

Evidence and Expertise in Genetic Nomenclatures

Colin Michael Egenberger Halverson, Center for Bioethics, Indiana University

ABSTRACT:

Clinicians and laboratory scientists use a number of different systems for naming genetic mutations in their daily activities. Based on participant observation at an American academic medical center and interviews with a variety of actors at American hospitals, this paper analyzes the use of these systems. I argue that their distribution corresponds to differences in professional regimes of responsibility. An examination of these often quite complex linguistic items reveals a correlation between the type of components (evidential versus epistemic modifiers) that constitute the names and the presupposed professional role of their intended audiences.

KEYWORDS: evidentiality, epistemic modality, medical genetics, professional responsibility, expertise

This is the author's manuscript of the article published in final edited form as:

Halverson, C. M. E. (2019). Evidence and Expertise in Genetic Nomenclatures. *Journal of Linguistic Anthropology*, 29(3), 314–331. <https://doi.org/10.1111/jola.12250>

Contemporary medical science asserts that genetics is implicated in almost all aspects of human health and disease (Nussbaum et al. 2016).¹ Medical centers in the United States are rapidly incorporating genetic testing into their standard healthcare practices for a wide range of diagnostic purposes (Manolio et al. 2013). To provide a diagnosis, though, scientists and clinicians must collaborate in order to identify a particular genetic mutation as the candidate source of a patient's disorder. They search for variation within an affected patient's genetic code – in the form of both large structural differences in the chromosomes as well as small mutations of single nucleotides within a gene. The ability to identify disease is therefore becoming increasingly linked to medical professionals' ability to locate and label different forms of genetic variation.

The question of how to name such variation has harassed scientists since the earliest years of their forays into the analysis of such mutations (Shafer et al. 2013). With disconnected genetics laboratories making separate – though occasionally simultaneous (Wain et al. 1999) – discoveries, unrelated and non-commensurate naming systems have arisen and come into competition (Halverson 2019). As genetics has increased in prominence, so too have concerns increased about the chaotic state of the existing systems (den Dunnen and Antonarakis 2000; Editorial 1997). However, “the issue is not a paucity of standards but, rather, too many to choose from” (Warner et al. 2016, 113). Today, there are as many as six major competing nomenclatures and many more minor or specialized systems that are used to label genetic variation (Shoenbill et al. 2014). In the hospital, one can hear professionals refer diversely to a unique individual mutation in a particular patient with multiple names, such as *the likely pathogenic variant, PMP22, a 17p deletion*, or one can read it printed on a laboratory report as *c.434delT*.

This seeming free variation of code choice struck me immediately at the onset of my fieldwork in a medical genetics clinic in the American Midwest, where I spent a year and two subsequent summers shadowing genetic counselors and other specialist clinicians, and working in a number of affiliated laboratories. I observed genetics talk in both the research setting among scientists and the healthcare delivery setting among practitioners. The specific clinic with which I was affiliated has the purpose of providing genetic tests for patients who have either a rare disease or an advanced cancer.²

At first, the diversity of genetic nomenclatures seemed commensurate with the diversity of expert personnel with whom I interacted. However, the variation in naming practices has also repeatedly been brought to my attention by my clinician interlocutors. They are enthusiastic to theorize extemporaneously about the meanings, systematization, and complexity of the various systems. Many make comments about the outwardly random nature of both their and their coworkers' choice of code. But, based on my observations, I argue that these folk intuitions result from the practitioners' language ideological (Kroskrity 2000; Schieffelin et al. 2000) reduction of a name's meaning to its referential function (Jakobson 1960; Silverstein 1976), that is, to its ability to pick out an individual genetic mutation. In this article, I present an analysis of how nomenclature systems are in fact much more regular and complex. I show that they track differences in professional regimes of responsibility and segregate according to the pragmatic demands of various medical contexts. I contend that what is seen as a sort of conceptual anarchy from the perspective of reference looks much more orderly from the perspective of indexicality.

In what follows, I link morphological segments of gene names to sociological systems of expertise and professional responsibility. The same gene has different aspects of its manifold existence described by different types of names. In particular, this paper focuses on segments of

names that convey meanings about the evidence and certitude with which a genetic mutation is described. Classic works in the sociology of knowledge production (Latour and Woolgar 1979; Latour 1987) contend that a scientific claim becomes increasingly “factual” as “modalities” – hedging, modulation, qualification, etcetera – are removed from its linguistic representation. Departing from these analyses, I argue that it is not merely that modalities are removed. Rather, I demonstrate that – following Michael Silverstein (2003b) – the information conveyed by gene names is *transduced*, changing in content and structure in some salient ways while maintaining the appearance of identity as it moves between alternative systems of representation.³ As a gene’s name is transduced across systems used by different classes of medical personnel, some information is lost while other information is gained. Moreover, I demonstrate that nomenclatural differentiation has more to do with the social and professional roles of the users of those systems than it does with perceptions and projections of “facticity.” By attending to the social identities of the users of the different nomenclature systems, we can see the relationship between the linguistic presentation of evidence and the sociological ascription of expertise. In what follows, we will see how language practice corresponds to different professional regimes of responsibility in the practice of medical genetics.

I argue that a combination of the imagined capacities of professional interlocutors and the goals of specific encounters determine an actor’s use of nomenclature system. I contend that this hinges particularly on the morphological structures employed by each system. That is, the morphological components that constitute gene and mutation names correlate with the professional roles and responsibilities presupposed of their intended contexts of use (Silverstein 2003a). The variable social values of the professional classes are reflected in the manners in which these classificational acts are undertaken (cf. Mauss and Durkheim 1963). I focus

particularly on the use of *evidential* and *epistemic* markers (cf. Aikhenvald and Dixon 2003; Aikhenvald 2004; Chafe and Nichols 1986) in these naming practices.

In brief, by evidential marker, I refer to a component of a gene or mutation name that details the means by which it has been discovered to exist – which test methodology has been used. By epistemic marker, I refer to a component of a name that evaluates the likelihood that that gene or mutation does in fact exist and is not merely an artifact of the technology used for its discovery. In this paper, I demonstrate that these two categories correspond to varying levels of ascribed scientific expertise: Evidential markers are used with expert audiences, while the use of epistemic markers tracks primarily with a relative professional distance from such realms of knowledge-production and interpretation.

The use of this kind of linguistic marking in scientific discourse has typically been analyzed – with limited morphological specificity – under the broad rubric of “hedges” (Hobbs 2003; Prince et al. 1982; Salager-Meyer 1994; Varttala 1999; see also Lakoff 1972, 195 for a discussion of hedging in general). This literature describes hedges variously as minimizing the threat to a speaker’s face, as a politeness strategy, and as (only occasionally) a qualification of a proposition’s accuracy (e.g. Salanger-Meyer 1994, 153). Such hedging has been found to be a common feature of academic and medical discourse in general (Crismore and Farnsworth 1990). Despite that, an examination of such linguistic practices with regard to the astoundingly complex ecology of genetic nomenclatures has been missing.

In order to fill this gap, I begin my analysis with an overview of medical genetics as a site of occasionally agonistic social life. I then turn to consider the ways scholars have described evidential and epistemic markers in scientific and vernacular language. Next, I provide an overview of the professional practices and actors who cross paths in the Department of Medical

Genetics at the Midwestern hospital at which I worked. Thereafter I analyze, in some detail, five competing and interacting systems of nomenclature used at different stages in the process of genetic testing. I specifically describe their use of evidential and epistemic markers and demonstrate the motivated relationship between these morphemes and the contexts in which the different systems are used. Finally, I conclude by analyzing the way the systems segregate and interact, demonstrating the relationship between expertise and evidential markers. In this way, drawing on the concepts of evidentiality and epistemic modality allows us to see the pragmatic rationale behind the confusing system of genetic nomenclatures.

Sociology of Medical Genetics

The primary contrast that my interlocutors make in the professional regimes of medical genetics is between research and therapy. Laboratory scientists and clinicians have distinct roles in academic medical centers, operating under different paradigms of professional responsibility (Beauchamp and Saghai 2012). Roughly put, the scientists with whom I have worked are relatively more interested in genetic test results insofar as they constitute generalizable knowledge about populations. They are also more critical of purported findings at the *genotypic* level, in terms of a test's ability to detect a mutation in the first instance. They consider whether and how well a particular test methodology can identify a particular type of mutation. On the other hand, my clinician interlocutors are responsible for determining how a result might inform the treatment of an individual patient in their care. They are more concerned with genetic findings at the *phenotypic* level, in terms of how those findings might relate to a particular health condition.

Along this dimension from laboratory to clinic, professional authority is granted incrementally over abstract and experimental knowledge-production toward one pole, and practical therapeutic intervention toward the other. Of course, reality is more complicated than this simple model suggests; hybrid clinician–scientists are common figures in the Department of Medical Genetics. Many of the laboratory heads with whom I have worked also hold medical doctorates, and many of the specialist clinicians collaborate intimately with laboratories in their research. In fact, that these two worlds collide is the very reason the practice is referred to as “translational medicine” in the first place (Mankoff et al. 2004): translating theory into practice. The Department of Medical Genetics stands as a sort of “trading zone” (Galison 1999), a site of conjuncture between two different professional paradigms. Clinicians and scientists bring with them different goals and different vocabularies, but in this shared space, they work together to bridge these ontological gaps.

Nonetheless, a clear division of clinic and laboratory is the ideological construct that structures interaction and undergirds local imaginations of professional roles. Characteristic comments from scientists on the topic include statements such as “I’m not sure all clinicians have a deep enough understanding of genetics [...] and how to use that information appropriately,” and “I’m not sure that *any* of us understand the nuances of [clinical] practice.” In interviews, clinicians have expressed parallel anxieties: that the laboratory scientists struggle to understand the medical reasons for ordering a test, and that they as clinicians do not have the expertise needed to scrutinize the applicability of the various methodologies and algorithms used in such testing.

Transducing Evidence for Non-Experts

When scientists and clinicians talk about nomenclature systems, they characterize the labels and descriptions of genetic mutations as “names.” While these *names* point to the same individual genetic variation, they typify it in different ways, according to the grammar of the given system. As my interlocutors in the clinic consider the words and phrases used in genetic nomenclatures to be *names* of these variants, they ignore the heteroglot circumstances in which they find themselves (Bakhtin 1981). Especially for clinicians, it seems that the different systems blend into each other in their emic representation of the total linguistic system. They do not always recognize how one name differs from another, nor can they necessarily determine to which system a name belongs. “It’s really confusing,” a pathologist warned me when I began my research into this topic. Even in discussions with highly respected experts, my linguistic examples often met surprise and self-contradiction. When they learned about my interests, many of my interviewees asked me, non-rhetorically, “Why *do* people call genes so many different things?” In this way, the discreteness of the nomenclatures and their systematicity can sit below users’ limits of awareness with regards to their own language practices (Silverstein 2001).

Specifically, my interlocutors reduce the functions of the names to their referential values, equating a single word with a single object, as different ways of saying the same thing. In this manner, they erase the meaningful distinctions in what the different systems communicate, which provides the background for local confusion over the seeming randomness of code choice. Moreover, for many of my clinician interlocutors, it is not even apparent that these ‘names’ can encode anything other than a transparent reference to a discrete genetic mutation. While they certainly know that other systems exist – and that those systems are used vaguely “in the lab,” as I was told when I first arrived at my field site, standing as a sort of second-order indexicality

(Johnstone, Andrus and Danielson 2006; Silverstein 2003a) – the possibility that they can relate something other than merely greater genotypic specificity is often not clear to them. In what follows, I attend to the features of the nomenclature systems that in this ideology are ignored in order to demonstrate the pragmatic differences between them.

I follow a number of linguists (e.g., Aikhenvald and Dixon 2003; Aikhenvald 2004; de Haan 1999; Michael 2012; Speas 2008) in making a categorical distinction between evidentiality and epistemic modality in my analysis of genetic nomenclatures. This contrasts with much of the discussion of “evidentiality” in English and scientific discourse, in which scholars either conflate the two functions or leave them in an ambiguous relationship (e.g., Atkinson 1999; Hobbs 2003; Prince, Frader and Bosk 1982). It is not that they do so without reason: Distinctions between the two functions are difficult to ascertain in languages such as English, in which they are not rigidly segregated or obligatorily marked (Fox 2001). In fact, some languages conflate the two in a single grammatical category (e.g., Cheyenne, see Murray 2017) or correlate the two forms of meaning in invariant ways (e.g., Kashaya, see Chafe and Nichols 1986, 43). However, I demonstrate that a strong distinction represents the semantic reality of the particular nomenclatures I analyze in this article, and I furthermore argue that this distinction correlates with the social contexts in which the nomenclatures are deployed. As such, it behooves us to consider in more detail what is meant by the two terms when they are kept distinct.

Alexandra Aikhenvald, in her classic introduction to the topic, defines evidential markers as those grammatical elements that express “a source of evidence for some information” (Aikhenvald and Dixon 2003, 1). In the current article, it is particularly relevant that evidential markers can specify the type of evidence used to back up an assertion. In more prototypical cases from linguistics, this includes marking whether a proposition is based on hearsay, inference, or

eyewitness, among other things. In the systems of nomenclature I discuss below, evidential markers signal, e.g., which type of test has been used to assert the existence of a particular genetic variant.

In order to distinguish epistemic modality from evidentiality, epistemic markers have been defined as those grammatical elements that express “the degree of commitment on the part of the speaker for his or her utterance” (de Haan 1999, 84). Thus, while evidential markers are assertions of source, epistemic markers are evaluations of reliability or confidence. Epistemic marking in genetic nomenclature systems signals, e.g., the modulation of certainty that a mutation exists or that it causes a particular disease.

Scholars of science and technology (e.g., Latour and Woolgar 1979; Latour 1987) have discussed similar phenomena in scientific fact-making under the rubric of “modalities,” qualifications about the specific circumstances under which a statement was constructed. These scholars argue that greater facticity is achieved by removing modalities from statements aspiring to be facts. In what follows, however, I argue that linguistic attention to the actual appearance and use of these “modalities” demonstrates a more complicated sociology of knowledge and knowledge production. In the practice of medical genetics, epistemic and evidential markers in genetic names illuminate the different types of expertise at play in the various contexts where clinicians and researchers interact.

Transductional Medicine: From Laboratory to Clinic

Genetic mutations receive their names through the process of genetic testing. The point of conducting a genetic test at my field site is first and foremost to provide a patient with a diagnosis. Both clinicians and scientists generally assume either that for the majority of patients,

disease is caused by a single genetic mutation, or that genetic testing is not yet advanced enough to be able to interpret the intersecting consequences of multiple genetic abnormalities.

(Sometimes clinicians put forward both of these assumptions at once, and many times they do not explicitly distinguish between the two possibilities.) Therefore, clinicians and scientists focus on finding a single variant in the patient's genetic code that could be responsible for his or her bad health. I refer to this as *the presumption of a single unified cause*.

Because of the relative novelty of both the science of genetics and its medical implementation, it only occasionally proves diagnostically successful for the patients at this specific clinic. A contemporary report from an unrelated institution demonstrated a 25% success rate in diagnosing patients using similar genetic tests (Yang 2014). This percentage is roughly analogous to the success of my interlocutors in determining the cause of an undiagnosed rare disease. If a diagnosis can be established, it is hoped that testing will also lead clinicians to a recommendation for a new or modified therapeutic regimen. The likelihood of this is, of course, even smaller than that of an initial diagnosis. (It is critical to note that the 'success' of genetic testing has nothing to do with the patient's actual health outcome but rather only with the test's ability to provide evidence in support of a diagnosis or treatment. That is, a patient can succumb to cancer while still being considered a success if testing reveals a new or more specific diagnosis.)

The divergent professional responsibilities of laboratory scientists and clinicians fall along these lines: Scientists are motivated to conduct genetic tests in order to discover new genotype–phenotype correlations and diagnoses relevant to a broad population. Clinicians, on the other hand, are primarily interested in genetic testing because of its ability to aid in prognostication and therapy for individual patients.

Among scientists, comments about genetic findings such as “We don’t feel it’s clinically relevant; it’s just for academic interest” can spark long conversations about the implications of the finding for generalizable genetic knowledge. Contrarily, when such “academic” concerns are raised in front of clinicians, they are regularly shut down in favor of issues with clear therapeutic value. For instance, two scientists at an interdisciplinary conference began discussing a particular patient’s mutation that one of them had dubbed “interesting from a biology point of view, but nothing clinically actionable.” The leader of the meeting, a clinician, quickly interrupted: “Folks, we’ve only got a few minutes left for this case.”

The two broad professional classes also differ in terms of their role in engaging reported genetic variants. One of the primary concerns for laboratory scientists is to determine whether a certain variant reported by a machine is “real” – that is, whether it actually exists in the patient’s genome or whether its report is simply the result of a computational error. Contrarily, the primary concern for clinicians is to determine whether a reported variant could in fact cause the particular medical problems the patient is experiencing.

The various systems of nomenclature used in the laboratory thus encode information relevant to the determination of whether a variant is “real,” while those used on the clinical side encode more information used to determine the likelihood that a variant causes disease. As the results from the tests circulate and accumulate medical interpretations, they change in form in order to bear (only) that information relevant to their rotating audiences. The semantic information encoded in the names is ‘transduced’ (Silverstein 2003b), changing in content and structure in some salient ways while maintaining the appearance of identity. And yet, the information load is not simply reduced, and it is not simply reformulated. Information that was once salient stands only as noise to a new audience and is therefore sloughed off. *At the same*

time, information that would not have been relevant to the previous audience is added for the benefit of the new audience. Both addition and reduction are important to the role of the transduction of genetic information as it is transposed into different nomenclature systems.

Rather than as pure and transparent representations of genotypic information, I argue that the classificational work of these nomenclature systems is motivated by the social meaningfulness of the variants within particular branches of contemporary medical genetics. This argument builds on work in the social studies of science (Bowker and Star 1999; Hacking 1986; Lampland and Star 2009; Star and Griesemer 1989) about how social and historical contexts structure such systems in ways that supersede transparent representation. In fact, some of the systems – especially those used by clinicians – work more to “facilitate action” rather than to “advance understanding” regarding the variants to which they refer (*contra* Mauss and Durkheim 1963, 81).

Sociolinguistic Overview

Having laid out the setting of the clinic at which I worked and briefly described the professional roles of the key classes of actors, I now return to the question of genetic nomenclatures. Below I analyze five prominent methods for referring to genetic mutations used in the clinic at which I conducted my fieldwork. Some of these systems create incredibly morphologically complex ‘names,’ while others yield quite simple results. Some names are ‘pronounceable’ and useful in verbal exchange, while others can only appear in written form. In order to understand how these systems relate to one another, some further information is also provided about the process of genetic testing itself. I roughly follow the actual chronology of the

test, beginning with the representation of the lowest-level or smallest-scale results from a laboratory test and progressing to the form in which those results are returned in the clinic.

I. FASTQ Nomenclature

In order to be sequenced, scientists extract DNA from a sample of a patient’s blood or other tissue. The DNA is chemically “amplified” – producing thousands or millions of copies in order for the sequencing machine to “read” the DNA more easily and more accurately.

Amplification requires that the strings of DNA be fragmented into segments of approximately 100 pairs of molecules called nucleotides – cytosine, guanine, adenine, and thymine (or C, G, A, and T, respectively). The duplicated segments are then reassembled according to the similarity of a particular segment’s sequence to the “reference genome,” a digital collection of DNA from multiple individuals considered to represent a “healthy” genotype. For many reasons, including the fact that this reassembly is not always perfect, the process can result in the emergence of errors in the reported sequence. The nucleotides assembled at a particular location do not necessarily represent the nucleotides at that location within the patient’s actual genome.

In order to make this genetic information discrete and legible, an algorithm transforms it into a FASTQ text file.⁴ An example of a segment of DNA represented in FASTQ nomenclature is presented in (1a-d) below.

```
(1a) @EAS139:136:FC706VJ:2:5:1000:12850 1:Y:18:ATCACG
(1b) GATTTGGGGTTCAAAGCAGTATCGATCAAATAG
(1c) +
(1d) h<BBCCCC?<A?BC?7@@@@@DBBA@@@@A@@@
```

Each DNA segment is given a particular sequence identifier, presented in the first line of the item (1a) following the at sign <@>. The second line (1b) consists of the actual sequence of nucleotides in the segment, represented by their abbreviations: A, C, T, G. This is called the

“short read,” as it merely represents a segment of DNA, or “read.” The third line (1c) always only contains a plus sign <+>, connecting the read with the “quality score” in the fourth line (1d). The quality score is made up of a number of characters equal to that of the short read, with each nucleotide paired with one of 93 ASCII characters. The quality-score character denotes the statistical probability that the nucleotide represented in the short read actually exists as such in the patient’s physical DNA sequence. The probability is determined based on the number of reassembled reads at a given location in the genome. The nucleotides with the lowest probability of existing are represented by non-alphanumeric characters, while those with the highest likelihood are represented by miniscule alphabetic characters. Thus, the first nucleotide in (1) is quite likely (h) to be truly a guanine molecule (G), while the second nucleotide is much less likely (<) to be an adenine molecule (A) – though it is still more likely to be adenine than any of the other three nucleotides.

FASTQ nomenclature therefore inflects two points of information: specific nucleotide (the short read, line 1b) and epistemic status (the quality score, line 1d). Scientists consult the FASTQ file in order to review the fine-grained epistemic information it encodes. If a mutation is selected subsequently as a candidate cause for a patient’s disorder, scientists can assure themselves that it is reasonably certain to exist by referring back to its FASTQ notation. One cancer biologist put it this way to me: “I would only look at [FASTQ] data if I am doubting what a coworker is doing with their interpretation.”

II. VCF Nomenclature

While most genetic nomenclature systems appear complex to both novices and experts alike, FASTQ is particularly arcane. Most human actors involved in genetic testing never or only

rarely refer to the FASTQ file; instead they encounter the genetic information in a more accessible (and more derived) form. After its FASTQ instantiation as bare code, genetic information is transduced into the Variant Call Format (VCF) by a computer algorithm.⁵ (A “call” is the term for the proposal that a specific nucleotide has been found at a given location in a patient’s genome.)

The FASTQ format conveys every nucleotide that has been sequenced, even those variants that match the reference genome and are thus considered healthy and not candidate causes for a patient’s disorder. The VCF file, on the other hand, consists of a representation of only those genetic variants that differ from the reference genome. Thus, the VCF file is in part a reduced version of the FASTQ file, making it more legible to audiences that are less expert in genetic science.

Yet, as a laboratory scientist told me, the VCF format is still “generally unintelligible to the general population,” meaning clinicians – who are not its intended audience. Clinicians regularly call VCF files “raw data,” as they are the first and most complex forms of genetic information that they typically encounter. I have, however, found some discrepancy in what counts as “raw data.” As a scientist explained, “it means different things to different people.” For clinicians, “raw data” refers to the VCF file, that is, to only those segments of DNA that differ from the healthy reference genome and thus are potential sources of a patient’s disease. Laboratory scientists, on the other hand, scoff at labeling the VCF file “raw data,” as it is already much reduced and reformatted (thus not “raw”) from the FASTQ format presented above.

That said, the VCF file is not simply a reduced version of FASTQ; the information it presents is both reordered and supplemented. An interlinear gloss of an exemplary VCF name is given in (2):

(2) *chr17* 43094464 *rs1799950* A G 29 PASS DP=100
 CHROM POS ID REF ALT QUAL FILTER INFO
 ‘Mutation from A to G on chromosome 17 at position 43094464, uniquely identified in dbSNP⁶ as “rs1799950,” called with high (29) certainty after having passed all filters, and with 100 reads covering this locus.’’

For the purposes of this article, it is important to note three key features of the VCF name. First, it only contains information regarding variation from the reference genome. Based on contemporary statistics, any given human being’s DNA is said to differ by merely 0.1% from the reference genome. Thus, a VCF file covers only a minute fraction of a patient’s complete sequence of nucleotides. It is much reduced in the information it bears.

Second, VCF names continue to exhibit the epistemic status of a particular nucleotide called at a particular location in the genome – giving it an explicit “quality score,” as in the FASTQ format. However, VCF names additionally provide a second marker – namely, the number of reads at a particular location (termed the “DP” or “combined depth across samples”). As the reader will recall, the more reads of the amplified DNA that return a particular nucleotide at a particular location, the greater the likelihood that this “call” does indeed accurately reflect the molecular reality of the particular sequence.

An understanding of how to convert “depth” into probability allows scientists to judge the reliability of the whole call. In general, a score between 30 and 50 reads at a particular location is considered an average return for a test. A score of 100 reads, as indicated in (2), is considered high. The DP segment thus represents a midway point between evidential and epistemic functions. While the pragmatic role for its inclusion in the name is to allow the reader to form a judgment about likelihood, it does not explicitly encode this information. The score has a conventional implicature (Grice 1989) of corresponding probability, but it requires some awareness and expertise in order to reach that pragmatic conclusion. An expert’s *interpretation*

of the DP segment coupled with the overt epistemic content borne by the quality score determines the confidence with which the call will be mobilized in clinical care.

Finally, the ID segment provides the first link between genotypic and phenotypic data. While not explicit in the VCF name itself, this unique identifier allows the reader to discover more medically relevant information about the variant. The ID can be entered into an online database in order to discover a mutation's "clinical significance." (In the above example, the database suggests that the variant is of uncertain clinical significance, but that it is located in a gene associated with breast cancer. Such data typically leads clinicians to propose further testing in order to determine whether any link exists between the particular variant and the genetic process that leads to the associated disease.) Thus, the VCF name begins to be relevant to the professional responsibilities of clinicians as well as to those of scientists.

III. Cytogenetic Nomenclature

The first two systems of nomenclature are used specifically with genetic sequencing, in which information is captured at the molecular level in the form of data about individual nucleotides and their ordering. Many genetic tests, however, work at a larger scale, looking instead at structural variation in chromosomes. In these instances, genetic variation is labeled using cytogenetic nomenclature. A standardized form of this system was first proposed in 1960 at a conference of genetic researchers (Robinson 1960). It underwent a number of subsequent changes over the next several years before coming under the jurisdiction of the newly formed International Standing Committee on Human Cytogenetic Nomenclature in 1976 (Shaffer et al. 2013, 15). The following year, the Committee produced the first complete system of nomenclature, called the International System for Human Cytogenetic Nomenclature (ISCN), a

system that (although since modified multiple times) is still used in contemporary naming practices. Unlike the systems studied in much earlier scholarship on classification (e.g., Lévi-Strauss 1962; Mauss and Durkheim 1963), genetic nomenclatures are not implicitly but rather explicitly developed and refined through reflective consensus for the purpose of addressing the social needs of their users, within their particular social contexts. Because of its origin, the ISCN system is particularly tailored to laboratory work and circulates primarily among cytogeneticists, scientists who study the structure and abnormalities of chromosomes. An interlinear gloss of a cytogenetic name is given in (3).

(3) 46, XX .ish ins(15;17) (q22;q21q21) (PML+,RARA+RARA+)
 CHROM SEX METHOD TYPE LOCATION PROBES
 ‘An insertion of the segment 17q21 from the long arm (q) of chromosome 17 into the 15q22 band of the long arm of chromosome 15, identified with probes for *PML* and *RARA* genes and using the chemical methodology of in-situ hybridization (*ish*).’

Of particular interest to the analysis at hand is that cytogenetic nomenclature provides both a chromosome- and a gene-scale representation of the mutation. In (3) the mutation described exists in an individual with 46 chromosomes (“normal”) and two X chromosomes (“normal female”). While this information is of little salience in determining the effects or location of the specific mutation, it provides a “global” context with which the mutation could interact. For instance, if the number of chromosomes or combination of sex chromosomes is atypical, one cytogeneticist explained to me, she would likely expect the specific mutation to be related to this abnormality. This is a part of the presumption of a single unified cause, as described above. If the chromosomes are disordered in some way, a scientist will anticipate that this genotypic disorder will prove the cause of the patient’s disease at the phenotypic scale as well. If the reported mutation is unrelated to the chromosome number, suspicions will be raised

as to the validity of the test. That is, if the disease is not related to a chromosomal abnormality, scientists may suspect that the reported chromosomal abnormality is artefactual rather than real.

The cytogenetic name describes a relatively more specific location of the mutation as well. It notates which chromosomes and which arms of those chromosomes are implicated in the mutation. For instance, in (3) there is an insertion of a segment of the long arm (q) of chromosome 17 into a segment of the long arm of chromosome 15. Thus, the name details information both about the “global” context of ambient features such as chromosome number as well as about the “local” context of the specific mutation.

Cytogenetic names are not complete without the addition of markers indicating evidential support for the mutation’s presupposed existence. Scientists consider certain methodologies to be unreliable or below the standard of practice. For example, certain methodologies are known to be unable to recognize particular types of genetic segments. That knowledge can lead scientists to dismiss such results outright, calling them “not real.” The markers denoting the probes (molecules used to detect specific segments of DNA) and methodology used in the test can provide the astute scientist with data by which to judge whether the mutation under consideration should be reported to clinicians as the candidate cause of a patient’s disorder.

For example, one cytogeneticist described a case in which a test had suggested a patient had DiGeorge syndrome, a disease caused by a genetic deletion on chromosome 22 and associated with heart defects and many other medical issues. Upon reviewing the report, however, he noticed that the evidence denoted in the mutation’s cytogenetic name did not conform with what he had expected. The information about the probes (disclosed as evidential markers) was atypical for what he as an expert knew about the methodology by which this particular mutation was normally discovered. “I was worried it could be an indication of failed

hybridization,” he told me. After redoing the test, he discovered that the reported mutation did not exist in the patient’s actual genome, and in fact a completely different mutation was the true cause of the patient’s symptoms.

On the other hand, a different laboratory scientist described a contrary case of unexpected evidential markers. During his graduate training, this young man had come across a probe (evidential) marker that did not match his expectations for the particular test results. He decided to run the test a second time. “It was good to re-check,” he noted, “to be sure the region [of the chromosome] was well captured and well covered” by the test, but in the end, he discovered that the first test had in fact been accurate. He attributed his misinterpretation to his youthful inexperience and lack of “connoisseurship,” in his own words, which he stressed is necessary for making these kinds of discriminations. This case demonstrates the cancelability of the epistemic implications of these evidential markers (cf. Speas 2008), further underscoring the important distinction between the two linguistic functions.

Scientific articles may begin by using the full cytogenetic name of a mutation under discussion. Thereafter, authors typically switch to a “shorthand” or abbreviated form, which points back anaphorically to the specificity of the referent provided by the full cytogenetic name. Thus scientists and clinicians say things such as “He has a 17p deletion,” or “The test confirms t(11;14).”⁷ That is, the location information is reformatted to demonstrate only on which arm – long (q) or short (p) – of which chromosome the mutation is found, rather than the highly specific location information encoded in the full name. The mutation type – typically a deletion, duplication, insertion, or translocation – is spoken either in full or using the abbreviation. These names do not convey all potentially salient information that a test has provided. For instance, they fail to denote the evidential information presented in the full cytogenetic name. These

abbreviated forms are the most common in the regular PowerPoint presentations of a patient's case to interdisciplinary audiences of specialists.

As with the VCF nomenclature, the morphological complexity of cytogenetic names means that they are not necessarily transparent in their meaning, even for specialists. For instance, while discussing nomenclature systems with me, one cytogeneticist admitted that she did not know what the methodology marker “enh” means. She then retrieved a copy of the 2013 edition of the *ISCN* handbook in order to look it up. “Here, *enhanced*, oh!” she said when arriving at the appropriate page. “Enhanced fluorescence. Okay, I haven't seen that one [before]. So it's talking about enhanced fluorescence and that methodology.” She admitted that unless pressed, she would have passed over the marker without attempting to decipher it. Not all markers are important for sufficient comprehension; they become salient due to context. However, because of the rules of syntax in these names, she could tell by its position that *enh* must denote the form of methodology used to discover the mutation (that is, that it was an evidential marker). Knowing what part of the name on which to concentrate, she said, “would depend on what you're looking at. [...] It would depend on what abnormality you're looking for.”

Scientists use systems like cytogenetic nomenclature because they are particularly interested in coming to conclusions about a reported mutation's existence. The use of evidential markers, as we have seen, allows such experts to make these determinations based on epistemic implicatures specific to their expert subculture. This calls to mind Susan Philips's (Philips 1992) discussion of the variability of “evidentiary standards” within the American legal system. The interpretations – and thus utility – of different forms of evidence are dependent on the presuppositions about their reliability as espoused by their professional groups.

The epistemic implicatures of these evidential markers only arise for a group of experts with scientific background knowledge about the relationship between test methodologies and the types of results they can and cannot produce. For instance, certain methodologies used in particular tests can be conventionally interpreted by experts to qualify the likelihood that a result is accurate. Because a scientist knows that, as noted above, a particular probe is incapable of reading a given mutation, that scientist may dismiss the report of such a mutation by that type of probe. The evidential marker describing the probe does not in itself or abstractly suggest that the result is unlikely; only in the particular context does it imply as much. Thus an evidential marker of how the result was found (what methodology) can be mobilized for epistemic purposes (confidence). Interactionally, the evidential can carry epistemic weight (Fox 2001; Kärkkäinen 2003).

However, again, these epistemic implicatures are cancelable because a result can still be considered “real” even if it does not match the expert’s evidential expectations. For example, in the case of the graduate student discussed above, the scientist initially discredited a result but upon further scrutiny, he realized the result was in fact accurate. The epistemic implicature he had attributed to the evidential marker (probe type) was canceled and he returned to assuming that the report represented his patient’s genetic reality. That is, the result was based on an unexpected methodology but it was still considered valid.

IV. Symbols

Clinicians, with their limited scientific expertise and greater interest in the care of individual patients, have much less use for evidential markers and for the related question of the existence of genetic mutations. More relevant for clinicians is the relationship of a genetic

mutation to the presentation of a disease. In order to explore this, I now turn to the most common nomenclature system employed in the Department of Medical Genetics, and the system that is most regularly used in verbal communication. This system consists of so-called “gene symbols.” Such symbols include lexical items such as *APOE4*, *EGFR*, and *BRCA1*.⁸ These names are spoken in case conferences and in “drive-by” consultations, they are written in articles, and some laboratory reports meant for patient readers use them as well. Symbols are relatively analytic or motivated, in that the letters that compose them often derive from the names of their gene products or the names of the diseases that they affect or effect. For instance, *APOE4* produces apolipoprotein E, and *EGFR* produces the EGFR protein, or epidermal growth factor receptor. (Symbols are traditionally italicized in print in order to distinguish the names of genes from the names of their products.) *BRCA1* is understood to be a major cause of breast cancer, hence its name relates to an even more derived or distal product of the gene. Many other genes – such as *PARK2* (Parkinson’s disease), *CFTR* (cystic fibrosis), and *PGL2* (paraganglioma, a rare type of tumor) – follow this scheme as well. In fact, as early as 1979 the Human Genome Organization officially recommended that “the name of the gene should describe the function of the enzyme” which it encodes (Wain et al. 1999, 162).⁹

Gene symbols are by far the most morphologically simple of the nomenclature systems, and semantically they appear relatively indeterminate. They are short, easily memorized, and quickly spoken – characteristics that promote their use in all forms of communication. Most important for clinicians is the explicit link they provide to their associated phenotypes. For clinicians, the utility of genetic testing is to determine potential healthcare interventions. Knowing a patient has a *BRCA* mutation, a clinician can immediately suspect breast cancer, or

hearing that a *RET* mutation has been found, an oncologist can recommend an inhibitor to stymie the mutated production of the RET enzyme.

However, symbols convey very little information. They merely demonstrate the gene in which a mutation exists rather than the specific variation it entails (which is what all the previously described systems of nomenclature do). Symbols' referents remain vague relative to the specificity provided by (and in fact required of) genetic testing. These names clearly do not convey all potentially salient information that a test has provided. For instance, they fail to describe what specific nucleotides are affected by the mutation they name. (Genes can be thousands – sometimes even millions – of nucleotides long.) There are also no evidential or epistemic markers. Symbols simply take the accuracy of the test result and the existence of the mutation for granted. They are facts without Latourian “modalities.”

While symbols like *BRCA* can refer to classes or types of genes – that is, the gene as it exists abstractly as a feature of all human genomes – in actual interaction, clinicians use symbols to pick out a particular mutation within a particular patient's genome. Thus, when a clinician says, “The patient tested positive for *BRCA*,” he or she is clearly not saying (merely) that the patient has a *BRCA* gene, but that the patient has one of a closed number of pathogenic mutations within that gene. The form is noticeably underspecified in that it does not in itself provide enough information to describe a unique individual mutation; however, it nonetheless acts as a name to refer to such an individual, rather than to a class. (Such specificity could be provided through the prior mention – either verbal or written – of a morphologically more complex name for the mutation.)

Certain gene symbols, like *BRCA1* and *BRCA2*, can be reduced to just their alphabetic portions. *BRCA* can refer equally to either gene, despite the former being located on chromosome

13 and the latter on chromosome 17 – and despite their producing totally different proteins. The genes are merely united in their eponymous link to breast cancer. This set of symbols therefore takes on a classificatory quality, not denoting a uniform segment of genetic material or its proximate gene products, but rather denoting the disease states associated with the genes. That is to say, such gene symbols are motivated primarily by their clinical relevance. Their ability to represent a variant’s social meaningfulness supersedes their ability to represent its genotypic specificity. As one clinician told me, “The value [of a genetic result] isn’t the raw data. It’s what the result says; it’s what it *means*.” The purpose of naming the result in the first place, in her opinion, was to demonstrate its relevance to clinical care.

V. Descriptions

Finally, I turn to the system that my clinician interlocutors most commonly refer to as “calls.” In this article, however, I prefer to adopt the term “descriptions,” which I take from the genetics literature. I use *description* in place of the favored local term in order to avoid confusion with the homonym *call* (described above), which refers to the scientist’s proposal that a particular nucleotide exists at a specific location in a patient’s genome. In contrast to this kind of *call*’s scientific interest in a mutation’s existence, descriptions are used when no genotypic information is required by the communicative situation. That is, descriptions encode only phenotypic information, clinical information about the likelihood that a particular mutation causes disease. The fact that clinicians refer to this system as “calls” betrays a locally imagined parallel between the scientist’s role in asserting a mutation’s existence (what one might distinguish as a scientific call) and the clinician’s role in asserting its link to disease (that is, a

clinical call). Because of their focus on phenotype, descriptions are the system preferred by clinicians in the Department of Medical Genetics.

Common descriptions used at the clinic under discussion include, for instance, *the likely benign variant* and *the variant of uncertain significance*. Descriptions explicitly encode epistemic information with modifiers such as *likely* and *uncertain*. What these epistemic markers qualify is the proposed link between genotype (mutation) and phenotype (disease). They do not qualify the likelihood that the named mutation exists in the first instance.

Systems of descriptions vary in localized fashion, in a dialect-like distribution across different institutions and laboratories. Typologically, they all constitute their names along two axes, namely, pathogenicity and certainty. (See Figure 1.) Any variation from the reference genome either causes disease (is pathogenic) or represents “normal variation” (is benign). A description links this aspect of the mutation with a qualification of an expert’s confidence in the characterization.

[Insert Figure 1 here.]

Local systems differ in terms of the number of points along this parabola and the specific lexical items used to denote those points. (For example, some ‘dialects’ use the term *deleterious* in place of *pathogenic*.) The system most commonly used in the clinic at which I worked consists of a five-point scale: pathogenic variant, likely pathogenic variant, variant of uncertain significance (VUS), likely benign variant, and benign variant. Some laboratories recognize six distinct points along the parabola, and a genetic counselor told me others recognize as many as seven, each with finer gradations of pathogenicity and certainty.

We can see that the system of descriptions is simultaneously the most heavily clinical and the most prominently epistemic of all the systems I have covered in this article. By the time

genetic test results are discussed in the clinic, they are presupposed to exist. That is, they are not under the same scrutiny as they are in the laboratory, where evidential and epistemic markers are used to hedge the reliability of a reported mutation's existence in the first instance. Descriptions, on the other hand, use epistemic modifiers to hedge the likelihood that a given mutation causes a clinical condition.

Descriptions represent a transduced version of the information presented in more scientist-oriented systems of nomenclature. While descriptions carry almost no content related to genotype (with regard neither to the genetic variant's location nor to its form), they include phenotypic information. To the interpretation of the mutation they add evaluations of how the variant under discussion relates to the presentation of a patient's disease.

Conclusion: Expertise and the Fate of Evidence

The Department of Medical Genetics at which I worked constitutes what Peter Galison calls a trading zone (Galison 1999). It is the site of exchange between laboratory scientists and clinicians, whose individual expertise and interactional goals differ significantly. Scientists seek to discover new mutations, new testing methodologies, and new forms of disease, to generalize that knowledge and apply it to populations. Clinicians, on the other hand, operate under a different regime of professional responsibility. They seek to intervene, to translate into practice those entities and mechanisms constructed by the scientists.

When they encounter each other in the Department, both clinicians and scientists work actively to reach an understanding of the purposes of their interlocutors. In order to 'trade,' they need to know how to market their goods, to what ends they will be put, and in what form they will be legible. As a molecular geneticist explained, the most important thing for scientists is to

learn “how the [clinicians] are managing their patients on a day-to-day basis.” That knowledge influences what information they choose to contribute.

The nomenclature systems presented above represent the differing concerns of their primary or stereotypical users. What can appear as an anarchic distribution of systems is explicable with a look beyond their referential function. Different systems encode different information in their morphology, corresponding to their users’ different social needs. As reports of mutations move closer to the clinic, their iterations begin to ‘reduce’ the scientific evidence they present. By the time a report faces non-specialist clinicians and patients, it has already been vetted and approved by a number of science-literate experts. The existence of the mutation is presupposed. Moreover, data such as test methodology and the specific location of a mutation within a gene are understood to be meaningless to the average clinician. This information is therefore removed. However, the transduction of nomenclature systems is not merely a simplification; it is a simultaneous incorporation of new and reformulated information relevant to the new audiences. For instance, as nomenclature systems move toward the clinic, they lose genotypic information about location and mutation type, but they also incorporate more phenotype-centric information such as predicted pathogenicity and information about correlated disorders.

Each nomenclature system provides the act of naming with a different array of semantic and pragmatic potentials. The systems described above obligatorily encode certain data while making other information merely optional or even impossible to encode. In nomenclatures used primarily by scientists (such as FASTQ and VCF), epistemic markers function to qualify the likelihood of a mutation’s existence. In these systems, evidential markers have conventional implicatures of epistemic evaluation, but this is not itself semantically encoded. In fact, these

implicatures only emerge within the specific context of expert evaluation. The relevance of such a marker for the interpretation of a test result is typically lost on the average clinician, but a laboratory scientist can use that evidential datum to discredit the result and even remove it from the list of candidate mutations.

Clinicians, on the other hand, are interested in the therapeutic relevance of a mutation. They have a more limited claim to scientific expertise and less of a pragmatic interest in genotypic information. Nomenclatures used primarily by clinicians therefore favor epistemic markers that qualify the likelihood that a particular mutation is the cause of a patient's disease.

We have seen that one of the key ways genetic nomenclature systems are distributed is along a scale according to whether a particular name encodes more genotypic or phenotypic information. I have further argued that this scale corresponds to relative expertise in laboratory science. The use of evidential markers correlates with the scientific competence to make epistemic judgments based on conventional implicatures specific to professional subcultures. Meanwhile, epistemic markers correlate with relative naivety in such science, relying instead on explicit and ready-made evaluations in order to make practical, therapeutic use of the genetic data that these names encode. These observations allow us to explain an otherwise random-seeming distribution of nomenclature systems. They also point to an intrinsic relationship between evidential and epistemic functions of sign systems and the social and hierarchical contexts in which they are used. I suggest that the pragmatic role of evidential marking for epistemic ends is likely to be found in a variety of other contexts of expertise and connoisseurship (e.g. declarations of *appellation* as a means of evaluating wine quality, see Silverstein 2003a).

It is difficult and not always problematic to ignore the boundaries between evidential and epistemic contributions of linguistic elements (e.g. Prince et al. 1982). However, important sociolinguistic insights can also be gained by attending to just such differences. I have presented an analysis of how different nomenclature systems segregate based on the professional responsibilities of different medical personnel. I have drawn out the variation between the systems in terms of the semantic and indexical content they encode and the ways that content is transduced and supplemented in translation between systems. By reducing the significance of mutation names to their referential function, much of their total meaning is lost. Instead, by observing how they function relative to the specific expert subcultures of their users, we discover that they are well-tailored and contextually contingent.

ACKNOWLEDGEMENTS

This paper is based on research funded by the Wenner Gren Foundation and the Hanna Holborn Gray Mellon Fellowship. It has benefited greatly from the close readings of Costas Nakassis, Chris Bloechl, Michael Chladek, Lindsey Conklin, Janet Connor, Catherine Egenberger, Britta Ingebretson, Adam Singerman, and Anna Weichselbraun. I am further indebted to the kind and helpful feedback from my discussant Summerson Carr and the attendees of the 2016 Michigan Conference, at which I presented a part of this analysis. I am also grateful to the anonymous reviewers for this journal and to its co-editor-in-chief, Chaise LaDousa, for their thoughtful comments and support. Finally, I would like to thank all of my clinician and scientist interlocutors for their generous consultations on this project. All errors, of course, remain my own.

NOTES

¹ While clinicians occasionally contend that genetics is absolutely determinative of disease, they nearly always consider it to be at least a factor in disease's expression.

² While testing is becoming increasingly common – and for many of my interlocutors, “mainstreaming” it as standard practice remains their goal – at the time of my fieldwork, it was still offered primarily in later stages of disease and in cases of more acute distress.

³ I rely on Michael Silverstein's definition of the term *transduction* throughout the paper. It is worth noting that this is distinct from a variety of other uses of the term in related literature, such as Stefan Helmreich (2007) and Stephen Black (2017), who use the word to refer to the conversion of sound to and from other media. *Transduction* is also in use in the genetics literature itself, referring to the transfer of genes between bacteria. My thanks go to my second reviewer for highlighting this potential point of confusion.

⁴ The name FASTQ derives from an older format for representing nucleotide sequences, namely, FASTA. The Q stands for *quality*, as in the quality scores, which are an addition to the older FASTA format.

⁵ In fact, most transductions between nomenclature systems are completed by computer programs. My interlocutors – both clinicians and scientists – are generally unaware of how different systems correspond. The only common transformations enacted on mutation names are the so-called “shorthand” forms discussed in section III, but even these represent set correspondences rather than active, conscious attempts at transducing between codes.

⁶ dbSNP is a popular database of genetic variation run by the National Institutes of Health.

⁷ These reduced forms mean, respectively, that the patient under discussion has (1) a deletion from the short arm (p) of chromosome 17 (associated with multiple myeloma), and (2) that the patient has a translocation of genetic material from chromosome 11 onto chromosome 14 (associated with lymphoma). Both of these particular mutations are common enough that such reduced names are anticipated to cause no confusion among specialists in the clinic.

⁸ All gene symbols are orthographically standardized (maintained by the HUGO (Human Genome Organization) Gene Nomenclature Committee), and most have entirely regular pronunciations. However, *BRCA* – which is one of the most commonly discussed mutations in the Department of Oncology – has a variety of acceptable phonological realizations. It is pronounced variously as [ˈbɪ.ækə], [ˈbɜːkə], or spelled out as [bi.ɑːsiˈeɪ], though the first pronunciation is the most common in the clinic at which I worked. Other symbols bear more limited variability. *RPTOR*, for instance, was most often pronounced like the word “raptor,” but the unapparent vowel in the first syllable was sometimes realized as [ɪ] instead. It is not obvious from orthography alone whether a gene symbol is spelled out or pronounced as a single word. *HLBA*, e.g., is always spelled out, but *FLT3* is pronounced “flit 3.”

⁹ However, many genes are multifunctional, and the first recognized role of the gene, for which it is named, is not necessarily its most clinically important. For instance, *NRAS* is named after the disease to which it was first linked, namely, neuroblastoma. (It gets the letters *RAS* because the gene produces Ras-type proteins, which were first discovered in a cancer-causing virus found in rat sarcomas.) *NRAS* has since, however, been most prominently implicated in a developmental disorder called Noonan syndrome (Cristea et al. 2010).

REFERENCES

- Aikhenvald, Alexandra and Robert Dixon, editors. 2003. *Studies in Evidentiality*. Philadelphia: John Benjamins Publishing Company.
- Aikhenvald, Alexandra. 2004. *Evidentiality*. Oxford: Oxford University Press.
- Atkinson, Paul. 1999. "Medical Discourse, Evidentiality and the Construction of Professional Responsibility." In *Talk, Work and Institutional Order: Discourse in Medical, Mediation and Management Settings*, edited by Srikant Sarangi and Celia Roberts, 75-108. Berlin: Mouton de Gruyter.
- Bakhtin, Mikhail. 1981. "Discourse in the Novel." *Dialogic Imagination*, 259-422. Austin: University of Texas Press.
- Beauchamp, Tom and Yashar Saghai. 2012. "The Historical Foundations of the Research-Practice Distinction in Bioethics." *Theoretical Medicine and Bioethics* 33:45-56.
- Black, Steven. 2017. "Anthropological Ethics and the Communicative Affordances of Audio-Visual Recorders in Ethnographic Fieldwork: Transduction as Theory." *American Anthropologist* 119:46-57.
- Bowker, Geoffrey, Susan Star. 1999. *Sorting Things Out*. Cambridge: MIT Press.
- Chafe, Wallace and Johanna Nichols, editors. 1986. *Evidentiality: The Linguistic Encoding of Epistemology*. Norwood, NJ: Ablex Publishing Corporation.
- Crismore, Avon and Rodney Farnsworth. 1990. "Metadiscourse in Popular and Professional Scientific Discourse." In *The Writing Scholar: Studies in the Language and Conventions of Academic Discourse*, edited by Walter Nash, 118-136. Newbury Park: Sage.

- Cristea, Ion, Kerstin Kutsche, Radovan Dvorsky, Lothar Gremer, Claudio Carta, Denise Horn, et al. 2010. "A Restricted Spectrum of *NRAS* Mutations Causes Noonan Syndrome." *Nature Genetics* 42:27-29.
- De Haan, Ferdinand. 1999. "Evidentiality and Epistemic Modality: Setting Boundaries." *Southwest Journal of Linguistics* 18:83-101.
- Den Dunnen, Johan and Stylianos Antonarakis. 2000. "Mutation Nomenclature Extensions and Suggestions to Describe Complex Mutations: A Discussion." *Human Mutation* 15:7-12.
- Editorial. 1997. "Obstacles of Nomenclature." *Nature* 389:1.
- Fox, Barbara. 2001. "Evidentiality: Authority, Responsibility, and Entitlement in English Conversation." *Journal of Linguistic Anthropology* 11:167-192.
- Galison, Peter. 1999. "Trading Zone: Coordinating Action and Belief." In *The Science Studies Reader*, edited by Mario Biagioli, 137-160. New York: Routledge.
- Grice, H. Paul. 1989. "Logic and Conversation." In *Studies in the Way of Words*, edited by Peter Cole and Jerry Morgan, 43-58. New York: Academic Press.
- Hacking, Ian. 1986. "Making Up People." In *Reconstructing Individualism*, edited by Thomas Heller, Morton Sosna and David Wellbery, 222-236. Stanford: Stanford University Press.
- Halverson, Colin. 2019. Standards and legacies: Pragmatic constraints on a uniform gene nomenclature. *Social Studies of Science* 49:432-455.
- Helmreich, Stefan. "An Anthropologist Underwater: Immersive Soundscapes, Submarine Cyborgs, and Transductive Ethnography." *American Ethnologist* 34:621-641.
- Hobbs, Pamela. 2003. "The Use of Evidentiality in Physicians' Progress Notes." *Discourse Studies* 5:451-478.

- Jakobson, Roman. 1960. "Linguistics and Poetics." In *Style in Language*, edited by Thomas Sebeok, 350-377. Cambridge, MA.
- Johnstone, Barbara, Jennifer Andrus, Andrew Danielson. 2006. "Mobility, Indexicality, and the Enregisterment of 'Pittsburghese.'" *Journal of English Linguistics* 34:77-104.
- Kärkkäinen, Elise. 2003. *Epistemic Stance in English Conversation*. Philadelphia: John Benjamins.
- Kesselheim, Aaro, Carolyn Treasure, Steven Joffe. 2017. "Biomarker-Defined Subsets of Common Diseases: Policy and Economic Implications of Orphan Drug Act Coverage." *PLoS Medicine* 14:e1002190.
- Kroskrity, Paul. 2000. *Regimes of Language: Ideology, Politics, and Identities*. Santa Fe: School of American Research Advanced Seminar Series.
- Lakoff, George. 1970. "Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts." *Journal of Philosophical Logic* 2:458-508.
- Lampland, Martha, Susan Star. *Standards and Their Stories*. Ithaca: Cornell University Press.
- Latour, Bruno and Steve Woolgar. 1979. *Laboratory Life: The Social Construction of Scientific Facts*. London: Sage.
- Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge: Harvard University Press.
- Lévi-Strauss, Claude. 1962. *Totemism*. Boston: Beacon Press.
- Mankoff, Stacey, Christian Brander, Soldano Ferrone and Francesco Marincola. 2004. "Lost in Translation: Obstacles to Translational Medicine." *Journal of Translational Medicine* 2:14.

- Manolio, Teri, Rex Chisholm, Brad Ozenberger, Dan Roden, Marc Williams, Richard Wilson, et al. 2013. "Implementing Genomic Medicine in the Clinic: The Future is Here." *Genetics in Medicine* 15:258-267.
- Mauss, Marcel, Émile Durkheim. 1963. *Primitive Classification*. Chicago: University of Chicago Press.
- Michael, Lev. 2012. "Nanti Self-Quotation: Implications for the Pragmatics of Reported Speech and Evidentiality." *Pragmatics and Society* 3:321-357.
- Murray, Sarah. 2017. *The Semantics of Evidentials*. Oxford: Oxford University Press.
- Nussbaum, Robert, Roderick McInnes, and Huntington Willard, editors. 2016. *Thompson & Thompson Genetics in Medicine*. Philadelphia: Elsevier.
- Philips, Susan. 1992. "Evidentiary Standards for American Trials: Just the Facts." In *Responsibility and Evidence in Oral Discourse*, edited by Jane Hill and Judith Irvine, 248-259. Cambridge: Cambridge University Press.
- Prince, Ellen, Joel Frader, and Charles Bosk. 1982. "On Hedging in Physician–Physician Discourse." In *Linguistics and the Professions: Proceedings of the Second Annual Delaware Symposium on Language Studies*, edited by Robert Di Pietro, 83-96. Norwood, N.J.: ABLEX Publishing Corporation.
- Robinson, Arthur. 1960 "A Proposed Standard System of Nomenclature of Human Mitotic Chromosomes." *Current Science* 29:337-340.
- Salager-Meyer, Françoise. 1994. "Hedges and Textual Communicative Function in Medical English Written Discourse." *English for Specific Purposes* 13:149-170.
- Schieffelin, Bambi, Kathryn Woolard and Paul Kroskrity. 1998. *Language Ideologies*. Oxford: Oxford University Press.

- Shaffeer, Lisa, Jean McGowan-Jordan and Michael Schmid. 2013. *ISCN 2013: An International System for Human Cytogenetic Nomenclature*. Basel: Karger.
- Shoenbill, Kimberly, Norman Fost, Umberto Tachinardi, and Eneida Mendonca. 2014. "Genetic Data and Electronic Health Records: A Discussion of Ethical, Logistical and Technological Considerations." *Journal of the American Medical Association* 21:171-180.
- Silverstein, Michael. 1976. "Shifters, Linguistic Categories and Cultural Description." In *Meaning in Anthropology*, edited by Keith Basso and Henry Selby, 11-55. Albuquerque: University of New Mexico Press.
- Silverstein, Michael. 2001. "The Limits of Awareness." In *Linguistic Anthropology: A Reader*, edited by Alessandro Duranti, 382-401. Malden: Blackwell.
- Silverstein, Michael. 2003a. "Indexical Order and the Dialectics of Sociolinguistic Life." *Language & Communication* 23:193-229.
- Silverstein, Michael. 2003b. "Translation, Transduction, Transformation: Skating Glissando on Thin Semiotic Ice." In *Translating Cultures: Perspectives on Translation and Anthropology*, edited by Paula Rubel and Abraham Rosman, 75-105. Oxford: Berg.
- Star, Susan, James Griesemer. 1989. "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39." *Social Studies of Science* 19:387-420.
- Varttala, Teppo. 1999. "Remarks on the Communicative Functions of Hedging in Popular Scientific and Specialist Research Articles on Medicine." *English for Specific Purposes* 18:177-200.
- Wain, Hester, Julia White, and Sue Povey. 1999. "The Changing Challenges of Nomenclature." *Cytogenetics and Cell Genetics* 86:162-164.

Warner, Jeremy, Sandeep Jain, and Mia Levy. 2016. "Integrating Cancer Genomic Data into Electronic Health Records." *Genome Medicine* 8:113.

Yang, Yaping, Donna Muzny, Fan Xia, Zhiyv Niu, Richard Person, Yan Ding, et al. "Molecular Findings Among Patients Referred for Clinical Whole-Exome Sequencing." *Journal of the American Medical Association* 312:1870-1879.

FIGURES:

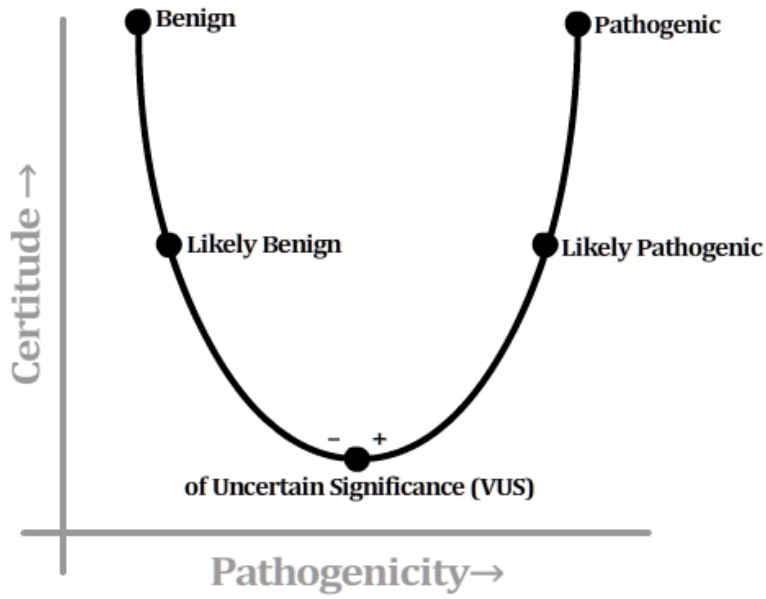


Figure 1: A five-point parabolic scale used to provide descriptions for genetic test results.