Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

WILEY

Shared decision-making and maternity care in the deep learning age: Acknowledging and overcoming inherited defeaters

Keith Begley PhD[1] | Cecily Begley PhD[2] | Valerie Smith PhD[2]

[1]Department of Philosophy, Trinity College Dublin, Dublin, Ireland

[2]School of Nursing and Midwifery, Trinity College Dublin, Dublin, Ireland

Correspondence
Keith Begley, Department of Philosophy, Trinity College Dublin, College Green, Dublin, Ireland.
Email: begleyk@tcd.ie

Abstract

In recent years there has been an explosion of interest in Artificial Intelligence (AI) both in health care and academic philosophy. This has been due mainly to the rise of effective machine learning and deep learning algorithms, together with increases in data collection and processing power, which have made rapid progress in many areas. However, use of this technology has brought with it philosophical issues and practical problems, in particular, epistemic and ethical. In this paper the authors, with backgrounds in philosophy, maternity care practice and clinical research, draw upon and extend a recent framework for *shared decision-making* (SDM) that identified a duty of care to the client's knowledge as a necessary condition for SDM. This duty entails the responsibility to acknowledge and overcome *epistemic defeaters*. This framework is applied to the use of AI in maternity care, in particular, the use of machine learning and deep learning technology to attempt to enhance electronic fetal monitoring (EFM). In doing so, various sub-kinds of *epistemic defeater*, namely, *transparent, opaque, underdetermined, and inherited defeaters* are taxonomized and discussed. The authors argue that, although effective current or future AI-enhanced EFM may impose an epistemic obligation on the part of clinicians to rely on such systems' predictions or diagnoses as input to SDM, such obligations may be overridden by *inherited defeaters*, caused by a form of *algorithmic bias*. The existence of *inherited defeaters* implies that the duty of care to the client's knowledge extends to any situation in which a clinician (or anyone else) is involved in producing training data for a system that will be used in SDM. Any future AI must be capable of assessing women individually, taking into account a wide range of factors including women's preferences, to provide a holistic range of evidence for clinical decision-making.

KEYWORDS
algorithmic bias, artificial intelligence, duty of care, electronic fetal monitoring, epistemic defeaters, shared decision-making

498 | **WILEY**— Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

BEGLEY ET AL.

1 | INTRODUCTION

In Begley et al[1] it was argued that *shared decision-making* (SDM) "takes the form of a dialogue within which the clinician fulfils their duty of care to the client's knowledge by making available their complete knowledge (based on all types of evidence) and expertise, including an exposition of any relevant and recognized potential defeaters."[1(p1119)] An *epistemic defeater*[2] is a truth such that, if one were aware of it, one would realize that one does not have knowledge in some particular case in which one had thought one did. For example, if it turned out that a clinician had recommended a treatment to me on the basis that they would be paid more rather than on the basis of medical evidence, even if they would have recommended the same treatment in both instances. In this case, I would not *know* that it was the best treatment, although I might nevertheless have a mere justified true belief that it was. Such cases are called *Gettier cases*.[3] Begley et al were concerned to show how fulfilling a duty of care to the client's knowledge helps to overcome epistemic defeaters resulting from biases and undue influence in clinical decision-making in maternity care, such as: clinicians' personal beliefs, concerns over litigation, lack of resources, private vs public insurance, clinicians' age and gender, etc.[1] If one were to give a broad label to these kinds of epistemic defeaters, perhaps they could be called "all too human." In the present article, we examine how this framework for SDM in maternity care also helps to address epistemic defeaters of a different kind, that is, those that are produced via the interaction of the "all too human" with the artificial, and perhaps "not human enough," in such a way that the defeaters associated with the former are *inherited* and disguised, obfuscated, or legitimized in the process and by way of the latter.

2 | AI, MACHINE LEARNING, AND DEEP LEARNING

It is important first of all to distinguish the meanings of three terms that are used a lot in such areas of discussion, but which are not always explained with sufficient care. The most well-known term, "Artificial Intelligence" (AI), is from the mid-twentieth century, although the idea has been around for hundreds of years. It has a number of distinct uses. It could be used in a specific sense to refer to a technology that does not yet exist, and might never exist, an AI that is comparable to a human being in terms of its cognition, intelligence, plasticity, and perhaps even sentience; the walking, talking AIs of science fiction. For the purposes of the present article, we leave aside this sense of the term and the philosophical issues pertaining to it.

The term is also used in a broad sense to refer to the ability of a piece of technology to perform some functions similar to those of cognisant and intelligent non-artificial creatures such as human beings. Early effective kinds of AI employed hard-coded, rule-based systems explicitly programmed to achieve certain ends. IBM's Deep Blue, which beat Kasparov in 1997, is a standard example of such AI. It was explicitly provided with the rules of chess and then merely calculated opportune moves to make by assessing the values of all the potential game boards resulting from those moves, that is, a brute-force approach.

"Machine Learning" (ML) is a subset of AI in the broad sense. Similarly, ML is not one technology but many technologies, some of which have been around for over 60 years. Broadly speaking an ML system is a system that "learns" or improves by iteratively evaluating and optimizing its representation of a problem implicitly determined by a data set, without the need for this to be explicitly hard-coded. The process repeats for perhaps thousands of iterations, until an optimum or near-optimum configuration is arrived at. The strategy being that such an algorithm will produce a near-optimum answer in a much shorter time than trying every configuration in a brute-force manner, or guessing, or hard-coding rules.

"Deep Learning" (DL) is a kind of ML that has taken off in recent times due to a confluence of improved employment of artificial neural networks, big data, and processing power (see Hinton[4] for a brief introduction written with health care in mind). DL systems employ training data to train a neural network by appropriately weighting connections between the nodes in the network to capture even weak correlations in data. This training isolates hidden "features" in the data, which allow a trained neural network to pick out similar features in future. That is, it allows such networks to pick out and appropriately weight such correlations in further instances that are not part of the original training data with which it is presented.

3 | TRANSPARENT, OPAQUE, AND UNDERDETERMINED DEFEATERS

Although some epistemic defeaters may be tacit, or withheld, they are nonetheless in principle relatively *transparent* in the sense that they are epistemically scrutable and available to the clinician, even if this would in some cases require the will, effort, integrity or introspection to realize. On the other hand, some epistemic defeaters may be *opaque*, unknown and unavailable to the clinician no matter how much effort is applied.

In a recent article, Bjerring & Busch put forward an argument to show that patient-centred decision-making (such as SDM), is undermined by what they call "black-box medicine," involving DL systems.[5] They begin by assuming that DL systems outperform, or with enough development would outperform, human practitioners. It follows from this, they argue, that there would be an epistemic obligation for practitioners to rely upon such DL systems, just as they would upon reliable experts. However, this is problematic for SDM.

> "The core reason is simple: since black-box AI systems do not reveal to practitioners how or why they reach the recommendations that they do, then neither can practitioners who rely on these black-box systems in decision-making—assuming that they honor their epistemic obligation—explain to patients how and why they give the recommendations that they do."[5(§4)]

The proximate reason for the client/patient's lack of knowledge in such a scenario is that even the practitioner would not know why they made a certain recommendation. The underlying problem is that such DL systems are *opaque* in the sense that the layers of hidden variables that they employ cannot be interpreted. That is, "we can literally fail to have a minimally sensible basic interpretation and explanation of the information that the algorithm employs for producing its recommendations."[5(§5)] Bjerring & Busch argue that this makes the opacity of such cases categorically different from the more usual cases of clinical decision-making, which are based on statistical correlation, experience, and practice rather than theory and causal explanation.[5]

Burrell[6] characterizes the kind of opacity involved in DL systems as being engendered by the sheer complexity of "the algorithm in action," the "interplay" between extremely large data sets and code, although each may be comprehensible by itself. That is, it is not merely due to a lack of technical understanding on the part of an interpreter or clinician, or, again, secrecy or obfuscation, as in the case of some of the "all too human" but relatively transparent defeaters mentioned earlier, and previously.[1] The technical reason for this complexity, and the resulting opacity, is that deep neural networks use "layers of learned, nonlinear features to model a huge number of complicated but weak regularities in the data."[4(p1102)] Thus, as Hinton further points out, these "features" have meaning only in relation to the complex and abstract interconnections contrived by the neural network.[4]

The problem of the epistemic opacity of DL systems, including its practical implications in many fields, is already well known, and solutions and alternatives are being actively developed, often under the name "explainable AI (XAI)" (see Gilpin et al[7] for an overview). However, it remains to be seen whether or not such methods will be viable and offer adequate solutions. Indeed, as Walmsley has argued,[8] *contestability* should instead form part of the training process even if explainability, or interpretability, etc., cannot be achieved.

There is also a further aspect to epistemic defeaters arising from DL systems, which stems from a problem that has occupied philosophers of science for at least the past century (and especially since van Fraassen[9] in 1980), that of *underdetermination*. Hinton clearly and succinctly explains the cause of this in DL systems:

> "[...] if the same neural net is refit to the same data, but with changes in the initial random values of the weights, there will be different features in the intermediate layers. This reflects that unlike models in which an expert specifies the hidden factors, a neural net has many different and equally good ways of modeling the same data set. It is not trying to identify the 'correct' hidden factors. It is merely using hidden factors to model the complicated relationship between the input variables and the output variables."[4(p1102)]

So, it would appear then that there is a deeper problem than merely the complexity of such models, namely, their *contrastive*

underdetermination; that is, their being underdetermined relative to alternatives. There is nothing to choose between one model and (at least potentially) an infinite variety of other models (trained neural networks) that would produce the *same* outcomes or predictions on the basis of the *same* data. Such models would thus be empirically equivalent. Indeed, there is nothing to say that one model as opposed to another should be considered a model of a portion of the real world at all, rather than being a model of a different abstract structure satisfying the same empirical constraints. To be merely empirically adequate there need not be, for example, a mother or fetus represented in these models, only an abstracted structure that happens to conform in various relevant ways with real world data about mothers and fetuses or, in another way of thinking about it, things empirically equivalent to them.

The problem that this underdetermination presents for SDM is that although DL systems might produce correct predictions, diagnoses, etc., the clinician relying on such predictions cannot claim that they *know* that these outputs relate to the client's case, rather than something that is merely empirically equivalent to the client's case in the relevant ways. Pushed far enough, this adjoins the debate between the broad philosophical positions of scientific realism and scientific antirealism.

Even if the defeaters associated with the opacity and underdetermination of DL systems were to be overcome, ameliorated, or (in the case of antirealism regarding underdetermination) accepted, there would nevertheless remain a prior issue, that of the comparatively transparent epistemic defeaters arising from various "all too human" factors. Furthermore, as we shall see, if the opacity and underdetermination of DL systems are not overcome or ameliorated, this has the effect of further obfuscating or disguising what we shall call the *inheritance* of epistemic defeaters deriving from those "all too human" factors.

4 | THE CASE OF ELECTRONIC FETAL MONITORING IN MATERNITY CARE

It has been recognized that the use of some diagnostic technology has the potential to lead to harm in a clinical setting. At first sight, this can seem like an odd situation to someone not acquainted with the area— How could having more information be disadvantageous?

A good example of this in maternity care is electronic fetal monitoring (EFM) in labour, where either an abdominal transducer placed on the woman's abdomen, or a small probe attached to the fetus' head measures the fetal heart rate and a second abdominal transducer measures uterine activity, presenting both as a graph tracing on paper or screen, the cardiotocograph, commonly referred to as the CTG. These traces are assessed visually by clinicians, using accepted country guidelines, or those published by the International Federation of Gynecology and Obstetrics (FIGO).[10] Initially designed as an assessment tool to aid clinical decision-making, EFM, based on trace outputs, has emerged as a clinical "decider" in and of itself. This may well be a legacy of the fanfare and excitement with which EFM was first

500 | **WILEY**—Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

BEGLEY ET AL.

introduced (some 50–60 years ago), with reports of over three-quarters of practising clinicians holding a genuine belief that the fetal monitor was one of obstetrics' "best inventions,"[11] and the majority of clinicians (96% and 63%, respectively) believing that EFM reduced perinatal mortality and morbidity and improved maternal and neonatal outcomes.[12,13] Clinicians seemed to believe that, with this technology, "the obstetrician may virtually eliminate intrapartum stillbirths and reduce morbidity associated with parturition,"[14(p33)] by virtue of being able to visualize the fetal heart rate continuously throughout a woman's labour.

Despite this initial excitement, high inter- and intra-observer variability of this visual assessment has been documented among clinicians in some areas,[15] and for some time,[16,17] although the use of the FIGO guidelines as a standardized approach to CTG interpretation may improve agreement between clinicians.[18] Distrust in EFM as a superior monitoring technology (to that of the traditional method of intermittently auscultating the fetal heart) also began to emerge in the wake of its widespread introduction, with later studies highlighting that clinicians held less trust in the CTG than over their own observations, had concerns for overreliance and overuse of EFM, and did not believe that EFM was essential for a successful, safe birth.[19-21] These changing views largely correspond with evidence that emerged after the introduction of EFM, which showed that CTG use may potentially cause harm in some cases. The Cochrane systematic review on the continuous use of CTGs in labour, for example, found that it tends to lead to higher caesarean section (CS) and instrumental birth rates, compared to other forms of monitoring such as intermittent auscultation of the fetal heart, without any improvement in rates of cerebral palsy, infant mortality or other assessments of neonatal wellbeing.[22] Accordingly, the guidelines on the care of women in labour, from the UK National Institute for Health and Care Excellence (NICE) now recommend "1.10.1 Do not offer cardiotocography to women at low risk of complications in established labour," "1.10.5 Do not offer continuous cardiotocography to women who have non-significant meconium if there are no other risk factors" and "1.10.6 Do not regard amniotomy alone for suspected delay in the established first stage of labour as an indication to start continuous cardiotocography."[23] However, in many maternity units EFM is routinely used, even in low-risk women, a practice that appears to be in conflict with the evidence and the (later) views of clinicians. Evidence from qualitative enquiry provides insight into this, noting that the contemporary use of EFM in maternity practice, is largely motivated by the ability of EFM to provide professionals with what they perceive as hard copy "proof" that a baby is not compromised while in their care and thus serves to guard against criticism and legal action should an adverse outcome occur.[24]

Concurrently, amidst this conflict, we see limited consideration for the value placed by women on the use of EFM as a clinical support and decision-making aid. For example, in a systematic review of 10 studies exploring women's views of fetal monitoring,[25] fear and anxiety associated with EFM were evident in women's narratives, and emphasized also a lack of understanding and knowledge that women had as to the technological functioning of the CTG machine; "*I thought I was going to be electrocuted. My water had broke. The cord of the machine was lying in the water*"[26(p351)]; "[I was] *worried the whole time*

that the baby's heart would stop if the machine stopped."[27(p2112)] Furthermore, labouring women experienced the CTG as a barrier to effective and personal communication; "*They all came with the machine and left with the machine,*"[27(p2112)] "*Everyone was just focused on this monitor and the heartbeat…. it was making me panic.*"[28(p401)] These narratives place emphasis on a lack of SDM associated with the use of EFM. For example, the evidence from Alfirevic's review shows that CS increases with the use of EFM compared to less technological methods of monitoring the fetal heart rate,[22] yet women are rarely informed of this in practice; they may not know how the technology works or the reason for its application in the first place—yet may accept this technological intervention on the basis of common practice, and a "doctor knows best" mentality, despite their fear and anxiety. Further emphasized in Smith et al's review[25] is the difference between women and clinicians regarding what they know or think, with the CTG acting as the conflict resolver; "*I was sure I was in labor, but the doctors didn't think so……I was glad the monitor was there to prove that I was really in labor.*"[29(p272)] This further demonstrates the point we made earlier that CTG technology, initially introduced as a clinical decision aid, has become a clinical "decider" in and of itself, providing information (a tracing of the uterine contraction pattern) that is perceived as proving what is actually happening physiologically.

In the context of considering how women might view an extension of EFM technology into DL or ML systems, it is salutary to reflect on the lack of acceptance already demonstrated by women in relation to more "modern" or "advanced" versions of EFM. For example, 11 women in Australia were asked their views of STAN (ST analysis), which combines standard CTG monitoring with simultaneous assessment of the fetal heart using ECG, with analysis of the ST wave and T wave to detect changes in the waveforms that may indicate myocardial hypoxia. Their views were cautious or negative, including "*Have they even used this before?…nah I think I would be sticking to the CTG*" and "*if it was just…everyone gets stuck to it I would probably think that it's not necessary.*"[30(p4)]

5 | ML-ENHANCED EFM AND EPISTEMIC OBLIGATION

The widespread use of the traditional CTG, despite obvious flaws, in particular the problem of inter- and intra-observer variability, which needs to be addressed if unnecessary CSs are to be avoided, is likely to continue because, as yet, there are no alternative, commonly available methods that can continuously monitor the fetal heart rate during labour. More advanced versions of the EFM technology, which employ ML and DL are, however, in development. Emerging systems have been tested with varying results. A systematic review and meta-analysis of nine studies found that inter-rater reliability between clinicians and AI interpretation of CTGs was only moderate and made no difference to neonatal outcomes.[31] More recent work gives some indication that ML[32] and DL, using an 8-layer deep convolutional neural network (CNN) framework,[33] both show higher levels of sensitivity and specificity than other modern methods.

For the sake of argument, let us consider a future DL-enhanced technology that by assumption analyses CTG charts in real time with a much higher degree of accuracy than human practitioners, thereby helping to eliminate, or at least reduce, the false positives that lead to higher intervention rates. In this case too there would seem to be an epistemic obligation to employ the technology, as we saw was argued for with regard to the general case by Bjerring & Busch.[5] These systems would nevertheless be *opaque* in that they give no explanation of the model they use to make their diagnoses. Furthermore, the model would be *underdetermined* in the sense that it is empirically equivalent to other such models, leaving open the possibility that it may not be one that truly models a portion of the real world at all.

In this case, it *seems* that we have exchanged one kind of epistemic defeater for another. We have exchanged the *transparent* but "all too human" defeaters for *opaque* and *underdetermined* defeaters. In neither case do we have knowledge, but perhaps the latter nonetheless leads to a better outcome for the patient/client, or clients on average. Are we then epistemically (not to mention ethically and clinically) obliged to rely upon DL systems in such cases (as suggested by Bjerring & Busch[5])? Does this then undermine SDM, and what emphasis should be placed on women's values in such a case?

Shaw et al[34] have suggested that, although beneficial, ML would be used mainly to *augment* rather than *replace* the work of clinicians providing health care. They believe that the current generation of ML capabilities work at the *task level*, and not at the level of conducting a complete *job*,[34] which would encompass addressing ethical, moral, legal and stakeholder standpoints. Similar reasoning has recently been employed by Di Nucci[35] in response to arguments put forward by McDougall.[36] That is, that there is a distinction between *advising* and *decision-making* and that we may quite reasonably decide to delegate an advising role to an ML system. However, this reasoning would not seem to be sufficient to address the force of the point regarding epistemic obligation. That is, we cannot simply dismiss such an obligation merely by pointing out that, in effect, it is up to us as delegator to decide what is best in the end. It should also be appreciated that, in the context of traditional CTG monitoring, the clinician might decide to do a CS, and in a short space of time, solely on the fetal heart rate pattern visible on the CTG trace with limited consideration for the entire clinical picture (gestation, parity, stage of labour, fetal blood oxygenation levels, etc.). Similarly, ML technology designed to augment or advise on clinical decisions is likely to influence the decision-maker unduly because other epistemic defeaters are at play (eg, fear of litigation), or perhaps the phenomenon of "algorithm appreciation."[37(cf8)]

McDougall has recently argued that AI systems pose a threat to patient autonomy, and thereby to SDM, through the rise of a neo-paternalistic "computer knows best" attitude that is not responsive to patient values and treatment goals.[36(cf5,n26)] Certainly in the area of EFM in labour, although an AI system might conceivably assess all the available facts and advise on a course of action, that advice does not necessarily take into account the woman's views and wishes. Often, during a very prolonged labour, clinical indications might suggest that a CS could or should be carried out. However, if the fetal heart tracing and other clinical indicators of maternal and fetal wellbeing are normal

and the woman prefers to continue in labour, there is no medical reason not to and clinicians caring for her may agree to that course of action (or may not agree, depending on the epistemic defeaters in that situation). Indeed, Bjerring & Busch ultimately answer their own challenge in a similar way, allowing that an epistemic obligation can be "overridden by other epistemic or non-epistemic factors,"[5(§1)] such as the *non-epistemic* factor of a conflict with patient/client values that they suggest.[5(§4)] In the next section, we will present an example of a kind of overriding *epistemic* factor that we shall call an *inherited defeater*.

6 | INHERITED DEFEATERS

In a general discussion of peer-disagreement in the context of medical AI, Grote & Berens[38] point out that it could be argued that "given that the algorithm is likely trained and validated on the opinions of several expert clinicians—deferring would seem like a reasonable choice, especially for a novice."[38(p207)] Indeed, there might seem to be an epistemic obligation to do so, as suggested by Bjerring & Busch.[5] However, it could be quickly responded that deferring merely to *opinions* is usually not a *reasonable* choice, and the expert opinions may not necessarily have been based on the best research evidence available.

Further, and more generally, it might not be so clear that the opacity and underdetermination of DL systems is always the underlying or ultimate source of an epistemic defeat. Since DL systems are often trained on data that has been influenced by the prior policies and actions of clinicians, that is, the training is "supervised" by them, and those clinicians are themselves influenceable and subject to biases,[39] it is likely that such epistemic defeaters are inherited by the trained DL systems, together with the combined "expertise" of those clinicians. This is a form of what is known more generally as *algorithmic bias*. In his response to McDougall[36] after relying on a mere hope, effectively an appeal to ignorance, regarding the neutrality of algorithms, Di Nucci concedes in a footnote that: "On the other hand, here we should be mindful of so-called *algorithmic bias*: algorithms are programmed by humans and we must be careful to avoid human bias being entrenched by being programmed into software."[35(p557: n.viii)] The fact that such policies and opinions have not been explicitly hard-coded ("programmed") by humans and the DL system in this case has instead been trained by data, makes no difference in this regard.

A good example of these inherited defeaters may be found in the case of DL-enhanced EFM that we discussed above. The current systems are being trained in a supervised manner to model clinicians' categorization of CTG charts.[40] Thus, they are effectively attempting to outperform clinicians in CTG assessment by relying on training data that is already known to produce a sub-optimal result for the clients involved. Switching all the machines off in this instance would often produce a better outcome with respect to the health of the client. As such, this would seem to be a deviant training case, in which epistemic defeaters are inherited through the emulation of a flawed model of maternity care. Just as human practitioners can fail to have a correct categorization of CTG readings, it is not surprising that DL systems

502 | WILEY— *Journal of* Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

BEGLEY ET AL.

can inherit a model of this categorization and prove not much more beneficial.

Moreover, there are other possible training modes to consider for such DL systems. A second possibility is that these DL systems be instead trained in an unsupervised mode, that is, by identifying features of CTG traces merely by their inherent regularities across these traces without input from clinicians. However, the resulting clusters of features would nevertheless need to be interpreted and given meaning by someone. Subjective decisions would again have to be made regarding where one cluster ends and another begins, a classic problem of vagueness. This again undoubtedly would introduce similar biases and the same or similar epistemic defeaters would be inherited by the system and subsequent decision-making.

A third possibility is a supervised mode of training in which clinicians are, for whatever reason, absent or make no intervention. In this case the DL system would be supervised merely in virtue of the recorded outcome for the mother and baby in each case, specific pathology or none. While this undoubtedly would be the best option for avoiding the defeaters already mentioned, there are perhaps other issues, complications, and defeaters to consider in such a case. First, it is uncertain whether enough quality data could or should be made available. Such data would only arise in situations in which continuous EFM was started, a CTG trace was recorded, and the clinician was absent so that labour concluded without any further intervention; a rare or unusual, if not non-existent situation. Secondly, should such a data set be produced, there would be ethical implications to consider, arising from its use. Thirdly, there is the fact that in such a scenario someone has already made the decision to start continuous EFM, for whatever reason, and that this in itself would bias the data. This could only be (partially) alleviated by some future advance toward an entirely non-invasive EFM technology. The bottom line here is that the outputs of such technology should always be put in the context of considering what would have happened if it were never used at all.

7 | CONCLUSION

Epistemic defeaters involved in SDM can be inherited by DL systems that attempt to model situations in which they are involved, or that are "supervised" by clinicians subject to such defeaters. So, it follows that the duty of care to the client's knowledge would extend to any situation in which a clinician (or anyone else) is involved in producing training data for a system that will be used in SDM, given that the system will eventually provide input to an intended SDM process. Although there are certainly inherent problems arising from opacity and underdetermination of DL systems, nevertheless such systems should not be used as black-boxes in which to obfuscate or legitimate our "all too human" problems, which are perhaps better addressed by adopting a duty of care to the client's knowledge. As with current EFM technology, we argue that DL systems may have a place in contemporary and future maternity care, yet it is where this place is and how these systems are or might be used that requires careful consideration. The excitement and awe that accompanied the introduction

of EFM, notwithstanding that EFM can provide information that leads to saving a baby's life, was replaced over time by evidence that it led to increased and unnecessary interventions in some cases, reduced trust in clinical observations in other cases, and afforded minimal consideration for the views of stakeholders on whom it was being used. Any future AI ought to have the capability of assessing women individually, taking into account a wide range of factors (smoker, diet, lifestyle, as well as the usual clinical factors of parity, gestation, medical and obstetric history, etc.) and combine these with client preferences or at least input, to provide a holistic picture when making clinical decisions. This possibility perhaps presents one of the greatest challenges to maternity care practice—a challenge about which one might wonder as to whether any AI would or could have the capability to meet.

ORCID

Keith Begley 🔟 https://orcid.org/0000-0003-1448-4344

REFERENCES

1. Begley K, Daly D, Panda S, Begley C. Shared decision-making in maternity care: acknowledging and overcoming epistemic defeaters. *J Eval Clin Pract*. 2019;25:1113-1120. https://doi.org/10.1111/jep.13243.
2. Chisholm RM. *Theory of Knowledge*. London: Prentice-Hall; 1989 [1966].
3. Gettier EL. Is justified true belief knowledge? *Analysis*. 1963;23(6): 121-123.
4. Hinton G. Deep learning—a technology with the potential to transform health care. *JAMA*. 2018;320(11):1101-1102. https://doi.org/10.1001/jama.2018.11100.
5. Bjerring JC, Busch J. Artificial intelligence and patient-centered decision-making. *Philos Technol*. 2020. https://doi.org/10.1007/s13347-019-00391-6.
6. Burrell J. How the machine 'thinks': understanding opacity in machine learning algorithms. *Big Data Soc*. 2016;3:1-12. https://doi.org/10.1177/2053951715622512.
7. Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L. *Explaining Explanations: an Overview of Interpretability of Machine Learning*. 2019. Available from https://arxiv.org/abs/1806.00069
8. Walmsley J. Artificial intelligence and the value of transparency. *AI Soc*. 2020. https://doi.org/10.1007/s00146-020-01066-z.
9. van Fraassen BC. *The Scientific Image*. New York, NY: Oxford University Press Inc.; 1980.
10. Ayres de Campos D, Spong CY, Chandraharan E. FIGO consensus guidelines on intrapartum fetal monitoring: cardiotocography. *Int J*

Gynecol Obstet. 2015;131:13-24. https://doi.org/10.1016/j.ijgo.2015.06.020.

11. Cranston CS. Obstetrical nurses' attitudes toward fetal monitoring. JOGN Nurs. 1980;9(6):344-347. https://doi.org/10.1111/j.1552-6909.1980.tb01342.x.

12. Birch L, Thompson B. Survey into fetal monitoring practices and attitudes. Br J Midwifery. 1997;5(12):732-734. https://doi.org/10.12968/bjom.1997.5.12.732.

13. Jackson JE, Vaughan M, Black P, D'Souza SW. Psychological aspects of fetal monitoring: maternal reaction to the position of the monitor and staff behaviour. J Psychosom Obstet Gynaecol. 1983;2(2):97-102. https://doi.org/10.3109/01674828309081266.

14. Filshie M. Intrapartum fetal monitoring. Br J Hosp Med. 1974;12:33-46.

15. Rhose S, Heinis AMF, Vandenbussche F, Drongelen J, Dillen J. Inter- and intra-observer agreement of non-reassuring cardiotography analysis and subsequent clinical management. Acta Obstet Gynecol Scand. 2014;93:596-602. https://doi.org/10.1111/aogs.12371.

16. Blix E, Sviggum O, Koss KS, Øian P. Inter-observer variation in assessment of 845 labour admission tests: comparison between midwives and obstetricians in the clinical setting and two experts. BJOG. 2003;110:1-5. https://doi.org/10.1046/j.1471-0528.2003.t01-1-02105.x.

17. Palomäki O, Luukkaala T, Luoto R, Tuimala R. Intrapartum cardiotocography–the dilemma of interpretational variation. J Perinat Med. 2006;34(4):298-302. https://doi.org/10.1515/JPM.2006.057.

18. Rei M, Tavarea S, Pinto P, et al. Interobserver agreement in CTG interpretation using the 2015 FIGO guidelines for intrapartum fetal monitoring. Eur J Obstet Gynecol Repro Biol. 2016;205:27-31. https://doi.org/10.1016/j.ejogrb.2016.08.017.

19. Altaf S, Oppenheimer C, Shaw R, Waugh J, Dixon-Woods M. Practice and views on fetal heart monitoring: a structured observation and interview study. BJOG. 2006;113:409-418. https://doi.org/10.1111/j.1471-0528.2006.00884.x.

20. Blix E, Ohlund L. Norwegian midwives' perception of the labour admission test. Midwifery. 2007;23(1):48-58. https://doi.org/10.1016/j.midw.2005.10.003.

21. McKevitt S, Gillen P, Sinclair M. Midwives' and doctors' attitudes towards the use of the cardiotocograph machine. Midwifery. 2011;27:e279-e285. https://doi.org/10.1016/j.midw.2010.11.003.

22. Alfirevic Z, Gyte GML, Cuthbert A, Devane D. Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour. Cochrane Database Syst Rev. 2017;2:CD006066. https://doi.org/10.1002/14651858.CD006066.pub3.

23. National Institute for Health and Care Excellence (NICE). Intrapartum care for healthy women and babies. Clinical guideline [CG190], 1.10 Monitoring in labour. Available from https://www.nice.org.uk/guidance/cg190/chapter/Recommendations#first-stage-of-labour

24. Smith V, Begley C, Clarke M, Devane D. Professionals' views of fetal monitoring during labour: a systematic review and thematic analysis. BMC Pregnancy Childbirth. 2012;12:165. https://doi.org/10.1186/1471-2393-12-166.

25. Smith V, Begley C, Devane D. Chapter 11: technology in childbirth–exploring women's views of fetal monitoring during labour: a systematic review. In: Church S, Downe S, Frith L, et al., eds. New Thinking on Improving on Improving Maternity Care: International Perspectives. London, UK: Pinter and Martin; 2017.

26. Beck C. Patient acceptance of fetal monitoring as a helpful tool. JOGN Nurs. 1980;9(6):350-353. https://doi.org/10.1111/j.1552-6909.1980.tb01343.x.

27. Shields D. Maternal reactions to fetal monitoring. Am J Nurs. 1978;78(12):2110-2112.

28. Barber V, Linsell L, Locock L, et al. Electronic fetal monitoring during labour and anxiety levels in women taking part in a RCT. Br J Midwifery. 2013;21(6):394-403. https://doi.org/10.12968/bjom.2013.21.6.394.

29. Starkman M. 1976Psychological responses to the use of the fetal monitor during labour. Pyschosom Med. 1976;38(4):269-277. https://doi.org/10.1097/00006842-197607000-00005.

30. Bryson K, Wilkinson C, Kuah S, Matthews G, Turnball D. A pilot exploratory investigation on pregnant women's views regarding Stan fetal monitoring technology. BMC Pregnancy Childbirth. 2017;17:446. https://doi.org/10.1186/s12844-017-1598-8.

31. Balayla J, Shrem G. Use of artificial intelligence (AI) in the interpretation of intrapartum fetal heart rate (FHR) tracings: a systematic review and meta-analysis. Arch Gynecol Obstet. 2019;300:7-14. https://doi.org/10.1007/s00404-019-05151-7.

32. Fergus P, Hussain A, Al-Jumeily D, Huang D-S, Bouguila N. Classification of caesarean section and normal vaginal deliveries using foetal heart rate signals and advanced machine learning algorithms. Biomed Eng Online. 2017;6(16,1):89. https://doi.org/10.1186/s12938-017-0378-z.

33. Zhao Z, Deng Y, Zhang Y, Zhang Y, Zhang X, Shao L. DeepFHR: intelligent prediction of fetal Acidemia using fetal heart rate signals based on convolutional neural network. BMC Med Inform Decis Mak. 2019;19:286. https://doi.org/10.1186/s12911-019-1007-5.

34. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. J Med Internet Res. 2019;21(7):e13659. https://doi.org/10.2196/13659.

35. Di Nucci E. Should we be afraid of medical AI? J Med Ethics. 2019;45(8):556-558. https://doi.org/10.1136/medethics-2018-105281.

36. McDougall RJ. Computer knows best? The need for value-flexibility in medical AI. J Med Ethics. 2019;45:156-160. https://doi.org/10.1136/medethics-2018-105118.

37. Logg JM, Minson JA, Moore DA. Algorithm appreciation: people prefer algorithmic to human judgment. Organ Behav Hum Decis Process. 2019;151:90-103. https://doi.org/10.1016/j.obhdp.2018.12.005.

38. Grote T, Berens P. On the ethics of algorithmic decision-making in healthcare. J Med Ethics. 2020;46:205-211. https://doi.org/10.1136/medethics-2019-105586.

39. Panda S, Begley C, Daly D. Clinicians' views of factors influencing decision-making for caesarean section: a systematic review and metasynthesis of qualitative, quantitative and mixed methods studies. PLoS ONE. 2018;13(7):e0200941. https://doi.org/10.1371/journal.pone.0200941.

40. Hoodbhoy Z, Noman M, Shafique A, Nasim A, Chowdhury D, Hasan B. Use of machine learning algorithms for prediction of fetal risk using cardiotocographic data. Int J App Basic Med Res. 2019;9:226-230. https://doi.org/10.4103/ijabmr.IJABMR_370_18.