



Breaking bad news in the era of artificial intelligence and algorithmic medicine: an exploration of disclosure and its ethical justification using the hedonic calculus

Benjamin Post^{1,2,3} · Cosmin Badea² · Aldo Faisal^{1,2,3,5} · Stephen J. Brett^{3,4}

Received: 20 June 2022 / Accepted: 12 October 2022
© The Author(s) 2022

Abstract

An appropriate ethical framework around the use of Artificial Intelligence (AI) in healthcare has become a key desirable with the increasingly widespread deployment of this technology. Advances in AI hold the promise of improving the precision of outcome prediction at the level of the individual. However, the addition of these technologies to patient–clinician interactions, as with any complex human interaction, has potential pitfalls. While physicians have always had to carefully consider the ethical background and implications of their actions, detailed deliberations around fast-moving technological progress may not have kept up. We use a common but key challenge in healthcare interactions, the disclosure of bad news (likely imminent death), to illustrate how the philosophical framework of the 'Felicific Calculus' developed in the eighteenth century by Jeremy Bentham, may have a timely quasi-quantitative application in the age of AI. We show how this ethical algorithm can be used to assess, across seven mutually exclusive and exhaustive domains, whether an AI-supported action can be morally justified.

Keywords Artificial Intelligence · AI ethics · Medical ethics · Utilitarianism · Felicific calculus · End of life

1 Introduction

A great deal of effort is currently being expended on developing risk prediction models for individuals and patient groups using a variety of approaches ranging from genomics and metabonomics through to socioeconomic phenotyping [1–6]. In the domain of healthcare, the expansion in predictive modelling research is paired with rapidly emerging concerns about the ethical use of such methods, particularly

artificial intelligence (AI) [7, 8, 9]. These include concerns around data privacy, algorithmic fairness, bias, safety, informed consent, and transparency [7, 9–12], for which the medical profession may be unprepared to navigate [12]. Accordingly, international bodies have started taking action to address concerns around medical AI and automation. Last year, in their extensive report regarding AI in healthcare, the World Health Organisation explicitly stated that ‘humans should remain in full control’ of medical decisions [13]. Article 22 of the European Union’s General Data Protection Regulation (GDPR) outlines the right of individuals not to be subject to decisions ‘based solely on automated processing’ [14] and thus decision-making is a crucial part of AI [15]. The importance of keeping humans involved in significant medical decisions needs to be carefully considered as statistical, machine learning and artificial intelligence models are increasingly aiding diagnostics, treatment decisions and outcome prediction [8].

The modelling of future life-threatening events and of death is one of the most common applications of predictive tools in healthcare, which well-encapsulates the ethical issues above, as well as the complexity of individualised prediction. Furthermore, these predictions may have significant

✉ Benjamin Post
b.post19@imperial.ac.uk

¹ Department of Bioengineering, Imperial College London, London, UK

² Department of Computing, Imperial College London, London, UK

³ UKRI Centre in AI for Healthcare, Imperial College London, London, UK

⁴ Department of Surgery and Cancer, Imperial College London, London, UK

⁵ Institute of Artificial and Human Intelligence, University of Bayreuth, Bayreuth, Germany

implications for patients and the communication of this highly sensitive medical information has its own ethical challenges. To assess whether complex actions are justified, *medical ethics* is highly integrated into healthcare practice. *Utilitarianism*, which prioritises maximising benefit to the greatest number of people, is a well-established paradigm within medical ethics that provides a tool to comprehensively consider an ethical dilemma—the *Felicific Calculus* [16]. This is a useful approach for contemplating complex medical interventions with the potential to impact many people [17, 18].

We thus set out to explore the consequences of disclosing bad news in the era of AI from a Utilitarian point of view. We provide background information on the current use of AI in healthcare, the complexity of disclosing bad news and an outline of the Felicific calculus. We move on to using this ethical tool to systematically investigate whether this disclosure can be deemed good or bad and finish by discussing the relevance of our findings in a rapidly evolving domain.

2 Background

2.1 AI predictions in healthcare

The current situation is that artificial intelligence (AI) has the potential to benefit patients using precise and individualised modelling [17–21], but new risks and complexity need to be carefully considered [13]. These AI-predictive models may be used to predict treatment outcomes, life-expectancy and progression of rehabilitation [22, 23]. Communicating these predictions will have an impact on patients, loved ones, carers, as well as healthcare workers. Importantly, complications can arise as most AI-based predictive models are trained on datasets comprised of numerous individuals' data, allowing for larger sample sizes, more training data and better population-level performance. However, when these models are applied at the *level of the individual* significant uncertainty may be carried forward [24, 25]. Therefore, how should a patient or clinician interpret a disease or outcome prediction made by a mathematical model, derived from many other peoples' healthcare data?

A complex relationship exists between the quality of a model's prediction and the ability or desire of healthcare providers to act on this knowledge. If an algorithm can accurately identify an individual at risk of a serious and preventable or treatable condition, our duty to disclose this information arguably increases. However, if diagnostic predictions are made that have poor precision, identify asymptomatic diseases or those that cannot be cured or medically acted upon, the moral imperative may be less clear.

We offer here a way forward, taking a view from medicine and artificial intelligence. We speculate that the outputs of

prediction modelling, in general, should be interpreted and presented as a *risk* or *probability* of an event occurring to an individual rather than a certainty [24–28]. Additionally, probabilistic methods and more empirical approaches such as calibration testing can provide a *confidence* (or uncertainty) around such predictions and arguably these should also be presented. Yet, every model is limited and biased by the data that were used to train it, with modelling assumptions and factors simply not captured or known at the time of development (e.g. due to persistently changing conditions). Furthermore, algorithms developed from specific patient cohorts may not translate well to populations in different parts of the world, with different demographics or baseline medical conditions [22, 25, 27–31]. These are complex concepts and not necessarily intuitive, even to experts in clinical and technical disciplines [9]. The ethical implications of this knowledge may be uncertain [125] and, hence, the appropriateness of disclosing this information may not be straightforward.

2.2 Disclosing bad news

In the healthcare context, 'bad news' can be defined as information that creates a negative view of a person's health [32] or reduces their choices in life [33, 34]. Historically, the protection of patients from potentially distressing news was regarded as reasonable and consistent with a physician's role [35]. This strongly paternalistic approach is generally no longer regarded as appropriate in contemporary medicine [36]; relatives may request withholding of information about impending death if nothing can be done to avert it—which can lead to tension in the *patient–doctor–family* relationship. In modern practice, disclosing bad news is central to the role of a medical professional [35–39] but may be considered one of the more challenging and stressful responsibilities [37]. Therefore, if a specific communication has potential negative consequences for the subject (e.g., stress induced in the healthcare professional) and the object (e.g. distress and reduced life choices for the patient, uncertainty) can it be justified?

Bad news encapsulates a plethora of scenarios and can range in significance from a delayed appointment to a terminal diagnosis. For the remainder of this manuscript, we use the term 'bad news' to refer to a hypothetical situation where impending death is a near certainty for the recipient. Finally, the disclosure of bad news can be analytically subdivided to include the *decision* to disclose and the *act* of disclosure. We choose to consider the disclosure of *bad news* to be the combined decision and action to communicate the knowledge of bad news to an individual.

2.3 The Felicific Calculus

Utilitarianism is one of the main branches of *consequentialism* within normative ethics (the philosophical discipline concerned with whether actions are morally right or wrong) [40]. Consequentialists maintain that the morality of an action is determined by its outcome and focus on the *consequences* of a moral act or set of rules [40]. In contrast to other consequentialist theories, *Utilitarianism* values the maximisation of pleasure (or happiness) for the greatest number of people. In healthcare there is a strong *consequence* heuristic, and utilitarianism has become a dominant ethical paradigm: at the public health level, this balances the importance of cost-effectiveness and maximising health benefits for the greatest number of people [17, 18]; and at an individual level explores the risks of a *theoretically* beneficial act, delivering an undesired or even catastrophic outcome. Accordingly, AI has been proposed as a potential moderator of rising healthcare costs along with having the potential to revolutionise population health [41]. As such, the use of a utilitarian ethical framework to assess the intersection between healthcare and artificial intelligence is valuable [9] and seems timely.

Put simply, Utilitarian ethicists believe that an action is only ‘good’ or ‘right’ if it is productive of the most utility (commonly interpreted as ‘happiness’ or ‘pleasure’) compared to its alternatives [16, 42]. Bentham focused his thinking on how the principle of utility could be used practically and considered ‘pleasure’ and ‘pain’ the chief considerations in evaluating happiness (hedonistic utilitarianism). Concisely, he held that the moral content of an act could be seen as a function of the balance between the pleasure and pain that it induces in the subject(s) considered. He devised the felicific (hedonic)

calculus to evaluate the balance between the degrees of pleasure and pain that a particular action may cause [16]. An action is assessed through seven different domains: *intensity, duration, certainty, propinquity, fecundity, purity and extent* (Table 1). Thus, the purpose of the felicific calculus is to assist in determining the moral status of an act.

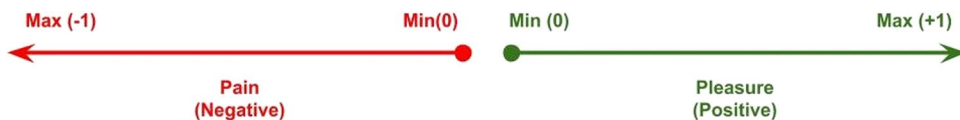
Bentham himself recognised the difficulties of implementing the felicific calculus as a practical tool and considered it most useful for the ethical deliberation of an act. The practical use of the felicific calculus was not intended to be calculative but rather to create a judgement of *good* or *bad* of the pains and pleasures of an action (immediately and subsequently). Nonetheless, attempts have been made to apply *numerical* values to the felicific calculus which has proven to be notoriously difficult [43] and, to the best of our knowledge, a practical working example has never yet been demonstrated. A graphical representation of this concept—the *hedonic scale* (See Fig. 1)—has been created for descriptive purposes. The numerical endpoints have been arbitrarily chosen to illustrate positive and negative utility values: ‘1’ representing maximum positive utility (happiness/pleasure), ‘0’ representing no utility value and ‘-1’ representing maximum negative utility (pain). At the end of each section, we declare a tentative score for each domain of the felicific calculus. This scale is mainly illustrative, to aid deliberations on some of the arguments below. The most important question, as per Bentham’s formulation of the calculus, is whether the balance between the pleasure and the pain produced by the action falls on the side of the former (> 0 on our scale), or the latter (< 0 on our scale).

The felicific calculus may not be able to provide a concrete mathematical ‘answer’ to every moral conundrum, however, it provides a useful framework for the deliberation of some philosophical scenarios. Informing an individual that they are dying (*or disclosing bad news*), specifically that their death is impending or imminent, is one such complex scenario. We recognise that a small number of people receiving such information may consider this as a *positive* experience, but for the purposes of this paper, we assume that most people will consider the receipt of this type of bad news an overall negative experience (and for the purposes of this exploration ignore the contrary response). We adopt this stance firstly because, in our opinion, the culturally accepted evaluation of finding out about one’s imminent mortality is predominantly negative (shock, dismay, sadness, grief, etc.). Moreover, one of the key responsibilities of Western medical practitioners is the preservation of human life [44]. Thus, informing a patient of the impossibility of such preservation may be seen as negative.

Table 1 The domains of the felicific calculus

Felicific variables	Definition
Intensity	The intensity of the pleasure or pain
Duration	How long the pleasure or pain will last
Certainty	The probability that the pleasure or pain will occur
Propinquity	How soon the pleasure or pain will occur
Fecundity	How likely the sensation (pleasure or pain) is to lead to more of the same sensation
Purity	How likely the sensation (pleasure or pain) is to lead to the opposite sensation
Extent	The number of people affected by the pleasure or pain

Fig. 1 Hedonic scale



In this paper, we use *the identification and disclosure of bad news* (impending death) as a test case for a utilitarian analysis in the context of cutting-edge prediction tools. We use this as a 'worked example', or step-by-step solution, for challenging ethical decision-making in modern healthcare.

3 Felicific Calculus – Intensity

Definition: The intensity of the pleasure or pain

” Do not go gentle into that good night.
Rage, rage against the dying of the light.”
- Dylan Thomas[45]

Arguably the knowledge of one’s own death (mortality salience) is one of the most defining features of the human condition. Despite the inevitability of death, many people find it difficult to discuss [44–48], particularly healthcare professionals [49], and even contemplating death leads to avoidance behaviours [50]. Of note, however, Walter argues that the discussion of death may not be as taboo as we collectively believe [51]. Learning of one’s own impending mortality may be one of the most psychologically traumatising events of a person’s life [52]. Dying is not 'easy' [53] and it is important to understand that the intensity of emotional pain may vary with time, depending on several factors [54]. Indeed, the behaviour of the agent (doctor), and the manner that news is imparted, can have a significant impact on the trajectory of this pain, but this behaviour only modulates pain and is unlikely to create pleasure [53–58].

We believe there is one important positive consequence of breaking bad news: the bestowal of *truth* upon the recipient of the news. After Bentham, George Edward Moore, an *ideal* utilitarian and critic of hedonistic utilitarianism, suggested that the ideals of truth and knowledge are as valuable as pleasure [59]. Therefore, despite truth not being explicitly covered by the felicific calculus, later proponents of utilitarian philosophy, such as Moore, did hold that the maximisation of truth is as important as the maximisation of pleasure [59]. Considering the above, the *intensity* of negativity induced by the knowledge of impending death (and potentially on the agent delivering that information) may be rated *toward* the most negative end of the pain scale (–1). We would note that there are a number of other significant life events (e.g. death of a spouse, divorce) that may induce more psychological suffering than personal illness [60]. In our clinical experience, we have met individuals who believed death would be a welcome relief from intractable chronic suffering such as pain, breathlessness, or loneliness. An interesting study in older people with multiple health conditions demonstrated that a significant number would choose to prioritise independence and pain relief *over* prolongation

of life [61]. As such, we have chosen not to assign a fully negative score to *intensity*, but one close to the end of the scale, as many individuals, whilst finding death undesirable, may find alternative burdensome states to be worse.

Felicific Score: – 0.9

4 Felicific Calculus—Duration

Definition: How long the pleasure or pain will last

Intuitively, estimating the duration of pain should be easy: *the pain lasts until death*. However, there are two major caveats to this premise. First, the grieving process is dynamic, and the intensity of grief may decrease or increase for each individual [62]. Indeed, positive feelings such as relief may even be experienced [63] and members of some religions will see death as a stepping stone to a desirable afterlife or rebirth [64]. Second, if attempts were made to minimise the *duration* of pain by delaying or avoiding the disclosure of bad news, the implications could be unpredictable. There is the possibility of 'blissful unawareness' where the patient’s life continues unperturbed, no knowledge of impending demise is experienced, and death is sudden, swift and painless. However, more likely is a negative scenario where the person suffers increasing symptomatology combined with an escalating sense of bewilderment and anxiety about their deterioration, with death prefaced only by discomfort, erosion of trust in their physician and eventual realisation. Accurately predicting survival in terminally ill patients is notoriously difficult [62]. Herein lies one, if not the most, significant of the utilitarian dilemmas—it is impossible to know, fully and accurately, the consequence of any action [65].

The stochastic nature of this probability state makes the assignment of a value on the hedonic scale difficult. Rule utilitarianism is one of the main branches of utilitarianism [66], and a rule utilitarian may take the view that one should aim to make the decision that would usually result in the greatest happiness. However, dying itself is a uniquely individual event that can only be experienced by the one dying, leaving aside for the moment secondary consequences on family and other loved ones. In our clinical experience, its trajectory, course, and duration are personal and unpredictable and thus there is no *usual* dying experience. Furthermore, the durations of the potential pain produced by the anxiety of approaching death and the potential pleasure produced by knowing the truth are both until death. As such we are unable to assign a hedonic value to *duration* in this context and must assume equipoise.

Felicific Score: 0

5 Felicitic Calculus—*Certainty*

Definition: The probability that the pleasure or pain will occur

Death is certain [67], but the accurate prediction of time to death (i.e., life expectancy) is difficult, even in those with the most extreme illness [66–71]. Under these circumstances, there are three distinct aspects to *certainty*: the certainty of a prediction (mathematical accuracy, precision, or error), the certainty of the expected outcome (the probability of the outcome occurring) and the certainty that this outcome will lead to pain or pleasure. In this section, we do not discuss the certainty of death (we presume death is imminent) but instead focus on the certainty of pleasure or pain emerging from an individual being confronted with the *knowledge* of their impending death, in accordance with Bentham's method.

The receipt of this sort of bad news is generally associated with a 'grief' response; the disclosure itself may also cause significant discomfiture for the person imparting the news. Thus, the default position is to assume this action is likely to result in pain. However, it is important to consider that more positive emotions, such as *relief*, may also occur [63, 72]. Indeed, positive and negative emotional reactions may be experienced in succession or concurrently [63]. Utilitarians remain in debate as to whether death is positive (e.g., removal of harm, end suffering, etc.) or negative (e.g., prevention of pleasures that would have otherwise been experienced) [71–75].

In practice, as with many human-centred theoretical concepts, matters are complicated. For any given individual, the measurement of utility can be difficult and biased. For example, in terminal illness, the desire to hasten death is not uncommon [76] but this desire can be confounded by several (potentially dynamic) factors, such as effectiveness of symptom control, mental illness and the dying individual's perceived burden on their family members [77, 78]. Physical pain may be prevalent towards the end of life [79] and the removal of this pain may be seen as 'positive utility'. However, this sensation is entangled with a range of other considerations (e.g., an individual's dignity at the end of life, their current financial affairs, the impact on their family, legal issues e.g., wills, etc.) [80]. For any given individual, each of these considerations may be generating positive or negative utility and each factor can be intercorrelated.

On rare occasions, a terminal diagnosis may be made in error [81]. One may think that realising a life-limiting illness has been incorrectly diagnosed would result in relief, celebration, and happiness. However, it can have significant negative ramifications including financial [80–84] and psychological [85, 86]. This suggests that an individual's

response to any life-changing news is deeply personal and based on a plethora of observable and unobservable factors.

Finally, we need to consider how *certainty* is affected by the statistical context of bad news; a terminal diagnosis can be communicated with varying levels of caveats. For example, major trauma can inflict injuries that are not compatible with life e.g., catastrophic brain injuries, so a prediction of imminent death can be accompanied by a high level of certainty. However, many terminal diagnoses (e.g., cancer) may be accompanied by a chance of survival within a given time. For instance, a '1-year survival rate of 50%' can best be understood as: 'Historical data suggest that half of people with this diagnosis will still be alive at 1 year'. However, for an individual receiving this news, it provides no personalised insight into their life-expectancy. The interpretation of this information will be wholly dependent on the individual—statistical knowledge, personality, outlook, individual experiences, social circumstances, etc. For instance, if a person is diagnosed with pancreatic cancer, their 5-year survival may be as low as 5% [87]. If a patient receives this information any of the following conclusions may be reached and none of them are incorrect:

- “I'm a fighter. I'll get through this and show the doctors I can beat these numbers. I'm going to be in the 5%.”
- “I've never been lucky; I'd be surprised if I survived more than one year.”
- “Wait, so I have a 95% chance of dying within 5 years? But that could be right now or in 5 years' time! How is this information helpful to me?”

Additionally, every non-deterministic model will have accompanying error rates associated with its predictions. These statistical caveats need to be considered when disclosing bad news, as for example, high false positive rates can lead to unnecessary psychological distress [88, 89]. Furthermore, statistical certainty can impact the certainty of pleasure or pain occurring as inaccurate predictions impact individuals differently [90]. At some point, the precision (or lack of it) of a 'prediction' must have an impact on the balance between whether an act of disclosure is a priori beneficial or harmful- does it induce *appropriate* or ultimately *inappropriate* pain.

Considering this statistical uncertainty and since neither the clinician nor the patient can have perfect statistical acumen, how can we rate certainty? Note that the 'certainty' we refer to in this section is not the certainty of death, but rather the certainty of the patient's positive/negative experience following the disclosure of bad news. Ultimately, as argued above, we believe the patient's experience will depend on their interpretation of the numbers as well as the numbers themselves. While the downstream utilitarian value is arguably indeterminable, we suggest that the probability

of pain being induced by the receipt of bad news is higher than the probability of pleasure. Thus, the felicific score is most likely to be negative, but both potential outcomes are feasible and may occur in the same individual.

Felicific Score: ≤ 0 (?)

6 Felicific Calculus – Propinquity

Definition: How soon the pleasure or pain will occur

As seen above, predicting the timeliness of death is challenging. Healthcare professionals can be asked to predict death on a continuous scale from minutes to years with varying degrees of informative patient-specific knowledge. The communication of medical information, especially if caveated by statistical parameters, is problematic and often leads to confusion [89–93]. As such, some individuals, when presented with an estimated survival time, may misunderstand this news and its implications. This added complexity can affect the course of grieving and impact the magnitude, duration and onset of pain or pleasure.

An established grieving paradigm, developed by Kübler-Ross in 1969, outlines stages of grieving that a patient may experience [94]. Kübler-Ross suggests that most people follow a linear journey through 5 stages of grieving: denial, anger, bargaining, depression, and, finally, acceptance. From a utilitarian point of view, this would suggest that in the early stages of grieving, pain and suffering occurs early with the potential for pleasure to occur later. Despite this model of grieving being rejected by modern psychologists [95], intuitively, our deduction that pain is likely to have high propinquity (occurring sooner) and pleasure likely to have a low propinquity (occurring later) after receiving bad news, seems reasonable.

For some, the goal of accurately predicting the point at which a person will die may seem perverse. Could not the benefits of this knowledge only serve to inform administrative tasks such as financial planning, provision of healthcare and resource allocation? Intuitively, the dying person's spiritual, psychological and emotional needs should take priority and it seems feasible that the knowledge of one's death may not serve this aim. We suggest that an individualised approach be adopted, allowing for an individual's preferences to be considered before a time-to-event prediction is made.

In summary, as with our conclusions in Sect. 4, the *propinquity* of a given hedonic sensation can be highly variable. However, we suggest that any suffering is likely to occur soon after the receipt of bad news, whilst pleasure, if it does occur, is much more likely to be delayed.

Felicific Score: ≤ 0 (?)

7 Felicific Calculus—Fecundity and Purity

Definitions:

Fecundity: How likely the sensation (pleasure or pain) is to lead to more of the same sensation.

Purity: How likely the sensation (pleasure or pain) is to lead to the opposite sensation

Once an individual has learned of their impending death, the resulting pain can be persistent and may even increase in severity [76, 96, 97]. However, pleasure can also be experienced during a well-managed end of life event [62, 63] and may, to some extent, be within the person's control [98, 99]. A person's dying experience is individual, with the possibility of experiencing both positive and negative emotions.

The concepts of *fecundity* and *purity* describe how the experience of one pleasurable or painful sensation impacts the likelihood of that same sensation (or the opposite sensation) occurring in the future. We feel that under these circumstances painful sensations are unlikely to lead to pleasure (and vice versa). Overall, it would be difficult to argue that the range of symptoms and emotions that are normally experienced during the dying process are positive or enviable, particularly as these factors tend to deteriorate over time requiring escalating medical intervention [100, 101]. As such, from a *fecundity and purity* perspective the assignment of a number to the hedonic scale must be negative and is likely to be approaching -1 .

Felicific Score: -1

8 Felicific Calculus—Extent

Definition: The number of people affected by the pleasure or pain

Humans are arguably one of the most successful species on Earth and this is likely, in part, due to our complex social and emotional connectivity and ability to cooperate [102, 103]. Accordingly, events that occur to one individual can have a ripple-like effect on other people, with those closest to that individual experiencing the greatest secondary effects. The dying process has dramatic and unpredictable implications for an individual and few other human events can cause such extensive, rapid and observable downstream effects. The impact on a dying person's family is profound [104]. The loss of a family member is one of the most stressful life events [60] and the resulting grief can increase after death [96, 105]. In fact, in addition to the emotional implications, bereavement can have financial implications [106], physical and psychological implications [107] and even increase the mortality rate in those affected [105–109]. Interestingly, a number of studies have shown that, when compared to an *expected* death, the duration and intensity of

grieving is worse in those who have *unexpectedly* lost a family member [110, 111]. This is a strong argument in favour of judging the breaking of bad news as consequentially good, as this action may directly reduce suffering. Put another way, while it is hard to argue that the disclosure of the imminent death of a loved one will deliver pleasure, removal of the shock of 'sudden death' can mitigate subsequent pain.

The wider impact of death also extends beyond those immediately affected. There is a significant fiscal impact to society [107, 112]. Healthcare utilisation and length of hospital stay increases [107] and these effects may be seen for years [113]. In fact, the people who deliver healthcare may suffer when delivering bad news [112–116] and even researchers can be affected by interacting with death [117]. Bereavement has a significant and long-lasting impact on society, but we know that this can be ameliorated, at least in part, by prior understanding and involvement of loved ones [118, 119]. The evidence outlined above suggests that utility to a patient's family and wider society may be maximised if an individual and those around them know that life is drawing to a close. In this context, *extent* should be assigned a highly positive value on the hedonic scale (approaching + 1). To clarify, this is not because this action itself implies pleasure, but rather because aiming for less *total* pain demands this action, our argument being a proof by contrapositive: *not* breaking bad news would result in much more pain. A similar example would be the prescription of a medication or vaccination with an unpleasant immediate side effect to prevent a catastrophic disease; occasionally, it has been argued that a justification for vaccination is that it delivers a very wide societal benefit. The action induces short-term pain but is directly responsible for avoiding more significant pain in the future.

In truth, assigning *extent* a hedonic score of + 1 may be a gross underestimate and arguably should be scaled to account for all the people who have avoided suffering. From a consequentialist point of view, the (potential) suffering of the dying individual is outweighed by the amelioration of pain experienced by others. As discussed in Sect. 2, this action has the secondary benefit of propagating truth across many individuals and institutions. Here, we can see the difficulties with using a scoring system such as ours, as intuitively we would opt to assign *extent* a large positive number, proportional to the number of people impacted. In algorithmic complexity terms, this could be considered of the order $O(n)$, where the previous domains would have been $O(1)$. As such, we assert that the positive numerical contribution from *extent's* score should overshadow any negative scores from the first 6 domains due to this qualitative difference, decisively tilting the scales towards the positive side of our evaluation of the rightness of breaking bad news. Thus, we can finally conclude that this act can indeed be justifiable under this normative framework.

It is important to note that it can be inelegant and precarious to justify the suffering of one individual (or group) to maximise the pleasure—or minimise the distress—of another (e.g., Robert Nozick's *Utility Monster* thought experiment [120]) and we maintain that minimising the dying individual's suffering must always be prioritised. Nonetheless, as we have shown herein, the use of a hedonistic framework can be used to demonstrate that breaking bad news is a justifiable act, based on the overwhelming positive impact observed through the lens of *extent*.

Felicific Score: $1 * [O(n)]$

9 Discussion

In this example, we have used Bentham's felicific calculus to demonstrate that disclosing the bad news of impending death clearly has negative and painful implications for the target individual, but this action maximises good to *society* by attenuating downstream suffering of others. In addition, the rapidly evolving world of AI, when used for prediction of death (or some similarly adverse or unpleasant event), offers new challenges with potentially unpredictable and harmful effects. The felicific calculus provides one potential framework for deliberating ethical challenges but has numerous practical drawbacks that may limit its use for more complex or uncertain scenarios. In addition, as shown in previous work [18], a utilitarian approach may be useful for healthcare decision-making at a population-level, but inelegant when applied to the individual.

Digital healthcare and medical prediction must always remain patient-centred, and this can only be achieved with trust. As demonstrated during the COVID-19 pandemic, loss of trust when conveying scientific information may lead to counterproductive outcomes or even harm [121]. All individuals associated with the development and communication of healthcare prediction should demonstrate trustworthiness and this is best achieved with 'expertise, honesty and good intentions' [122]. Specific consideration must be given to the way in which we communicate scientific knowledge so as not to mislead our target audience or deprive them of crucial desired additional information [88, 122]. Communicating uncertainty about our knowledge and predictions is vital and can decidedly improve patient trust [90, 120–124]. The creation of ground-breaking, complex, and highly accurate AI is futile if the target beneficiaries do not trust its integration in their medical care or comprehend its value and limitations. The reliance on symbol-based media for communication may be error-prone and interpretation-based, especially in the context of AI, where trust very much revolves around such interpretations of language (see the so-called *Interpretation problem*) [125].

Members of the artificial intelligence community are acutely aware of the need to tackle growing concerns over trust and the ethical implications of widespread adoption of this fast-moving technology [126]. However, some have argued that AI itself may be able to resolve its own trust concerns; AI-generated characters have been successfully developed to support education and wellbeing [127]. Steps have even been taken to test AI in highly complex medical communication, with one group deploying virtual agents for end-of-life planning [128]. The conversations with the AI agent were very well-received and the study group of 44 older adults were even comfortable discussing their spiritual preferences with the virtual agent. For some patients, a virtual agent may even be preferred to a real person under some circumstances. In their 2014 paper, Lucas and colleagues demonstrated that some individuals felt more comfortable discussing sensitive issues with a virtual agent, in particular because they felt their responses were not being judged [129]. Whilst much more research needs to be done, early studies suggest that human–AI interactions may be more welcome than might be anticipated.

The pathways for approval of AI for human-use in health-care is governed by the regulatory category of *Software as a Medical Device* [130]. The most important goal of this process is to minimize use-related hazards and risks so these devices may be used safely and effectively. A key component of this process is *human factors engineering* which includes clarity and precision of communication involving the device. Here, we have shown an approach of how to consider nuanced ethical factors, influencing human interactions with medical AI devices, using the example of predicting and disclosing *bad news*.

9.1 Limitations, wider applicability and future work

9.1.1 Quantitative vs qualitative

The assignment of numerical values to a moral or ethical decision rapidly increases in complexity, even if only arbitrarily quantitative once the impact beyond the central individual is considered. With regards to our chosen scoring system, we are limited to an interval of -1 to 1 and thus, cannot capture numerically the expressive power associated with *extent*. This point demonstrates the limitations of numerical analysis for such decisions (compared with qualitative analysis)—a common desire in automating decision-making system [40].

9.1.2 Plurality through a coefficient matrix

A more comprehensive use of the hedonic scale could include a matrix of hedonic scores for all people affected

by an event, with an accompanying coefficient matrix to account for moderator variables such as their 'distance' from an event, numbers of people and individual ages. This would allow a utilitarian value judgement to be assigned to the many people impacted by a dying person and then scaled according to the direct emotional impact to those people the duration of benefit (e.g., younger people may have more suffering ameliorated). However, it could be countered that a coefficient matrix unnecessarily complicates the process as all people will eventually die. Thus, this serves as a natural normalising effect over the course of a person's existence. Regardless, the collection and aggregation of all these values would be extremely complicated and not likely to be available at the point of decision to disclose [40], highlighting the difficulties of practically implementing the felicific calculus.

9.1.3 Probabilistic considerations

We chose to apply the hedonic calculus to a specific set of hypothetical scenarios where two key factors were assumed to be *certain*: the pleasure or harm occurring *and* death occurring. However, this does not reflect reality where the prediction of death is never guaranteed and is accompanied by a level of probability. The same is true for the probability of an outcome (pleasure/harm) occurring. At best, our understanding of future events can only be assigned a probability with an associated distribution of uncertainty. As a result, two layers of (un)certainly then arise: first, an *epistemic* probability of a future patient outcome (death in this case, but other 'non-death' outcomes clearly exist) occurring and second, a *consequential* probability of a particular pain or pleasure occurring on learning of this future event. To further complicate matters, these two uncertainties are inherently linked, as the probability and certainty of dying will certainly affect the probability and certainty of harm or pleasure occurring. Importantly, in many circumstances the balance of benefit and harm may be difficult to anticipate as a *probability* clearly acknowledges that a future event may not occur; the distress around receipt of an ultimately incorrect prediction delivers will, however, be harmful ('pain' only). At a population level, prediction-failure (bad to the individual) must be balanced against prediction success (maybe good overall). This is not a problem unique to AI-based approaches [15], but many AI-based researchers have an intimate familiarity with probability-based models which acknowledge uncertainty and missing data throughout. Future work should consider how the inherent uncertainty, associated with event prediction, can be accounted for in philosophical analyses.

9.1.4 Non-binary prediction outcomes, personalisation and the individual

Our analysis has considered the implications of bad news where the patient is told they may die, thus the outcome considered is binary—imminent death versus prolonged survival. Other types of bad news (e.g., the diagnosis of a new severe medical condition or a likelihood of a deterioration in health or functional status) would have nonbinary outcome measures that would increase the complexity of a calculus-based ethical analysis. However, it is much harder to envisage, a priori, how non-disclosure would be feasible or justifiable under many such circumstances. Individual patient-level preference would need consideration as each person attributes different importance to aspects of their life (e.g., access to their family, independence, mobility). Of note, many standardised quality of life assessments such as the short form 36 (SF-36) [131] and EuroQol-5 Dimension (Eq-5D) [132] instruments do not allow individuals to assign an *importance* to each quality-of-life domain or even identify unspecified 'domains' which hold importance to them.

As discussed above in Sect. 4, *Certainty* plays a complicated role in our hedonic calculations. When we consider outcomes other than imminent death, the role of *certainty* is likely to change. For example, disclosing an imperfect prediction of a diagnosis with profound consequences could be catastrophic to an individual. Recent magnetic resonance imaging (MRI) research from the University of Singapore showed that Alzheimer's disease can be diagnosed 4 years before the onset of symptoms, with an accuracy of approximately 80% [133]. For an individual given this information, they would receive a prediction about a disease, with a margin of error, for a disease several years away. In addition, at this point in time, no therapeutic options would be available. Thus, predicting an undesirable future state that is unactionable and uncertain would need careful ethical consideration.

Individuals may choose not to share their bad news with loved ones or may be solitary people without any close personal relationships [134]. Under these circumstances our conclusion, that disclosing bad news is justifiable due to the importance of *extent*, could be challenged. The prevalence of this phenomenon is not well-researched, however, our personal experience as clinicians is that this is uncommon, particularly as *dying alone* is a key fear around death for most people [135]. As such, at a population level, this would only apply to a small percentage of individuals.

Given the current acceleration in the development of individual outcome prediction models based on biomarkers, genomic attributes and 'big data' derivations, we feel there is a need for further exploration in this domain and future work should consider the real-world application and assessment of the Felicific Calculus or analogous frameworks for breaking bad news. In addition to its practical feasibility, a

patient-centred and clinician-centred assessment would need to be undertaken to ensure positive clinical impacts. Note that individualised health (including prediction) is a major focus of health policy and AI development internationally [8, 9, 136].

9.1.5 Alternative moral paradigms

The felicific calculus itself needs interrogation to assess its validity as a framework to judge clinical ethical queries. Other philosophical and ethical systems (e.g., deontology, virtue ethics [40]) should be used to appraise the ethical justification for breaking uncertain bad news. Western medical ethics have conventionally been based on Principlism (autonomy, non-maleficence, beneficence, justice) [137]. However, this approach is not always well received, especially in cultures where the *family* is the central unit of identity rather than the *individual* [138, 139]. Instead, *Narrative Ethics* is increasingly being discussed as a more nuanced approach to breaking bad news [140, 141]. Its central tenet being that every ethical dilemma is unique and thus a set of universal ethical rules or principles cannot be applicable to every situation [142]. As we move to more globalised medicine, particularly as technology and information can be rapidly shared internationally, our approach to bioethics may need to change. Non-western countries and cultures can have different foundational moral paradigms, from which locally applicable approaches to bioethical questions arise [143]. Indeed, this subject can be highly complex as ethical paradigms may arise from completely different causal interpretation of disease, healthcare priorities, and cultural preferences [141–145]. However, despite different foundations, there can be significant overlap in moral conclusions between western and non-western philosophies [145].

10 Conclusion

We have used the felicific calculus as a framework to discuss the utilitarian implications of telling someone they are dying. Communicating bad news will always be challenging and the decision to disclose this information should always be made after careful consideration of each individual and the potential impact on others. The first six elements of the felicific calculus focus on a detailed description and analysis of the pain or pleasure of the target *individual*. As demonstrated above, for each of these *hedonic* elements, our arguments along with knowledge gained from the existing literature indicate that breaking bad news either induces suffering (intensity, fecundity, purity) or an unpredictable outcome (certainty, duration, propinquity). If we only considered the individual, using these six elements in isolation, the action of telling

someone they are dying may not be justifiable under a hedonistic framework. However, once the seventh and final element of the felicific calculus, *extent*, is accounted for, we could reach a different conclusion:

Breaking bad news can be viewed as a *good* act when other people and wider society are considered, in addition to the individual, due to the powerful impact of *extent*.

We hope that our analysis will both inform future developments in ethical AI and serve as a proof-of-concept for solving other ethical conundrums in healthcare.

Funding BP was supported by the UKRI CDT in AI for Healthcare <http://ai4health.io> (Grant No. P/S023283/1). AAF holds a UKRI Turing AI Fellowship (Grant No. EP/V025449/1). This research was also funded by the NIHR Imperial Biomedical Research Centre (BRC), and by Research Capability Funding. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. For the purpose of open access, the author has applied a 'Creative Commons Attribution (CC BY)' licence to any Author Accepted Manuscript (AAM) version arising.

Declarations

Conflict of interest None of the authors have any conflicts of interest to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Kulmanov, M., Hoehndorf, R.: DeepPheno: predicting single gene loss-of-function phenotypes using an ontology-aware hierarchical classifier. *PLoS Comput. Biol.* **16**(11), e1008453 (2020). <https://doi.org/10.1371/journal.pcbi.1008453>
- Lello, L., Raben, T.G., Yong, S.Y., Tellier, L.C.A.M., Hsu, S.D.H.: 'Genomic prediction of 16 complex disease risks including heart attack diabetes, breast and prostate cancer.' *Sci. Rep.* **9**(1), 15286 (2019). <https://doi.org/10.1038/s41598-019-51258-x>
- Deelen, J., et al.: A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. *Nat. Commun.* **10**(1), 3346 (2019). <https://doi.org/10.1038/s41467-019-11311-9>
- Antcliffe, D., Gordon, A.C.: Metabonomics and intensive care. *Crit. Care* **20**, 68 (2016). <https://doi.org/10.1186/s13054-016-1222-8>
- Hill, W.D., et al.: Genome-wide analysis identifies molecular systems and 149 genetic loci associated with income. *Nat. Commun.* **10**(1), 5741 (2019). <https://doi.org/10.1038/s41467-019-13585-5>
- von Stumm, S., et al.: Predicting educational achievement from genomic measures and socioeconomic status. *Dev. Sci.* **23**(3), e12925 (2020). <https://doi.org/10.1111/desc.12925>
- Murphy, K., et al.: Artificial intelligence for good health: a scoping review of the ethics literature. *BMC Med. Ethics* **22**(1), 14 (2021). <https://doi.org/10.1186/s12910-021-00577-8>
- Hindocha, S., Badea, C.: Moral exemplars for the virtuous machine: the clinician's role in ethical artificial intelligence for healthcare. *AI Ethics* (2021). <https://doi.org/10.1007/s43681-021-00089-6>
- Bolton, W.J., et al.: Developing moral AI to support antimicrobial decision making. *Nature Machine Intelligence* (accepted). (2022). Preprint available at <https://doi.org/10.48550/arXiv.2208.06327>
- Martinez-Martin, N., et al.: Ethical issues in using ambient intelligence in health-care settings. *Lancet Dig. Health* **3**(2), e115–e123 (2021). [https://doi.org/10.1016/S2589-7500\(20\)30275-2](https://doi.org/10.1016/S2589-7500(20)30275-2)
- Gerke, S., Minssen, T., Cohen, G.: Ethical and legal challenges of artificial intelligence-driven healthcare. *Artif. Intell. Healthc.* (2020). <https://doi.org/10.1016/B978-0-12-818438-7.00012-5>
- Rigby, M.J.: Ethical dimensions of using artificial intelligence in health care. *AMA J. Ethics* **21**(2), 121–124 (2019). <https://doi.org/10.1001/amajethics.2019.121>
- 'Ethics and governance of artificial intelligence for health'. <https://www.who.int/publications-detail-redirect/9789240029200> Accessed 06 Oct 2021
- Regulation (EU) 2016/679 of the EU Parliament and of the Council of 27 April 2016 (General Data Protection Regulation (GDPR))
- Badea, C., Gilpin, L.: Establishing Meta-Decision-Making for AI: An Ontology of Relevance, Representation and Reasoning. In *AAAI 2021 Fall Symposium FSS-21*. (2021). Also available at: <https://doi.org/10.48550/arXiv.2210.00608>.
- Sinnott-Armstrong, W.: 'Consequentialism', in *The Stanford Encyclopedia of Philosophy*, Summer 2019., E. N. Zalta, Ed. Metaphysics Research Lab, Stanford University, 2019. 15 Accessed Feb 2020. <https://plato.stanford.edu/archives/sum2019/entries/consequentialism/>
- Marseille, E., Kahn, J.G.: Utilitarianism and the ethical foundations of cost-effectiveness analysis in resource allocation for global health. *Philos. Ethics Humanit. Med.* **14**(1), 5 (2019). <https://doi.org/10.1186/s13010-019-0074-7>
- Felzmann, H.: Utilitarianism as an approach to ethical decision making in health care. In: Scott, P.A. (ed.) *Key concepts and issues in nursing ethics*, pp. 29–41. Springer International Publishing, Cham (2017)
- Elfiky, A.A., Pany, M.J., Parikh, R.B., Obermeyer, Z.: Development and application of a machine learning approach to assess short-term mortality risk among patients with cancer starting chemotherapy. *JAMA Netw. Open* **1**(3), e180926 (2018). <https://doi.org/10.1001/jamanetworkopen.2018.0926>
- Rajkomar, A., Dean, J., Kohane, I.: Machine learning in medicine. *N. Engl. J. Med.* **380**(14), 1347–1358 (2019). <https://doi.org/10.1056/NEJMr1814259>
- Avati, A., Jung, K., Harman, S., Downing, L., Ng, A., Shah, N.H.: Improving palliative care with deep learning. *BMC Med. Inform. Decis. Mak.* **18**(Suppl 4), 122 (2018). <https://doi.org/10.1186/s12911-018-0677-8>
- Morgenstern, J.D., et al.: Predicting population health with machine learning: a scoping review. *BMJ Open* **10**(10), e037860 (2020). <https://doi.org/10.1136/bmjopen-2020-037860>
- O. Lambercy *et al.*, 'Neurorehabilitation From a Distance: Can Intelligent Technology Support Decentralized Access to Quality Therapy?', *Frontiers in Robotics and AI*, 2022. Doi: <https://doi.org/10.3389/frobt.2021.612415>

24. Pate, A., Emsley, R., Ashcroft, D.M., Brown, B., van Staa, T.: The uncertainty with using risk prediction models for individual decision making: an exemplar cohort study examining the prediction of cardiovascular disease in English primary care. *BMC Med.* **17**(1), 134 (2019). <https://doi.org/10.1186/s12916-019-1368-8>
25. Gianfrancesco, M.A., Tamang, S., Yazdany, J., Schmajuk, G.: Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA Intern Med* **178**(11), 1544–1547 (2018). <https://doi.org/10.1001/jamainternmed.2018.3763>
26. Trevena, L.J., et al.: Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. *BMC Med. Inform. Decis. Mak.* **13**(2), S7 (2013). <https://doi.org/10.1186/1472-6947-13-S2-S7>
27. Naik, G., Ahmed, H., Edwards, A.G.: Communicating risk to patients and the public. *Br J Gen Pract* **62**(597), 213–216 (2012). <https://doi.org/10.3399/bjgp12X636236>
28. Y. Li, M. Sperrin, M. Belmonte, A. Pate, D. M. Ashcroft, and T. P. van Staa, ‘Do population-level risk prediction models that use routinely collected health data reliably predict individual risks?’, *Sci Rep*, vol. 9, no. 1, Art. no. 1, Aug. 2019, doi: <https://doi.org/10.1038/s41598-019-47712-5>.
29. Parikh, R.B., Teeple, S., Navathe, A.S.: Addressing Bias in Artificial Intelligence in Health Care. *JAMA* **322**(24), 2377–2378 (2019). <https://doi.org/10.1001/jama.2019.18058>
30. K. N. Vokinger, S. Feuerriegel, and A. S. Kesselheim, ‘Mitigating bias in machine learning for medicine’, *Commun Med*, vol. 1, no. 1, Art. no. 1, Aug. 2021, doi: <https://doi.org/10.1038/s43856-021-00028-w>.
31. Swami, N., Corti, C., Curigliano, G., Celi, L.A., Dee, E.C.: Exploring biases in predictive modelling across diverse populations. *The Lancet Healthy Longevity* **3**(2), e88 (2022). [https://doi.org/10.1016/S2666-7568\(21\)00307-X](https://doi.org/10.1016/S2666-7568(21)00307-X)
32. Buckman, R.: Breaking bad news: why is it still so difficult? *Br Med J (Clin Res Ed)* **288**(6430), 1597–1599 (1984)
33. Bor, R., Miller, R., Goldman, E., Scher, I.: The meaning of bad news in HIV disease: Counselling about dreaded issues revisited. *Couns. Psychol. Q.* **6**(1), 69–80 (1993). <https://doi.org/10.1080/09515079308254494>
34. Sparks, L., Villagran, M.M., Parker-Raley, J., Cunningham, C.B.: A Patient-centered Approach to Breaking Bad News: Communication Guidelines for Health Care Providers. *J. Appl. Commun. Res.* **35**(2), 177–196 (2007). <https://doi.org/10.1080/00909880701262997>
35. Weiss, G.B.: Paternalism modernised. *J Med Ethics* **11**(4), 184–187 (1985). <https://doi.org/10.1136/jme.11.4.184>
36. J. F. Childress, *Paternalism in Healthcare and Health Policy*. Oxford University Press, pp. 38–52. Accessed: Jan. 26, 2022. [Online]. Available: <https://oxfordmedicine.com/view/https://doi.org/10.1093/med/9780199798483.001.0001/med-9780199798483-chapter-3>
37. Fallowfield, L., Jenkins, V.: Communicating sad, bad, and difficult news in medicine. *The Lancet* **363**(9405), 312–319 (2004). [https://doi.org/10.1016/S0140-6736\(03\)15392-5](https://doi.org/10.1016/S0140-6736(03)15392-5)
38. Schildmann, J., Cushing, A., Doyal, L., Vollmann, J.: Breaking bad news: experiences, views and difficulties of pre-registration house officers. *Palliat Med* **19**(2), 93–98 (2005). <https://doi.org/10.1191/0269216305pm9960a>
39. J. Payne, ‘Breaking Bad News’, *patient.info*, Nov. 24, 2014. <https://patient.info/doctor/breaking-bad-news> (accessed Dec. 30, 2020).
40. C. Badea, ‘Have a break from making decisions, have a MARS: The Multi-valued Action Reasoning System’, *arXiv:2109.03283 [cs]*, Sep. 2021, Accessed: Nov. 09, 2021. [Online]. Available: <http://arxiv.org/abs/2109.03283>
41. Bajwa, J., Munir, U., Nori, A., Williams, B.: Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthc J* **8**(2), e188–e194 (2021). <https://doi.org/10.7861/fhj.2021-0095>
42. J. Bentham, *An Introduction to the Principles of Morals and Legislation*. 1780. Accessed: Feb. 18, 2020. [Online]. Available: <https://www.bookdepository.com/Introduction-Principles-Morals-Legislation-Jeremy-Bentham/9780486454528>
43. R. Rzepka and K. Araki, ‘Possible Usage of Sentiment Analysis for Calculating Vectors of Felific Calculus’, in *2013 IEEE 13th International Conference on Data Mining Workshops*, Dec. 2013, pp. 967–970. doi: <https://doi.org/10.1109/ICDMW.2013.70>.
44. GMC, ‘Treatment and care towards the end of life: good practice in decision making’. May 20, 2010. Accessed: Mar. 08, 2021. [Online]. Available: https://www.gmc-uk.org/-/media/documents/treatment-and-care-towards-the-end-of-life---english-1015_pdf-48902105.pdf?la=en&hash=41EF651C76FDBEC141FB674C08261661BDEFD004
45. Thomas, Dylan, *Do not go gentle into that good night*. New Directions, 1952.
46. Cancer Research UK, ‘Talking about dying | Dying with cancer | Cancer Research UK’. <https://www.cancerresearchuk.org/about-cancer/coping/dying-with-cancer/coping-with-the-news/talking-about-dying> (accessed May 08, 2021).
47. American Psychological Association, ‘Speaking of Psychology: Making talking about death easier’, <https://www.apa.org>. <https://www.apa.org/research/action/speaking-of-psychology/talking-death> (accessed May 08, 2021).
48. Age UK, ‘Why we should all be encouraged to talk about death and dying’, *Age UK*. <https://www.ageuk.org.uk/discover/2018/why-we-should-all-be-encouraged-to-talk-about-death-and-dying/> (accessed May 08, 2021).
49. Royal College of Physicians, ‘Talking about dying: How to begin honest conversations about what lies ahead.’ RCP, Oct. 2018. [Online]. Available: <http://www.rcplondon.ac.uk/talking-about-dying>
50. J. Greenberg, S. L. Koole, and T. A. Pyszczynski, *Handbook of Experimental Existential Psychology*. Guilford Press, 2004.
51. Walter, T.: Modern Death: Taboo or not Taboo? *Sociology* **25**(2), 293–310 (1991). <https://doi.org/10.1177/0038038591025002009>
52. Coyle, N.: The existential slap – a crisis of disclosure. *Int J Palliat Nurs* **10**(11), 520–520 (2004). <https://doi.org/10.12968/ijpn.2004.10.11.17130>
53. Coyle, N.: The Hard Work of Living in the Face of Death. *J. Pain Symptom Manage.* **32**(3), 266–274 (2006). <https://doi.org/10.1016/j.jpainsymman.2006.04.003>
54. C. Jones, ‘Right to die: There were times when I would have ended my life if it’, *The Independent*, Jul. 13, 2014. <http://www.independent.co.uk/life-style/health-and-families/features/right-to-die-there-were-times-when-i-would-have-ended-my-life-if-it-were-legal-coming-out-the-other-9602300.html> (accessed Feb. 13, 2020).
55. von Gunten, C.F., Ferris, F.D., Emanuel, L.L.: Ensuring Competency in End-of-Life Care: Communication and Relational Skills. *JAMA* **284**(23), 3051–3057 (2000). <https://doi.org/10.1001/jama.284.23.3051>
56. Kerr, J., Engel, J., Schlesinger-Raab, A., Sauer, H., Hölzel, D.: Communication, quality of life and age: results of a 5-year prospective study in breast cancer patients. *Ann. Oncol.* **14**(3), 421–427 (2003). <https://doi.org/10.1093/annonc/mdg098>
57. M. Schmid Mast, A. Kindlimann, and W. Langewitz, ‘Recipients’ perspective on breaking bad news: How you put it really makes a difference’, *Patient Education and Counseling*, vol. 58, no. 3, pp. 244–251, Sep. 2005, doi: <https://doi.org/10.1016/j.pec.2005.05.005>.

58. Schofield, P.E., Butow, P.N., Thompson, J.F., Tattersall, M.H.N., Beeney, L.J., Dunn, S.M.: Psychological responses of patients receiving a diagnosis of cancer. *Ann. Oncol.* **14**(1), 48–56 (2003). <https://doi.org/10.1093/annonc/mdg010>
59. Hurka, T.: ‘Moore’s Moral Philosophy’, in *The Stanford Encyclopedia of Philosophy*, Summer 2021., E. N. Zalta, Ed. Metaphysics Research Lab, Stanford University, 2021. Accessed 08 May 2021. <https://plato.stanford.edu/archives/sum2021/entries/moore-moral/>
60. Holmes, T.H., Rahe, R.H.: The social readjustment rating scale. *J. Psychosom. Res.* **11**(2), 213–218 (1967). [https://doi.org/10.1016/0022-3999\(67\)90010-4](https://doi.org/10.1016/0022-3999(67)90010-4)
61. Fried, T.R., Tinetti, M.E., Iannone, L., O’Leary, J.R., Towle, V., Van Ness, P.H.: Health outcome prioritization as a tool for decision making among older persons with multiple chronic conditions. *Arch. Intern. Med.* **171**(20), 1854–1856 (2011). <https://doi.org/10.1001/archinternmed.2011.424>
62. Stroebe, M.S., Stroebe, W., Hansson, R.O.: *Handbook of bereavement: theory, research, and intervention*. Cambridge University Press, Cambridge (1993)
63. NHS, ‘Coping with a terminal illness’, *nhs.uk*, May 29, 2018. <https://www.nhs.uk/conditions/end-of-life-care/coping-with-a-terminal-illness/>. Accessed 13 Mar 2021
64. San Filippo, D.: *Religious interpretations of death, afterlife & NDEs*, Faculty Publications, 2006. https://digitalcommons.nl.edu/faculty_publications/32
65. Mulgan, T.: How should utilitarians think about the future? *Can. J. Philos.* **47**(2–3), 290–312 (2017). <https://doi.org/10.1080/00455091.2017.1279517>
66. Internet Encyclopedia of Philosophy, ‘Utilitarianism, Act and Rule’. <https://iep.utm.edu/util-a-r/>. Accessed 08 May 2021.
67. Bullock, C.: *The Cobbler of Preston: A Farce*. publisher not identified, 1757. <https://books.google.co.uk/books?id=A4fzHgAACAAJ>
68. Christakis, N.A., Lamont, E.B.: Extent and determinants of error in doctors’ prognoses in terminally ill patients: prospective cohort study. *BMJ* **320**(7233), 469–473 (2000)
69. Cooke, C.R., Hotchkin, D.L., Engelberg, R.A., Rubinson, L., Curtis, J.R.: Predictors of time to death after terminal withdrawal of mechanical ventilation in the ICU. *Chest* **138**(2), 289–297 (2010). <https://doi.org/10.1378/chest.10-0289>
70. Downar, J., Goldman, R., Pinto, R., Englesakis, M., Adhikari, N.K.J.: The “surprise question” for predicting death in seriously ill patients: a systematic review and meta-analysis. *CMAJ* **189**(13), E484–E493 (2017). <https://doi.org/10.1503/cmaj.160775>
71. Munshi, L., Dhanani, S., Shemie, S.D., Hornby, L., Gore, G., Shahin, J.: Predicting time to death after withdrawal of life-sustaining therapy. *Intensive Care Med.* **41**(6), 1014–1028 (2015). <https://doi.org/10.1007/s00134-015-3762-9>
72. Williams, L.: *Relief after a death: the unspoken grief emotion*. *Whats your Grief*, 2016. <https://whatsyourgrief.com/relief-after-a-death-the-unspoken-emotion/>. Accessed 13 Mar 2021
73. Blatti, S.: Death’s distinctive harm. *Am. Philos. Q.* **49**(4), 317–330 (2012)
74. Vinding, M.: The harm of death by magnus vinding. <https://www.utilitarianism.com/magnus-vinding/harm-death.html>. Accessed 18 Feb 2020
75. Gazzaniga, M.: *The Animals Issue: moral theory in practice*. Cambridge University Press, 1992. <http://faculty.philosophy.umd.edu/pcarruthers/AI-4.htm>. Accessed 15 Feb 2020.
76. Guy, M., Stern, T.A.: The desire for death in the setting of terminal illness: a case discussion. *Prim. Care Companion J. Clin. Psychiatry* **8**(5), 299–305 (2006)
77. Kelly, B.J., Burnett, P.C., Pelusi, D., Badger, S.J., Varghese, F.T., Robertson, M.M.: Association between clinician factors and a patient’s wish to hasten death: terminally ill cancer patients and their doctors. *Psychosomatics* **45**(4), 311–318 (2004). <https://doi.org/10.1176/appi.psy.45.4.311>
78. Wilson, K.G., et al.: Mental disorders and the desire for death in patients receiving palliative care for cancer. *BMJ Support. Palliat. Care* **6**(2), 170–177 (2016). <https://doi.org/10.1136/bmjspcare-2013-000604>
79. Smith, A.K., et al.: The epidemiology of pain during the last two years of life. *Ann. Intern. Med.* **153**(9), 563–569 (2010). <https://doi.org/10.1059/0003-4819-153-9-201011020-00005>
80. Woo, J.A., Maytal, G., Stern, T.A.: Clinical challenges to the delivery of end-of-life care. *Prim. Care Companion J. Clin. Psychiatry* **8**(6), 367–372 (2006)
81. Rees, W.D., Dover, S.B., Low-Beer, T.S.: Patients with terminal cancer” who have neither terminal illness nor cancer. *Br. Med. J. (Clin. Res. Ed.)* **295**(6593), 318–319 (1987)
82. ‘Briton misdiagnosed with terminal cancer sues after spending life savings’, *Associated Press*, 2015. <https://www.foxnews.com/story/briton-misdiagnosed-with-terminal-cancer-sues-after-spending-life-savings>. Accessed 13 Feb 2020
83. ‘Gran blows life savings after being misdiagnosed with terminal illness’, *Starts at 60*, Feb. 22, 2018. <https://starts60.com/health/health-issues/dementia/gran-blows-life-savings-after-being-misdiagnosed-with-terminal-illness>. Accessed 13 Feb 2020
84. Watson, L.: ‘Man with “nothing to lose” splashed out £40,000 on bucket list after being diagnosed with terminal cancer... before being told doctors made a mistake’, *Mail Online*, Jun. 17, 2012. <https://www.dailymail.co.uk/news/article-2160620/Man-lose-splashed-40-000-bucket-list-diagnosed-terminal-cancer--told-doctors-mistake.html>. Accessed 13 Feb 2020
85. Agencies, ‘Mother who planned her funeral after terminal cancer diagnosis told “it was a mistake”’, Jul. 25, 2014. Accessed 13 Feb 2020. <https://www.telegraph.co.uk/news/health/news/10990220/Mother-who-planned-her-funeral-after-terminal-cancer-diagnosis-told-it-was-a-mistake.html>
86. Solicitors, T.: ‘Mum given wrong diagnosis for terminal cancer compensated by hospital’, *Thompsons Solicitors*. <https://www.thompsons.law/news/news-releases/medical-negligence-news/mum-given-wrong-diagnosis-for-terminal-cancer-compensated-by-hospital>. Accessed 13 Feb 2020
87. Bengtsson, A., Andersson, R., Ansari, D.: The actual 5-year survivors of pancreatic ductal adenocarcinoma based on real-world data. *Sci. Rep.* **10**(1), 16425 (2020). <https://doi.org/10.1038/s41598-020-73525-y>
88. Forbes, L.J., Ramirez, A.-J.: Offering informed choice about breast screening. *J. Med. Screen* **21**(4), 194–200 (2014). <https://doi.org/10.1177/0969141314555350>
89. Rasmussen, J.F., Siersma, V., Malmqvist, J., Brodersen, J.: Psychosocial consequences of false positives in the Danish Lung Cancer CT Screening Trial: a nested matched cohort study. *BMJ Open* **10**(6), e034682 (2020). <https://doi.org/10.1136/bmjopen-2019-034682>
90. Simpkin, A.L., Armstrong, K.A.: Communicating uncertainty: a narrative review and framework for future research. *J. Gen. Intern. Med.* **34**(11), 2586–2591 (2019). <https://doi.org/10.1007/s11606-019-04860-8>
91. Freeman, A.L.J.: How to communicate evidence to patients. *DTB* **57**(8), 119–124 (2019). <https://doi.org/10.1136/dtb.2019.000008>
92. Ha, J.F., Longnecker, N.: Doctor-patient communication: a review. *Ochsner J.* **10**(1), 38–43 (2010)
93. Institute for Healthcare Communication, ‘Impact of Communication in Healthcare’, 2011. <https://healthcarecomm.org/about-us/impact-of-communication-in-healthcare/>. Accessed 15 Feb 2020

94. Kübler-Ross, E.: On death and dying: what the dying have to teach doctors, nurses, clergy and their own families, p. 260. Collier Books/Macmillan Publishing Co, New York (1970)
95. Stroebe, M., Schut, H., Boerner, K.: Cautioning health-care professionals. *Omega (Westport)* **74**(4), 455–473 (2017). <https://doi.org/10.1177/0030222817691870>
96. Bonanno, G.A., Malgaroli, M.: Trajectories of grief: Comparing symptoms from the DSM-5 and ICD-11 diagnoses. *Depress. Anxiety* **37**(1), 17–25 (2020). <https://doi.org/10.1002/da.22902>
97. Houston, R.E.: The angry dying patient. *Prim. Care Companion J. Clin. Psychiatry* **1**(1), 5–8 (1999)
98. Arndt, J., Schimel, J., Goldenberg, J.L.: Death can be good for your health: fitness intentions as a proximal and distal defense against mortality salience. *J. Appl. Soc. Psychol.* **33**(8), 1726–1746 (2003). <https://doi.org/10.1111/j.1559-1816.2003.tb01972.x>
99. Schroepfer, T.A.: Critical events in the dying process: the potential for physical and psychosocial suffering. *J. Palliat. Med.* **10**(1), 136–147 (2007). <https://doi.org/10.1089/jpm.2006.0157>
100. Cartwright, J.C., Hickman, S., Perrin, N., Tilden, V.: Symptom experiences of residents dying in assisted living. *J. Am. Med. Dir. Assoc.* **7**(4), 219–223 (2006). <https://doi.org/10.1016/j.jamda.2005.09.011>
101. Goodkin, K., Kompella, S., Kendell, S.F.: End-of-life care and bereavement issues in HIV/AIDS. *Nurs. Clin. N. Am.* **53**(1), 123–135 (2018). <https://doi.org/10.1016/j.cnur.2017.10.010>
102. Boyd, R., Richerson, P.J.: Culture and the evolution of human cooperation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**(1533), 3281–3288 (2009). <https://doi.org/10.1098/rstb.2009.0134>
103. Nunney, L.: Group selection, altruism, and structured-deme models. *Am. Nat.* **126**(2), 212–230 (1985)
104. Coelho, A., de Brito, M., Barbosa, A.: Caregiver anticipatory grief: phenomenology, assessment and clinical interventions. *Curr. Opin. Support. Palliat. Care* **12**(1), 52–57 (2018). <https://doi.org/10.1097/SPC.0000000000000321>
105. Pohlkamp, L., Kreicbergs, U., Sveen, J.: Factors during a child's illness are associated with levels of prolonged grief symptoms in bereaved mothers and fathers. *JCO* **38**(2), 137–144 (2019). <https://doi.org/10.1200/JCO.19.01493>
106. Martikainen, P., Valkonen, T.: Do education and income buffer the effects of death of spouse on mortality? *Epidemiology* **9**(5), 530–534 (1998)
107. Stephen, A.I., et al.: The economic cost of bereavement in Scotland. *Death Stud.* **39**(3), 151–157 (2015). <https://doi.org/10.1080/07481187.2014.920435>
108. Cory, S.: MyGrief.ca. *Ann. Palliat. Med.* **7**(Suppl 1), AB017 (2018). <https://doi.org/10.21037/apm.2018.s017>
109. Woof, W.R., Carter, Y.H.: The grieving adult and the general practitioner: a literature review in two parts (Part 1). *Br J Gen Pract* **47**(420), 443–448 (1997)
110. Christopher, S.: Sudden versus unexpected death: Its complications and consequences. *Ambulance UK*, vol. 22, (2007)
111. Lundin, T.: Long-term outcome of bereavement. *Br. J. Psychiatry* **145**(4), 424–428 (1984). <https://doi.org/10.1192/bjp.145.4.424>
112. Maercker, A., et al.: Diagnosis and classification of disorders specifically associated with stress: proposals for ICD-11. *World Psychiatry* **12**(3), 198–206 (2013). <https://doi.org/10.1002/wps.20057>
113. Luo, W., et al.: Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J. Med. Internet Res.* **18**(12), e323 (2016). <https://doi.org/10.2196/jmir.5870>
114. Hulsman, R.L., Pranger, S., Koot, S., Fabriek, M., Karemaker, J.M., Smets, E.M.A.: How stressful is doctor–patient communication? Physiological and psychological stress of medical students in simulated history taking and bad-news consultations. *Int. J. Psychophysiol.* **77**(1), 26–34 (2010). <https://doi.org/10.1016/j.ijpsycho.2010.04.001>
115. Ptacek, J.T., Fries, E.A., Eberhardt, T.L., Ptacek, J.J.: Breaking bad news to patients: physicians' perceptions of the process. *Support Care Cancer* **7**(3), 113–120 (1999). <https://doi.org/10.1007/s005200050240>
116. Studer, R.K., Danuser, B., Gomez, P.: Physicians' psychophysiological stress reaction in medical communication of bad news: A critical literature review. *Int. J. Psychophysiol.* **120**, 14–22 (2017). <https://doi.org/10.1016/j.ijpsycho.2017.06.006>
117. Knopke, E.: Touching the dead: Autoethnographical reflections about the researcher's body in the field of death, dying, and bereavement. *Death Stud.* **42**(10), 640–648 (2018). <https://doi.org/10.1080/07481187.2018.1426656>
118. Rossi Ferrario, S., Cardillo, V., Vicario, F., Balzarini, E., Zotti, A.M.: Advanced cancer at home: caregiving and bereavement. *Palliat. Med.* **18**(2), 129–136 (2004). <https://doi.org/10.1191/0269216304pm8700a>
119. Yamamoto, S., et al.: Decision making regarding the place of end-of-life cancer care: the burden on bereaved families and related factors. *J. Pain Symptom Manag.* **53**(5), 862–870 (2017). <https://doi.org/10.1016/j.jpainsymman.2016.12.348>
120. Nozick, R.: *Anarchy, state and Utopia*. Blackwell, Oxford (1975)
121. Petersen, M.B.: COVID lesson: trust the public with hard truths. *Nature* **598**(7880), 237–237 (2021). <https://doi.org/10.1038/d41586-021-02758-2>
122. Blastland, M., Freeman, A.L.J., van der Linden, S., Marteau, T.M., Spiegelhalter, D.: Five rules for evidence communication. *Nature* **587**(7834), 362–364 (2020). <https://doi.org/10.1038/d41586-020-03189-1>
123. van der Bles, A.M., van der Linden, S., Freeman, A.L.J., Spiegelhalter, D.J.: The effects of communicating uncertainty on public trust in facts and numbers. *Proc. Natl. Acad. Sci. USA* **117**(14), 7672–7683 (2020). <https://doi.org/10.1073/pnas.1913678117>
124. van der Bles, A.M., et al.: Communicating uncertainty about facts, numbers and science. *R. Soc. Open Sci.* **6**(5), 181870 (2019). <https://doi.org/10.1098/rsos.181870>
125. Badaea, C., Artus, G.: 'Morality, Machines and the Interpretation Problem: A value-based, Wittgensteinian approach to building Moral Agents', [arXiv:2103.02728 \[cs\]](https://arxiv.org/abs/2103.02728), 2021, Accessed 26 Jan 2022. <http://arxiv.org/abs/2103.02728>
126. 'AI: using trust and ethics to accelerate adoption'. <https://www.ukri.org/blog/ai-using-trust-and-ethics-to-accelerate-adoption/>. Accessed 13 Mar 2022
127. Pataranutaporn, P., et al.: AI-generated characters for supporting personalized learning and well-being. *Nat. Mach. Intell.* **3**(12), 12 (2021). <https://doi.org/10.1038/s42256-021-00417-9>
128. Utami, D., Bickmore, T., Nikolopoulou, A., Paasche-Orlow, M.: Talk about death: end of life planning with a virtual agent. *Intell. Virtual Agents* (2017). https://doi.org/10.1007/978-3-319-67401-8_55
129. Lucas, G.M., Gratch, J., King, A., Morency, L.-P.: It's only a computer: Virtual humans increase willingness to disclose. *Comput. Hum. Behav.* **37**, 94–100 (2014). <https://doi.org/10.1016/j.chb.2014.04.043>
130. Health Center for Devices and Radiological, 'International Medical Device Regulators Forum (IMDRF)', *FDA*, Mar. 16, 2021. <https://www.fda.gov/medical-devices/cdrh-international-programs/international-medical-device-regulators-forum-imdrf>. Accessed 13 Sept 2022
131. Brazier, J., Roberts, J., Deverill, M.: The estimation of a preference-based measure of health from the SF-36. *J. Health Econ.* **21**(2), 271–292 (2002). [https://doi.org/10.1016/s0167-6296\(01\)00130-8](https://doi.org/10.1016/s0167-6296(01)00130-8)

132. EuroQol Group: EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* **16**(3), 199–208 (1990). [https://doi.org/10.1016/0168-8510\(90\)90421-9](https://doi.org/10.1016/0168-8510(90)90421-9)
133. Qiu, A., Xu, L., Liu, C.: Predicting diagnosis 4 years prior to Alzheimer’s disease incident. *NeuroImage Clin.* **34**, 102993 (2022). <https://doi.org/10.1016/j.nicl.2022.102993>
134. ‘Why some patients keep their cancer secret—and how oncologists guide them through that chosen seclusion’, *Fred Hutch*, Aug. 29, 2016. <https://www.fredhutch.org/en/news/center-news/2016/08/why-some-cancer-patients-keep-diagnosis-secret-how-oncologists-guide-privacy.html>. Accessed 21 Apr 2022.
135. Davidson, S., Gentry, T.: Age UK end of life evidence review 2013. https://www.ageuk.org.uk/globalassets/age-uk/documents/reports-and-publications/reports-and-briefings/health-wellbeing/rb_oct13_age_uk_end_of_life_evidence_review.pdf. Accessed 21 Apr 2022
136. Mathur, S., Sutton, J.: Personalized medicine could transform healthcare. *Biomed Rep* **7**(1), 3–5 (2017). <https://doi.org/10.3892/br.2017.922>
137. McCarthy, J.: Principlism or narrative ethics: must we choose between them? *Med. Humanit.* **29**(2), 65–71 (2003). <https://doi.org/10.1136/mh.29.2.65>
138. Launer, J.: Breaking the news. *QJM* **98**(5), 385–386 (2005). <https://doi.org/10.1093/qjmed/hci056>
139. Candib, L.M.: Truth telling and advance planning at the end of life: Problems with autonomy in a multicultural world. *Fam. Syst. Health* **20**(3), 213–228 (2002). <https://doi.org/10.1037/h0089471>
140. Lagay, F.L.: The ethical force of stories: narrative ethics and beyond. *AMA J. Ethics* **16**(8), 622–625 (2014). <https://doi.org/10.1001/virtualmentor.2014.16.8.jdsc1-1408>
141. Ahlén, R.: Narrativity and medicine: some critical reflections. *Philos. Ethics Humanit. Med.* **14**(1), 9 (2019). <https://doi.org/10.1186/s13010-019-0078-3>
142. Brody, H., Clark, M.: Narrative ethics: a narrative. *Hastings Cent. Rep.* **44**(s1), S7–S11 (2014). <https://doi.org/10.1002/hast.261>
143. Nortjé, N., Jones-Bonofiglio, K., Sotomayor, C.R.: Exploring values among three cultures from a global bioethics perspective. *Glob. Bioeth.* **32**(1), 1–14 (2021). <https://doi.org/10.1080/11287462.2021.1879462>
144. Chattopadhyay, S., De Vries, R.: Bioethical concerns are global, bioethics is Western. *Eubios J. Asian Int. Bioeth.* **18**(4), 106–109 (2008)
145. Tai, M.C.-T.: Western or Eastern principles in globalized bioethics? An Asian perspective view. *Tzu Chi Med. J.* **25**(1), 64–67 (2013). <https://doi.org/10.1016/j.tcmj.2012.05.004>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.