**Defining, detection, and dealing with outliers**

**requires values-laden decisions with unforeseen consequences**

David Linnard Wheeler

1. **Summary**

The objective of this essay is to show that defining, detecting, and dealing with outliers requires value-laden decisions. To motivate this topic I will expose the subtle and pernicious consequences that arise from different definitions, detection, and handling procedures. To start, I will show that the different conceptions we use to categorize outliers arouse controversies that are hard to predict. Next I will describe the different sources of outliers and the mirages encountered when ascribing provenance. To underscore the categorical/hypothetical imperatives introduced when confronted with outliers, I will show that the hubris and or expedience that often animates outlier omission is a special case of a more fundamental problem - that reality is how it appears. Further, I will show that systematic outlier omission could snowball and stifle scientific advancements by suppression of anomalies - the things that, as Kuhn argued, often precipitate paradigms shifts. Finally, I will show that both inclusion and exclusion of outliers from data sets introduce biases and conflicts between our fiduciary and moral responsibilities. From these points, I will conclude that confrontation with outliers can introduce ethical issues that are not always obvious and demand critical examination, caution, and actions that may be at odds with near-term analytical duties. It is therefore incumbent upon analysts to (i) be explicit about the value-laden decisions they use to navigate outlier definitions, detection, and handling procedures, and (ii) balance both analytical and moral obligations.

1. **Introduction:**

Outliers arouse a diversity of emotions in people who analyze and interpret data. They are a source of fear, anxiety, suspicion, or even excitement. They can introduce bias/alter conclusions but also signify rare and important events (Anguinis et al. 2013; **CITATIONS**). They are both the objects of interest and despair. Both noise and signal.

Emotional responses to outliers are likely as old as empirical inference itself. In the earliest written mention of outliers, Bernoulli (1777) expressed a sense of frustration with the practice of outlier removal without a priori justification:

“But is it right to hold that the several observations are of the same weight or moment, or equally prone to any and every error? Are errors of some degrees as easy to make as others of as many minutes? Is there everywhere the same probability? Such an assertion would be quite absurd,... I think each and every observation should be admitted whatever its quality.”

Further, Bernoulli (1777) sympathized with the rejection of outliers when observers encounter justifiable reasons before inspecting the actual data:

“I do not condemn in every case the principle of rejecting one or other of the observations, indeed I approve it, whenever in the course of observation an accident occurs which itself raises an immediate scruple in the mind of the observer, before he has considered the event and compared it with the other observations.”

Thus, since the first published record of people wrestling with outliers, context seems to govern their valence. In some contexts, we welcome outliers as harbingers of fraud we want to detect. In other contexts, we disdain them as obstacles to overcome. In all cases, our feelings about outliers are conditional on our objectives. Outliers are contextual. Their definitions, identification, and treatment are all dependent on other observations (can outliers exist when n=1?) as well as the goals and capabilities of the analysts.

Thus, outliers are a contentious and potentially contested topic. The controversy likely dates back at least to Bernoulli in 1777 but shows up extensively in the 1800s. At one end of the spectrum are the “rejectors” who favored the rejection of outliers (Legendre, 1805). At the other end of the spectrum are the “retainers” who favored the retention of outliers (Bessel and Bauer, 1838). In between these two poles is a rich array of nuanced options.

What is the *right* thing to do?

This essay will not answer this question. Instead I will argue that our conceptions, detection strategies, and treatment of outliers require value-laden decisions that have unforeseen consequences. Decisions about outliers are thus often fraught with unforeseen repercussions. My goal is just to expose the nontrivial reverberations that arise from seemingly trivial analytical decisions.

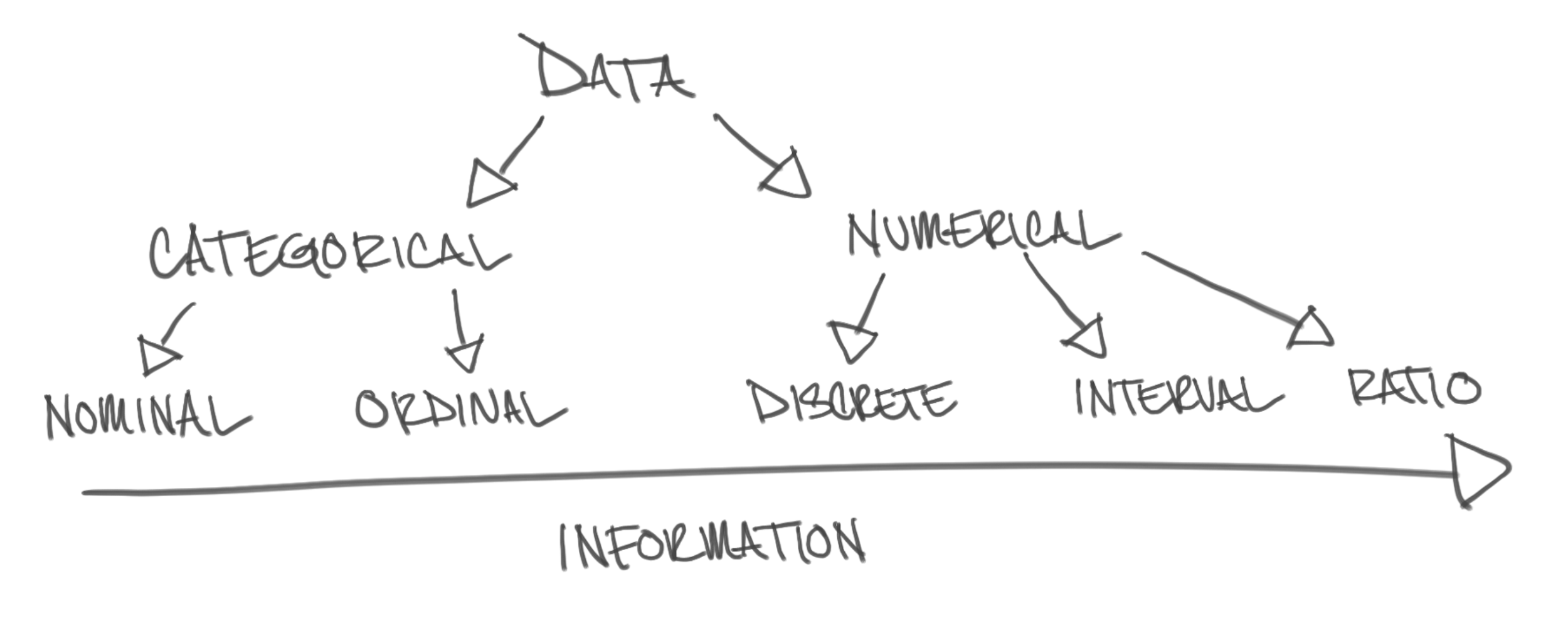
To begin, I will first show that attempts to define outliers require value-laden decisions. Further, these decisions have real world consequences.

1. **What is an outlier?**

Outliers, anomalies, aberrants, contaminants, discordant observations, and stragglers have all been used, sometimes interchangeably, to describe more or less the same thing: observations that are markedly different from other observations.

Over time, many have tried to define outliers. Some of these definitions are ambiguous and therefore difficult to operationalize. For example, Grubbs (1969) defines an outlier as an observation that “appears to deviate markedly from other members of the sample in which it occurs”. Other definitions, for example one from Anscombe and Guttman (1960), are more technical, precise, and easy to operationalize: “[a]n observation with an abnormally large residual…”. Notice, however, that this conception of outliers fails to include categorical data. Still other conceptions, like this one from the Dorland’s Illustrated Medical Dictionary, define outliers with brazen and prescriptive language: “an observation so distant from the central mass of data that it is considered an obvious mistake that should be removed from the data whether or not a cause of deviation can be found” (Anderson et al. 1998).

And these conceptions only really capture univariate numerical data types (**Figure 1**). Categorical outliers and those in multivariate spaces have received less attention. For the latter case, outliers are generally conceived of as “rare data objects” (Pang et al. 2016; Suri and Athithan, 2019) or data objects with frequency occurrences that are “exceptionally typical or un-typical within the distribution of frequencies occurrences of any other attribute value” (Angiulli et al. 2020).



Hence, there is a diversity of conceptions of outliers. To be more precise, a literature review by Anguinis et al 2013 identified 14 mutually exclusive definitions of outliers. Similarly, a small (n=23), non-random, survey of Master of Information and Data Science (MIDS) students at Berkeley revealed conflicting conceptions of outliers (**supplemental document 1**).

The variability present in different conceptions of outliers may just seem like semantics, however; much more is at stake. Categorization of outliers and non-outliers is an essential value-laden judgement. Just as Kraemer et al. (2010) argued that algorithms often require value-laden decisions, one can not demarcate outliers from non-outliers *a priori* without encountering ethical issues.

**III.I Outlier Definitions Require Value-Laden Decisions**

For example, take the case of *Hadlum* v. *Hadlum* (Barnett, 1978). Mr Hadlum left home for military service in August of 1944. His wife, Mrs. Hadlum delivered a baby nearly a year later, 349 days after Mr. Hadlum left home. Upon his return, Mr. Hadlum filed for divorce. He argued that, since the baby was born weeks after the average gestation period of 280 days, Mrs. Hadlum *must* have committed adultery.

Is Mr. Hadlum right? Did Mrs. Hadlum deliver another man’s child? To answer this question we must define the gestation time of 349 days as an outlier or not (and assume no infidelity before Mr. Hadlum left for military service). To do this is to make a value-laden judgement, fraught with marital consequences.

Moreover, to argue for or against Mr. Hadlum’s claim of adultery is to claim that we know how the data *should* be distributed. That is, to claim that the 349 day gestation period is an outlier is to claim that no babies can possibly be born beyond 348 days. Conversely, to claim it does not constitute an outlier is to claim that babies can be born beyond a 348 day gestation period. Each case invokes the problem with induction, as discussed by Hume (1779) and others. Just because most previous human gestations are less than 349 days does not mean that all future gestation periods are less than 349 days. In either case the stakes are high - one could make or break the marriage depending on one's definition of an outlier.

The court ruled in favor of Mrs. Hadlum. The gestation period, they decided, was unlikely but biologically possible. The marriage was preserved.

Cases like *Hadlum v. Hadlum* underscore the primacy of outlier classification. This case shows that implicit but operational definitions about outliers (the courts later defined outlying gestation periods as those longer than 360 days (Barnett, 1978)) require value-laden decisions.

Moreover, this case illustrates how outlier classification systems conjure ethical issues raised by the Belmont Report (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978). For example, if we imagine that Mrs. Hadlum was unfaithful to Mr Hadlum and still found not-guilty, justice, à la The Belmont Report, would not be served. Alternatively, even if Mrs. Hadlum was faithful, it is hard to imagine that their marriage was the same after the failed divorce case. In either case distribution of benefits and burdens seems far from just.

Furthermore, the Hadlum case is not the only example of a case where our definitions of outliers matter and trespass ethical norms. Examples of ethical issues that follow from attempts to define numerical and categorical outliers are abound.

Consider the case, for example, of the biased facial analysis softwares discovered by Joy Buolamwini (Hardesty, 2018). One putative source of the bias in these softwares is the dearth or absence of underrepresented minorities in the training sets (Hardesty, 2018). How does this relate to definitions of outliers? If truly a source of bias, then the failure to include certain demographics, for example black women, is tantamount to defining these demographics as outliers, even if implicitly. Again, if the absence of minority demographics truly constitutes a source of bias, then the programmers who designed the algorithms implicitly drew a line in sand. All people beyond the line *are* outliers.

The choice to include or exclude certain demographics from training sets for algorithms again constitutes a value-laden decision. The consequences, like the case of *Hadlum v. Hadlum*, include infringements on both beneficence and justice, as conceived by the Belmont Report.

The above cases illustrate that outlier definitions demand value-laden decisions. These decisions, in turn, can yield unforeseen consequences that violate ethical principles. This problem requires a solution.

One solution could be to define outliers in theoretical terms that are universal across data types and operational terms within data types (**Figure 1**). This latter task, however, is not trivial. Given the diversity of data types and definitions of outliers in each, this task is beyond the scope of this project. For, to operationalize outlier definitions for each type of data is to recast theoretical definitions as mathematical expressions that are deployable by the masses. Thus, for now I will just submit a candidate theoretical definition from (Barnett and Lewis, 1994): “we shall define an outlier in a set of data to be an observation (or a series of observations) which appears to be inconsistent with the remainder of that set of data”.

**III.II Application of Bowker and Stars Framework for Classification Systems**

To test the performance of this definition in classifying outliers, I will now apply Bowker and Star’s (1999) framework for classification systems. If the definition above furnishes comparability, visibility, and control across data types and domains then perhaps it will serve as a valuable classification system.

For a classification system to be comparable it must enable and facilitate communication, by regularization of semantics, across entities (Bowker and Star, 1999). Thus the candidate definition must serve all data types and domains. From a bird’s eye view, this definition seems to serve all data types. That is, it is easy to imagine instances of both categorical and numerical data that are inconsistent with the rest of the data. For example, imagine a dataset of people with eye color and height. From these data, categorical (nominal) outliers might constitute subjects with abnormal eye colors (e.g. black). Similarly, numerical (ratio) outliers might constitute subjects with abnormal or impossible heights (e.g. -2 feet or 11 feet).

So far so good, right? Maybe not. The fidelity of this candidate definition to outliers from all data types is likely a mirage. Upon closer inspection we can see that this definition fails to capture all categorical outliers, for example. Although it catches local cases of unexpected categories, like black eye color, it does not capture cases where the frequency occurrences of expected categories are anomalous. For instance, if we collected eye color data from an island populated with almost entirely brown eyed people but a fraction of blue eyed people, this definition would not catch the blue eyed subjects even though the frequency occurrences of these subjects might be anomalous. They would be “invisible” or residual. This brings us to the next criterion: visibility.~~Thus this candidate definition is not entirely comparable across all data types - it is biased towards numerical data.~~

For categories within a classification system to be visible, they must be classifiable (Bowker and Star, (1999). To be invisible is to be unclassifiable or residual. Thus our candidate definition provides visibility for those outliers that are discernible within its scope. Some outliers, like those that differ not necessarily in the distance from other observations but in their frequency occurrences, will slip under the radar. They will be “invisible”.

Finally, we have control. Objects of a classification system, like outliers, are subject to more or less control by those who analyze data. The candidate theoretical definition presented above does not exercise much control over instances of outliers in each data type or domain. That is, the classification system itself is not constrained by this definition. The onus is on the analyst to decide and justify what exactly it means for observations to be inconsistent. This is both good and bad. The ambiguity of the definition endows each analyst with the license to exercise their own due diligence. The cost, however, is of comparibility and visibility. It is possible, for example, for some analysts to implement a localized incarnation of the candidate definition that is not comparable across data types or domains and or invisible to consumers of their work.

In summary of this section about outlier definitions, I hope the following is clear. A diversity of definitions of outliers are present in the literature and among data scientists within the MIDS program at UC Berkeley. To define outliers is to erect classification systems that demarcate outliers from non-outliers. This act requires value-laden judgements that are often implicit. These value-laden decisions have ethical consequences. The severity of the ethical consequences can be qualified to some extent by the Belmont Report. For example, the cases of *Hadlum v Hadlum* and facial analysis softwares show us that, from our conceptions of outliers and the requisite value-laden judgments that animate them, arise consequences that are not obvious when the definitions are first conceived. These consequences can violate notions of justice and beneficence, as conceived by the Belmont Report (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978). A theoretical candidate definition, formalized by Barnett and Lewis (1994), was presented. This definition appears to enable comparability across most data types. However, this comparability property, as predicted by Bowker and Star (1999), comes at a cost of control.

Before I show that detection of outliers also requires value-laden decisions, I will briefly describe the origin of outliers.

1. **Where do outliers come from?**

Outliers arise from several sources. Most authors who write about such topics (e.g. Anscombe 1960; Barnett, 1978; Grubbs 1969; Smiti, 2020) agree that outliers can be attributed to:

1. Inherent variability
2. Execution error
3. Measurement error

Inherent variability is just the variability “that would be observed in the population even if all measurements were perfectly accurate” (Anscombe 1960). This source of variability cannot be modified without intervention of the underlying population from which the samples were collected. An example of inherent variability includes Mrs. Hadlum’s child that was born after a 349 day gestation period. In many cases, the parameters about the underlying population are fixed and unknown without exhaustive censuses.

Execution errors are introduced when there is a “discrepancy between what we intend to do and what is actually done, other than error in the use of measuring instrument” (Anscombe 1960). Some authors further parse execution errors into sub-categories containing (i) sampling error and (ii) mis-reporting errors (Smiti, 2020). For the sake of precedence, parsimony, and Bowker and Star’s (1999) compatibility criterion, I clump these sub-categories into the larger execution errors category. Execution errors are absent in a world free from mistakes. An execution error would occur if, in an attempt to determine the gestation period for Mrs. Hadlum child, they unintentionally asked another person - not Mrs. Hadlum. To determine if an outlier arose from an execution error, we would likely need an exhaustive and searchable record of the procedures that proceeded data collection. For most researchers, this information is likely not available.

Errors that arise from “measuring instruments” are measurement errors (Anscombe 1960). If the things with which we collected data were precise and accurate, then measurement errors would not exist. For example, if Mrs. Hadlum’s gestation period was measured by a fallible tool that did not reflect her true gestation period, a measurement error would have occurred. Measurement errors are likely present in most data collection environments. They become outliers recognizable by the candidate definition above and detectable by the methods discussed below when they breach above baseline errors associated with other observations. Like execution errors, it can be exceedingly difficult to attribute an outlier to measurement error unless scrupulous records are preserved, searched, and cross-referenced.

The provenances of outliers as described above does not appear to be controversial. Most authors who write about the subject agree on the mechanisms by which outliers arise. The *attribution* of provenance, however, can be controversial.

How do we know the true source of the outliers? The question becomes harder to answer with less and less information. That is the accuracy and indeed our confidence about the answer to this question *should* be inversely proportional to the availability of information with which we can attribute provenance. In practice, answering this question can be next to impossible. One can compile records to support hypotheses about provenance. However, such evidence is often a limiting resource. Moreover the absence of such evidence is not the evidence of absence of provenance (i.e. *argumentum ad ignorantiam*). And this is just the start of it. ~~Under these surface level issues is a bigger and broader issue - that appearances equal reality.~~

To attribute provenance to outliers is again to confront value-laden decisions. Most examples of such value-laden decisions seem to involve accusations directed towards individuals on the basis of the assumption that the outlier(s) in question have arisen from inherent variability. Below are a couple familiar examples.

For example, take the familiar case of *Hadlum v. Hadlum* (Barnett 1978)*.* The source of controversy here was not just the definition of an outlier but also the provenance of the outlier. In accusing Mrs. Hadlum of infidelity Mr. Hadlum assumed the outlying gestation period was due to inherent variability. This is a value-laden decision because, in making this assumption, he confronted an ethical problem: infidelity.

Other examples, where value-laden decisions are embedded in the attribution of provenance to outliers, are available in other domains. Another example is from Costanza-Chock’s (2018) account of their experience with TSA agents after being flagged as anomalous by the millimeter wave scanner. Constanza-Chock is a “nonbinary, transgender, femme presenting person” Costanza-Chock’s (2018). While in TSA, their groin was flagged by the millimeter wave scanner as anomalous. Since millimeter wave scanners are designed to detect concealed objects the anomalous label warrants completion of further security protocols. Hence, they were then subjected to further searches by a TSA agent.

This encounter required layers of value-laden decisions, some of which will be discussed further below. For now, I want to emphasize the value-laden decision entailed by the accusation of concealment. The accusation assumed that the object(s) that triggered the anomalous label were due to inherent variability. In this case, the assumption is likely valid. Nonetheless, the attribution of provenance to the outlier(s) here involved value-laden decisions because one could not make such a decision without introducing values.

Lastly, these value-laden decisions illicit real ethical implications. In the first case of *Hadlum v Hadlum*, it is hard to see how justice, as conceived by the Belmont Report (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978), is satisfied by assuming that the outlying gestation period is due to inherent variability. To be just the burdens, and accusations that arise therefrom, should be uniformly distributed across all candidate sources of outliers, not just inherent variability. To use Bayesian language, our prior for the putative provenance of the outlier in question should be uniform until and unless evidence to the contrary is discovered. Deviations from the uniform prior are not just because, in this case, they result in accusations of infidelity that might also be explained by other sources.

Likewise, for the TSA agents to be just to Constanza-Chock (2018) they should have assigned tentative blame to each potential source of outliers before accusing Constanza-Chock of something that could have been explained by other phenomena. To explore all explanations for an anomalous label is different from first accusing someone of something without due diligence.

The cases above appear to represent a convergence of Bayesian, deontological, and consequentialists arguments. If we accept the claim that both Mrs. Hadlum and Constanza-Chock were harmed by injusttice then each of these frameworks arrive at the same destination. This convergence of disparate philosophies to the same conclusion might create the illusion that they are all equally *right.* Since we can not predict outcomes *a priori*, however, it follows that the conservative framework - the one that minimizes harms by exploration of all possible sources of outliers and the value-laden decisions they entail - is the Bayesian-deontological hybrid.

Up until now I have shown that value-laden decisions are implicit in both defining and attributing provenance to outliers. I will next argue that value-laden decisions seep into detection of outliers.

1. **How do we detect outliers?**

* Operationalize definition in uni- and multivariate space

TSA issues: Design Justice, A.I., and Escape from the Matrix of Domination (Costanza-Chock)

1. **How do we deal with outliers?**
   1. Premise 3
      1. Plato’s cave
   2. Premise 4
      1. Taha
   3. Premise 5
   4. Premise 6

**~~Premise 2: Outliers arise via from three major sources~~** ~~(Anscombe 1960, Grubs 1969, Barnett; 1979): measurement errors, execution errors, and intrinsic variability within the population of interest.~~

* ~~In this section, I will show that outliers that arise via measurement and execution error can be difficult to differentiate between those that arise from intrinsic variability. I will use simulations and historical examples (e.g. the space shuttle Challenger disaster~~**~~,~~***~~Hadlum v. Hadlum~~***~~,~~** ~~and medical cases (Papadimos and Marco, 2004)) to show that outliers arise from intrinsic variability within widely invoked distributions (e.g. Gaussian), especially when the sample sizes are small. I will note that the failure to find evidence that outliers arise from measurement or execution error does not equate to evidence that the outliers in question did not arise from either type of error (i.e~~ *~~argumentum ad ignorantiam~~*~~). Implications for observations/subjects that arise from each source will be briefly discussed.~~

**~~Literature:~~**

* ~~Data fundamentalism:~~

**Premise 3: We mostly do not know or see the processes by which the data we collect are generated.**

* Like Plato’s prisoners in the cave allegory, we see only *shadows* or samples when we collect data. We don’t know or see the objects that cast the shadows - the population from which the data arise. This is often the motivation to collect data. The simulations described under premise 2 may or may not be revisited or extended here to show that, even under ideal conditions, outliers arise. Examples where estimates from samples (aka statistics) differ from those of the population (parameters) will be briefly discussed. Lastly, I should reconsider the placement of this premise- perhaps it is better suited as the first premise presented?

**Literature:**

* Positivism (Jurgenson)
* Bernoulli: “ I see no way of drawing a dividing line between those that are to be utterly rejected and those that are to be wholly retained; it may even happen that the rejected observation is the one that would have supplied the best correction to the others.”
* Plato’s cave and Bertrand Russel’s appearance vs reality
* Hume’s PUN

**Premise 4: Some outliers contain signals and are of primary interest to researchers and the civilians they serve.**

* The premise functions to further support the conclusion by presenting cases where outliers are of primary interest. Cases where outliers are used to detect fraudulent voting activities, malfunctions in manufacturing, medical issues, and cyber security will be presented.
* Cases where outlying observations lead to Kuhnian paradigm shifts will be presented (<https://www.uky.edu/~eushe2/Pajares/Kuhn.html>).

**Literature:**

* Outliers are valued more than other observations: Valuating Privacy (Huberman et al. 2005).
* Kuhn

**~~Premise 5: Outliers are removed by a~~ *~~non-trivial~~* ~~proportion of researchers.~~**

* ~~This is probably the most controversial premise. This premise will be supported by a literature review. A representative survey of practitioners would be ideal but, in lieu of the capital required to issue a survey, a literature review appears to be the next best solution.~~
* ~~Data fundamentalism/crawford: data are objectives, aberrations are not, therefore remove aberrations.~~

**~~Literature:~~**

* ~~Bernoulli: “ I see no way of drawing a dividing line between those that are to be utterly rejected and those that are to be wholly retained; it may even happen that the rejected observation is the one that would have supplied the best correction to the others.”~~
* ~~consequences of this classification scheme will be discussed through the prism of the literature from week 4, Plato, Hume, and Bertrand Russel~~

**Premise 6: Outlier inclusion and exclusion both introduce substantial biases in data analysis and interpretation.**

* This will be the crux of the argument. Historical cases will be presented to show that biases can be introduced when outliers are included and excluded. The effect of selection bias on this premise (i.e. we don’t really hear about algorithms that perform well when outliers were removed) will be discussed and used to temper the conclusions.

**Literature:**

* Data is biased:
  + Hidden Biases in Big Data (Crawford)
* Exclusion:
  + Big Data and Its Exclusions (Lerman)
  + Unfair or deceptive acts (FTC section 5)
  + Challenger
  + Farman et al. 1985, (<https://undsci.berkeley.edu/article/0_0_0/ozone_depletion_09>)
* Inclusion:
  + privacy violations (Jared)
  + Taxonomy of privacy (Solove)
* Inclusion and exclusion:
  + conflicts with FIPPS (Fair Information Practice Principles (FIPPs) in the Information Sharing Environment)
  + Conflicts with the OECD privacy framework
  + Repercussion for algorithms: The Relevance of Algorithms. (Tarleton Gillespie)
* Semantics: Privacy is an essentially contested concept: a multi-dimensional analytic for mapping privacy (Mulligan et al. 2016)
* “In the final analysis it would seem that the rejection or the retention of a discordant observation reduces to a question of common sense. Certainly the judgement of an experienced observer should be allowed considerable influence in reading a decision.” Rider (1933)
* “I see no way of drawing a dividing line between those [observations] that are to be utterly rejected and those that are to be wholly retained” (Bernoulli, 1777)

**Premise 7: Remedial measures are available to secure unbiased parameter estimates when outliers are included in a data set.**

* A brief summary of the available remedial measures, and use-cases, will be presented to show analytical obligations can still be satisfied if outliers are included in data sets.

**Literature:**

**Recommendations:**

* Be explicit

**Conclusion:** Outliers arise in data from three major sources. We can not always determine the source of the outliers. Some outliers are of primary interest to researchers, even if only retrospectively. Some researchers remove outliers without justification. Remedial techniques are available for satisfying analytical duties while including outliers. Consequences for outlier inclusion and exclusion are mostly well-documented. The consequences of outlier exclusion in *some* cases eclipse in magnitude the consequences of inclusion. Therefore, unless defensible occasions arise or the consequences of omission are predictable, outliers should, at the very least, be critically evaluated before being excluded. In short, the moral and analytical obligations of analysts are paramount to expedient data analyses.

* Conflicts with belmont report
* Classification of outliers is a value-laden decision: Is there an ethics of algorithms? (Felicitas Kraemer, Kees van Overveld, and Martin Peterson)
* Outlier removal is a hypothetical or categorical imperative and or imperative of right( Is there an ethics of algorithms? (Felicitas Kraemer, Kees van Overveld, and Martin Peterson)/ Kant

**Discussion:** Outlier inclusion and exclusion creates challenges and opportunities. Decisions about inclusion and exclusion test our fidelity to the data and the data’s fidelity to reality - *the* objects of which the prisoners in Plato’s allegory of the cave see only shadows. These decisions can create conflicts between competing obligations to satisfy duties as analysts and moral citizens. Such conflicts will be explored and solutions will be proposed from several perspectives (e.g. utilitarianism, Aristotelian, Kantian, and deontological, consequentialist). The risks associated with outlier exclusion in certain cases discussed in premise 6 surpass in importance those encountered when outliers are included. Given the costs associated with inclusion and exclusion, data analysts are obliged to, at the very least, think critically about the costs of outlier removal before clicking buttons.

that outliers create challenges and opportunities. Decisions about outliers test our fidelity to the data, the data’s fidelity to reality, and our assumptions about the processes that generate data. Much like Plato’s allegory of the cave, we see only the shadows of the objects that cast them. Our duty is to determine what to do with the shadows that don’t conform to our expectations.

These decisions can create conflicts between competing obligations to satisfy duties as analysts and moral citizens. Such conflicts will be explored and solutions will be proposed from several perspectives (e.g. utilitarianism, Aristotelian, Kantian, and deontological). The risks associated with outlier exclusion in certain cases surpass in importance those encountered when outliers are included. Given the costs associated with inclusion and exclusion, data analysts are obliged to, at the very least, think critically about the costs of outlier removal before clicking buttons.

**References**

1. Anderson DM, Keith J, Novak PD, Elliot MA. 1998. Dorland’s Illustrated Medical Dictionary. Saunders, Philadelphia, PA.
2. Angiulli F, Fassetti F, Palopoli L, and Serrao C. 2020. Detecting and Explaining Exceptional Values in Categorical Data. SEBD. http://ceur-ws.org/Vol-2646/29-paper.pdf
3. Anguinis H, Gottfredson RK, Joo H. 2013. Best-practice recommendations for defining, identifying, and handling outliers. 16:270-301.
4. Anscombe FJ, and Guttman I. 1960. Rejection of outliers. Technometrics. 2: 123-147
5. Barnett V. 1978. The study of outliers: purpose and model. Journey of the Royal Statistical Society. 3:242-250.
6. Barnett V and Lewis T. 1994. Outliers in statistical data. John Wiley & Sons. 3rd Edition
7. Bernoulli D. 1777. The most probably choice between several discrepant observations and the formation therefore of the most likely induction. In C.G. Allen (1961), *Biometrika*, 48:3-13.
8. Bessel FW and Baeuer JJ. 1838. Gradmessung in Ostpreussen und ihre Verindung mit Presussischen und Russischen Dreiecksketten. Berlin. Reprinted in Adhendlungen von FW Bessel
9. Bowker GC and Star SL. 1999.. What a Difference a Name Makes - The Classification of Nursing Work. Chapter 7 of *Sorting Things Out: Classification and its Consequences*. MIT Press. <https://github.com/UC-Berkeley-I-School/w231/blob/master/Readings/Bowker%20and%20Star.%20Sorting%20things%20Out%20ch7.pdf>
10. Constanza-Chock S. 2018. Design Justice, A.I., and the Escape from the Matrix of Domination. JoDs. <https://doi.org/10.21428/96c8d426>
11. Grubbs FE. 1969. Procedures for Detecting Outlying Observations in Samples. Technometrics. 11: 1-21.
12. Hume D. 1779. An enquiry concerning human understanding. In D. Hume, Essays and treatises on several subjects, Vol. 2. Containing An enquiry concerning human understanding, A dissertation on the passions, An enquiry concerning the principles of morals, and The natural history of religion (p. 3–212). Unknown Publisher.<https://doi.org/10.1037/11713-001>
13. Kraemer F, van Overveld K, Peterson M. 2011. Is there an ethics of algorithms? Ethics Inf Technol. 13: 251-260.
14. Legendre AM. 1805. Nouvelles Méthodes pour la Determination des Orbits des Cométes. Courcier, Paris.
15. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1978. *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. [Bethesda, Md.]: The Commission.
16. Hardesty L. 2018. Study finds gender and skin-type bias in commercial artificial-intelligence systems. MIT News. <https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212>
17. Pang G, Cao L, and Chen L. 2016. Outlier Detection in Complex Categorical Data by Modelling the Feature Value Couplings. Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. 1902-1908.
18. Rider PR. 1933. Criteria for rejection of observations. Washington University Studies - New Series, Science and Technology - No. 8, St. Louis.
19. Ranga Suri N.N.R., Murty M N., Athithan G. 2019. Outlier Detection: Techniques and Applications. Intelligent Systems Reference Library, vol 155. Springer, Cham. <https://doi.org/10.1007/978-3-030-05127-3_5>
20. Smiti A. 2020. A critical overview of outlier detection methods. Computer Science Review. <https://doi.org/10.1016/j.cosrev.2020.100306>