Published as: Wylie, C.D. (2019). Overcoming the underdetermination of specimens. *Biology and Philosophy, 34(24)*. https://doi.org/10.1007/s10539-019-9674-2

Overcoming the underdetermination of specimens

Caitlin Donahue Wylie (wylie@virginia.edu)

ABSTRACT: Philosophers of science are well aware that theories are underdetermined by data. But what about the data? Scientific data are selected and processed representations or pieces of nature. What is useless context and what is valuable specimen, as well as how specimens are processed for study, are not obvious or predetermined givens. Instead, they are decisions made by scientists and other research workers, such as technicians, that produce different outcomes for the data. Vertebrate fossils provide a revealing case of this data-processing, because they are embedded in rock that often matches the fossils' color and texture, requiring an expert eye to judge where the fossil/context interface is. Fossil preparators then permanently define this interface by chiseling away the material they identify as rock. As a result, fossil specimens can emerge in multiple possible forms depending on the preparator's judgment, skill, and chosen tools. A prepared fossil then is not yet data but *potential* data, following Leonelli's (2015) relational framework in which data are defined as evidence that scientists have used to support a proposed theory. This paper draws on ethnographic evidence to assess how scientists overcome this *underdetermination of specimens*, as potential data, in addition to the underdetermination of theories and of data, to successfully construct specimen-based knowledge. Among other strategies, paleontology maintains a division of labor between data-makers and theory-makers. This distinction serves to justify the omission of preparators' nonstandard, individualized techniques from scientific publications. This separation has benefits for both scientists and technicians; however, it restricts knowledge production by preventing scientists from understanding how the pieces of nature they study were processed into researchable specimens.

Keywords: Underdetermination, scientific practice, paleontology, material culture, specimens

The underdetermination of theory by data is a near-ubiquitous claim in philosophy of science, in the broad sense that data alone are an insufficient basis for constructing or choosing a theory (Okasha 2010; Cleland 2002). But what about the data themselves? Scientific data are selected and processed representations or pieces of nature, with their surrounding context, i.e., "background" or "noise," deemed less important. How then do scientists distinguish data from so-called noise? Vertebrate fossils are a revealing example of how nature is made into data. The fossil record is notoriously incomplete and the few bones that do fossilize are fractured, fragile, and warped by geological processes. They are known to be problematic data sources about past life, although they provide valuable—and sometimes the sole—information about their species and environment. In addition, distinguishing coveted fossilized bone from useless rock requires the tacit knowledge of expert visual judgment and manual skill to reveal the fossil by destroying the rock (i.e., the process of fossil preparation). Because of this, the same starting rock can result in a variety of different specimens. Potential data from fossils, therefore, are *underdetermined* by the surrounding rock.

Following Sabina Leonelli's (2015) relational framework for defining data as evidence used to support a knowledge claim, fossils are not data until scientists use them to support an argument. I argue that specimens are merely *potential* data, in that they can become data through a series of processes that select, prepare, organize, and share them in ways that then enable them to be deployed as data, if a scientist references them to argue for a knowledge claim (Leonelli 2015). We know that the work of theorizing is not limited to the construction of scientific theories; it also happens when evidence is collected in the field (Latour 1999; Holtorf 2002; Chapman and Wylie 2016) and—less well studied—when evidence is prepared for research. In the underdetermination of theory, a single fossil can support multiple theories. For example, based on the same *T. rex* skeletons, scientists have proposed that the animal was primarily a savage hunter and primarily a passive scavenger, two incommensurable theories. Furthermore, like data underdetermines evidence, *potential* data underdetermines data, in that there is far more potential data (e.g., specimens) than that which scientists actually study. The countless fossils that have been collected but never cited in a paper attest to the selection process by which scientists promote only some potential data to data. To go one step further, I argue that *potential* data are also underdetermined. After all, potential data must be selected from their vast not-data context (e.g., their background, surroundings, noise) and processed. For example, a bone can emerge as many possible physical forms as a result of what a fossil preparator designates as "rock" and how she removes it. How then do scientists overcome the underdetermination of potential data, in addition to the other layers of underdetermination, to construct accepted knowledge claims? This paper investigates the least-studied layer: the underdetermination of potential data by context.

Philosophers and social scientists have identified many strategies that scientists implement to overcome the underdetermination of theories, i.e., how scientists make a theory convincing in spite of evidence that supports multiple possible theories. Examples of these strategies include scientists' reputation, rhetoric, quantified results, multiple trials of experiments, and use of different methods in search of congruent evidence (Latour and Woolgar 1986; Latour 1987; Grande and Rieppel 1994; Porter 1995). Based on historic and contemporary case studies of the work of making fossils into potential data, I show that fossils are underdetermined by rock. Next I ask how scientists manage to propose credible knowledge claims despite the nonexclusive distinction of potential data from context. In paleontology, practitioners currently solve this problem by constructing separate social categories of work they call "research" and "preparation." Preparators produce potential data, which scientists then select from and analyze to make data and, accordingly, theories. Research workers (i.e., scientists) vs. preparation workers (i.e., technicians, such as fossil preparators) have different training, tasks, ways of communicating, and status in the scientific community.

These findings emerged from an ethnography I conducted of several vertebrate paleontology laboratories in the United States and United Kingdom in 2008-2012, with follow-up visits until 2017 (Wylie 2009, 2013, 2015, 2016b, 2018). My methods draw from the field of laboratory studies, which takes today's research spaces and communities as the locus of investigation to understand scientific knowledge production (e.g., Latour and Woolgar 1986; Traweek 1988; Doing 2009). Based on observations and interviews of paleontologists and preparators, as well as other workers, I investigated the complex social structure they construct through everyday interactions. In particular, I was struck by the deep distinction workers described and enacted between scientists and preparators, despite their shared specimens and interdependent work. How could it be that the people who physically define scientists' data sources do not typically appear in scientists' papers, write their own papers, or otherwise document their work? Following inductive analysis and grounded theory (in which ethnographers pursue research questions and propose theory drawn from their observations [Creswell 2007]), I took this invisibility in print as a starting point to illuminate the social and epistemic work of separating such closely related work and workers.

By separating data-makers from knowledge-makers and by omitting the work of data-making from publications, paleontologists promote the idea that prepared fossils are powerful evidence for their knowledge claims. This is a long-established and coherent strategy, in the sense that it has enabled successful further research on fossils. For example, paleontologists' theories, such as definitions of phylogenetic groups and that ceratopsians lived in herds, are based on fossils prepared by different people and a variety of techniques. When diversely-made data support a single interpretation, "convergent objectivity" and therefore powerful epistemic strength is achieved (Douglas 2004, 457-8). However, paleontologists' strategy is problematic because it makes data preparation invisible in print. As Heather Douglas explains, "convergent objectivity require[s] examination of the experimental process (or more generally, human-world interactions) to find the process markers that support ascription of objectivity" (2004, 458). Paleontologists prevent each other from assessing how specimens were shaped by obscuring that information.

To address the problem of an undocumented process of preparing potential data, I propose an approach that preserves the separate cultures and responsibilities of scientists and technicians while improving the groups' understanding of each other's practices. Specifically, I suggest that preparators create records of their work as part of each specimen's metadata, which would be accessible in institutions' databases but not in print. By evaluating the relative success of existing and proposed strategies to overcome underdetermined potential data, we can better understand what scientific evidence is and how it is constructed, as one of many steps on the path to making knowledge.

How Specimens Are Underdetermined

The underdetermination of theory versus that of data and potential data may seem like a distinction between the epistemic (i.e., theory) and the physical (i.e., data and data sources). After all, fossilization is not a method of theory selection or knowledge production. But fossil preparation *is.* The decisions and actions of a fossil preparator derive from her knowledge about fossils, tools, and paleontologists' interests and priorities. As Alison Wylie has documented in the case of archaeology (e.g., 1997; Chapman and Wylie 2016), values and assumptions have a ubiquitous impact on the production of scientific knowledge, down to the level of specimen preparation. But we don't know how archeologists' epistemic and practical "scaffolding" (Chapman and Wylie 2016, 84-5) affects how artifacts are cleaned or repaired for study.

The same infiltration of expectations happens during the processing of objects as during their collection and study. For example, it is sometimes not knowable whether a material is *specimen*, such as when a fossil matches the color, texture, and density of the rock around it. Preparators can judge this visual and tactile boundary better than anyone else can, including scientists, yet for some specimens the resemblance is too close to be sure. A powerful example of this phenomenon is trace fossils, such as impressions of an animal's skin, feathers, footprints, or burrows. These traces are preserved as marks on rock surfaces, rather than as objects in rock, thus further complicating the distinction between specimen and not-specimen because both in this case have the same material composition. Several preparators and scientists told me that they suspect trace fossils are rare in part because preparators focus on bones, not the surrounding rock, and therefore don't notice the trace fossils. Of course, preparators focus on bones because *scientists* are interested in bones, as the most commonly preserved traces of ancient life and thus the most scientifically-documented evidence. As a result of this bone bias, scientists and preparators suspect that preparators tend to unconsciously dismiss trace fossils as rock, i.e., as not-specimen, and either ignore them or destroy them to expose underlying bones. If true, this selection process would be an example of the high-impact and irreversible implications of the decisions of defining specimen from not-specimen, i.e., potential data from irrelevant background. Peter Galison describes the processes of defining potential data and defining noise as inseparable: "The task of removing the background is not ancillary to identifying the foreground—*the two tasks are one and the same*" (1987, 256, original emphasis). Thus, identifying the rock defines the fossil, and vice versa. The physical existence of fossils is therefore epistemically irrelevant until they are deemed to be data and used to support a theory. This is also true in fieldwork, in that a fossil that no one collects does not contribute to knowledge about the history of life and Earth.

In most sciences, once something is considered worthy of study then technicians follow standard methods to prepare it, in accordance with the belief that consistently-processed evidence is more reliable.[1] For example, like fossils, archeological artifacts are potential data that are underdetermined by their context. A post-hole or a worked stone can require expert judgment to recognize as potential data and therefore as different from the naturally dark soil or many other stones that might surround them. Archeological work is typically separated into excavation, done by professional and volunteer fieldworkers applying widely-followed protocols of collecting and documenting artifacts, and theory-building, done by academic archeologists who may never have seen the excavation site (Chapman and Wylie 2016, 72-81). By defining how fieldworkers produce data, archeologists argue that that data can stand alone, thereby justifying archeologists' knowledge-making despite their absence from the field site.

In comparison, fossil preparators select and design their own techniques, which are not standardized or widely disseminated. Scientists expect and trust preparators to use the methods they deem best to achieve the general goal of making fossils researchable, i.e., accessible visually and physically. This work usually entails removing rock, gluing broken bones, strengthening fossils with chemical adhesives so they can be handled without breaking, and sometimes reconstructing missing parts. Typically, scientists—and many kinds of science technicians—are somewhat standardized themselves, through shared training in techniques and paradigms (Traweek 1988; Kuhn 1996; Barley et al. 2016). But fossil preparation has no universal training program, credential, or manual. Instead, preparators learn on-the-job, through trial and error with guidance from other preparators. This lack of the typical gatekeepers for scientific work, such as standard methods and academic degrees, could open fossils to suspicions of subjective—or even fraudulent—preparation.

What then does it mean for science if potential data, as well as data and theory, is underdetermined? How do scientists get around this problem? Many sciences rely on repeated trials and controlled variables to strengthen their evidence against accusations of inaccuracy or fraud. Replicability and controls are less achievable for observing and measuring the relatively small number of vertebrate fossil specimens, though scientists do quantify fossil data by building databases for phylogenetic analysis and modeling (Sepkoski 2005, 2012). Yet neither scientists nor preparators call for the standardization of fossil preparation (Wylie 2015, 2018), perhaps because the current diversity and flexibility of techniques allows a tailored approach that

[1] This belief relates to the trust in machines or machine-like standardization as less-biased ways to produce evidence, which Daston and Galison (2007) call mechanical objectivity.

matches the uniqueness of vertebrate fossils. To broaden Adrian Currie's notion of "methodological omnivory" (2015, 188; 2018, 157-161), i.e., paleontologists' openness to use and invent a variety of research methods to triangulate across kinds of methods and data, I suggest that preparators share this openness to techniques in pursuit of the best available (potential) evidence. Working with the wide-ranging morphological forms of vertebrate fossils, each with a unique preservation history and resulting physical and chemical composition, certainly benefits from and may even require adaptable techniques as well as the tacit knowledge to design and apply them.

One common strategy to make research credible is detailing data preparation methods in publications, thus ostensibly inviting others to replicate the results. Particularly in cases for which flexible methods are more effective than standardized ones, detailed explanations of those methods in print can serve to justify their use and reliability. It seems logical that paleontologists would meticulously document how fossils came to look as they do, because relatively few vertebrate fossils are found and as a result they are carefully stored and re-studied over centuries. How preparators freed fossils of matrix and stabilized them with adhesives, for example, can shape generations of fossil-based knowledge claims. In addition, explaining how specimens were processed would help other scientists interpret the specimens more accurately and in relation to broader questions, as Leonelli (2008, 2016) suggests with regards to biological datasets. But this is not the case. Paleontologists do not describe preparation methods in print, and few preparators publish or even keep records of their techniques. The "methods" section of fossil-based research papers describes the paper's *research* methods, such as imaging techniques and chemical analyses. How those fossils were made into researchable specimens is missing. This nonstandard, undocumented preparation work seems like a threat to the reliability of fossils as data sources and therefore for the credibility of scientists' fossil-based knowledge claims. How then are fossils made into reliable potential evidence?

How Specimens Are Made

Paleontologists and preparators know that specimens are underdetermined. They deal with that reality every day, in quotidian tasks and decisions as well as in rare cases of preparation mistakes and, rarer still, forgeries. The tacit knowledge that underlies how preparators prepare and scientists interpret specimens in response to underdetermination is evident in the following examples, the first two drawn from my ethnographic research and the third from published sources. First I follow a preparator, whom I call Jay,[2] through typical projects of preparing and repairing fossils. His work illustrates how preparators decide and achieve one of many possible appearances for each fossil. Next, a case of a preparator's destructive error makes visible the delicate balance of omitting fossil preparation in print while providing enough information for scientists to interpret a fossil accurately. Finally, although fraud is rare in specimen preparation (despite the media attention it receives, such as the notorious cases of Piltdown Man and Archaeoraptor), an accusation of fraud against a preparator concerning the famous *Archaeopteryx lithographica* type specimen reveals how scientists and preparators defend their work.[3] These three cases illustrate unwritten and sometimes tacit strategies that scientists and technicians use to compensate for the underdetermination of potential data.

Good preparation

Delineating fossil from rock is a fundamental part of preparators' expertise and daily practices. Jay, an experienced preparator, often does this in the form of microsorting, i.e., searching for tiny fossils in a small box of fine-grained sediment under a microscope. In this case, he was hoping to find teeth that belonged to a fossil mammal jaw found in that sediment. As I watched, Jay triumphantly lifted out a "frag" (fragment) of fossilized enamel, grasped carefully in his tweezers. He inspected it and guessed that the frag might "bridge the gap" between two other pieces he'd found, thus making it possible to assemble all three pieces into one reconstructed tooth. By deciding that the object was a tooth fragment, Jay could next consider how to make that tooth look as it had when its owner was alive. Preparing fossils is a process of defining them in their current state, while repairing and reconstructing fossils involves, as the prefix "*re*" implies, returning a specimen to its previous appearance, i.e., in life. That is impossible to fully achieve, but making a tooth look complete instead of shattered is considered a step closer to seeing that tooth as it looked when it was part of an animal.

---

[2] All names are pseudonyms.

[3] A " type" or "holotype" specimen is an individual that scientists choose to be the permanent defining example of its species.

Then Jay decided to stop microsorting because the sediment contained only “tooth dust”: pieces of enamel that he judged too small to reconstruct or be scientifically informative. He deemed the tooth dust to be part of the animal, but not useful enough to be labeled as “specimen” and accordingly worthy of the careful study and preservation that specimens receive. Thus, Jay exercised the impressive power of deciding what is and is not potential data. Hypothetically, what if another preparator had spotted different tiny tooth fragments from the ones Jay saved? The jaw could take multiple possible forms based on which material a preparator deems to be useable fragments. These decisions are necessary because specimens are underdetermined by their surroundings. Most of preparators’ decisions are irreversible, such as gluing or discarding frags, and they affect how a fossil looks and therefore what scientists learn from it. Furthermore, Jay did not explain his reasoning process or his decision to anyone (except me, because I was there and asked him). No future researcher or preparator can know that there had been “tooth dust” preserved alongside this jaw. This is a loss of specimen and of metadata about that specimen that someone someday might miss; for example, perhaps future research techniques would be able to derive potential data from tooth dust.

Repairing fossils is another theory-laden activity that relies on preparators’ expectations of how a specimen looks. This was evident when Jay worked on the type specimen of a small ancestor species of modern horses, which had deteriorated in the many decades since it had been first prepared. Several bone fragments had broken off the skull, and glue from previous repairs obscured their broken edges, making it difficult to tell where they once had fit. Jay said to me, “I don’t see any obvious joins. But that doesn’t mean there aren’t any.” Preparators’ power over a specimen’s appearance includes their ability to *not* apply techniques, such as by not identifying or attaching broken pieces. They are well aware that their judgment is fallible. When I suggested that one frag belonged to the skull’s eye socket, Jay disagreed, replying with a smile, “Put it there if you want it to be a *new* type,” meaning a new—and invented—species. Similarly, Jay suggested gluing an unattached tooth on the end of the skull’s nose like a horn. These jokes highlight preparators’ power to decide how a specimen looks, such as by their judgment about whether and how broken bones fit together. These jokes may also serve to lighten the mood in the face of preparators’ concern about whether their work is true to the animal’s anatomy. That is arguably unknowable; nonetheless, preparators try and scientists trust their efforts, and this system has produced centuries of paleontological knowledge. Despite the diversity of specimens and preparation techniques, prepared fossils successfully undergo a variety of research methods and can answer diverse research questions, similar to the repurposing of biological and archeological data for studies beyond the data collectors' original intent (Leonelli 2008, 2016; Wylie 2017). Thus, scientists can overcome the underdetermination of potential data by context, despite the apparent liability of nonstandard data-making workers and methods.

After a few days of struggling to match the loose pieces with the skull's broken edges, Jay decided to consult the original published description of the specimen for hints. Based on the paper's text and photographs, Jay reattached many of the frags. He also found traces of glue, indicating a former join, on a canine tooth fragment and a matching glue-lined edge on the skull's jaw. So he adhered the tooth there, even though it was not mentioned or pictured in the publication. There is a kind of circularity in this story, in that a preparator prepared this specimen, then a researcher described it in a publication, and then a later preparator (Jay) repaired the specimen to match the publication. However, Jay also added an unpublished feature, the canine tooth, based on someone else's undocumented repair work after the paper was published. He chose his own judgment of the specimen's completeness over unquestioned preservation of the published description. It was a power play of sorts between a publication and a preparator's knowledge of tooth anatomy, visual judgment of previously-glued joins, and trust in an earlier, unknown preparator who had glued the tooth. Jay's judgments are even more important in this case because this skeleton, as a type specimen, physically defines its species. By gluing on unrecorded fragments, Jay altered how scientists might understand that species as well as that specimen.

Bad preparation

Jay's microsorting and repair work may seem like guesswork, but it involved careful thought and experienced judgment. The challenge of fossils' underdetermination by rock can also cause preparators to make *bad* judgments. For example, a paleontologist I call Henry told me about a case of preparator error:

> I'm currently studying a specimen that was first exposed by a volunteer at another museum, fortunately not ours, and when the volunteer uncovered the specimen he made a couple of really nice big holes in it. [laughs]

The museum's staff had asked a volunteer preparator to work on this rock-embedded specimen, which scientists later decided represented a new species of dinosaur. Unfortunately, the volunteer was not a skillful preparator and mistook the side of the animal's snout for rock, which he gouged with a metal tool. I later heard Henry present this specimen at a conference and he explained, with a laugh, that a volunteer preparator "added some foramina" to the specimen's skull. "Foramina" (singular, foramen) is the scientific term for naturally occurring holes in a bone, such as those through which blood vessels and nerves pass. Henry was making a joke by referring to human-made holes by a word that means natural holes. He likened the dinosaur's holes to extant animals' "foramen of Winchester" —the bullet hole in zoological specimens' skulls from collectors' Winchester guns. These fake "foramina," portrayed light-heartedly by Henry as unfortunate side effects of specimen preparation, are shocking permanent damage to specimens. This is also an example of how "bad" work is made explicit, unlike the usually tacit conceptions of "good" work.

Furthermore, as a new species and the sole representative of that species, no one knew whether this dinosaur actually *had* foramina on its snout. This uncertainty seems unresolvable, calling attention to the unsettling resemblance between natural features and changes made by people. How scientists decided these holes were mistakes rather than anatomical characteristics is that, luckily, the volunteer had not prepared the snout's opposite side, on which more-skillful preparators found no matching holes. Without the preserved mirror-image of the skull's opposite side (and the assumption of bilateral symmetry), scientists could have believed that this species had snout foramina. To skilled eyes, tool-made holes can look different from natural bone openings, but the distinction is ambiguous. Scientists don't typically mention fossil preparation in research papers and talks, and it is not mentioned in Henry's published conference abstract; but in the later publication describing this specimen and naming its species, Henry and his coauthors explain that a volunteer had damaged one side of the skull. In this rare documentation of preparation in print, the scientists probably intended to prevent other scientists from mistaking the preparator's holes for anatomical foramina.[4] This case shows that scientists are sometimes willing to describe specimen-making in print, at least in cases of misleading potential evidence.

[4] This case matches Shapin's observation (1989) that technicians and technical work are only mentioned in print in situations of questionable research results, for which researchers then blame the technicians.

Suspicious preparation

The volunteer preparator's hole-carving was an unintentional mistake rather than purposeful alteration of a specimen. Despite their underdetermined specimens and their nonstandard work and backgrounds, preparators are rarely accused of fraud. One unusual case of an attempt to discredit a fossil by blaming its preparator sheds light on how scientists and technicians defend themselves and their work (Wylie 2016a; also see Bokulich 2018). In the 1980s, two respected astrobiologists, Fred Hoyle and Chandra Wickramasinghe, claimed that the famous *Archaeopteryx lithographica* type specimen was a forgery (1986a; Watkins et al. 1985a, 1985b). The reptile-like bird challenged their pet theory of panspermia, that life on earth originated as organic particles from outer space and that terrestrial evolution is driven by occasional genetic exchanges with extraterrestrial viruses delivered by comets (e.g., Hoyle and Wickramasinghe 1982, 1986a, 1986b). (Specifically, *Archaeopteryx* predates the late-Mesozoic arrival of the supposed comet and its stowaway viruses that Hoyle and Wickramasinghe claim sparked dinosaurs' evolution into birds.) To criticize *Archaeopteryx* and promote panspermia, Hoyle and Wickramasinghe pointed to several changes made to the specimen over its long history at the British Museum (Natural History)[5] as evidence of fraud. The most recent example, of the many the authors proposed, concerned the specimen's partially preserved skull:

> In 1862 the "brain-case" was said to be that of a bird … As time went on, and in the opinion of paleontologists the skeleton became ever less bird-like, the thought of the brain case being like a magpie must have seemed unattractive … So what to do? Simply remove the brain case, as was done a year or two ago. (Hoyle and Wickramasinghe 1986a, 97)

The authors suggest that as scientific theories about the phylogeny of dinosaurs vacillated between birds and reptiles, preparators physically altered *Archaeopteryx* accordingly. The authors describe this skull removal as "a whole chunk of the main slab now hacked away," in "a vigorous assault that was mounted on this unfortunate creature, with extensive chisel marks to be seen" (1986a, 99).

But, in a rare publication about preparation methods, Peter Whybrow (1982), head of the museum's fossil preparation laboratory, had described his work of separating the brain case from the specimen's main slab as done meticulously and without damage. He explained that the

[5] This London museum is now named the Natural History Museum.

museum had agreed to have the brain case prepared at the request of two paleontologists who were studying the evolution of birds (Whybrow 1982, 84). As the only uncrushed *Archaeopteryx* brain case ever found, it was a singular source of information about the skull's shape and thus the species' phylogeny. But it could not be studied while half-embedded in rock; hence Whybrow's work to remove it. Hoyle and Wickramasinghe criticized Whybrow's description of his work: "How it was possible to write of not causing the *slightest* damage to the fossil in the wake of the gross defacement … we leave to the reader's judgement" (1986a 99, original emphasis). This claim invites the reader to see Whybrow as a liar who protected his forgery by denying it in print.

In response to this insult against *Archaeopteryx* and its caretakers, several of the museum's paleontologists and Whybrow published a powerful rebuttal article in *Science* (Charig et al. 1986). They systematically documented the impossibility of Hoyle and Wickramasinghe's many accusations, particularly their claim that the specimen's feather impressions were fake, by thin-sectioning *Archaeopteryx*'s rock slab to show that it has a consistent composition with no irregularities and is therefore authentic. They highlight the strength of their arguments and the weakness of the accusers' by describing their *Science* paper as "using a sledgehammer to crack a rather trivial nut" (Charig et al. 1986, 233), thereby also hinting at scientists' widespread rejection of the theory of panspermia as ridiculous, or, as Hoyle and Wickramasinghe themselves gently call it, "eccentric" (1986b, 509). But the museum authors explain the need to "put the record straight" as necessary due to Hoyle's fame and the threat of creationists (1986, 623). They do not, however, mention *Archaeopteryx*'s brain case, perhaps because geological analyses cannot show that the bone had been removed without damage and in good faith. Because of this omission, maybe, or because he was so offended by the accusation of being a forger, Whybrow further defended his preparation work in a letter to *New Scientist* (1986). In it, Whybrow explained that he was not concerned about wrongful interpretations: "I do not really care if Hoyle and Wickramasinghe believe *Archaeopteryx* to be a forgery. They can hold sincere views, however mistaken" (1986, 62). Instead, he was angry that they had questioned the legitimacy of his preparation methods and his professional credibility, because they "reject … the authenticity of the London *Archaeopteryx* by stating that devious, deceitful and unprofessional methods have been recently used" (1986, 62).

Here we see exposed for a moment the soft underbelly of rock-based science. Interpreting the fossil is not Whybrow's concern; its honest and skillful preparation is. Debates over knowledge claims don't worry him—they're normal, after all—but debates about the truthfulness of his work are unacceptable. Scientists across disciplines recognize the underdetermination of *knowledge* by data: they know that multiple interpretations of data are possible, and that, as a result, their interpretations will be debated. But publicly reminding them that potential data is itself underdetermined by raw material—e.g., that specimens take different forms and yield different potential data depending on how people process them—is more dangerous. It threatens the trustworthiness of the objects that are the foundation of research.

In no way am I saying that scientists and technicians are wrong, or "devious," "deceitful," or "unprofessional," as Whybrow felt accused of. Underdetermination is not specific to fossils or fossil-based theories; it is an underlying fixture of knowledge and, I suggest, of all research objects and other forms of potential data. The fascinating thing is how scientists and technicians develop ways to make credible knowledge from always-incomplete information, as I discuss next.

How Specimens Are Made Reliable

Here I assess the strategies scientists use to overcome underdetermination, in order to identify problems and propose a solution. One widespread way to promote a preferred theory is to question or disregard evidence that opposes that theory, which Thomas Kuhn (1996) calls anomalies. Skepticism about anomalies is a powerful and somewhat necessary way to promote a theory, because, after all, data will always support multiple theories. One way to deal with anomalies besides denying them is to claim that they are the responsibility of another field, such as technical work, and thereby deem them irrelevant to science. Distinguishing theory selection from the messy work of producing first potential evidence and then evidence can help scientists disregard that work in print and perhaps in practice (Wylie 2016b).

Many sciences follow this strategy. For example, 20th-century physicists preferred to hire workers with no training in physics as bubble chamber operators, to prevent them from imagining—and thereby inventing—desirable results (Galison 1997, 199-200; also Traweek 1988, 28-9). Many of these operators were women with little education but significant control over physicists' data. Likewise, the Manhattan Project found that women were more efficient and effective workers in uranium enrichment plants than Ph.D.-holding scientists (Nichols 1987 as cited in Kiernan 2013, 109-110). Project leaders believed that the women, who were not told what the plant was producing or why, followed instructions better than the scientists, who knew the end goals and deviated from instructions to try to improve the process. Thus, scientific knowledge is sometimes perceived as a barrier to good data preparation. These cases of unskilled labor differ from the skillful work of preparing specimens; however, the importance lies not in how skillful the work is but rather in how much—or, more accurately, how little—the workers know about the relevant science. Distinguishing data producers from data interpreters can prevent interpreters' paradigmatic beliefs from informing data, regardless of producers' skill.

Once this labor is divided and assigned unequal importance, it is easy to ignore the lower-status one, i.e., anything considered not science. By "deleting" or "abstracting" data preparation from research, scientists create distance between it and their knowledge claims (Star and Strauss 1999; Law 1994). Accordingly, paleontologists and preparators consider themselves separate fields, and institutions deem them unequal through salary and job rank. These groups do not share training or mindsets. Few preparators have formally studied anatomy, morphology, or phylogeny; therefore they can't recognize new species or never-before-seen fossil features. Accordingly, scientific theories and debates only minimally inform how preparators prepare a specimen, preventing their inadvertent construction of desired characteristics, as Galison's physicists feared about bubble chamber operators. Thus, it is likely that preparators don't know enough about anatomy to sculpt a creature that could convince scientists of its authenticity. In theory, this separation of kinds of knowledge could strengthen fossils' credibility as data sources. The problem is that this separation is not made explicit in print. Instead, I recommend that scientists explicitly frame the separation of preparation from study as a way to protect fossils from intentional or unintentional sculpting into scientifically interesting specimens. Rather than apparently hiding preparators' influence over fossils, scientists can portray it as a purposeful strategy to reduce their own bias over their potential data.

I am not arguing that preparators should write methods sections for scientists' publications or write their own publications about their methods, because most preparators are not interested in authorship and because invisibility grants preparators remarkable and valuable autonomy over their work. Unlike the women workers at Oak Ridge, preparators do not receive detailed instructions or unquestionable commands. Instead, scientists ask them to prepare particular fossils, sometimes with attention to certain valued features, and preparators decide how to achieve those goals. As a result, preparators have established their own priorities as a community, such as removing matrix by "following the fossil" closely to avoid inadvertently carving new features or destroying unexpected ones, keeping associated bones together (e.g., with glue or in storage boxes) to prevent losing pieces, and making particular aspects visible based on researchers' requests. Authorship could threaten this autonomy, by forcing preparators to conform more to scientists' conceptions of good work than to their own .

As one potential solution, the practice of "reflexive excavation" on archeological sites brings together excavators and theory-makers by sharing their discoveries and interpretations *during* fieldwork, not after (Hodder 2000 as cited in Chapman and Wylie 2016, 82-83). The intention is to adapt excavation practices in response to real-time interpretations of finds, rather than strictly following protocols without regard for what the evidence might mean. This "deliberate reintegration of interpretation into fieldwork" (Chapman and Wylie 2016, 82) seems like a useful mode of training non-scientists to find and prepare artifacts and specimens to please scientists, a practically important achievement because the non-scientists typically are employed by the scientists. However, it could also degrade the reliability of the excavation and preparation of potential data by matching the non-scientists' expectations to scientists', thereby recreating the scientists' assumptions and biases rather than balancing them with the non-scientists'. For example, fossil preparators typically know only the basics of paleontological "scaffolding," learned informally through conversations with scientists. If scientists and preparators discuss what is scientifically interesting about a specimen during its preparation, then scientists' theories and expectations will inevitably be incorporated into preparators' decisions about that specimen. In the end, the prepared specimen will probably better match the scientist's intended purposes for it, but it may be more informed by scientists' theories than a specimen prepared under the current system of separation. Revealingly, archeologists don't seem concerned about this issue, perhaps because, as Chapman and Wylie point out, "in practice, however, the aims of integration and democratization are not always realized" (2016, 83). The practical challenges of overcoming long-established hierarchies and differences in training are likely causes for the relative rarity of reflexive excavation, though it's possible that the undesirable epistemic risk of inserting scientists' paradigms into the work of producing potential data also creates a barrier.

I propose a compromise approach that preserves the distinction between scientists' and preparators' expertise and domains of power while making preparators' work more transparent to scientists. A published methods description written by a specimen's preparator would be the gold standard of transparency, by explaining how an expert judged what the fossil looks like as a credible way of overcoming that specimen's underdetermination. However, scientists might expect such a description to be standardized and replicable, as is the norm in most scientific disciplines but neither of which is true of fossils or their preparation. Preparators' flexible, innovative methods in print might invite paleontologists' criticism that they wouldn't warrant in everyday lab interactions. For this approach to successfully preserve prepared fossils' reliability as data sources, scientists' general expectations of methods descriptions would have to change, which is a significant and unlikely transformation.

Instead, I suggest that preparators write unpublished records of how they prepared each fossil, such as a list of the tools, materials, and techniques with the preparator's name. Some preparators create these records for their own use, and a few preparators have succeeded in getting their records added to institutions' collection databases. Museums and universities keep these preparation records for other kinds of specimens, such as pressed plants, animal skins, and geological samples, stored in archives and often in databases. Why fossils' records don't include preparation information when other specimens' records do is a mystery to me, but clearly extending this established record-keeping to fossils would preserve and disseminate scientifically relevant information. For example, which chemicals a preparator added to a fossil (because additives such as glue interfere with the results of geochemical analyses) and whether scratches on the bone were caused by preparation tools (and not long-dead scavengers as a scientist might surmise) would alter how scientists study and understand that specimen. Specimen databases tend to have a public-facing version with limited information and an internal version with complete information; I suggest adding prep records to the internal version. That location would provide access to scientists who want more metadata about a specimen, in that they could visit an institution or ask institution staff to look up how, when, and by whom a specimen was prepared. Long or unknown time spans between the collection, processing, and dissemination of data can complicate how scientists interpret it (Leonelli 2018), and long, complex, and undocumented shelf lives are common for fossil specimens. Naming the preparator creates the opportunity for a scientist to ask the preparator for more details (assuming, of course, that the preparator and scientist are alive at the same time, which is not always the case). It also grants the preparator responsibility for that specimen (i.e., credit for good work or blame for poor work). Ideally, a publication would include a reference to each fossil's preparation record, thereby inviting other scientists to seek more information about the specimen while distancing that information from scientists' primary product, i.e., published knowledge claims. This approach would provide scientists with epistemically important information about a specimen's appearance that is currently missing from publications, while preserving the separate autonomies of specimen-makers and theory-makers. Separating fossil preparation from fossil research, then, offers scientists stronger claims to credibility and grants technicians more credit for their work while still largely protecting technicians' diverse methods from scientists' scrutiny.

Of course, documenting preparators' work at all invites some accountability and visibility, which may make preparators more vulnerable to the influence of scientists and scientists' paradigms. However, this visibility would be less than authorship or published methods descriptions and is therefore likely to preserve enough autonomy for preparators that their decisions remain reliable. This system would thereby preserve metadata for future scientists to consult, though with the extra step of visiting a collection or requesting information from an institution. Accessing the records through publications or online databases, in comparison, would be easier, but that extra exposure could threaten the professional and epistemic separation that scientists and preparators have long preserved. Thus, this compromise is better for scientific knowledge writ large, as compared with the status quo and with the more extreme solutions of published methods descriptions or authorship for technicians.

This approach resembles Sabina Leonelli's analysis of biological databases, in that "facts travel well when stripped of everything but their content and means of expression; on the other hand, the reliability of facts cannot be assessed except by reference to the way in which facts are produced" (2008, 12). In response, database "curators" upload data and information about the methods, tools, and people who produced the data, but as linked entries rather than a single entry so that researchers can view the data with or without their metadata. This approach is part of the work of "packaging" data for travel across epistemic contexts (Leonelli 2008, 3). Similarly, judging a fossil's reliability requires understanding how it was prepared, among the myriad possible ways it could have been prepared. The absence of that information in publications limits the extent to which scientists can understand a fossil's appearance. Providing links to specimen records would help fill that gap.

Conclusion: Making Theories Requires Making (Potential) Evidence

I have proposed that underdetermination occurs beyond that of theory by evidence, in that defining evidence itself can take many possible forms, from selecting which potential data best support a knowledge claim to distinguishing potential data from their context. As theories must be argued for and defended as the best interpretation, so must evidence. I propose that there are *three* moments of underdetermination: that of the theory, the data that scientists choose to serve as evidence for that theory (following Leonelli's [2015] relational framework), and the potential data that scientists could cite as evidence but haven't yet. Vertebrate fossils provide a particularly clear instance of what I suggest is a broad and perhaps ubiquitous reality of the underdetermination of potential data. After all, gravity waves, microbes, and human behavior must be similarly defined and processed by practitioners' decisions and actions before they can be referenced in support of a theory.

There are many possible ways for a fossil to look as a result of collection and preparation, just as there are many possible theories to explain that fossil's phylogeny, behavior, and role in its ecosystem, among many other research questions. Scientists generally deal with the underdetermination of evidence by standardizing how evidence is prepared, to argue that all specimens, samples, observations, and other traces processed by the same methods are comparable and valid. Vertebrate paleontology does not have standardized methods for preparing specimens. Instead, paleontologists portray fossil evidence as reliable by separating its preparation from research, such as by outsourcing preparation to non-scientists and omitting descriptions of this work from publications. This approach hides the variety, skill, and subjectivity of preparation work, while also inaccurately implying that it is irrelevant to research. In response, I propose that preparators write records that scientists then reference in their papers, a convention that could preserve the established distance between data preparation and interpretation while providing information that is crucial to understanding specimens.

It may appear that underdetermination weakens both theory and (potential) evidence, by requiring them to be the results of decisions and arguments rather than obvious facts. However, Leonelli astutely recognizes the value of evidence that applies to multiple theories with regards to open-access biological datasets: "Underdetermination in this case is a strength of scientific research. Data produced with high efforts and costs can and arguably should be used as evidence for a variety of claims about phenomena" (2008, 35). Likewise, drawing on the same potential evidence, such as fossils that require time-intensive and skilled preparation work, to serve as evidence for a variety of studies maximizes the objects' epistemic contributions. A specimen's underdetermination can make that specimen relevant to more research questions and more theories. But it may then be more difficult to distinguish among those theories, making knowledge claims accordingly more difficult to defend. Acknowledging that scientists' and technicians' assertions about what is a specimen and what is data are actually evidence-based arguments would invite richer discussions about how specimens and data are made and how that work shapes how scientists interpret them.

To create knowledge in spite of the underdetermination of potential data, scientists and other data-makers develop experiential know-how for judging potential data vs. not-data, skillful ways of separating the two (e.g., preparing fossils, "cleaning" experimental data by discarding problematic trials, creating algorithms to select certain kinds of data points), trust in each other's work, and divisions of scientific and non-scientific labor. By obscuring the work of making fossils researchable and by defending preparators' skill and integrity in rare cases of blame, scientists portray fossils as a trustworthy source of evidence for our knowledge about evolution and Earth history. But this approach limits specimens' potential uses and interpretations as scientific evidence by omitting information about how they were processed. Creating records of the work and workers who shape specimens would improve those specimens' epistemic value and thus the quality of the knowledge claims they inspire.

Acknowledgements

I am grateful for thoughtful feedback from two reviewers as well as Adrian Currie, Emily McTernan, Robin Andreasen, Jonathan Birch, and attendees of the 2017 Philosophy of Paleobiology workshop at Dinosaur Provincial Park.

References

Barley SR, Bechky BA, Nelsen BJ (2016) What do technicians mean when they talk about professionalism? An ethnography of speaking. Res Sociol Organ 47:125–160. doi: 10.1108/S0733-558X20160000047017

Bokulich A (2018) Using models to correct data: paleodiversity and the fossil record. Synthese 1–22. doi: 10.1007/s11229-018-1820-x

Chapman R, Wylie A (2016) Evidential Reasoning in Archaeology. Bloomsbury Academic, London

Charig AJ, Greenaway F, Milner AC, et al (1986) Archaeopteryx is not a forgery. Science (80-) 232:622–6

Cleland CE (2002) Methodological and Epistemic Differences between Historical Science and Experimental Science. Philos Sci 69:447–451

Creswell J (2007) Qualitative inquiry and research design: choosing among five approaches, 2nd edn. Sage, Thousand Oaks, CA

Currie A (2015) Marsupial lions and methodological omnivory: function, success and reconstruction in paleobiology. Biol Philos 30:187–209. doi: 10.1007/s10539-014-9470-y

Currie A (2018) Rock, Bone, and Ruin: An Optimist's Guide to the Historical Sciences. MIT Press, Cambridge, MA

Daston L, Galison P (2007) Objectivity. Zone Books, Brooklyn, NY

Doing P (2009) Velvet Revolution at the synchrotron. MIT Press, Cambridge, Mass

Douglas H (2004) The irreducible complexity of objectivity. Synthese 138:453–473. doi: 10.1023/B:SYNT.0000016451.18182.91

Galison P (1987) How experiments end. University of Chicago Press, Chicago

Galison P (1997) Image and logic: a material culture of microphysics. University of Chicago Press, Chicago

Grande L, Rieppel O (eds) (1994) Interpreting the hierarchy of nature. Academic Press, San Diego

Hodder I (ed) (2000) Towards reflexive method in archaeology: the example at Çatalhöyük. McDonald Institute for Archaeological Research, Cambridge

Holtorf C (2002) Notes on the life history of a pot sherd. J Mater Cult 7:49–71. doi: 10.1177/1359183502007001305

Hoyle F, Wickramasinghe C (1982) Proofs the life is cosmic. Mem. Inst. Fundam. Stud. 1:1–155

Hoyle F, Wickramasinghe C (1986a) The case for life as a cosmic phenomenon. Nature 322:509–511

Hoyle F, Wickramasinghe C (1986b) Archaeopteryx, the primordial bird: a case of fossil forgery. Christopher Davies Publishers, Swansea

Kiernan D (2013) The Girls of Atomic City. Simon and Schuster, New York

Kuhn TS (1996) The Structure of Scientific Revolutions, 3rd edn. University of Chicago Press, Chicago

Latour B (1987) Science in action: how to follow scientists and engineers through society. Harvard University Press, Cambridge, Mass

Latour B (1999) Circulating reference: sampling the soil in the Amazon forest. In: Latour B (ed) Pandora's Hope: Essays on the Reality of Science Studies. Harvard University Press, London, pp 24–79

Latour B, Woolgar S (1986) Laboratory life: the construction of scientific facts, 2nd edn. Princeton University Press, Princeton, NJ

Law J (1994) Organizing modernity: social ordering and social theory. Blackwell, Cambridge, Mass

Leonelli S (2008) Circulating Evidence Across Research Contexts: The Locality of Data and Claims in Model Organism Research. London

Leonelli S (2015) What Counts as Scientific Data? A Relational Framework. Philos Sci 82:810–821. doi: 10.1086/684083

Leonelli S (2016) Data-Centric Biology. University of Chicago Press, Chicago

Leonelli S (2018) The Time of Data: Time-Scales of Data Use in the Life Sciences. Philos Sci 85:741–754. doi: 10.1086/699699

Nichols JD (1987) The Road to Trinity. Morrow, New York

Okasha S (2000) The underdetermination of theory by data and the “strong programme” in the sociology of knowledge. Int Stud Philos Sci 14:283–297. doi: 10.1080/026985900437782

Porter TM (1995) Trust in numbers: the pursuit of objectivity in science and public life. Princeton University Press, Princeton, NJ

Sepkoski D (2005) Stephen Jay Gould, Jack Sepkoski, and the “quantitative revolution” in American paleobiology. J Hist Biol 38:209–237. doi: 10.1007/s10739-004-2084-5

Sepkoski D (2012) Rereading the fossil record: the growth of paleobiology as an evolutionary discipline. University of Chicago Press, Chicago

Shapin S (1989) The invisible technician. Am Sci 77:554–563

Star SL, Strauss A (1999) Layers of silence, arenas of voice: the ecology of visible and invisible work. Comput Support Coop Work 8:9–30

Traweek S (1988) Beamtimes and lifetimes: the world of high energy physicists. Harvard University Press, Cambridge, Mass

Watkins R, Hoyle F, Wickramasinghe C, et al (1985a) Archaeopteryx: a photographic study. Br J Photogr 132:264–6

Watkins R, Hoyle F, Wickramasinghe C, et al (1985b) Archaeopteryx - a further comment. Br J Photogr 132:358–9

Whybrow PJ (1982) Preparation of the cranium of the holotype of Archaeopteryx lithographica from the collections of the British Museum (Natural History). Neues Jahrb Fur Geol Und Palaontologie 3:184–192

Whybrow PJ (1986) Rare controversy. New Sci 62

Wylie A (1997) The Engendering of Archaeology: Refiguring Feminist Science Studies. Osiris 12:80–99

Wylie A (2017) How Archaeological Evidence Bites Back: Strategies for Putting Old Data to Work in New Ways. Sci Technol Hum Values 42:203–225. doi: 10.1177/0162243916671200

Wylie CD (2009) Preparation in action: paleontological skill and the role of the fossil preparator. In: Brown MA, Kane JF, Parker WG (eds) Methods in Fossil Preparation: Proceedings of the First Annual Fossil Preparation and Collections Symposium. pp 3–12.

Wylie CD (2013) Invisible technicians: a sociology of scientific work, workers, and specimens in paleontology laboratories. University of Cambridge

Wylie CD (2015) “The artist’s piece is already in the stone”: constructing creativity in paleontology laboratories. Soc Stud Sci 45:31–55. doi: 10.1177/0306312714549794

Wylie CD (2016a) Overcoming underdetermination. In: Extinct. http://www.extinctblog.org/extinct/2016/4/11/overcoming-underdetermination. Accessed 10 Oct 2018

Wylie CD (2016b) Invisibility as a mechanism of social ordering: Defining groups among laboratory workers. In: Bangham J, Kaplan J (eds) Invisible Labour. Max Planck Institute for the History of Science, Berlin. https://www.mpiwg-berlin.mpg.de/sites/default/files/migrated/book_invisibility_and_labour_in_the_human_sciences_preprint_.pdf. Accessed 18 January 2019.

Wylie CD (2018) Trust in Technicians in Paleontology Laboratories. Sci Technol Hum Values 43(2), 324-348. doi: 10.1177/0162243917722844