Ambiguous Yet Polarizing: An Ideographic Approach to Analyzing the Variation in Usages and Conflicts of Values That Have Arisen From the Term "Algorithmic Fairness"

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Nuzaba Nuzhat

Spring 2020

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Kathryn A. Neeley, Associate Professor of STS, Department of Engineering and Society

Introduction

In the current age of rapid automation fueled by software advancements, machines have acquired a skill long deemed fit only for human intelligence – high risk decision-making. Powered by big data, machine learning algorithms are increasingly augmenting or replacing human decision makers in fields as varied as criminal justice, employee screening, college admission, loan approvals (Castelluccia & Le Métayer, 2019). The result has been an unprecedented increase in speed, consistency, and cost-saving. However, at the same time, ethical concerns regarding the fairness, accountability, and transparency of such systems have come into light (Rainie et al., 2017). The focus of this paper is on the first of these concerns, a phenomenon commonly alluded to in the literature as "algorithmic fairness".

So, what is algorithmic fairness? Unfortunately, both data researchers and legal experts have struggled to come up with a definition that can satisfy the needs of most stakeholders. The problem, as many have pointed out, is that the different definitions prioritize different values, such as, equality or social justice or individual fairness or greater good, and more often than not, the conflict that ensues presents no straightforward compromise (Friedler et al., 2016). According to some scholars, to resolve this current state of disagreement, what is necessary is not a newer, better definition, but rather a reorientation of focus. In the words of Selbst and his fellow researchers (2019):

> Many papers have been written proposing definitions of fairness…...Almost all of these papers bound the system of interest narrowly. They consider the machine learning model, the inputs, and the outputs, and abstract away any context that surrounds this system. We contend that by abstracting away the social context in which these systems will be

> deployed, fair-ML researchers miss the broader context, including information necessary to create fairer outcomes, or even to understand fairness as a concept. (p. 59)

The goal of this paper is to shed light on this broader social context by examining how various users refer to the term “algorithmic fairness” in private and public discourse. To aid this process, the paper employs the framework of ideographs. Originally conceived by McGee (1980), ideographs are rhetorical tools that help shape public consciousness regarding topics deemed indispensable to societal values, without providing clear cut definitions of what those values are. This lends them a flexibility that can be utilized to defend one course of action over another, even those that might seem highly questionable. I argue that by treating “algorithmic fairness” as an emerging ideograph, we can better understand and appreciate the values that the term engenders, as well as those that are in conflict with it. This enhanced understanding should help us not only to make better informed decisions about what we mean by the term when we allude to it, but also make clear to us why we strive for “algorithmic fairness” in the first place.

Part I: Literature Review

The Current State of Algorithmic Bias and Discrimination

The European Parliament’s Panel for the Future of Science and Technology defines algorithms engaged in decision-making as Algorithmic Decision Systems (ADS): systems that depend on “analysis of large amounts of personal data to infer correlations or, more generally, to derive information deemed useful to make decisions” (p. 3). To this date, they have been used to “improve climate forecasts, detect diseases, or discover new viruses”, “to help make predictions, recommendations or decisions in various areas such as information, finance, planning, logistics etc.” and “to provide autonomy to physical objects by limiting human supervision” (p 4-5).

However, since these algorithms are mimicking imperfect human decisions, based on datasets that are often inaccurate, biased, and incomplete, the results they produce often disadvantage one group of people over another, and that disadvantaged group is often a demographic minority. The discrimination that is caused by such imperfect algorithms can be illegal or ethically questionable and cause harm that can limit opportunities for economic and social growth (see Figure 1 for examples).

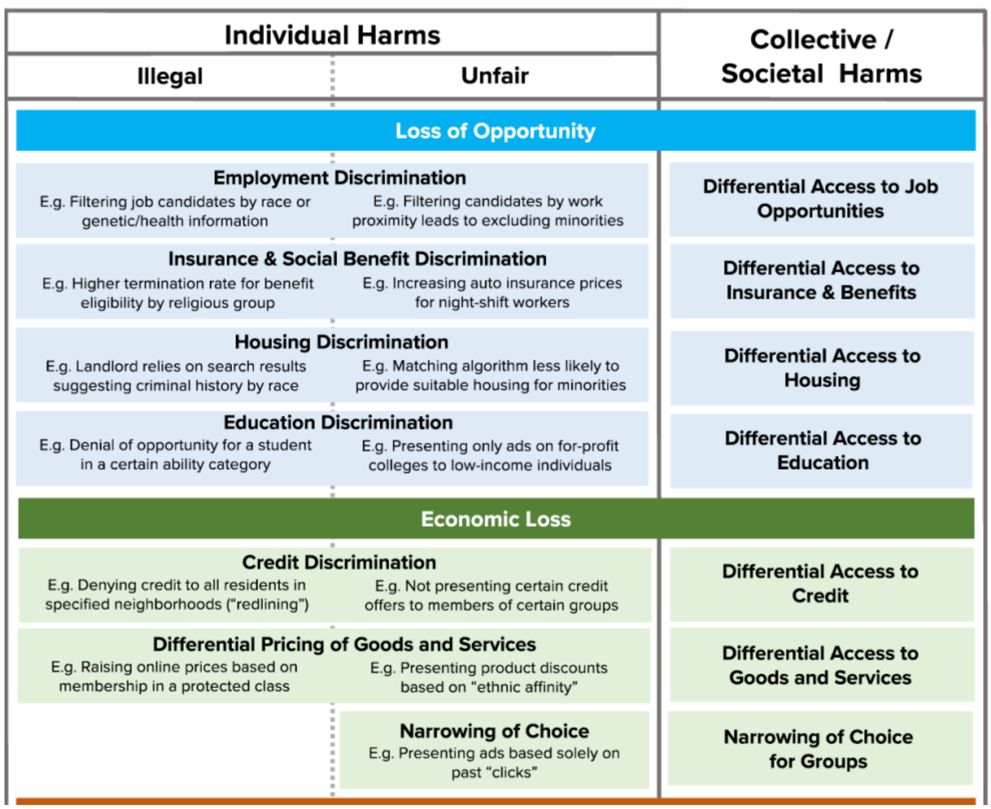

Figure 1: Some examples of the potential harm caused by biased automated decision making (Castelluccia & Le Métayer, 2019, p. 19)

These are not just hypothetical situations. In 2016, ProPublica audited a machine learning software named COMPAS, that assigned recidivism scores to current offenders, and found that "Black defendants were still 77 percent more likely to be pegged as at higher risk of committing a future violent crime and 45 percent more likely to be predicted to commit a future crime of any kind" (Larson et al., 2016). In 2018, Amazon scrapped a data tool that it thought would be the "holy grail of hiring", as it was shown to be biased against women (Dastin, 2018). In 2019, Goldman Sachs came under fire for assigning a woman a credit limit 20 times lower than her husband, who had a lower credit score than her (Natarajan & Nasiripour, 2019). It is cases like this and others that have led to widespread call for ensuring algorithms that are not just efficient in their decision making, but also adequately fair.

Algorithmic Fairness: The Solution Without a Definition

In recent years, discourse surrounding algorithmic fairness has been on a gradual rise. Public leaders from all levels have called attention to this (*"Guidance for Regulation of Artificial Intelligence Applications",* 2019; Algorithmic Accountability Act, 2019). The machine learning community has had entire conferences where this is a focal point (AAAI/ACM Conference on AI, Ethics, and Society, FAT* '20: Conference on Fairness, Accountability, and Transparency), and almost every company that deals in this space now has a page on their website that talks about how "fair" their algorithm is. (*Bias, AI Ethics, and the HireVue Approach*, n.d.; *ZAML Fair | Don't Sacrifice Accuracy for Fairness*, n.d.)

And although there is consensus regarding the need for algorithmic fairness, there unfortunately is not much consensus regarding its definition. Scholars in machine learning and law have taken up the challenge and come up with dozens of convincing definitions. When Verma & Rubin (2018) tried to consolidate them, they found "more than twenty different notions of fairness proposed in the last few years", but "no clear agreement on which definition to apply in each situation" (p. 1). And although there may be multiple definitions that may be appropriate for any given situation, often the goals underlying those definitions are "fundamentally incompatible with one another" (Friedler et al., 2016). For example, one definition of fairness advocates for “fairness of process” (procedural fairness), which could entail not feeding the algorithm protected attributes, such as gender or race. However, for college admission scenarios, if protected attributes are not accounted for, black students may constitute a disproportionately low fraction of the student body, since on average their SAT score is lower. In that case, if “fairness of outcome” (distributional fairness) is to be achieved, the algorithms need to be fed protected attributes to achieve demographic equality, which would require, at least to some extent, “fairness of process” to be sacrificed.

Not only that, these definitions are also very often in conflict with other goals an individual or group values. For example, Corbett-Davies et al. (2017) in their research established that when it came to the COMPAS algorithm cited earlier, “satisfying common definitions of fairness means one must in theory sacrifice some degree of public safety” (p. 802). Figure 2, based on research by Hardt, Price & Srebro (2016), illustrates these conflicts fairly well. The x-axis is the percentage of non-offending borrowers who received a loan because their FICO score crossed a certain threshold, and the y-axis presents the various goals that were used to arrive at the results. As we can see not only do the various fairness goals lead to differing outcomes between the various racial groups, the outcomes also differ the most when profit maximization is the primary goal, instead of some notion of fairness.

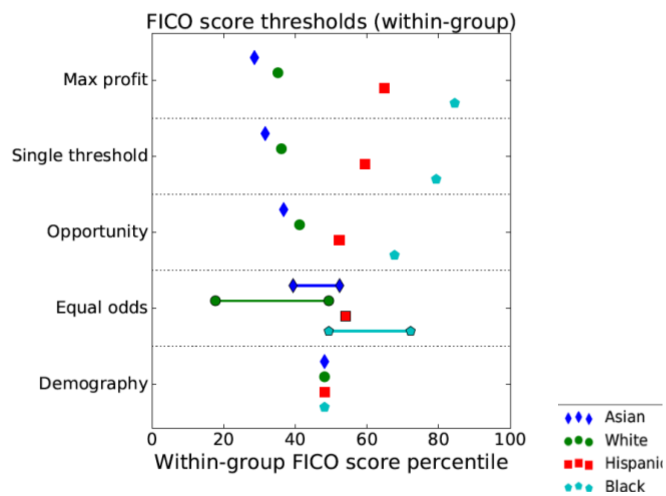

Figure 2: A group by group breakdown of the percentile of non-defaulters who will cross the required FICO score threshold for loan approval (x-axis), which is set by the various criterion (y-axis). The 'Single threshold' criterion uses a race-blind algorithm. The 'Opportunity' criterion ensures "the fraction of non-defaulting group member who are approved is the same across groups". The 'Equal odds' criterion ensures "fraction of non-defaulters that qualify for loans and the fraction of defaulters that qualify for loans to be constant across groups". The 'Demography' criterion ensures that "across each group the same percentage of applicants get approved". As seen, when the goal is profit maximization, only black borrowers at the 80th percentile of their racial group are approved, compared to the 50th percentile in the demographic parity case. (Hardt et al., 2016, p. 19)

The "impossibility theorem" and the "inherent trade-off" of algorithmic fairness

Wong in his 2018 article draws attention to these two polylemmas; the first he terms as "impossibility theorem" (since it is impossible to satisfy more than two of the many definitions of algorithmic fairness at once (Narayanan, 2018)) and the second as the "inherent trade-off between fairness and performance". According to him, this two-dimensional disagreement stems from "the fact of reasonable pluralism" inherent in all demographic societies, and thus can never be resolved, but only reduced. However, doing so effectively requires a better understanding of all relevant stakeholder views on the matter. The current literature does a great job informing us what the experts takes on these perplexing issues are, but when it comes to non-experts' views on the matter, it has little to offer. This paper contributes to this gap in literature, through deploying the concept of ideographs: a rhetorical tool often used for political discourse analysis, which is ideal for such two-dimensional conflicts of meaning and value, especially when the actors involved are not as meticulous in their framing of the issue.

Part II: Methods

The Characteristics of an Ideograph That Lend it its Force

In the world of linguistics, ideographs are graphical symbols, occurring in ancient languages, that convey abstract ideas without indicating the sounds required to say them, much like photographs communicate ideas of physical objects without spelling them out. Michael McGee, a rhetorical theorist, through his 1980 article "The 'Ideograph': A Link Between Rhetoric and Ideology" used this concept of more-than-meets-the-eye to give this word a new meaning, which continues to prevail today in political, historical, philosophical analysis.

According to McGee, "Human beings are 'conditioned' not directly to belief and behavior, but to a vocabulary of concepts that act as guides, warrants, reasons, or excuses for behavior and belief" (p 5). These concepts are what he calls ideographs and their examples include terms such as "law", "liberty", "tyranny", or "trial by jury". They represent "collective commitment to a particular but equivocal and ill-defined normative goal" (p. 5). They possess a certain "intrinsic force" that enables them to elicit strong responses within individuals, even without specification of context. A firm belief in their value is a precondition to membership within a certain group, and a lack of such belief is often “penalized” to prevent alternate ideographs from taking hold (p. 8-9).

It is impossible to pinpoint a “pure” meaning of such ideographs that “is unpolluted by historical, ideographic usages” (p. 9), and it is through this variation of usages, instead of their “alleged idea-content”, through which McGee believes ideographs earn their significance (p. 10). Thus, to understand how ideographs control and shape public consciousness, it is imperative to examine how their usages evolve and conflict with one another, a process which McGee terms as an ideograph’s “vertical structuring”. In this process, usages of ideographs accumulate with the passage of time, and earlier usages act as precedents for latter ones. As an example, he analyzes the changing meaning of the ideograph “equality”. During the segregation era, equality of education was thought of as just the “equal opportunity” to being educated, whether it be in a black-only school or a white-only school. But during the civil rights era, a new meaning of the ideograph emerged - equal “access to education”, meaning quality of education had to be equal across racial demographics.

And although differing usages of the same ideograph over time can lead to conflict, (as was the case with the two meanings of equality in *Brown vs. Board of Education*) and thus create illusions of power, vertical structuring is not how ideographs earn their rhetoric force. For that, we have to turn to what McGee calls horizontal structuring of ideographs. As McGee describes, no ideograph operates in isolation, instead they exist in an interconnected mesh with other ideographs; when no conflict exists within such a group of ideographs, the usages of the individual ideographs are relatively stable, and even reinforce one another. However, often an event disrupts this “consonance”, and ideographs that were “synchronically” positioned next to one another, are suddenly fighting for greater value in the collective consciousness. For example, when it comes to proper governance, “equality” and “liberty” together ensure that citizens are treated fairly by their government. Yet, when they are used in “ideological arguments”, the relationship changes, as suddenly one meaning of the ideograph “equality” – “equal access to education” - is more in consonance with the ideograph “liberty” than is the other – equal rights to “being educated”. Conflicts arise as groups form around their allegiance to one ideograph over another, which in turn reconfigure an ideograph’s “standing” among its neighboring ideographs (p. 13) and alter the power it wields over public consciousness.

Potential of the Ideographic Lens in Sociotechnical Analysis: Examples from STS

Although ideographs are best known for their usage in political analysis, they have also been frequently employed to analyze the discourse that surrounds promising new technology. Liao (2012) in one such paper, where he categorizes the promises made regarding Augmented Reality (AR), uses ideographs to understand “how people are speaking for the technology, (de)-legitimizing the technology, making the technology (un)-necessary, and building coalitions to make their visions a reality” (p. 2). For example, when it came to AR’s ability to display data on top of a user’s visual fields, the proponents deployed the ideographs “obtaining knowledge”, “education”, “time saving” to tout its utility, while opponents made references to ideographs such as “misinformation”, “pornography” to quell some of the avidity (p. 12).

Another example of such ideographic analysis, where a budding technology is under the limelight, is Bos, Walhout, Peine, and Lente's 2014 paper: “Steering with big words: articulating ideographs in research programs”. Here the authors observe the usage of three positive ideographs: “sustainability”, “valorization”, and “responsible research initiative (RRI)”, as they relate to the Dutch nanotechnology research and innovation consortium NanoNextNL. By comparing and contrasting the way these ideographs are deployed across policy, management, and researcher level, the authors were able to connect “the logical and hierarchical structuring of terms” with “the sociological structuring of research practices” (p. 152).

For instance, when the authors tried to understand what “sustainability” means in regards to nanotechnology research, i.e. how the term is vertically structured, they found actors in the policy and management level using it in relation to water and energy projects. Researchers working on solar panel projects, also found it easy to reference sustainability in their work. However, other researchers, working on the Nano sensors project, had less agreement regarding whether they could bend the term sustainability to connect it to their research. As for the horizontal structuring of the same term, Bos et al. found that higher in the hierarchy “sustainability” was easily referenced next to ideographs related to other societal goods, such as “public health”, “environment”, and “safety”. However, lower in the hierarchy, researchers sometimes found “sustainability” less important than some of these ideographs. For example, a few researchers thought ensuring industry funding to be more important than finding societal relevance for their work, and thus for them an ideographical clash existed between “valorization” and “sustainability”.

	Vertical structuring	Horizontal structuring
Policy level	Energy, water, environment	No clashes, but structuring alongside other societal goals
Management	Energy, water, environment	No clashes, but structuring alongside other societal goals
Research project	In solar projects a fixed funnel of articulation is present, which is consequently referred to. In sensor projects more diverse referrals	Structuring versus valorization and health care, and it is considered a strategic move

Figure 3: Horizontal and vertical structuring of the ideograph "sustainability" among nanotechnology researcher. As a reminder, vertical structuring refers different usages of an ideograph that compile over time, whereas horizontal structuring refers to the concurrent co-existence of multiple ideographs that either coalesce or conflict with one another. (The "fixed funnel" terminology is not relevant for the purposes of this paper) (Bos et al., 2014)

Through this discussion, three reasons for using ideographic analysis to understand the "impossibility theorem" and "inherent tradeoff" of "algorithmic fairness" across multiple stakeholder groups emerge. First, from McGee's definition of ideographs, it is easy to see how the "impossibility theorem" we discussed in Part I actually stems from the vertical structuring of the ideograph "algorithmic fairness", while the "inherent tradeoff" is a direct result of horizontal structuring. Second, examples of how STS researchers used the ideographic lens to analyze discussions surrounding promising technologies such as AR, and nanotechnology proves the framework's suitability for researching "algorithmic fairness", since its source technology (ADS) is also fairly novel. Finally, Bos et al.'s usage of the term to analyze how the different stakeholder groups horizontally and vertically structure ideographs, can be easily extended to what I plan to do for this research: understanding how actors other than researchers are ideographically using the term "algorithmic fairness" to support their cause and courses of action, while bringing down that of their challengers.

Part III: Results

Why "Algorithmic Fairness" Can be Considered an Emerging Ideograph

Before we can conduct an ideographic analysis of the term "algorithmic fairness" we must first verify if the term is really an ideograph. Fortunately, McGee, in his summary of results, listed quite a few identifying characteristics of ideographs that we can compare our candidate ideograph against (p. 15). The first of these is that it must be prevalent in common language use. Granted "algorithmic fairness" is not very common in public discourse yet, but the constituent term "fairness" most definitely is. And with the rate at which algorithmic decision systems in pervading every aspect of our life, it is not hard to imagine "algorithmic fairness" or some other synonymous term to enter the public's vocabulary in the near future.

Another character of an ideograph is that it must act as the "representation of a collective commitment to a particular but equivocal goal". As we have seen in Part I, scholars have deemed "algorithmic fairness" important enough to organize conferences around the topic and write numerous papers on it, all while disagreeing on the goals it should represent. The same is true for non-scholars. For instance, Woodruff, Fox and Rousso-Schindler, in 2018, had conducted interviews with people from disadvantaged backgrounds to learn about how they perceive the controversies surrounding algorithmic unfairness. Most of the interviewees had replied they had not heard of the term, but "learning about it elicited strong negative feelings", with some calling it "modern incarnation of familiar forms of discrimination" (p. 6).

McGee also draws attention to an ideograph's ability to "warrant the use of power" and "excuse behavior" that may be deemed "eccentric or antisocial". As we will see later on, a few government bodies have passed legislation that will allow public agencies to inspect software vendors' algorithms for biases - an act which, without the legitimizing force of "algorithmic fairness", would have been highly controversial. McGee also posits that ideographs are often "culture bound", with signification of a certain term varying across cultures. This is also apparent in the case of "algorithmic fairness", when we categorize the various papers on this topic by their country of origin (Figure 4). As we can see, much more scholarly attention is paid to this concept in the US, compared to other European countries, since among the western nations US has the most vibrant software innovation culture, along with a complex history of struggling to fend off discrimination.

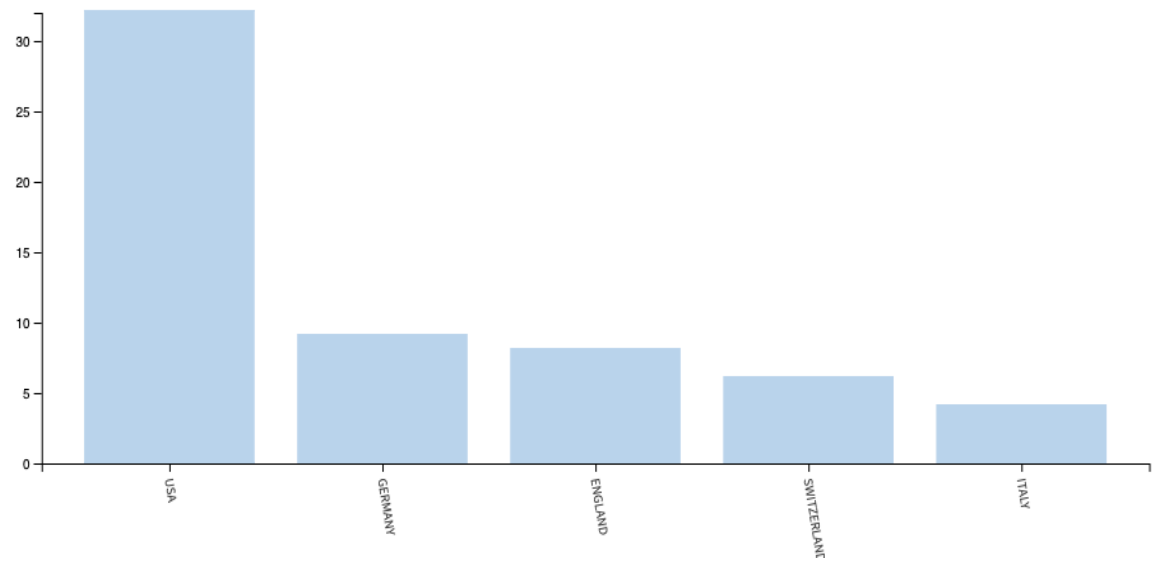

Figure 4: The number of papers that discuss "algorithmic fairness", grouped by their country of origin. United States is the clear front runner, with more than a third of all papers. (Created by author using the "Analyze results" tool at webofknowledge.com.)

Since "algorithmic fairness" more or less fulfills all the requirements set forth by McGee, we can quite reasonably label it as an emerging ideograph and move onto analyzing how exactly it is shaping public consciousness and our collective reality.

How the Ideograph "Algorithmic Fairness" is Vertically and Horizontally Structured

The analysis of the vertical and horizontal structuring of "algorithmic fairness" here will include the viewpoints of two different stakeholder groups: government, and the lay public. Among the two groups, the former has received almost no attention from the research community about how they perceive "algorithmic fairness", while the latter group, despite being the subject of quite a few papers, still provide plenty of opportunity for original observations.

Stakeholder Group 1: The US Government

In 2020, under a directive from the white house, the Office of Management & Budget (OMB) released a memorandum titled "Guidance for Regulation of Artificial Intelligence Applications". Here the term "Fairness and Non-discrimination" appears as follows:

> When considering regulations or non-regulatory approaches related to AI applications, agencies should consider, in accordance with law, issues of fairness and non-discrimination with respect to outcomes and decisions produced by the AI application at issue, as well as whether the AI application at issue may reduce levels of unlawful, unfair, or otherwise unintended discrimination as compared to existing processes (p. 5)

The definition of "algorithmic fairness" provided here is quite broad, but it does hint that the goal is to ensure outcome parity across demographic attributes. So that accounts for the vertical structuring; as for horizontal structuring, there are quite a few instances of ideographs that are in harmony with "algorithmic fairness", such as "Public Trust", "Scientific Integrity and Information Quality", "Disclosure and Transparency" etc. However, based on the following directive: "Federal agencies must avoid regulatory or non-regulatory actions that needlessly hamper AI innovation and growth", there is also potential for an ideographic clash between “algorithmic fairness” and “innovation and growth” and if that potential is actualized, the white house would probably uphold “innovation and growth” over “algorithmic fairness”.

In the legislative branch, the most prominent piece of legislature so far has been the “Algorithmic Accountability Act of 2019”, which advocated requiring all who employ algorithms for decision making to make available their algorithms and data so that “impact assessments” related to “accuracy, fairness, bias, discrimination, privacy, and security” can be performed. In the case of vertical structuring, the language of the document as well as the pro-diversity stance of the Democratic senators who introduced the bill makes it clear that by “fair algorithms” they mean algorithms which are fair when it comes to the treatment protected groups receive. As for horizontal structuring, the investigatory power the bill endows upon government agencies, is highly likely to draw criticism from companies protective of their trade secrets. This may eventually lead to a conflicting horizontal structuring between “corporate liberty” and “algorithmic fairness”, with each side touting the value of their respecting ideograph over the other.

So, as we can see here, although both the Republican executive branch, and the Democratic legislative branch have vertically structured the term in a similar manner, they have taken different routes in terms of their horizontal structuring, with party ideology causing the split. The results are summarized in the first row of Figure 5.

Stakeholder Group 2: The General Public

As we have already established based on Woodruff et al. (2018)’s research, the public may not be familiar with the term “algorithmic fairness”, but are aware of the consequence of its absence. The participants of the aforementioned study not only had strong opinions about this matter, as is natural with ideographs, they were also quick to "draw connections between algorithmic unfairness and national dialogue about racial injustice and economic inequality", which is proof of consonant horizontal structuring in the public domain.

In another research that sheds light on the vertical structuring of "algorithmic fairness", participants had to decide between 3 definitions of fairness in the case of deciding how to distribute loan among candidates (Saxena et al., 2019). The first one tried "treating similar individuals similarly", the second "never favored a worse individual over a better one", and third had the money "divided up in proportion to their repayment rates". The results showed that when provided with no demographic attributes of the borrowers, people consistently preferred the third approach, however when attributes such as race were introduced, people would change the preference towards an affirmative action approach, exemplifying how is it is to switch between vertical structurings of a term.

Another study (Marcinkowski, Kieslich, Stark & Lünich, 2020) that looked at how students perceived ADS's role in regards to college admission, found that participants thought both the "procedural" and "distributive" aspects of fairness important (see Part I for definitions of terms). However, when asked which mattered most when it came to their university's reputation, they chose the distributive aspect, whereas when it came to voicing their concerns about algorithmic unfairness, the procedural aspect played a more salient role. Again, we see the role context plays in the vertical structuring of a term.

The students also thought that, when compared to human decision makers, ADS were fairer, both in terms of their procedure and outcome. A different study which also looked at these two aspects of algorithmic decision making, however, found the opposite results, with participants preferring the "human touch" in decision making (Binns et al., 2018). Asked about scenarios in which ADS made decisions related to loan approval, promotion, car insurance etc., participants called such decision systems as “impersonal”, and “mean” (p. 6). One participant when asked “was the decision-making fair” answered: “It’s not fair at all really, but it’s understandable from the perspective of the business” (p. 11). Thus, this presents two cases of conflicting horizontal structuring: between “algorithmic fairness” and “human touch” as well as between “algorithmic fairness” and “business interest”. The results are again summarized in Figure 5, this time in row 2.

Field	Examples of Vertical Structuring	Examples of Horizontal Structuring
Government	Bias-free, non-discrimination, Equal treatment of disadvantaged population	No clash: Public Trust, Scientific Integrity, Transparency Clash: Innovation and growth, Corporate liberty
General Population	Process Fairness, Distributive Fairness	No clash: Racial Justice, Economic equality Clash: Human touch, Business interest

Figure 5: The various way the ideograph “algorithmic fairness” is vertically and horizontally structured across stakeholder groups.

Conclusion

In this paper, I have drawn attention to the emergence of a new ideograph in the field of ADS: “algorithmic fairness”. While the research community continues to be locked in a never-ending cycle of inventing, and optimizing definitions of “algorithmic fairness”, this ideograph, in true ideographic manner, has started “shaping public consciousness”, “controlling power”, and “influencing, if not determining, the shape and texture of each individual’s reality” (McGee, 2008, p. 4). Specifically by observing how two very different stakeholder groups – governing bodies, and the general public - are vertically and horizontally structuring the term for their own purposes, the paper has illuminated how much the criterion of fairness can very across groups and contexts, and how ideographs as wide-ranging as “corporate liberty”, “transparency”, and “human touch” are entering the scene in order to reinforce or weaken the rhetoric force of “algorithmic fairness”.

For the technically inclined, the results of this research should reinforce the need to look at the broader social context when trying to solve a technical problem, and inform non-technical actors to the best of one’s ability. For the responsible policy maker, the result of this result should lead them to be more cognizant of the rhetoric force an ideograph such as “algorithmic fairness” yields, and when they wield their governing power, to take into account the range of vertical and horizontal structuring of the term that are applicable in each specific context. And for the general public, heeding the lessons of this paper should result in paying better attention to the layers of meanings underlying the term “algorithmic fairness”, which should then allow them to better separate the individual mind from the collective, and evaluate the merits and demerits of this emerging ideograph on their own.

References

Algorithmic Accountability Act of 2019, S.1108, 116th Congress, 15 (2019).

Bias, AI Ethics, and the HireVue Approach. (n.d.). HireVue. Retrieved March 12, 2020, from https://www.hirevue.com/why-hirevue/ethical-ai

Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., & Shadbolt, N. (2018). “It’s reducing a human being to a percentage”; Perceptions of justice in algorithmic decisions. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ’18*, 1–14.

Bos, C., Walhout, B., Peine, A., & Lente, H. van. (2014). Steering with big words: Articulating ideographs in research programs. *Journal of Responsible Innovation*, *1*(2), 151–170.

Castelluccia, C., & Le Métayer, D. (2019). Understanding algorithmic decision-making: Opportunities and challenges. *Panel for the Future of Science and Technology., European Parliamentary Research Service*.

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic decision making and the cost of fairness. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 797–806.

Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

Friedler, S. A., Scheidegger, C., & Venkatasubramanian, S. (2016). On the (im)possibility of fairness. *ArXiv:1609.07236 [Cs, Stat]*. http://arxiv.org/abs/1609.07236

Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. *ArXiv:1610.02413 [Cs]*. http://arxiv.org/abs/1610.02413

Larson, J., Mattu, S., Kirchner, L., & Angwin, J. (2016). How we analyzed the COMPAS recidivism algorithm. *ProPublica (5 2016)*, *9*.

Liao, T. (2012). A framework for debating augmented futures: Classifying the visions, promises and ideographs advanced about augmented reality. *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*, 3–12.

Marcinkowski, F., Kieslich, K., Starke, C., & Lünich, M. (2020). Implications of AI (un-)fairness in higher education admissions: The effects of perceived AI (un-)fairness on exit, voice and organizational reputation. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 122–130.

McGee, M. C. (1980). The "Ideograph": A link between rhetoric and ideology. *Quarterly Journal of Speech*, *66*(1), 1.

Narayanan, A. (2018). *Tutorial: 21 fairness definitions and their politics, 2018*.

Natarajan, S., & Nasiripour, S. (2019, November 9). Viral tweet about apple card leads to goldman sachs probe. https://www.bloomberg.com/news/articles/2019-11-09/viral-tweet-about-apple-card-leads-to-probe-into-goldman-sachs

Rainie, L., Anderson, J., & Page, D. (2017). Code-dependent: Pros and cons of the algorithm age. *Pew Research Center*, *8*.

Request for Comments on a Draft Memorandum to the Heads of Executive Departments and Agencies, "Guidance for Regulation of Artificial Intelligence Applications." (2020, January 13). Federal Register. https://www.federalregister.gov/documents/2020/01/13/2020-00261/request-for-comments-on-a-draft-memorandum-to-the-heads-of-executive-departments-and-agencies

Saxena, N. A., Huang, K., DeFilippis, E., Radanovic, G., Parkes, D. C., & Liu, Y. (2019). How do fairness definitions fare?: Examining public attitudes towards algorithmic definitions of fairness. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society - AIES '19*, 99–106. https://doi.org/10.1145/3306618.3314248

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68. https://doi.org/10.1145/3287560.3287598

Verma, S., & Rubin, J. (2018). Fairness definitions explained. *2018 IEEE/ACM International Workshop on Software Fairness (FairWare)*, 1–7.

Woodruff, A., Fox, S. E., Rousso-Schindler, S., & Warshaw, J. (2018). A qualitative exploration of perceptions of algorithmic fairness. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14.

ZAML Fair | Don't Sacrifice Accuracy for Fairness. (n.d.). Zest.Ai. Retrieved March 12, 2020, from https://www.zest.ai/zaml-fair