RESEARCH ARTICLE

From satisficing to artificing: The evolution of administrative decision-making in the age of the algorithm

Thea Snow 🔟

Centre for Public Impact, Melbourne, Australia and New Zealand
Corresponding author. E-mail: theasnow@gmail.com

Abstract

Algorithmic decision tools (ADTs) are being introduced into public sector organizations to support more accurate and consistent decision-making. Whether they succeed turns, in large part, on how administrators use these tools. This is one of the first empirical studies to explore how ADTs are being used by Street Level Bureaucrats (SLBs). The author develops an original conceptual framework and uses in-depth interviews to explore whether SLBs are ignoring ADTs (algorithm aversion); deferring to ADTs (automation bias); or using ADTs together with their own judgment (an approach the author calls "artificing"). Interviews reveal that artificing is the most common use-type, followed by aversion, while deference is rare. Five conditions appear to influence how practitioners use ADTs: (a) understanding of the tool (b) perception of human judgment (c) seeing value in the tool (d) being offered opportunities to modify the tool (e) alignment of tool with expectations.

Policy Significance Statement

This research offers policymakers a new framework for analyzing and understanding how practitioners use Algorithmic Decision Tools (ADTs) to inform their decision-making. The analysis also offers some suggestions around the conditions that appear to contribute to different approaches to using ADTs. At present, the literature offers very little in terms of helping us to understand what drives different use types. This research helps to fill that gap by offering insights into the different conditions that appeared to be associated with different use-types. Understanding which conditions support a constructive human–machine interaction, and which do the opposite, is critical if these tools are going to succeed in supporting better decision-making in the public sector.

1. Main Text

Understanding how administrators make decisions is critical to understand how and why governments function as they do. The recent introduction of algorithmic decision tools (ADTs) to public sector agencies is changing how decisions are being made, particularly by Street Level Bureaucrats (SLB)—public sector workers who interact directly with citizens (Bovens and Zouridis, 2002). While there is significant research on the impact of technology on SLB discretion (Ellis, 2011; Hoybye-Mortensen, 2015; Busch and Henriksen, 2018), insufficient research considers how SLBs use ADTs to inform their decision

processes (Green and Chen, 2019). This paper contributes to this nascent field of research by exploring how an archetypal form of SLB—social workers—are using ADTs to guide their decision-making.

Previous research in the field of algorithmic decision-making has focused on either algorithm aversion, whereby practitioners ignore ADTs (Dietvorst et al., 2015), or automation bias, whereby practitioners defer to ADTs (Skitka et al., 2000). Yet these approaches have neglected the ways in which practitioners *combine* ADTs with their own intuition and expertise. This article introduces the new concept of "artificing" to better understand such situations. The introduction of artificing bridges the previously posed divide between aversion and automation bias and offers a new framework for analyzing and understanding how practitioners use ADTs to inform their decision-making.

To validate this framework, explore the prevalence of each use-type, and develop possible explanations for why different use types might be more or less common under different conditions, the author carried out in-depth interviews with practitioners across three separate ADTs—one in the United States (US), and two in the United Kingdom (UK). One of the ADTs was a "rule-based" tool, which takes a sequence of human instructions and applies those rules to analyze real-world data (Fry, 2018). The other tools examined are Artificial Intelligence (AI) tools, meaning that tool is given a goal, fed data, given feedback, and then left to achieve the specified end goal, without being programmed to do so (Bishop, 2006, 2).

Interviews revealed that artificing was the most common use-type, followed by aversion, while deference to ADTs was rare. In addition, common conditions were identified, which appear to influence how practitioners interact with ADTs: (a) understanding of the tool (b) perception of human judgment (c) seeing value in the tool (d) being offered opportunities to modify the tool (e) alignment of tool with expectations. These findings highlight the nuanced ways in which ADTs interact with human intuition and expertise, as well offering insights into which conditions support, and which inhibit, different use-types. Understanding this is critical if ADTs are to succeed in supporting better decision-making in the public sector.

2. Background

Over the last decade, ADTs have been introduced to support "more timely and accurate decision-making" (Shafiq, 2016) and address known shortcomings of human decision-making, for example, inconsistency (Parada et al., 2007) and bias (Kahneman and Klein, 2009; Gillingham, 2016).

In the field of children's services, actuarial decision-tools have already been used for several decades (Gillingham, 2011; Cuccaro-Alamin et al., 2017). What is much newer is the use of AI tools, which draw on vast quantities of data and use computational methods to "identify… patterns in datasets that, previously, were too large and complex to analyse" (Gillingham and Graham, 2017) and "learn" to calculate the relationship between those variables and a defined response variable (e.g., likelihood of child maltreatment).

Despite the increasing prevalence of algorithmic tools in a public sector context (Andrews, 2019; Vydra and Klievink, 2019; Kuziemski and Misuraca, 2020; Vogl et al., 2020), little research to date has explored how they are used by SLBs in practice.

This study focuses on three ADTs being used across four sites,[1] summarized in Table 1.

3. Conceptual Framework

In 1947, Herbert Simon introduced the idea that administrators make "boundedly rational" decisions (1997, 88), and "satisfice rather than maximise" (1997, 120). That is, they analyze the relevant and accessible facts in the time they have available and then use intuition to find their way to a satisfactory, yet not necessarily optimal, decision. Simon's work influenced the thinking of Lipsky, who suggested that SLBs rely on intuition, simplification, and routinization of tasks to manage the complex task of making decisions (2010, 83).

---

[1] Interviews were agreed on the basis of anonymity. Descriptions therefore offer enough information to provide a broad understanding of the tool and its context, while remaining vague enough to ensure anonymity is preserved.

Table 1. *Overview of ADTs being used across four interview sites.*

	Country	Tool type	Tool input	Tool output	Organizational context	No. interviews	Length of time tool in place
Tool 1	UK	Rule-based (non-AI)	Datasets include (but not limited to): school attendance, free school meals, statutory assessment results	Data dashboard —raises risk flags when children trigger pre-defined criteria	Used by individual early help/ intervention practitioners within a Local Authority as part of daily practice	4	Approx 10 months
Tool 2	US	AI tool	Datasets include (but not limited to): age, previous interactions with child protection, interactions with criminal justice system, previous case notes	Risk score	Used by intake teams working for a US County (responsible for investigating abuse and neglect allegations) as part of weekly team meetings to inform decisions about which cases warrant further investigation	4	Approx 9 months
Tool 3a	UK	AI tool	Datasets include (but not limited to): previous interactions with child protection, previous case notes, education, social benefits, council tax	Two-page brief for practitioners including risk summary, chronology, and guidance re: appropriate interventions	Used by individual practitioners within a Local Authority working in early intervention teams *as* part of daily practice	2	Approx 12 months Note that neither practitioner was actively using the tool (because it wasn't mandated) Practitioner reflections are theoretical only

Table 1. *Continued.*

	Country	Tool type	Tool input	Tool output	Organizational context	No. interviews	Length of time tool in place
Tool 3b	As above	As above	As above	As above	Used by individual practitioners working in multi-agency safeguarding hub teams, based in a Local Authority, as part of daily practice	3	Due to delays, practitioners were not actively using the tool when interviewed. Practitioners had received training, were part of a "champions group" providing ongoing feedback about the tool and were anticipating imminent roll-out. Practitioners' answers are therefore well-informed, but nonetheless, still theoretical

Abbreviation: AI, Artificial Intelligence.

Practitioners in the field of children's services represent a canonical form of SLB operating under conditions of bounded rationality. Practitioners are often making decisions under intense time pressure and with incomplete information (Kirkman and Melrose, 2014; Capatosto, 2017) meaning it is not unreasonable to assume that they are satisficing (Lipsky, 2010); in other words, analyzing the information that is available, in the time they have, and using that analysis, combined with their intuition, to come to a satisfactory, but not necessarily optimal, decision (Simon, 1997, 119). This study proceeds on the assumption that SLBs satisfice.

Scholars differ on the role of intuition in satisficing behavior. Simon belongs to a school of thought known as Naturalistic Decision-Making (NDM) theorists, who argue that humans tend to rely on "expert intuition" (Kahneman and Klein, 2009, 515) to satisfice. The NDM approach grew out of early research on master chess players (De Groot, 1978) and later decision made by firefighter commanders (Klein et al., 1986)—professionals who are able to win games, and successfully put out fires, without having to calculate all possible contingencies. According to NDM practitioners, humans draw on a "repertoire of patterns" (Kahneman and Klein, 2009, 515), compiled over the course of their careers, to make intuitive, but highly informed and effective, decisions.

In contrast, proponents of the Heuristics and Biases (HB) school are skeptical about the notion of expert intuition (Kahneman and Klein, 2009). They argue that human intuition tends to be characterized not by expertise, but by biases and heuristics (Moynihan and Lavertu, 2012). This means that relying on intuition to guide decisions "…sometimes yield reasonable judgment and sometimes lead to severe and systematic errors" (Kahneman and Tversky, 1973).

The key difference between NDM and HB practitioners turns on whether expert intuition or HB is seen to be more dominant. NDM theorists believe that intuition is *generally* expert, while HB theorists believe the opposite to be true (Kahneman and Klein, 2009). For this reason, NDM theorists tend to distrust algorithmic tools, while HB researchers "are predisposed to recommend the replacement of informal judgment by algorithms whenever possible" (Kahneman and Klein, 2009, 523).

The introduction of ADTs in recent years has been justified, at least in part, by a desire to reduce the extent to which public administrators satisfice; that is, rely on intuition in their decision-making (Schwartz et al., 2017). This is because the tools can very quickly and accurately capture, process, and spot patterns in the "buzzing, blooming confusion that constitutes the real world" taking into account hundreds of relevant factors, as opposed to the "just a few," that humans are able to analyze (Simon, 1997, 120). ADTs are also being embraced because studies demonstrate that algorithms tend to outperform humans in forecasting tasks (Grove et al., 2000).

However, while there is evidence that algorithms outperform humans at prediction tasks, most studies have compared *pure* clinical judgment to *pure* algorithmic judgment, rather than exploring what happens when human judgment (of which intuition forms a part) is used *in conjunction* with algorithmic tools. This insight, combined with the highly sensitive and complex nature of children's services, the challenges of algorithmic bias and incomplete datasets, and—in some jurisdictions—legislative requirements,[2] mean that designers of ADTs in children's services are not suggesting that the tools should completely replace human decision-making; instead, designers of ADTs suggest they will achieve optimal results when practitioners employ ADTs in combination with clinical judgment, and use them to augment, rather than supplant, decision-making processes (Cuccaro-Alamin et al., 2017; Celia, 2018).

This approach means that an element of human intuition endures as part of the decision-making process, even after the introduction of algorithmic tools. The author calls this process "artificing"—the form of satisficing which persists following the introduction of ADTs.

Taking into account the discussion above, it appears that two forms of artificing likely exist: expert artificing—when insights from the tool are combined with *expert* intuition; and biased artificing—when insights from the tool are combined with HB.

---

[2] Article 22(1) of the EU General Data Protection Regulation (2018) limits the ability to make solely automated decisions.

It is also important to highlight that practitioners do not always use the tools as intended by their designers (Gillingham, 2016). As such, there are likely additional ways—beyond artificing—that practitioners are using the tools. Previous literature suggests two additional use-types are likely.

First, practitioners could be largely or completely ignoring the tool; this is known as algorithm aversion (Dietvorst et al., 2015). Second, practitioners could be deferring to the tool rather than using it to augment their own professional judgment—a phenomenon known as "automation bias" (Skitka et al., 2000). Automation bias manifests in both errors of commission (following incorrect algorithmic advice) and omission (failing to act because the algorithmic tool does not prompt the practitioner to do so) (Goddard et al., 2012).

4. Methodology

This was a theory-generating exploratory study, using a small number of participants. This approach privileged depth over breadth, and focused on revealing insights which are difficult to achieve with numeric data (Ospina et al., 2018), although future research could build on these insights using survey approaches.

Thirteen practitioners were interviewed across four sites. As outlined in Table 1, two of the sites were actively using the tool, while two were not yet actively applying it as part of their everyday practice, meaning that practitioners' answers are well-informed, but hypothetical. Eight interviews were conducted by phone (some video-link; some audio only), while five were conducted in person. Each interview lasted approximately 1 hr.

The interviews were designed to reveal how practitioners were using ADTs in their daily practice according to the conceptual framework outlined above. By conducting semi-structured interviews at length, the author was able to explore the complexity and nuance of both *how* and *why* practitioners were using tools in particular ways to support their decision-making (Adams, 2015).

In addition to questions, each interview included three true-to-life vignettes, which were adapted for the different tools and were designed to offer insights into the way in which HB might play into the decision-making process. All subjects were relayed the three scenarios and asked to talk through how they would make a decision given that particular set of circumstances. This approach offers a unique way to explore how practitioners balance multiple factors when making complex decisions (Taylor, 2005; Stokes and Schmidt, 2012).

In order to generate insights into the conditions associated with different use-types, the author qualitatively coded all interviews to identify common conditions, and then assessed which conditions were present amongst artificing practitioners as compared with those who were ignoring or deferring to the tools.

5. Findings and Analysis

Practitioners were classified as demonstrating one or more use-types based on how they described themselves using the tool. Once classified within a category, the author analyzed the practitioner's description of how they were using the tool as well as the factors that contributed to their approach to using the tool—whether directly stated or inferred.

In this section, the author analyses the prevalence of each use-type; describes what each looks like in practice; and offers possible explanations for why practitioners adopt different approaches to using ADTs under different conditions.

6. Artificing

Participants were classified within this category if they demonstrated a willingness to reconsider or change a decision based on the tool's insights:

[The tool has] definitely led us to change a decision… (Tool 2, Interview 2)

Nine out of the 13 participants fell into this category.[3] While participants in this category indicated that they *may* change a decision because of the advice of the tool, they stressed that they would not *always* do so. Rather, they described taking the advice of the tool into consideration, combining it with their own judgment, and then coming to a decision based on a synthesis of that information. For example, one participant explained, "I'm not going to take what the tool says as gospel… I will explore [whatever the tool brings up], discuss it with the family and young person, and then come to a conclusion" (Tool 3b, Interview 3).

6.1. Distinguishing between expert and biased artificing

While biased artificing and expert artificing are presented as distinct use-types in the conceptual framework, in practice, these categories appear to converge in the vast majority of cases. Of the nine participants classified as "expert artificing," seven also displayed elements of "biased artificing." This came to light through the use of vignettes.

The most common form of bias was recall bias—where people allow their judgment to be strongly influenced by their ability to recall a similar example or event (Kirkman and Melrose, 2014). Participants were presented with the third vignette which describes the case of a young boy who is identified by the tool as low risk, apart from some concerns about poor school attendance. Practitioners were told that the previous week an almost identical case had been incorrectly assigned for no further action (the child was, in fact, being abused) and asked what they would do when presented with the current case.

A number of participants acknowledged that, even though the tool was identifying the child as low risk, they would treat it with greater caution as a result of the previous week's case:

> Would I consider what happened previously? Yes. You do. You do naturally. (Tool 3b, Interview 2)

While it is not possible to definitively exclude Bayesian updating as an influencing factor, the vignette was designed to elicit an emotional, rather than rational response from practitioners. Indeed, a number of practitioners acknowledged that, in response to this vignette, their decision was based on being "worried" or "overly sensitive" (Tool 2, Interview 4) and that the past experience would be "subconsciously" affecting their decision-making (Tool 3b, Interview 2). On this basis, the author's interpretation was that this was an example of recall bias.

Another bias revealed through the vignettes was confirmation bias—the phenomenon whereby people only look for evidence that confirms their pre-existing views (Kirkman and Melrose, 2014). When presented with the third vignette, a number of practitioners explained that they would again treat the case as high-risk even though the tool was suggesting otherwise; not because of the prior week's case, but because they believed that poor school attendance is associated with negative outcomes. For example, one participant explained, "my specialism is domestic abuse and… attendance is one of the factors [associated with domestic abuse]" (Tool 1, Interview 3). These practitioners were placing an irrationally high focus on school attendance as a risk factor because of pre-existing beliefs and preconceptions.

These findings highlight that the majority of practitioners who were artificing were drawing on both expert intuition *and* HB to supplement the tool's insights, highlighting that biased artificing and expert artificing coexist more often than not.

6.2. What does artificing look like?

There does not appear to be a uniform way of "artificing." Some participants in this category described using the tool to help them prioritize cases (Tool 1, Interviews 2 and 4); one person described using the tool like the advice of an experienced colleague (Tool 3a, Interview 2); and a number described using the tool to trigger earlier conversations (Tool 1, Interview 4; Tool 2, Interview 2; Tool 3b, Interviews 2 and 3). Many practitioners described using the tool as a check-and-balance:

---

[3] Total figures do not add to 13 because some subjects were classified in more than one category.

If it's a case I've decided we don't need to respond on, and then I see a risk score of 20, it makes me think again a little bit about it. Maybe not make me change my mind… but it helps me… go back through and take a look at it again… are we still comfortable with the decision we've made? (Tool 2, Interview 4)

A number of participants also described the tool as helping to guard against complacency:

When you've worked in social work for a long time, you build a high tolerance for neglect and unacceptable behaviors. Tools help you to be more reflective, so it's no longer just your view. (Tool 1, Interview 2)

Public sector organizations, who are being advised by tool designers to encourage artificing, might find this variability confusing, because it makes artificing harder to define and model as the preferred use-type. However, it does not appear to be possible, or desirable, to attempt to codify artificing for two reasons.

Firstly, AI tools are not as good as predicting rare events and unusual combinations of circumstances as they are common ones, because there are less data available to train them on (Church and Fairchild, 2017; Pryce et al., 2018; Munro, 2019). For this reason, it may be appropriate for practitioners to determine the extent to which they rely on the tool in any given case, depending on whether the circumstances presenting are commonplace or appear to be more unusual. Avoiding a tight definition gives practitioners licence to exercise this flexibility in their approach. And determine the extent to which they rely on a tool in any given case.

Secondly, a key reason that designers recommend that the tool be used in conjunction with professional judgment is to ensure that uniquely human skills, such as empathy and intuition, continue to inform the decision-making process (Munro, 2008). Being overly prescriptive in how to use the tool is likely to undermine practitioners' ability to express the very humanness that is seen to be an important counterbalance to the tool.

Thus, artificing is a dynamic process which varies from practitioner-to-practitioner, and case-to-case.

Interviews revealed five conditions that appeared to be present when practitioners were artificing. These emerged inductively on the basis of the interviews through a process of qualitative coding.

6.3. Solid understanding of the tool

Practitioners who were classified in this category demonstrated a solid understanding of the tool, meaning they:

- were able to identify many of the tool's data streams;
- demonstrated a basic understanding about how the tool works
- understood the strengths of ADTs, for example, their ability to "pick up on patterns which may not have been picked up before" (Tool 3b, Interview 3); and
- were able to specify the response variable that the tool was using to inform its analysis (relevant for AI tools only).

Interviews did not reveal why these practitioners had a strong understanding of the tool. The training that practitioners received in how to use a tool did not appear to be a critical factor and neither did the tool interface. There were several examples of practitioners using the same tool—having received the same training—who exhibited vastly different understandings of the tool—some very strong, and others very weak. Future studies could focus on this question.

It is also important to note that, in addition to understanding the strengths of these tools, participants in this category also recognized their limitations. For example, one practitioner explained:

The tool is going to let you know what school attendance is, but it's not going to let you know the reasons behind that… (Tool 3b, Interview 2)

However, artificers did not see those limitations as a reason to reject the tool; rather, limitations were seen as a reason to use the tool in conjunction with their "professional curiosity" (Tool 3b, Interview 1) or expert intuition:

It's about using the data but putting the human factor in as well. Using the two together. (Tool 3a, Interview 2)

There are three reasons why practitioners in this category might have been willing to tolerate the tools' limitations.

Firstly, a number of participants in this category described being able to make adjustments to the tool. Dietvorst et al. (2018) have reported that people will use imperfect algorithms if they can (even slightly) modify them. This is discussed as a distinct condition below but appears to be an important reason why practitioners in this category were willing to work with the tool despite knowing its shortcomings.

Secondly, many practitioners in this category felt confident to adapt how they were using the tool in order to overcome its limitations. For example, one practitioner working with the non-AI tool recognized that the data were out of date and recognized this as a shortcoming. However, she had designed a way of using the tool which compensated for this shortcoming, meaning that she was happy to continue using the tool. She used the tool to prioritize children and then flagged these priority children with partner organizations 2 weeks in advance to give them time to gather up-to-date data on the child ahead of their meeting (Tool 1, Interview 4). This suggests that when people can adapt how they're using the tool, they will be more tolerant of its limitations.

Finally, it may be that participants in this category were willing to tolerate algorithmic error because they did not see themselves as having to make a stark choice between their own judgment and the advice of the tool. In Dietvorst et al.'s studies (2015, 2018), people were asked to choose between the prediction of a human *or* a machine. Perhaps when people are able to identify the third option—humans working *together* with machines—they are more tolerant of algorithmic error because they feel more confident that they will be able to compensate for the tool's deficiencies.

6.4. *Acknowledgment of human limitations and bias*

Another feature which was present when participants were artificing was an acknowledgment of the limitations of human decision-making.

The tool was seen to be something which could help overcome these limitations. In particular, there was an acknowledgment that the tools could process a lot more information than practitioners were able to. One participant described the tool as being able to "see what has happened over a long period of time." As a result, he explained, the tool may be able to identify and highlight relevant history which is not apparent to the practitioner who has limited information, and limited time in which to process it (Tool 3b, Interview 3).

Practitioners also identified the existence of cognitive biases, and the role that ADTs could play in helping to address these:

When you see the same families come through again and again you start to think, "oh, it's this family again… and we're just going to screen it out"… but the tool is a reminder to stop and think… am I being biased here? (Tool 2, Interview 4)

6.5. *Value of the tool clear*

Building on the finding above, practitioners who were artificing saw significant value in the tool; they felt that the tool would help them to "do a better job" (Tool 3b, Interview 3) by helping them to work faster and reduce the chance of missing something.

A number of practitioners in this category also explained that the tool provides them with a greater sense of confidence, both because the tool makes more information available, but also because the tool is

more reliable than the advice of another professional. As one participant explained, without the tool, her data-gathering relies on the advice of other professionals, which is often given over the phone. This creates a sense of unease because "they can turn around and say, 'I never said that'" (Tool 2b, Interview 1). In contrast, if the tool provides a particular data-point about a child, this feels more concrete.

6.6. *Opportunity to modify*

A number of practitioners who were artificing had been given the opportunity to provide feedback on the tool and suggest how it might be improved. This ability to provide feedback on the tool likely contributed to their willingness to use and trust it.

A good illustration of this was borne out by conversations about the non-AI tool, where the data which the tool depended upon were often out-of-date. Where a practitioner felt that her feedback that the data were out-of-date was ignored, she ignored the tool as a result (Tool 1, Interview 1). In contrast, where another practitioner using the same tool was invited to be involved in a pilot to improve data currency, she *was* comfortable using the tool, and was, in fact, artificing, despite being aware of its shortcomings (Tool 1, Interview 4).

6.7. *Tool verifying experience and expectations*

Finally, a number of participants in this category suggested they were comfortable using the tool because it was largely confirming their own views and expectations:

> When I'm going to score someone, I can look at the history and say, this person's going to be [very high]. And usually I'm right. (Tool 2, Interview 4)

Unlike many of the other conditions listed above, this does not appear to be a sound justification for using the tool; rather, it appears to be a form of confirmation bias (Munro, 2008, 21).

7. Algorithm Aversion

Participants were classified within this category if they described the tool as having no impact on their decision-making:

> If I assign something, it's because I believe it meets criteria and everything else for me is null and void. I won't change my mind… I would never un-assign something because of a [low] score [on the tool]. (Tool 2, Interview 1)

In addition, participants in this category suggested that their practice had not changed, or they didn't anticipate it changing, as a result of the introduction of the tool. For example, one participant observed that if a practitioner in her team reported concerns about a child's school attendance, her approach would not change, even if the tool indicated that attendance was 100%

> …in all honesty, I'd still want to pick up the phone. I'd still pick up the phone [and call the school]. (Tool 3b, Interview 2)

Practitioners in this group described relying more on traditional sources of information like the referral report, case-note systems, and existing databases to inform their decision-making, with the tool sitting at the bottom of the hierarchy.

Five out of the 13 practitioners fell into this category.[4] Interviews revealed a number of common conditions that appeared to be present when participants displayed algorithm aversion.

---

[4] Total figures do not add to 13 because some subjects were classified in more than one category.

7.1. *Poor understanding of the tool*

Participants in this group had a poor understanding of the tool. Many identified knowing very little—if anything—about the data streams that the tool was drawing on to generate its output:

> I don't know if it draws [data from] other systems… I think it's just ours. (Tool 2, Interview 3)

A number of practitioners within this category also misunderstood how the tool functioned, suggesting that it drew on static historical data whereas, in fact, the data are regularly updated.[5] For one participant, this misapprehension that the data were purely historical was a key reason why she wouldn't change a decision based on the insights provided by the tool:

> I don't want to make a decision simply based on everything from the past. (Tool 2, Interview 3)

Participants in this category also had very little comprehension of how the tool worked. For example, when asked whether she knew how the tool generates a risk score, one participant answered, "I have no clue" (Tool 2, Interview 3).

7.2. *Belief in superiority of human judgment*

Unlike artificers, who highlighted the limitations of human decision-making, practitioners in this category saw human judgment as being superior to ADTs:

> If I'm having a face-to-face with somebody and they're telling me… this is the information, I'm going to trust that, rather than technology. (Tool 3b, Interview 2)

This was in part because participants in this category saw observation as being a uniquely human skill, which a machine could not emulate:

> Observations… are fundamental…looking around the home environment [you can assess]… How does it seem? Have they got electricity on? Have they got food in the fridge? Is the carpet sticky?… It's what you see and how people respond to questions; how quickly they're talking, their eye contact… that sort of thing… that may not show up on a database. (Tool 3a, Interview 1)

Practitioners were also reluctant to use the tools because they believed that the tools could not incorporate qualitative data into their analysis (Grove and Meehl, 1996). In fact, both of the AI tools incorporate case-note information to inform their output. This misperception that the tools work with purely quantitative data was seen as a significant shortcoming:

> Often a home visit tells you a lot about a family's situation. Is it messy? It is cluttered? Is it clean? Is it overly clean?… How do the parents present? It's those types of observations that you make… [that] inform your lines of inquiry…and you don't always get that information, I wouldn't have thought, from a data-source. (Tool 3a, Interview 1)

Unlike their peers in the "artificing" category, who accepted that the tool would naturally have some limitations, practitioners in this category saw the limitations of the tool as being unacceptable. In fact, this group appeared to hold the tool to impossibly high standards (Dawes, 1979)—standards that far exceed what humans might be reasonably expected to achieve:

> I just don't know that it's 100% accurate every time. (Tool 2, Interview 1)

This phenomenon of holding machines to a higher standard than humans is again explored in the work of Dietvorst et al. (2015) whose study showed that humans display "greater intolerance for error from

---

[5] The non-AI tool updated its data far less regularly than the AI tools.

algorithms than from humans. People are more likely to abandon an algorithm than a human judge for making the same mistake" (124).

In addition, a number of participants in this category described having had negative experiences with the tool previously, which likely contributed to their mistrust of the tool. One participant told a story about visiting a school to discuss the circumstances of a child who had been flagged as high risk, only to be told "that the child had died. And that's very embarrassing" (Tool 1, Interview 1).

The fact that these negative experiences resulted in practitioners rejecting the tools supports the findings of previous studies, which demonstrate that people tend to avoid algorithms after seeing them make mistakes.

7.3. Value of the tool not clear

In contrast to artificers, practitioners in this category did not see the tool as adding any value to their practice:

> If we didn't have [the tool], I'm not sure it would make a lot of difference to our decision-making… I'm not sure it would be sorely missed. (Tool 1, Interview 1)

Practitioners described feeling comfortable and confident with their practice, and did not see the tool as having anything to offer:

> I feel that our approach and our assessment is quite holistic. We can cover the areas we need to cover there and then. I'm not sure what [the tool] would add. (Tool 3a, Interview 1)

The fact that practitioners in this category did not tend to acknowledge the limitations of human decision-making, but rather, focused on the superiority of human judgment, meant that the value of the tool was far less clear than it was for artificers.

7.4. No opportunity to modify

Unlike practitioners who were artificing, practitioners in this category felt that they had no opportunities to modify, or provide feedback, on the tool's performance. This was a particular issue for the non-AI tool, where data currency was identified as a significant shortcoming. One practitioner felt very frustrated by the data being out-of-date, and felt she had no means of recourse:

> We're repeatedly feeding back the frustrations… [but] I can't see it changing at all, which is unfortunate. (Tool 1, Interview 1)

This suggests that the *inability* to provide feedback on the tool's performance can contribute to algorithm aversion.

7.5. Tool not verifying experience and expectations

The final condition that appeared to contribute to algorithm aversion was the tool *not* verifying practitioners' previous experience or expectations. One participant described having mixed feelings about the tool precisely because there had been a number of times where the score was vastly different to her own assessment:

> …there were sometimes where it has pulled information and gave (sic) a family a pretty high score that just really didn't make sense. (Tool 2, Interview 3)

Rejecting the tool because it does not conform to expectations is another example of confirmation bias. While confirmation bias is characterized by a tendency to look for information that supports existing beliefs, is also characterized by having "a blind spot when it comes to evidence that contradicts these views" (Munro, 2008, 21).

8. Automation Bias

Participants were classified within this use-type if they described relying entirely on the tool to make their decisions. Notably, only one participant was classified in this category.

The one observed instance of automation bias was, interestingly, by someone using the non-AI tool (generally speaking, fears around automation bias have largely focused on AI tools). Early in the interview, the participant claimed, "even if a family had five flags, or no flags, that wouldn't be the thing that guides my decision" (Tool 1, Interview 3). Nevertheless, the discussion elicited through the vignettes suggested that she *would* defer to the tool in certain instances.

The practitioner was presented with the first vignette, which describes a scenario where the tool indicates that a child is high risk, but the practitioner's previous personal interactions with the child had indicated there were no causes for concern. When asked what she would do in this situation, the practitioner responded that she would investigate immediately:

> With five flags, that's very significant. There's something going on for that child. (Tool 1, Interview 3)

Here we see that, despite saying that the tool wouldn't be *the* key factor in her decision-making, when presented with a scenario where the tool is raising five risk-flags, the practitioner explains that she would investigate. She accepts the information offered by the tool, without tempering it with her own experience with the child.

It is important to note that this participant also exhibited signs of artificing. Given this, it may be that the deferential behavior was driven by more risk aversion than true automation bias.

Practitioners working in children's services are (unsurprisingly) notoriously risk averse (Kirkman and Melrose, 2014). It is highly plausible, then, that the reason the participant relied so heavily on the tool in this particular instance was because the tool was suggesting that there were significant risks. It would perhaps be easier to confidently classify behavior as "deferential" where a practitioner decided—against his or her own best judgment—to *not* investigate a child because of a low-risk score. There were no examples of this.

9. Relationship Between Practitioner Personal Characteristics and Use-Type

As part of the interview process, participants were asked a range of questions about their personal characteristics to explore whether there was any association between these features and practitioners' approach to the tool. As illustrated in Table 2, and recognizing the very small sample sizes, there were no meaningful associations. Unlike previous studies, which have identified that "novices" are more likely to defer to decision-tools, while experts are more likely to artifice or ignore the tool (Fook, 2000; Gillingham, 2011; Kirkman and Melrose, 2014), here there was no apparent association between years' experience and use-type.

In addition, practitioners' sex, tool type, age, qualification level, attitudes toward technology, seniority, and field of work all appeared to have very little bearing on use-type.

As mentioned above, the limitations of this study mean that these conclusions should not be overstated. Nevertheless, they provide an interesting counterpoint to previous studies, which would benefit from further exploration.

10. Discussion

The findings above show that practitioners are using the tools differently, supporting findings that practitioners do not necessarily use the tool as instructed (Gillingham, 2016; Gillingham and Graham, 2017). This discussion will explore the conditions associated with different approaches, as well as the implications of different use-types.

11. Conditions Associated with Different Use-Types

There are five observed conditions that influence a practitioner's approach to using the tool.

Table 2. *Practitioners' personal characteristics and relationship to use-type.*

Variable	Artificing (expert/biased)	Automation bias	Algorithm aversion	Total
Sex, *n* (%)				
Female	6 (54.5)	1[a] (9.1)	4[a] (36.3)	11 (100)
Male	1 (50)	0	1 (50)	2 (100)
No. years' experience working in children's services, *n* (%)				
<5	1 (50)	0	1 (50)	2 (100)
5–10	2 (66.7)	0	1[a] (33.3)	3 (100)
>10	4 (50)	1[a] (12.5)	3 (37.5)	8 (100)
Tool type, *n* (%)				
Non-AI	2 (50)	1[a] (25)	1 (25)	4 (100)
AI	5 (55.5)	–	4[a] (44.5)	9 (100)
Age of participant, *n* (%)				
20–30	1 (100)	–	–	1 (100)
30–40	2 (66.7)	–	1 (33.3)	3 (100)
40–60	4 (44.5)	1[a] (11)	4[a] (44.5)	9 (100)
Qualification, *n* (%)				
Master's degree	1 (33.3)	1[a] (33.3)	1 (33.3)	3 (100)
Bachelor's degree	4 (57.2)	–	3[a] (42.8)	7 (100)
Certificate level qualification	2 (66.7)	–	1 (33.3)	3 (100)
Attitude toward technology, *n* (%)				
Tech optimism	6 (50)	1[a] (8.3)	5[a] (41.7)	12 (100)
Tech pessimism	1 (100)	–	–	1 (100)
Seniority within the organization, *n* (%)				
Management	4 (50)	1[a] (12.5)	3[a] (37.5)	8 (100)
Non-management	3 (60)	–	2 (40)	5 (100)
Field of work within children's services, *n* (%)				
Early help	4 (57.1)	1[a] (14.3)	2 (28.6)	7 (100)
Statutory services	3 (50)	–	3[a] (50)	6 (100)

Abbreviation: AI, Artificial Intelligence.
[a]Single subject also displayed elements of artificing.

The findings also show that the conditions present when people are artificing are an inversion of the conditions present when people exhibit algorithm aversion, shown below in Figure 1.

While these conditions have been presented as distinct, many of them are synergistic. For example, someone who recognizes the flaws in human decision-making is more likely to value the tool. In contrast, someone who believes that human decision-making is infallible is unlikely to see the tool's value.

It is also important to acknowledge that there were observed inconsistencies in the conditions that were present when people were artificing or ignoring the tool. What the findings above have attempted to do is draw out common, rather than universal, conditions which help explain why some participants were artificing while others were ignoring the tool. The cross-over and inconsistencies highlight that this is an area which would benefit from further research.

In addition, given the small number of interviews, future studies could also seek to verify the veracity of these conditions and explore whether other conditions—for example, decision-type—also influence the way practitioners use the tool.
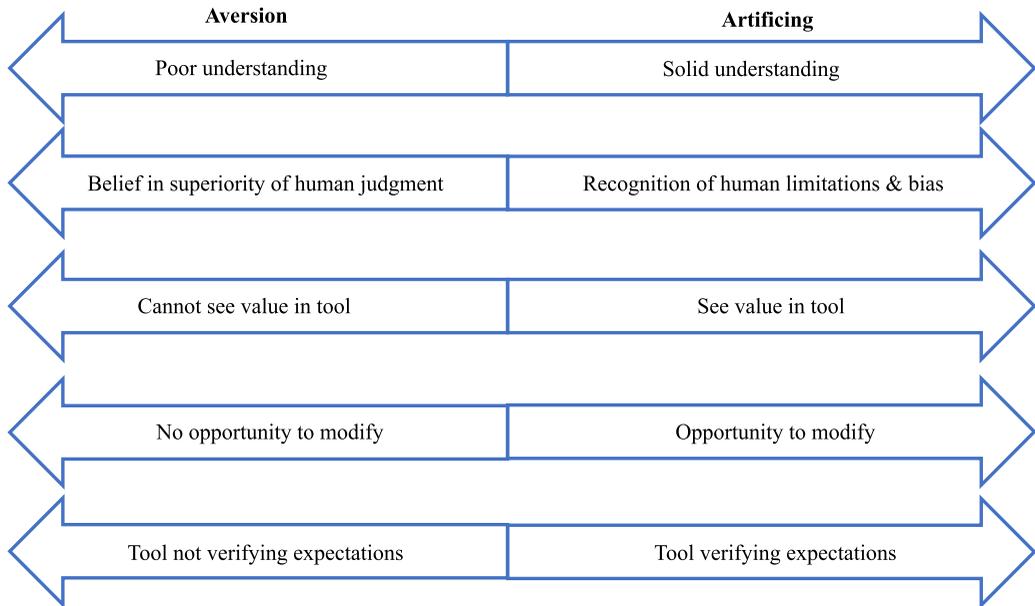
Figure 1. *Conditions affecting use-type—a spectrum.*

There were also instances of people displaying elements of more than one use-type, with one participant displaying elements of artificing *and* aversion, and one participant displaying elements of artificing *and* automation bias. This highlights that while there might be some practitioners who sit consistently at either the left or the right of every arrow, others may show more variability. Further research is needed to explore this in greater depth.

Given that only one participant demonstrated elements of automation bias, the conditions which support this use-type were not observable from this study. Future studies could explore whether the same conditions observed for artificing and aversion are present when people are exhibiting automation bias.

12. Implications of Different Use-Types

12.1. Artificing

Nine out of 13 practitioners were expert artificing, meaning that they were using the tool as recommended by tool designers. However, of these nine practitioners, seven were also biased artificing, meaning they were drawing on elements of expertise, *together with elements of HB*, to inform the intuitive element of the decision-making process.

This finding challenges a core claim used to justify the introduction of these tools; namely that they will minimize bias and increase objectivity in human decision-making (Shlonsky and Wagner, 2005; Church and Fairchild, 2017; Schwartz et al., 2017; Pryce et al., 2018). It also highlights that ADTs should not replace investment in professional education programs designed to minimize bias and develop professional expertise (Fook et al., 2000) amongst children's services practitioners.

12.2. Aversion

More than one in three practitioners were ignoring the tool. This is significant because "even the most predictive, psychometrically sound instrument, if not used, will be ineffective in the field" (Shlonsky and Wagner, 2005).

This finding also has implications for evaluating and understanding the tool's impact. Tool impact assessments which *assume* that people are using the tools as instructed will produce misleading results if

people are, in fact, ignoring the tool. For example, the authors of the recent impact assessment of the Allegheny Tool in the United States by Stanford University note that 39% of practitioners were not referring cases for further assessment when mandated to by the tool. Despite this, the authors did not take potential algorithm aversion into account when conducting the evaluation (Allegheny County Department of Human Services, 2019), meaning the results give an incomplete understanding of the tool's effects in practice.

Incorporating the framework of artificing, automation bias and aversion into future impact assessments would help to paint a clearer picture of the true impact that tools are having.

13. Automation Bias

As highlighted above, automation bias did not feature strongly as a use-type. A key reason may be that deference is being explicitly discouraged in tool training. One practitioner described that she and her colleagues had been, "directed that we shouldn't be basing our decision solely off the tool" (Tool 2, Interview 1).

However, while automation bias is not yet an observed feature in how the majority of practitioners are using these tools, a number of practitioners expressed concerns that deference *will* emerge as an issue once tools become more embedded in practice:

> [I'm nervous about reaching a point where] we take what's written down [by the tool] as gospel and stop speaking to people. (Tool 3b, Interview 1)

This is certainly something to continue monitoring, as an overly deferential approach is something to be avoided because, while ADTs are very good at certain things—for example, processing and synthesizing vast quantities of data and spotting patterns—they also have significant limitations (Shlonsky and Wagner, 2005). In particular, they:

- are not good at predicting rare events (Pryce et al., 2018);
- "cannot be programmed to predict for every single event that may occur at any point in the future" (Elish, 2019, 10);
- are often trained on incomplete data (Munro, 2019);
- are often trained on biased data, resulting in discriminatory tools (Munro, 2019).

For these reasons, it is critical to cultivate an environment where SLBs feel empowered to question, scrutinize, and challenge ADTs; treating tools less like an objective master and more like any other source of power (Fry, 2018).

Having described the implications of the different use-types, I will now explore the conditions that appear to support different approaches to use the tools.

14. Conclusion

This paper has examined how practitioners working in children's services are using ADTs to support their decision-making. Drawing on Herbert Simon's theory of satisficing, together with NDM and Heuristic and Biases theories, the author introduces a new conceptual framework for understanding how SLBs use ADTs.

Interviews reveal that artificing and algorithm aversion are the most common use-types. Only one practitioner displayed elements of automation bias. However, there are fears that deference may become more prevalent as tools become more established. This is something to continue monitoring, as keeping a "human in the loop" (Elish, 2019, 20) is critical given the known fallibilities of ADTs.

Interviews also revealed that, while the distinction between expert and biased artificing is conceptually quite neat, in practice, the lines are blurred. When practitioners rely on intuition to guide decisions, they are—in most cases—drawing on a combination of biased and expert intuition. This highlights that HB endures as a feature of decision-making despite the introduction of ADTs.

The existing literature offers very little in terms of helping us to understand what drives different use types (Dietvorst et al., 2015). This research helps to fill that gap by offering insights into the different conditions that were present when people were artificing and ignoring the tool. Understanding which conditions support a constructive human–machine interaction, and which do the opposite, is critical if these tools are going to succeed in supporting better decision-making in the public sector.

There are a number of key limitations to this study. Firstly, the insights gleaned from this research are based on semi-structured interviews only; an ethnomethodological approach (Garfinkel, 1984) would offer deeper insights. Secondly, the number of subjects is small. This was due to the nascent nature of the tools, as well as a general reluctance by Local Authorities to be interviewed due to intense media scrutiny of the use of ADTs in children's social care (see e.g., "Thurrock Council slammed for using analytics to profile residents and predict traits", Thurrock Gazette, 2018; Pegg and McIntyre, 2018). However, given the nature of this study was theory generating, rather than theory testing, the limited number of subjects was seen to be acceptable.

This study also lays foundations for future research. While artificing is being promoted by tool designers as the optimal way to use the tool, there is, as yet, no empirical evidence to demonstrate the superiority of this approach. The framework outlined in this study could be used to empirically validate the hypothesis that a combination of human *and* machine intelligence results in the best outcomes for families and children.

Data Availability Statement.   The Appendix (uploaded separately) to this paper contains interview protocols, questions, and vignettes used to conduct the interviews. In order to conduct in-depth and frank interviews, interviewees were assured anonymity both for themselves and for their organizations and work sites, and so research ethics concerns preclude making transcripts available.

Supplementary Materials.   To view supplementary material for this article, please visit http://dx.doi.org/10.1017/dap.2020.25.

References

Adams WC (2015) Conducting semi-structured interviews. In Newcomer KE, Hatry HP, and Wholey JS (eds), *Handbook of Practical Program Evaluation*. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 492–505. https://doi.org/10.1002/9781119171386.ch19.

Allegheny County Department of Human Services (2019) *Impact Evaluation Summary of the Allegheny Family Screening Tool.* Stanford University. Available at https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Impact-Evaluation-Summary-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-5.pdf (accessed on 21 May 2019).

Andrews L (2018) Public administration, public leadership and the construction of public value in the age of the algorithm and "big data". *Public Administration 97*(2), 296–310. https://doi.org/10.1111/padm.12534

Bishop CM (2006) Pattern recognition and machine learning. In *Information Science and Statistics*. New York: Springer 2.

Bovens M and Zouridis S (2002) From street-level to system-level bureaucracies: how information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review 62*(2), 174–184. https://doi.org/10.1111/0033-3352.00168

Busch PA and Henriksen HZ (2018) Digital discretion: a systematic literature review of ICT and street-level discretion. *Information Polity 23*(1), 3–28. https://doi.org/10.3233/IP-170050

Capatosto K (2017) *Foretelling the Future: A Critical Perspective on the Use of Predictive Analytics in Child Welfare.* Research Report. Kirwan Institute. Available at http://kirwaninstitute.osu.edu/wp-content/uploads/2017/05/ki-predictive-analytics.pdf (accessed on 5 May 2019).

Celia H (2018) *Building Capacity Through Data and Analytics to Improve Life Outcomes*. Presented at the UK Authority Data4Good, 16 October 2018. Available at https://www.ukauthority.com/events/event-hub-ukauthority-data4good-2018/ (accessed on 28 April 2019).

Church CE and Fairchild AJ (2017) In search of a silver bullet: child welfare's embrace of predictive analytics. *Juvenile and Family Court Journal 68*(1), 67–81. https://doi.org/10.1111/jfcj.12086

Cuccaro-Alamin S, **Foust R**, **Vaithianathan R and Putnam-Hornstein E** (2017) Risk assessment and decision making in child protective services: predictive risk modeling in context. *Children and Youth Services Review 79*(August), 291–298. https://doi.org/10.1016/j.childyouth.2017.06.027

Dawes RM (1979) The robust beauty of improper linear models in decision making. *American Psychologist 34*(7), 571–582. https://doi.org/10.1037/0003-066X.34.7.571

De Groot AD (1978) *Thought and Choice in Chess. 2nd ed. Psychological Studies 4*. The Hague: Mouton.

Dietvorst BJ, **Simmons JP and Massey C** (2015) Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General 144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dietvorst BJ, **Simmons JP and Massey C** (2018) Overcoming algorithm aversion: people will use imperfect algorithms if they can (even slightly) modify them. *Management Science 64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Elish, Madeleine Clare (2019) Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction (pre-print). *Engaging Science, Technology, and Society* (pre-print), Available at SSRN: https://ssrn.com/abstract=2757236 or http://dx.doi.org/10.2139/ssrn.2757236.

Ellis K (2011) "Street-level bureaucracy" revisited: the changing face of frontline discretion in adult social care in England: "street-level bureaucracy" revisited. *Social Policy & Administration 45*(3), 221–244. https://doi.org/10.1111/j.1467-9515.2011.00766.x

Fook J, **Ryan M and Hawkins L** (2000) *Professional Expertise: Practice, Theory and Education for Working in Uncertainty*. London: Whiting and Birch.

Fry H (2018) *Hello World: Being Human in the Age of Algorithms*, 1st Edn. New York, NY: W.W. Norton & Company.

Garfinkel H (1984) *Studies in Ethnomethodology*. Cambridge, UK: Polity Press.

Gazette T (2018) *Thurrock Council Slammed for Using Analytics to Profile Residents and Predict Traits*, 17 September 2018. Available at https://www.thurrockgazette.co.uk/news/16884467.thurrock-council-slammed-for-using-analytics-to-profile-residents-and-predict-traits/. [Please note that this is from the Thurrock Gazette - a newspaper - with no identified author. Citation will need to be adjusted to reflect this] (accessed 16 June 2019).

Gillingham P (2011) Decision-making tools and the development of expertise in child protection practitioners: are we "just breeding workers who are good at ticking boxes"? Decision-making tools and expertise. *Child & Family Social Work 16*(4), 412–421. https://doi.org/10.1111/j.1365-2206.2011.00756.x

Gillingham P (2016) Predictive risk modelling to prevent child maltreatment and other adverse outcomes for service users: inside the "black box" of machine learning. *British Journal of Social Work 46*(4), 1044–1058. https://doi.org/10.1093/bjsw/bcv031

Gillingham P and Graham T (2017) Big data in social welfare: the development of a critical perspective on social work's latest "electronic turn". *Australian Social Work 70*(2), 135–147. https://doi.org/10.1080/0312407X.2015.1134606

Goddard K, **Roudsari A and Wyatt JC** (2012) Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association 19*(1), 121–127. https://doi.org/10.1136/amiajnl-2011-000089

Green B and Chen Y (2019) The principles and limits of algorithm-in-the-loop decision making. *Proceedings of the ACM on Human-Computer Interaction 3*(CSCW), 1–24. https://doi.org/10.1145/3359152

Grove WM and Meehl PE (1996) Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical-statistical controversy. *Psychology, Public Policy, and Law 2*(2), 293–323. https://doi.org/10.1037/1076-8971.2.2.293

Grove WM, **Zald DH**, **Lebow BS**, **Snitz BE and Nelson C** (2000) Clinical versus mechanical prediction: a meta-analysis. *Psychological Assessment 12*(1), 19–30. https://doi.org/10.1037//1040-3590.12.1.19

Hoybye-Mortensen M (2015) Decision-making tools and their influence on caseworkers' room for discretion. *British Journal of Social Work 45*(2), 600–615. https://doi.org/10.1093/bjsw/bct144

Kahneman D and Klein G (2009) Conditions for intuitive expertise: a failure to disagree. *American Psychologist 64*(6), 515–526. https://doi.org/10.1037/a0016755

Kahneman D and Tversky A (1973) On the psychology of prediction. *Psychological Review 80*(4), 237–251. https://doi.org/10.1037/h0034747

Kirkman E and Melrose K (2014) *Clinical Judgement and Decision-Making in Children's Social Work: An Analysis of the "Front Door" System.* Research Report RR337. Department for Education.

Klein GA, **Calderwood R and Clinton-Cirocco A** (1986) Automation bias and errors: are crews better than individuals? *Rapid Decision Making on the FireGround 30*(6), 576–80. https://doi.org/10.1177/154193128603000616

Kuziemski M and Misuraca G (2020) AI governance in the public sector: three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications Policy 44*(6), 101976. https://doi.org/10.1016/j.telpol.2020.101976

Lipsky M (2010) *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. 30th anniversary expanded ed. - New York: Russell Sage Foundation.

Moynihan DP and Lavertu S (2012) Cognitive biases in governing: technology preferences in election administration. *Public Administration Review 72*(1), 68–77. https://doi.org/10.1111/j.1540-6210.2011.02478.x

Munro E (2008) *Effective Child Protection*. 2nd ed. Los Angeles: Sage Publications.

Munro E (2019) *Predictive Analytics in Child Protection.* CHESS Working Paper No. 2019-03, April.

Ospina SM, **Esteve M and Lee S** (2018) Assessing qualitative studies in public administration research. *Public Administration Review 78*(4), 593–605. https://doi.org/10.1111/puar.12837

Parada H, **Barnoff L and Coleman B** (2007) Negotiating "professional agency": social work and decision-making within the Ontario child welfare system. *The Journal of Sociology & Social Welfare 34*(4), 35–56.

Pegg D and McIntyre N (2018) Child Abuse Algorithms: From Science Fiction to Cost-Cutting Reality. *The Guardian*, 16 September 2018. Available at https://www.theguardian.com/society/2018/sep/16/child-abuse-algorithms-from-science-fiction-to-cost-cutting-reality?CMP=Share_iOSApp_Other (accessed on 16 June 2019).

Protection Regulation (2018). Available at https://gdpr-info.eu/

Pryce J, **Yelick A**, **Zhang Y and Fields K** (2018) Using Artificial Intelligence, machine learning, and predictive analytics in decision-making. In *White Paper*. Florida Institute for Child Welfare (Florida University). Available at https://pdfs.semanticscholar.org/d8b6/b80d41d34c460ddfb351ca95ff2c2965ead4.pdf (accessed on 20 May 2019).

Schwartz IM, **York P**, **Nowakowski-Sims E and Ramos-Hernandez A** (2017) Predictive and prescriptive analytics, machine learning and child welfare risk assessment: the Broward county experience. *Children and Youth Services Review 81*(October), 309–320. https://doi.org/10.1016/j.childyouth.2017.08.020

Shafiq W (2016) Shadow of the Smart Machine: Algorithm Guided Decision Making in the Public Sector. *Shadow of the Smart Machine: Algorithm Guided Decision Making in the Public Sector* (blog). January 29, 2016. Available at https://www.nesta.org.uk/blog/shadow-of-the-smart-machine-algorithm-guided-decision-making-in-the-public-sector/ (accessed on 28 April 2019).

Shlonsky A and Wagner D (2005) The next step: integrating actuarial risk assessment and clinical judgment into an evidence-based practice framework in CPS case management. *Children and Youth Services Review 27*(4), 409–427. https://doi.org/10.1016/j.childyouth.2004.11.007

Simon HA (1997) *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations*. 4th Edn. New York: Free Press.

Skitka LJ, **Mosier KL**, **Burdick M and Rosenblatt B** (2000) Automation bias and errors: are crews better than individuals? *The International Journal of Aviation Psychology 10*(1), 85–97. https://doi.org/10.1207/S15327108IJAP1001_5

Stokes J and Schmidt G (2012) Child protection decision making: a factorial analysis using case vignettes. *Social Work 57*(1), 83–90. https://doi.org/10.1093/sw/swr007

Taylor BJ (2005) Factorial surveys: using vignettes to study professional judgement. *British Journal of Social Work 36*(7), 1187–1207. https://doi.org/10.1093/bjsw/bch345

Vogl TM, **Seidelin C**, **Ganesh B and Bright J** (2020) Smart technology and the emergence of algorithmic bureaucracy: Artificial Intelligence in UK local authorities. *Public Administration Review 80*, 946–961. https://doi.org/10.1111/puar.13286

Vydra S and Klievink B (2019) Techno-optimism and policy-pessimism in the public sector big data debate. *Government Information Quarterly 36*(4), 101383. https://doi.org/10.1016/j.giq.2019.05.010