying that I am a big fan of yours, though admittedly in a selective way that favors my own positions about taxonomy but mostly ignores many of your other accomplishments and occasional failings. Therefore I will not waste anyone's time with remembering the latter, but instead will let you in on some post-millennial developments in the field of taxonomy that you have prepared for us so brilliantly. It's a personal view on what your system of nomenclature means and where we are heading today.

As a matter of provision, I will take your taxonomic practice to best represent your more foundational positions about nomenclature. (I have some reason to think that this will not offend you too much; let's not revisit the theory of reproductive systems you used to create those 24 'plant' classes.) For example, I understand that a large number of your species descriptions were associated with particular designated specimens that were later reinterpreted as types. The 'type method' did not become a formal requisite until the mid-19th century, yet in some sense one could claim—as I prefer to—that you were an early practitioner. Similarly, when you classified fleas (which are secondarily wingless) closer to other winged insects than to ancestrally wingless insects, one could argue that you had an intuitive handle on such concepts as global parsimony ("what is the most plausible placement given all available character evidence?") and evolutionary reversal (acquisition, then loss of wings), even though there were no theories available at the time for you to properly accommodate these judgments. Undoubtedly some will say that this line of interpretation is giving you too much credit. I am sure there's a sophisticated reply to that charge. But just between us—giving you lots of credit is what big fans will do.

Fast forward to the 21st century. Near the turn of the millennium I began studying insect systematics for my doctoral degree in Ithaca, New York. Before that my understanding of nomenclature was basic, though I was fortunate to have had several years of Latin in high school[1]. As an incoming student at Cornell University I was instantly attracted to a particular school called phylogenetic systematics or

The weevil *Cleonis piger* (Scopoli, 1763) was described as *Curculio sulcirostris* by Linnaeus in 1767

cladistics. Cladists had come to identify themselves as a group some 25 years ago based on ideas that had more to do with the proper recognition of natural taxa than with rules of nomenclature. Beyond that they seemed to share a disposition (turned youth recruitment strategy) for talking tough, both to members and non-members, with varying motivations and results.

I soon realized that systematists are a dynamic and sometimes pugnacious bunch! There were a seemingly infinite number of historical and contemporary issues of contention to catch up on. How can we postulate natural groups? What are species? What *were* these authors smoking? I was sorting through the particular mix of posturing and solid argumentation that was going on in each case, attempting to identify my own mix. In retrospect my early attempts at defending principled stands about particular systematic methods were both fun and pathetic[2]. Naturally some positions changed through practice and experience. The initial cladistic chip on my shoulder morphed into a perspective where considerations of theoretical plausibility could win over a strict adherence to methodological purity.

There are two points I'd like you to take from this. First, my generation of systematists was trained by one that fought passionately for the recognition of their school. On many counts they've succeeded, and as a result cladistic methods are now widespread. The new generation is perhaps less combative, enjoying a certain status and confidence that weren't there for many years. This is not to say that all is rosy; new challenges pop up at every turn (as you will see). But I presume you would be pleased to see that we are still describing taxa in ways that are in line with the foundations you've laid, and that lately our field is reclaiming its vital status in biology. Second, although phylogenetic systematists radically opposed the taxonomic groupings of competing schools, at the time all schools accepted the system of Linnaean nomenclature. The applicability of your system to cladistics in particular is rooted in a shared use of hierarchical, tree-like summaries reflecting the character arrangement and relationships among perceived natural taxa. It would be hard to believe that Linnaean names survived so many episodes of bloodletting in systematics unless you had gotten something profoundly right about how we should describe and name natural groups.

Or perhaps not? Some 20 years ago a group of systematists launched a critique of numerous aspects of Linnaean nomenclature. These authors ultimately proposed an alternative under the heading of phylogenetic nomenclature. This letter will not allow me to do full justice to this alternative and its various refinements, nor will I have sufficient space to represent the diverse arguments, ranging from foundational

to practical, that defend the Linnaean system against this newer proposal. Instead I will cherry-pick some bits and pieces from both sides to help set the stage for the main points that will follow.

From a certain perspective, our classifications have changed much over the past 250 years, and for too many groups there is no end in sight! Proponents of phylogenetic nomenclature have criticized many aspects of established nomenclatural practice, including a perceived lack of referential stability, the use of ranks (genus, species, etc.), the use of type species and specimens, and the use of phenotypic features to define species and higher taxa. Some criticisms are on theoretical grounds; viz. the alleged failure to differentiate between classes (which have essential properties) and individuals (which merely are parts of a whole), or the failure to define taxa in terms of phylogenetic ancestry. On the other hand, arguments against Linnaean ranks and other seemingly destabilizing rules for adjusting names are of a more practical nature.

So what's really going on? Surely, if we choose to abandon a working language, we had better be right about the nature of the problem, or risk throwing the baby out with the bathwater. Few systematists—or biologists, for that matter—think that the Linnaean system is perfectly suited for all communication needs. Advocates of any alternative system can thus exploit a certain lack of content with the *status quo*. But this does not free them from the task of finding out what shortcomings really matter, and how to overcome them.

It's a pretty safe bet that Linnaean nomenclature is *not* fundamentally incompatible with modern evolutionary or philosophical reasoning. It's relatively easy to see why. In defining taxa we traditionally make reference to particular phenotypic features; e.g., spiders have spinnerets. On the surface this may look as though we are thereby stipulating that taxa must have universal and immutable characteristics, which would amount to a non-evolutionary view. But perhaps we are just sloppy with our use of language. When we say "spiders are defined by their spinnerets" in a systematic treatment, what do we actually mean? Suppose a young lineage of spiders lost its spinnerets. We would still consider these taxa as spiders, adding in our minds a qualifying phrase to the original definition, such as "unless the spinnerets were lost in an event which—according to our most reliable phylogenetic estimates—occurred later on in a particular subgroup of spiders". What if spinnerets evolved elsewhere, say, in ants? Then we might clarify in this manner: "the term 'spinnerets' in spiders really refers to a series of (inferred, historical) transformations at the molecular level that are expressed phenotypically as silk-producing glands in a lineage whose sister group includes whip-scorpions, though not wasps". In the same manner, we could account for the loss of spinnerets at the molecular

level by stipulating the origin of a secondary genetic transformation that inhibits the expression of spinnerets.

With the proper semantic modifications, the term 'spinnerets' can thus refer precisely to that series of historical evolutionary events that resulted in an ancestral spider species which subsequently underwent diversification into some 38,000 descendant species. Similarly, we can say that snakes are defined by their loss of legs, where 'loss of legs' ultimately refers to a secondary event at the molecular level that inhibits the expression of legs, thus making it a condition unique to a particular lineage of tetrapods and distinct from ancestrally legless vertebrates such as lampreys. The point is, while these referential refinements seem too cumbersome for everyday use, they clearly rely on modern concepts of evolutionary change and common ancestry. Indeed, many Linnaean definitions of taxa may benefit from such refinements[3]. There is no need, however, to regard feature-based definitions as incompatible with evolutionary thinking when the main villain is linguistic laziness.

What about abandoning Linnaean name definitions for philosophical reasons? As the argument goes, taxa are evolving and are therefore more like historical individuals than classes with stable properties. In truth, taxa are somewhere in the middle of these two categories. Their individual or class-like nature is more or less relevant depending on the kinds of inferences one intends to make in a particular context. The challenge for proponents of phylogenetic nomenclature is thus twofold. They must prove that (1) nearly all of the most critical contexts in which taxa are mentioned concern their nature as individuals; and (2) still more importantly, that applying the philosophical theory of classes versus individuals is more appropriate than using competing philosophical concepts that readily accommodate established practice, e.g., by relaxing the criterion of immutability in classes.

Let's look at an example of a property-free definition of passerine birds[4]: "Passeri are the most inclusive clade containing *Passer domesticus* and any extant species and including *Corvus monedula* but not *Tyrannus tyrannus, Pitta sordida, Furnarius rufus,* and *Thamnophilus doliatus*." Even though in principle this definition is independent of a tree diagram, we need to visualize some sort of tree to understand what "Passeri" means. Yet even then, what *can* we understand given such a phrase? Can we recognize a

The House Sparrow, *Passer domesticus* L. (Passeridae) from John Gould's *Birds of Europe* (1837)

member of the Passeri in a collection or in nature? Can we explain what Passeri are to colleagues, students, or to children? Can we associate with this name any prominent adaptations and relate these to other genetic, ecological, or behavioral information we have come to learn about passerine birds? If the answers are negative, then we have proposed a definition that is largely devoid of cognitive content and unable to support the kinds of inferences that biologists intend to make! At that point, when taxonomic names are no longer the primary vehicles for the causal phenomena we wish to learn and talk about—a situation that would seem unique in any science or language in general—it is time to question the philosophical wisdom that got us there.

Phylogenetic definitions such as the one above may hold stable across multiple succeeding taxonomic perspectives. Yet again, what are the pay-offs of such nomenclatural stability when the associated age, referential extension, and evolutionary properties of a name can change significantly from one phylogeny to the next? As a general guideline, nomenclatural rules should strive to bring out as much stability as possible in succeeding classifications, but should never be designed to obscure different hypotheses about phylogenetic relationships and thus weaken the name/taxon link over time. If a system of nomenclature is to remain stable in light of significantly diverging views about how nature is organized, then that system can no longer represent progress in systematics.

The Jackdaw, *Corvus monedula* L. (Corvidae) from John Gould's *Birds of Europe* (1837)

Let's move along. The consistent use of Linnaean ranks often leads to difficult real-life decisions. Seemingly minor adjustments in phylogenetic arrangement may require a series of inconvenient changes in nomenclature. Yet again, any practical shortcomings need to be weighed against the immense inferential benefits of forming a match between temporally succeeding evolutionary events and hierarchically nested taxa and names. For example, by knowing that a spider is a member of the Salticidae, one can infer more than 158 million three-taxon statements of mutually exclusive groupings among spiders alone[5]. As inductively working animals, we simply like these kinds of cognitive shortcuts too much to abandon them for a language that requires infinitely more acts of unconnected memorization.

You see, Carl, your system has come a long way. I venture to say that this was not an accident, or a product mainly of convention. Your proposal to embed hierarchical information into Linnaean names, to root these names in both physical specimens (ostension) and in perceived phylogenetically relevant properties (intension), is fundamentally sound. It is sound and successful not because it follows the tenets of an appealing philosophy, but because it is particularly well suited for human learning about nature and for accommodating the kinds of inferences that are of primary interest to biologists. Accommodation is the capacity of a linguistic tradition to align itself with the causal structure of the world and thus enable efficient communication about relevant natural phenomena. A successful language must excel at accommodation, a quality that emerges gradually over time.

Linnaean names have an impressive track record of facilitating scientific progress across all biological disciplines. This list certainly includes evolution, as it turns out that evolutionary biologists focus much of their research on the distribution, temporal sequence of transformation, and biological significance of *organismal properties*! To uphold their practice over time these biologists must rely on the responsiveness of taxonomic names to newer and more accurate hypotheses about the phylogeny of a group of organisms and their characteristics. In other words, the process of 'voting' for or against a system occurs in no small measure outside of the context of systematics and philosophy. Linnaean names had to pass many external and independent trials of accommodation in order to establish and maintain their current standing.

Still, all is not well today. Clearly, we should have a more thoroughgoing philosophical account explaining why and to what extent Linnaean names and their semantic components succeed at accommodation. That account is limited, however, by how much we understand about successful reference in general. We know for instance that Linnaean names are of a hybrid nature that tends to yield more than the sum of its parts. Accepted naming practice involves an event of baptism, e.g., the type designation for a perceived taxon. Such an event can trigger a causal chain of speakers who may understand each other even though everybody misjudges what characteristics pick out the taxon among its relatives (= reference in spite of misdescription). Similarly, the property-based description of a taxon may be incomplete or imprecise but performs sufficiently well in a number of relevant situations (= partial reference). And so the acts of pointing and characterizing jointly refine the meaning of taxonomic names.

Unfortunately, that's not all. My limited knowledge of philosophy of language has taught me that humans have a knack for understand-

ing each other even when prominent philosophical theories say they shouldn't! Success, partial success, and failure in communication are profoundly situational. Each outcome depends in part on whether one speaker makes the right assumptions about the other speaker's training background and present usage of a taxonomic name. To make matters worse, we rarely provide more contextual information than we assume is needed, preferring instead to adopt certain usages of terms until a misunderstanding is apparent. This ubiquitous habit of ours presents a challenge for philosophers, who must account for contextuality and variable underlying assumptions in order to explain why reference can work in some cases though not in others. It also constitutes a real-life problem for systematics, especially in taxonomic groups with a history of significant rearrangements.

Which leads me to my main point. The issues we're tackling today with Linnaean names are not really rooted in the naming process *per se*. I think that proponents of phylogenetic nomenclature correctly sensed that there was a problem, but got the diagnosis mostly wrong. The real issues arise through a combination of (1) how the naming of taxa is legally regulated (through the Codes, etc.), (2) how these rules are implemented and supplemented with additional information, and (3) how these two processes interact over time. Many users who are unsatisfied with 'the system' primarily feel that there is a lot of baggage in taxonomy. It's difficult to impossible to sort through that baggage, leading to linguistic imprecision or even paralysis in certain taxonomic groups. These users have a point, though the problem is more likely rooted in a history of inadequate systematic inferences and poor linguistic implementation than nomenclatural rules.

Let's be honest, Carl, by the time your later editions of the *Systema Naturae* were published, you thought you had a solid handle on a large chunk of nature. Well, allow me to say that in the case of weevil species, you finished short of the mark by a factor of more than 2,000! So with less than 1/2000 or 0.0005% of the total species-level diversity on hand, we had a system in use. It seems preposterous, in a way, but then not much has changed. Even today, we have a tendency to advertise systematic works as definitive (see 'the tree of life' or other 'synthesis' projects) when in reality we're still stepping in the dark. That's how science works in a competitive world where the first and the loudest reap the most benefits for their program. The Linnaean system is hardly at fault here. And in case we're closer to the light in terms of adequate taxon sampling, we might still have inadequate methods for phylogenetic inference. Remember, the most powerful concepts and tools for generating phylogenetic trees are less than four decades old. Many groups have not yet been subjected to these methods. Then there's poor taxonomic work or judgment in spite of

good sampling and methodology. It happens—with or without using Linnaean names.

In short, much nomenclatural baggage over the years is due primarily to work that looked sufficiently decent at the time but was ultimately too far off the mark to remain in use today. We'd be hard pressed to find any practical remedy for this phenomenon. New rules for naming won't change the way we (over)confidently regard and sell our products. There's also no point in holding off too long with a new system. We simply can't get from nearly 100 recognized weevil species (*Systema Naturae* in 1758) to 62,000 species (today) to 220,000 species (recently estimated) unless we split up the task and recognize the utility of small, incremental gains in phylogenetic knowledge. It's a sensible risk/reward strategy—something wins out over nothing—but naturally we pay a price for the resulting taxonomic and nomenclatural baggage.

So while we shouldn't abandon the ground rules and can't seem to escape the costly strategy of gradual increments, I think there's plenty of room for *strengthening the semantic ties* among multiple succeeding classifications. If systematists can't guarantee stability in meaning then we should at least offer more transparency. As experts we can make explicit our underlying assumptions, new insights, and differences with former systems on a much more regular basis. I will give you an example. Two years ago I published a phylogenetic revision of a weevil tribe (Derelomini Lacordaire) that now includes some 40 genera and 270 species. In that work I transferred 11 genera into the tribe that previously were placed in 4 other tribes. I also transferred 6 genera that had been previously part of the tribe into 4 other tribes. In all, 17 generic rearrangements were made, and 7 tribes were affected including the tribe I revised. Now, when I say "previously", I mean the classification presented in a particular weevil catalogue that was published less than a decade before. That much was made clear in my revision. I also attempted to highlight the proposed synapomorphies for the tribe, including reversals in select nested lineages, and avoided presenting them along with phylogenetically less relevant diagnostic features. So I was aiming for some level of semantic precision and transparency.

But then I knew so much more that remained unmentioned. For example, I had studied the taxonomic history of virtually every genus and species in the tribe, and could have rigorously traced its taxonomic placement in any major revision from 1798 to 2006. I also understood the relevant property-based definitions of the tribe published from 1866 onwards, and could have specified the extent to which they overlap among each other and with my perspective. This sort of information would help tremendously in terms of reconciling past and present taxonomic perspectives, thereby reducing the semantic

A tray of weevil specimens from the Linnaean collections at the Linnean Society of London

inconsistency and ambiguity that has characterized the taxonomic history of the tribe. Alas, this is almost never done. Furthermore, I should have presented a more formalized taxonomic update of the six non-focal tribes that underwent rearrangements in my study, by presenting side by side the constituent genera of each tribe according to the previous and my revised perspective. For further semantic disambiguation I could have added statements such as "tribe X [previous perspective] *corresponds* to tribe X [present perspective], *minus* genus Y, and *plus* genus Z". There are ways to express this so that virtually any user or specialized computer software can infer the full set of similarities and differences between the two perspectives.

We're at a juncture in systematics when more precise phylogenetic estimates are published at an increasing rate. There is a concomitant trend to archive the results in networked repositories intended to serve as the primary 'hubs' for systematic information[6]. Both the systematic and the computer science community seem to have bought into this vision. However it is likely that each community underestimates just how much we need to adjust our linguistic habits in order to achieve long-term integration of systematic products. Computer scientists use a formal language (description logic) to build highly structured networks (ontologies) that may include classes, instances, parts, properties, relationships, and other components and qualifiers. Once the structure is in place then powerful algorithms can 'reason' about the constituent elements, connect them to other ontologies created for related subject areas, and so on.

As computer scientists learn about systematics they must initially see a strong match between an ontology and a published taxonomy. However, as we've seen, a classification is never entirely comprehensible in isolation, and instead represents a complex mosaic of previous and new elements with implicit identities and relationships to each other. Too often such expert-made classifications are only comprehensible to other expert speakers, i.e., persons who share an intimate understanding of the contextuality of the new system and are thus able to make explicit the implicit semantic links to previous systems. So I predict that early generations of ontologies for systematics will either permit a very limited number of automated inferences, or will rely very heavily on expert input in order to reason among multiple succeeding classifications. In short, computer ontologies and systematic practice are not yet ready for each other.

Why have systematists relied so much on painstakingly acquired, implicit assumptions about the taxonomic history of particular groups when presenting their new classifications? I believe the reason is neither some form of elitism ("take that, users!") nor a lack of self-

esteem ("who wants to read about all these subtle similarities and differences?"). More likely, it's simply human habit—we make things just as explicit as we think is needed at the moment—paired with the similarly human notion that the latest perspective is really the one that's going to last for a long time, in spite of all historical evidence to the contrary. And so we pass the burden of full semantic resolution, both looking backward and forward, on to future specialists.

Let me try to sum up. The semantic problems we are confronting today in systematics are the result of a complex and long-winded interaction between accepted Linnaean practice for naming taxa and the particularities of taxonomic work that have piled up over centuries. The solution lies primarily outside of the Linnaean system of nomenclature as implemented today. The latter has served us mightily in accommodating inferences about the perceived properties of taxa. It generally mandates a hybrid model of intensional and ostensive definitions which, if properly interpreted, are compatible with evolutionary thinking as well as modern theories in the philosophy of language. Adhering to Linnaean ranks has allowed us to learn and communicate about nature in a way that suits our mental capacities as well as the causal structure of the natural world. We can only thank you for that, Carl.

However, the Linnaean system is *not* capable of capturing the *entirety* of semantic adjustments that occur when a previous classification is revised in light of new evidence. In fact it was purposefully designed to respond to some kinds of taxonomic rearrangements but not others[7]. Instead of abandoning the Linnaean system, this observation should lead us to express more clearly and more consistently what we mean when presenting a new classification. We must invent ways to semantically map each component of the new system to the corresponding component of a relevant predecessor, stating all intensional and ostensive similarities, differences, and apparent ambiguities. In doing so, we will reduce the sense of baggage in systematics and make progress towards a full semantic integration of the taxonomic process (not just individual snapshots) via ontology-driven services. At the human level, this requires that we routinely acknowledge the ephemerality of our latest insights, spend more time comparing our perspective to a previous one that we no longer think holds true, and generally pay more attention to the context in which we use taxonomic names. Efforts to achieve this are presently underway and are summarized under the term 'taxonomic concept approach'[8]. If we *supplement* the Linnaean system with these conventions, there will be more linguistic transparency and less mistaken urgency to purge the idiosyncrasies of the past or legislate a wrong consensus. So let's start to take our semantic supplements; they're all that's really needed to successfully use Linnaean names for the next 250 years!

In deep admiration,

Nico Franz

Notes:

[1] I suppose that most people take for granted the balance of beauty and information of your binomials. I have never heard the phrases "Linnaeus awkwardly named . . ." or "Linnaeus inexplicably named . . ." Part of the Linnaean success story is the cognitive and esthetic appeal of the actual names you've created.

[2] Unfortunately, none of this relieved me from the difficult task of working with weevils. You may be surprised that we now recognize 5,800 genera and 62,000 species of these little snouted beetles, up from 2 genera and 94 species described by you in the *Systema Naturae.*

[3] There are limits to this practice. As the number of convergent characters and evolutionary reversals increases within a lineage, the referential ambiguity will increase as well. Property-free definitions are perhaps the best option for some lineages such as bacteria, where high rates of transformation and horizontal inheritance obscure phylogeny.

[4] Adopted and slightly simplified from Sereno, P.C. 2005. Systematic Biology 54: 595–619.

[5] See Platnick, N.I. 2001. http://www.systass.org/archive/events-archive/2001/platnick.pdf

[6] Some authors, typically working outside of the systematics community, have suggested that systematists should stipulate 'consensus classifications'. We'll grant that wish right after there's a unified and stable view on pressing topics in ecology or other biological disciplines.

[7] Some measure of semantic ambiguity is desirable in a language that is reflective of small and often short-lived increments in knowledge.

[8] For a not-very-humble reference see Franz *et al.* 2008. In Wheeler, Q.D. (Ed.): *The New Taxonomy*, Systematics Association Special Volume Series 74. Taylor & Francis, Boca Raton, FL; pp. 63–86.