

## **Title:**

### **Should the Decisions of Institutional Review Boards be Consistent?**

#### **Introduction**

A good deal has been written, mostly in complaint, about the functioning of Institutional Review Boards (IRBs)<sup>1</sup> and their judgements about the proposed research that they consider. One clearly discernible line of complaint has been variation or, more normatively, inconsistency between committees. Such inconsistency has been documented both across IRBs and within IRBs, and has been the source of mounting criticisms from researchers who require approval from these committees before they can begin their research. This paper attempts to make progress on the problem of inconsistency in IRB decision-making. While we acknowledge that consistency is an important value in research ethics and oversight, we argue that achieving it fully may come at too great of a cost.

In the first section, we begin by mapping out the issue of inconsistency, including the different kinds of inconsistency seen in IRB decision-making and the responses they have generated. The second section introduces a distinction between procedural and content consistency, describes how fairness provides the ethical grounding of consistency, and considers the relationship between fairness and the two types of consistency. The third section lays out what we call the decision-maker's dilemma. The dilemma suggests that, on one horn, decision makers can adopt an algorithmic decision making procedure to ensure consistency, and risk excluding morally relevant factors from the decision-making process, or on the other horn, they can embrace a discretionary decision-making procedure, and risk including morally irrelevant factors in the decision-making process. In the final section, we consider the question of which horn of this dilemma systems of research ethics ought to embrace. We argue that the second horn, that of discretionary decision making, is preferable because of important

---

<sup>1</sup> An IRB in the US is effectively the same as a Research Ethics Committee (REC) or a Research Ethics Board (REB). There are some important systemic differences across countries that depend on regulatory, historical and cultural differences. These do not have a significant impact on the arguments here and are beyond the scope of this paper.

parallels between the justice system of systems of research ethics. Finally, we suggest several important safeguards should be put in place to increase both fairness and perceptions of fairness in research governance.

## **I. Inconsistency in Research Ethics Review**

Concerns regarding the inconsistency of IRBs have been mounting for decades, as has evidence for the preponderance of inconsistencies in the decision-making of IRBs (Trace and Kolstoe 2017, Abbott and Grady 2011, Klitzman 2015, Edwards, Stone, and Swift 2007, Stark 2011, Goldman and Katz 1982). Inconsistency in IRB functioning takes various forms. Initial concerns revolved primarily around the experiences of research teams conducting multi-site research, who were required to seek approval from several different IRBs (McWilliams et al. 2003, Caulfield, Ries, and Barr 2011, Driscoll et al. 2008). These IRBs would often request different, sometimes conflicting, changes to be made to the same research protocol, which frequently would consume a considerable amount of time and funding, and, at times, would decrease the value of the research (e.g. through reducing generalizability) (Petersen et al. 2012).

Fortunately, several countries have implemented centralized systems of research ethics review, in order to reduce the redundancy and variation seen across local IRBs within multi-site research. In the United Kingdom, the introduction multi-centre research ethics committees (MRECs) ensures that research conducted on patients in the NHS is only reviewed by one IRB (Department of Health 1997). The United States' system of research oversight has also undergone changes in order to reduce the burden of multi-site review, with the release of the Final NIH Policy on the Use of a Single Institutional Review Board for Multi-Site Research (The National Institutes of Health 2016). Australia has taken similar measures, so that investigators involved in multi-site research are required to submit a national common application form that is recognized across the country (Needham, Kapadia, and Offringa 2015). Several jurisdictions in Canada are moving towards centralized review as well, although not all have done so (Page and Nyeboer 2017).

Despite these changes, inconsistency in IRB functioning is still a widely reported issue, both within and across IRBs. Single IRB are sometimes found to be inconsistent with themselves, such as when a committee finds fault with a research protocol that they previously approved, even though no changes have been made to it, or when a committee treats two similar protocols differently. Additionally, empirical researchers have continued to examine the differences across IRB responses to a single protocol. While the final decisions of IRBs tend to converge (e.g. whether to approve the protocol), the way the committees reach these decisions and the number and types of modifications they request along the way tend to differ substantially, illuminating discrepancies between their decision-making processes (Vadeboncoeur et al. 2016). Inconsistencies across IRB processes that have been documented include differences in the amount preparation done before review, the average time to approval of a protocol, and costs of review (Hirshon et al. 2002, Khan et al. 2014, Byrne et al. 2006, Vick et al. 2005, Dziak et al. 2005). Inconsistencies have also been found across IRBs in terms of the type of review assigned to a particular protocol (e.g., exempt, expedited), whether informed consent waivers were deemed acceptable, and the ways in which regulations are interpreted (Hirshon et al. 2002, Silverman, Hull, and Sugarman 2001, Khan et al. 2014, Petersen et al. 2012). For example, an investigation by Shah et al. found that while federal regulations in the United States permit children to be enrolled in research only when risks are determined to be minimal or a minor increase over minimal, or when the research leads to direct benefits, IRBs vary considerably with regards to what they consider to be ‘minimal risk’ or ‘direct benefit’ (Shah et al. 2004). These inconsistencies can appear between IRBs in the same city and even within the same institution (Klitzman 2011), suggesting that local context is not always the source of the differences.

Critics point out that inconsistency is harmful in that it can lead to wasted resources (e.g. Green et al. reported 4,680 hours of staff time dedicated towards IRB process over 19 months), frustrated investigators, and the delay of valuable research (Green et al. 2006, Whitney et al. 2008, Stark and Greene 2016). Inconsistencies within IRBs are likely to lead to frustrations among researchers and reductions in confidence in IRB functioning, while inconsistencies across IRBs may increase the practice of IRB shopping, so that IRBs perceived as faster and more permissible will receive more

protocols for review. Reasons behind inconsistencies within and across IRBs are likely to be multiple and varied. Evidence suggests that IRB members often have different degrees of knowledge about existing regulations (Mhaskar et al. 2015), different senses of their role in relation to research participants and investigators (e.g. protectionists, expeditors) (Sayers 2007), as well as different concerns, past experiences, and personalities (Klitzman 2011, 2015), all of which may contribute to the variation seen in IRB decision-making. Furthermore, regulations that provide guidance for IRBs are frequently vague or inconsistent (e.g., the FDA, Common Rule, and HHS all offer different definitions of and protocols for responding to conflicts of interest), making it all the more likely that IRBs will interpret or apply regulations differently (Friesen, Redman, and Caplan 2018, Emanuel et al. 2004). Stark's examination of IRB practices also suggests that individual IRBs tend to develop local precedents when novel ethical issues arise and are resolved; these precedents may increase consistency within committees as they tend to guide responses to similar protocols, but may reinforce inconsistency across committees (Stark 2011). Barriers to increasing consistency within and across IRBs are substantial. Such barriers that have been flagged by research ethicists, who point out that moral disagreement is ubiquitous and inevitable, there is no algorithm available on which to produce the 'right' ethical output of committees, and there is little incentive for IRB members to try to increase consistency (Klitzman 2011, Sayers 2007, Angell et al. 2007, Edwards, Ashcroft, and Kirchin 2004).

In light of these barriers, some go as far as to defend inconsistency in IRB decision making. The most detailed defence of inconsistency is offered by Edwards et al., who argue that we ought "to reject the view that we should strive for complete consistency" (Edwards, Ashcroft, and Kirchin 2004, 416). They offer three arguments in support of this thesis. The justice argument maintains that local context, and in particular, cultural differences, often justify differences in IRB functioning. The moral pluralism argument makes the case that disagreement between IRBs is both legitimate and desirable. Finally, their due process argument holds that the process of moral deliberation is often more important than the outcome of the deliberation (Edwards, Ashcroft, and Kirchin 2004). Similarly, McGuinness argues that inconsistency in IRB decision making is legitimate as long as it is based on a fair procedure (McGuinness 2008). This emphasis on process closely relates to the distinction

between procedural and content consistency that we outline in the following section, and we are sympathetic to the conclusions they reach. Sayers also defends the inconsistency of IRBs, arguing that the goal of consistency is not in fact a desirable when one looks closer. Sayers rhetorically suggests that one way to reach such a goal would be to screen IRB members in advance to determine their moral views, or to tell them how to think, leading to an Orwellian state of affairs (Sayers 2007). Qualitative research with IRB members has also revealed that IRB members sometimes defend the inconsistencies between committees as well, although with little apparent justification (Klitzman 2011).

In line with Edwards et al., McGuinness, and Sayers, we hope to argue that research ethics review should not aim for complete consistency. Rather than focusing on acceptable forms of inconsistency, as Edwards and colleagues do, we defend the importance of consistency in IRB decision-making, and in line with Sayers' warning that inconsistency is the lesser of two evils, we argue that embracing some inconsistency in IRB decision-making may be the best approach available, given the nature of the decisions being made. In the next section, we outline an important distinction between two types of consistency, the moral basis of consistency in decision-making, and the connection between them.

## **II. Examining Consistency**

### **Procedural and Content Consistency.**

Two kinds of consistency in decision-making can be distinguished: procedural and content consistency<sup>2</sup>. While procedural consistency is concerned with the process by which a decision is reached, content consistency is concerned with the outcome of the decision. These can come apart quite cleanly in principle, although as will be discussed in more detail below, they rarely do in practice. In the context of research ethics review, mere procedural consistency could occur in the case

---

<sup>2</sup> This distinction relates closely to one drawn by McGuinness in her discussion of IRB deliberations, which refers to procedural and substantive issues in deliberation (McGuinness 2008), as well as to Rawls' distinction between procedural and substantive justice (Rawls 1971).

of an IRB that made all decisions on the basis of a coin flip. Regardless of the outcomes of these decisions, they would be procedurally consistent, because their process for making decisions is always the same. On the other hand, content consistency results only from agreement between the substantive judgements made by committees, and not the process. If two committees approve the same research protocol and require no modifications, one as the result of reading and discussing the protocol in detail, and the other because they always give a full pass to protocols that are submitted on a Tuesday, they would be consistent in content, but not in procedure. Note that in neither of these cases of mere procedural or content consistency is there any reference to an independent standard of what counts as 'getting it right'.

### **Fairness and Consistency.**

We often care about consistency because it represents fairness. To be consistent is to treat equals equally, ensuring that individuals are not treated differently when there is no good reason to do so. Of course, there are important limits on when and where we expect the content of decisions to be consistent. In many situations, we agree that there are important differences which ought to prevent us from treating two cases in the same way. In these cases, to ensure fairness, it is important for decision makers to be procedurally consistent, to decide on the same basis, taking into account the same kinds of reasons in each case.

Content consistency, in which the same decision is reached within two cases that are the same in factors relevant to the decision being made, is not always required for fairness. Individuals accept differences in content consistency when they feel that the process has been fair, which in some decisions, can involve a process relying on chance. If all options are equal at the outset, using a system based on chance ensures that no differential treatment is given when it is not justified. So if one person wins in a lottery and another does not, even though there is no content consistency, there is no sense of injustice, because each person was given an equal chance in the lottery process.

Sometimes, however, fairness depends on their being the right kind of relationship between the procedure and the content of the decision, and in these cases, procedural and content consistency are closely related. This occurs when there are relevant factors that ought to inform the decision, but that would be excluded if the decision was left to mere chance. In such cases, the procedure ought to take into account the right kinds of reasons in the right way, in order to be fair. Take the allocation of a scarce resource, such as organs, for example. If one person had been waiting for a liver for several years and was likely to die if they didn't receive one in the next 6 months, while another had just signed up for the organ registry and was not in dire shape, it would not be fair to flip a coin between these two to decide who should receive the next available liver. The reason for this is that there are morally relevant reasons that ought to be taken into consideration (e.g. time on the waiting list, prognosis) within a decision-making process regarding the allocation of organ donations, and flipping a coin fails to recognize these as reasons. Systems that have developed in order to ensure the just allotment of organs (e.g. MELD scores) aim to connect the procedure of decision-making with the content of the decision, taking into account the right reasons in the right way<sup>3</sup>.

### **Fairness and Consistency in IRBs.**

The decisions of IRBs are more like decisions related to the allocation of organs than decisions based on the results of a lottery. There are important differences between the risks and benefits contained with different research protocols that ought to be taken into account during the review process. This suggests that the kind of consistency that we should be concerned with when it comes to IRB decision making is not only consistency in terms of the final decisions that IRBs make (e.g., approve, approve with modifications, reject) (Angell et al. 2006, Stair et al. 2001, Stark, Tyson, and Hibberd 2010), but in terms of the reasons or values given as support the decisions that they reach (Goldman and Katz 1982). Of course, if the same kinds of reasons or values are guiding the approaches of different IRBs, it is all the more likely that they will reach the same decisions about different protocols. This points to

---

<sup>3</sup> This is not an argument that these systems have been successful in selecting the right criteria to take into consideration, but only an acknowledgement that they intend to.

an important way in which procedural and content consistency are related: the more procedural consistency one has, the more likely one is to have content consistency (except in cases in which the procedure is based on chance).

As with other types of consistency, fairness underlies the importance of consistency in IRB decision-making, both for the sake of investigators and potential research participants. It is unfair to treat two identical protocols differently if they are being proposed for the same context (so there are no relevant local differences), because this allows one investigator to proceed with her proposed research, while another is required to wait. Alternatively, one researcher may substantially improve her research protocol through engagement with the committee's requests, while another may not benefit from this engagement. Likewise, potential participants can be either unfairly benefitted or harmed by inconsistencies in research ethics review. If a protocol with risks that outweigh its benefits is approved by one committee and not another, the patient population in the jurisdiction of the first IRB is protected while the second is not. However, a risk-adverse IRB may unfairly disadvantage a local population by turning down a research protocol that includes the opportunity for patients to enrol in a clinical trial, while another committee may approve the same protocol, thereby benefiting potential participants in a particular area<sup>4</sup>.

### **Perceptions of Fairness.**

Importantly, perceptions of fairness shape how individuals interact with a system. In many contexts, an increase in perceived injustice leads to an increase in deviant behaviour (Keith-Spiegel and Koocher 2005). For example, research investigating the link between perceptions of injustice and workplace behaviour suggest that retaliatory behaviour (e.g. theft, calling in sick when not ill) is moderated by feelings of justice (Skarlicki and Folger 1997). Extending these findings to the realm of

---

<sup>4</sup> Interestingly, those writing about inconsistencies in IRBs have tended to focus on fairness to either investigators or research participants, but not both (e.g. Sayers writes primarily about injustice to researchers, while Edwards et al. discuss unfairness to participants (Sayers 2007, Edwards, Ashcroft, and Kirchin 2004)).



IRBs, Keith-Spiegel and colleagues have suggested that “the perception of unfairness will motivate some investigators to engage in subterfuge and misconduct designed to ‘level the playing field’” (Keith-Spiegel, Koocher, and Tabachnick 2006, 78). Keith-Spiegel and Koocher argue that particular features of IRBs make them especially susceptible to being perceived as unfair (e.g. flexibility in interpretation of guidelines, pressure to protect the institution, a lack of training/ professionalization), making it likely that, paradoxically, their behaviours may contribute to unethical behaviour in scientists (Keith-Spiegel and Koocher 2005).

Others have also pointed to the ways in which perceived injustice within IRB decision-making can alienate researchers from the ethics review process and lead them to cut corners or leave morally relevant information out of applications (DuBois 2004, Klitzman 2011). Research suggests that the majority of complaints from investigators regarding IRBs relate to issues of fairness, and when polled, researchers report, somewhat surprisingly, that procedural justice (how decisions are made) and interactional justice (interpersonal sensitivity and justification) are more important to them in IRBs than competence, the absence of bias, or protecting the rights of participants (Keith-Spiegel, Koocher, and Tabachnick 2006). This suggests that in addition to being unfair, inconsistency within and across IRBs may have harmful downstream effects, as it may lead to dishonesty within research protocols, IRB shopping, and damage to the relationships between IRB members and researchers. This points to the importance of having buy in from researchers within a system of research governance, or risk encouraging behaviour that poses a direct threat to the purpose of having ethical review of research. This is a tall order; not only is a fair system of research governance required, but a system that is both fair and perceived as fair.

Perceptions of fairness also depend on the availability of information regarding the decision-making process. If a decision-making process is unknown to those impacted by it, content consistency is all that is available and so is central to perceptions of fairness. For example, if two people with the same terminal illness wish to gain access to an investigational drug and only one is granted access, but they have no knowledge of how the decision was reached, the decision-making process may well be

perceived as unfair, regardless of whether it was. Similarly, when investigators are unaware of how decisions about research proposals are made within IRBs, researchers may be inclined to assume the process is unfair, given the variety of experiences they have heard about and the lack of justification given for the modifications requested, regardless of what actually took place within IRB discussions. This suggests that some form of transparency of process can play an important role in perceptions of fairness, a point which we will return to later.

### **III. The Decision-Maker's Dilemma**

Sometimes, a decision is complex enough that the procedure cannot be specified in advance, because the relevant factors are too many or interact in unpredictable ways. In these cases, individual or community discretion is often valued (e.g. interviewing candidates for a job). This is particularly the case in normative decisions, because the ethical criteria that are relevant are often under dispute and resist easy measurement (e.g. determining the sentence for a crime committed). The task of IRBs, in reviewing research protocols in order to evaluate their ethical and scientific quality is both complex and normative, making it the kind of decision-making process in which discretion is valued<sup>5</sup>. As a result, most IRBs consist of a diverse group of individuals with a range of experiences and areas of expertise.

However, the more complex and normative the decision that must be made, the more likely that factors that are thought to be irrelevant to the decision making process will slip in to a discretionary decision-making process. This results from both the complexity, in that the many variables involved that are difficult to specify in advance, and the normativity, in that moral judgements across individuals tend to be diverse and resist easy consensus. Take decisions regarding promotions at work, for example. While managers may have every intention of narrowing the pay gap between men

---

<sup>5</sup> There is a nest of issues in here about the relationship between scientific and ethical considerations that are beyond the scope of this paper but which are important to acknowledge (Freedman 1987, Dawson and Yentis 2007, Angell et al. 2008).

and women, there are various ways in which gender stereotypes with regards to men and leadership positions can infiltrate the process of deciding who to promote. In the many cases in which women are paid less than men while working the same job, it appears that a morally irrelevant factor (e.g. gender) has slipped into the decision-making process, creating an unfair system of inconsistent rewards. Similarly, there is little doubt that irrelevant factors enter the decision-making processes of IRBs. Qualitative research by Klitzman reveals that idiosyncratic features of individuals and communities can play a substantial role in shaping feedback given by IRBs to investigators, both as a result of the complexity of decisions involved and differences in terms of moral judgements (Klitzman 2011, 2015). Uncertainty also tends to exacerbate the difficulties related to decision making that is complex and normative; the less that is known about which factors are relevant (e.g., the risks and benefits that might be involved in a particular research protocol) and the relationship between different values (e.g., how to weigh patient autonomy and the potential knowledge that might be gained against each other), the more less likely that the only the right factors will enter the decision-making process. This inevitably leads to inconsistencies within the process.

In decision-making systems in which such inconsistencies begin to appear, those responsible for decision-making are faced with a dilemma. One horn of the dilemma contains the option of specifying in advance how factors relevant to the decision should impact it, fixing the content to ensure consistency and prevent irrelevant information from entering the decision-making process. This can be thought of as *algorithmic decision-making*. This leads to both procedural and content consistency. In response to the gender pay gap, a boss might decide that rather than offering promotions on the basis of his perception of their merit, individuals will receive promotions on the basis of how long they have been with the company. While this new system will ensure that men and women are not treated differently on the basis of their gender, other factors that ought to contribute to decisions regarding promotions will be left out (e.g. performance). With regards to complex, normative decisions, consistency comes with a cost. While irrelevant factors tend to be excluded from the decision, some relevant ones tend to go too, both increasing and decreasing the fairness of the decision. With algorithmic decision-making, the bathwater is gone, but the baby is too.

On the other horn, a decision-maker can embrace the importance of discretion, and work towards creating a more consistent procedure of decision-making, without fixing the content of decision in relation to relevant factors. This can be thought of as *discretionary decision-making*. This allows all relevant factors that ought to be taken into account at least potentially, to be included. So, a manager might be aware that he tends to promote men more often than women, but he might also see the benefit of allowing other features to contribute to decisions about promotions. In this way, he embraces inconsistency, ensuring that a significant range of relevant factors can be considered in any one decision about a promotion (e.g. maternity leave), but he also inevitably allows irrelevant factors (e.g. shared interests) to enter the decision-making process. Perhaps he will try to patch up the discretionary procedure by reducing the amount of irrelevant information that enters the decision-making process (e.g. taking a course in implicit bias, including women on the hiring team), hoping this will contribute to content consistency. While these measures may do some good, a discretionary process of decision-making will always be limited in terms of both procedural and content consistency, since more than just morally relevant factors will enter into the process of decision-making. With discretionary decision-making, the baby is safe, but the bathwater is here to stay.

#### **IV. Embracing Inconsistency**

How should systems of research governance respond to this dilemma? We argue that IRBs should embrace inconsistency by choosing the second horn of the dilemma and continuing to embrace a discretionary process of decision-making. Rather than attempting to specify in advance precisely what factors should contribute to decisions regarding research protocols, they should aim to improve content consistency by way of procedural changes<sup>6</sup>. This choice involves recognizing that consistency is merely one value among many within research ethics; by choosing discretionary decision-making, some consistency is sacrificed in order to improve the likelihood that the broad range of relevant

---

<sup>6</sup> By unpacking this dilemma, we do not intend to suggest that any research ethicists are arguing for an entirely algorithmic process of decision-making, but merely to illuminate the issues that are likely to arise if one leans too far towards either horn of the dilemma.

ethical issues (e.g. novel risks and benefits) that can arise in research protocols will be identified by an IRB and appropriate modifications will be requested. In favour of the second horn, we discuss here several features that are shared between systems of justice and systems of research governance. Because of these similarities, we argue that systems of research ethics ought to follow the courts in recognizing the importance of discretion and therefore accepting limitations with regards to consistency. We also discuss several safeguards that systems of research governance could adopt in order to counter the risks that arise from embracing the second horn and adopting a discretionary process of decision-making.

### **The Justice System.**

Many research ethicists have noted the similarities of research ethics systems and the justice system in brief remarks, but the connection has not been explored in detail (Angell et al. 2007, Klitzman 2011, Hirshon et al. 2002)<sup>7</sup>. We think there is real weight and importance to these comparisons. As with systems of research oversight, consistency is a constant concern in systems of justice. For example, in the United States, those who have been convicted for the murder of a white female are 10 times more likely to be executed than those who have been convicted for the murder of a black male (Baumgartner et al. 2015). This violates the principle of parity, a grounding principle in sentencing which holds that those who have committed the same offence should, all things being equal, receive equal punishment, is considered to be foundational in law. As Bagaric explains, “at the core of the principle of parity is the concept of equal justice and the patent unfairness in dealing differently with offenders who are similarly placed” (Bagaric and Pathinayake 2013). Just as fairness underlies the importance of consistency in IRB decision-making, the principle of parity underlies the importance of consistency in criminal sentencing.

---

<sup>7</sup> An exception is Schneider who offers a detailed argument in favor of replacing the current system of research governance with a system of tort law and self-regulation, but this is a different connection than the one we emphasize here (Schneider 2015).

While ensuring fairness through consistency is thought to be a central goal of systems of justice, the principle of parity is also generally accepted as an impossible goal. Several features of the justice system contribute to this impossibility, each of which are shared by systems of research governance. The first feature is that there is no available ground truth that can provide a foundation for consistent moral decision-making. While both systems have legislation, regulations, guiding documents, and principles, no universal moral system has been agreed upon as a guide<sup>8</sup>. As Angell and colleagues have put it with regards to IRBs, there is no “final moral authority” (Angell et al. 2007, 13). This contributes to the second feature, which is continuous evolution when it comes to ethical agreement. That which is considered right or wrong in systems of justice and research ethics changes over time, as do interpretations of regulations and principles. While in some cases, these changes will be instantiated in changes to standards and regulations, in other cases, they occur gradually within processes of decision-making. This makes consistency over time a very difficult, and arguably undesirable, goal<sup>9</sup>. The third feature is one that relates to this constant state of moral change: the importance of context. Both within legal systems and within research governance, the details of a case are inseparable from an ethical examination of it. This rules out the possibility of embracing the first horn of the dilemma by developing an algorithmic decision-making procedure so that the content of decisions can be derived from the factors relevant to that decisions. Given the complexity of context within sentencing and research protocols, the factors are too many and too varied, and they interact in unpredictable ways. Consider the role of aggravating and mitigating circumstances in sentencing as an illustration of the importance of the details of a case from its evaluation. While certain features are intuitively classed as either aggravating (e.g. past convictions) or mitigating (e.g. remorse), these features tend to express themselves differently in different contexts, resisting any attempts to quantify them. The remorse of an offender with a long list of past convictions takes on a different hue than the remorse of a first time offender, suggesting that to assign each of these features a particular weight in

---

<sup>8</sup> Note that this is arguably the case for all moral problems, and yet, as discussed above (e.g. organ donations), in less complex cases, the most important morally relevant features can be selected and built into an adequate system.

<sup>9</sup> Of course, evolution within ethical reasoning does not justify inconsistencies within or across IRBs in the present day, but adopting an algorithmic process of decision-making would restrict the potential for such evolution to take place.

all circumstances would be to miss out on relevant information within sentencing (Maslen 2015). Similarly, potential risks and benefits change shape within research settings, depending on the context, and would be very hard to pin down in a settled system of measurement. A risky protocol submitted by an experienced and conscientious primary investigator (PI) may well be less risky than the very same protocol submitted by a hands-off PI who spends most of her time overseas <sup>10</sup>.

A strong case can be made for the importance of choosing discretionary decision-making in systems of research ethics, given these shared features with systems of justice. In recognizing these features, those grappling with the complex, normative decisions that must be made within sentencing should embrace the second horn of inconsistencies, recognizing that to map out the correct relationship between every case and every outcome is an unachievable and undesirable aim. In embracing discretionary decision-making, the justice system risks leaving room for morally irrelevant features to enter these decisions, but ensures that many of those that are morally relevant, and the intricate relationships between them, are included. As systems that are also tasked with decisions that are evolving, ungrounded, and contextual, research oversight systems are faced with the same dilemma. Given the similarities between systems of justice and systems of research ethics, we argue that like courts, IRBs ought to maintain discretionary decision-making processes. However, as will be discussed below with the case of sentencing guidelines, there are safeguards that can be put into place in order to prevent too many morally irrelevant details from entering decisions within the justice system.

It is worth noting that while systems of justice have had centuries to evolve into the shape they are in today, systems of research governance are, by comparison, very young. In order to embrace the first horn of the dilemma and fix the relationship between moral factors and decisions with regards to proposed research, one must be confident that the moral factors one has identified as inputs, the decisions one has selected as outputs, and the relationships between the two, are the best they can be.

---

<sup>10</sup> Interestingly, it appears that IRB members' confidence in their risk assessments decreases as the riskiness of research increases (Grinnell et al. 2017).

We are a far cry from this stage within research governance. Consider the shifts that have taken place in research ethics in just the past several decades. Single-site, institutionally funded trials have ballooned into multi-site, international, industry funded trials, while technologies such as body cameras and genetic tests have arrived on the scene, introducing novel ethical issues that IRBs must grapple with. The priorities of the field of research ethics have shifted as well, moving away from an initial emphasis on protectionism within research towards one of increasing and encouraging participation in research (Friesen et al. 2017). Discussions regarding dissolving the boundary between research and practice, the importance of granting researched communities ownership over research design and data, and the growth of industry-led investigations which fall outside the current domain of research oversight are all taking place, each of which could fundamentally impact the shape of research governance (Faden et al. 2013, Quigley 2001, Puschmann and Bozdag 2014). In a rapidly shifting realm such as this one, to attempt to specify the process of deliberation within research ethics too closely would be a significant misstep.

One might object to this argument, which paints a picture of justice systems as having chosen the second horn of the dilemma, by pointing towards several efforts to increase consistency within sentencing that look quite a bit like an embrace of the first horn of the dilemma. In the United States, the Federal Sentencing Guidelines, introduced in 1984 and mandatory until 2005, sought to reduce the influence of morally irrelevant factors in sentencing practices by specifying the criteria which ought to shape an offenders' sentence. Two inputs (criminal history and the severity of an offense) were combined to produce a score, which indicated the sentence an offender should receive (Mustard 2001). Similarly, in Australia, a sentencing calculus was introduced, in order to reduce inconsistencies and increase fairness (Bagaric and Pathinayake 2013). However, while these measures included suggested limitations with regards to sentencing, discretion still guided decisions at the end of the day, particularly with the incorporation of aggravating and mitigating factors into a sentence. In the United States, this use of discretion was so significant that in 2002, 35% of sentences fell outside the federal guidelines (Department of Justice 2006). While this use of discretion probably occurred because the sentencing guidelines were unable to capture all morally relevant features, it also certainly allowed for



morally irrelevant criteria, such as race and income, to impact sentencing as well (Mustard 2001). As Bagaric writes of the calculus in Australia, “The complexity of the sentencing calculus, stemming from the large number of aggravating and mitigating factors and the reluctance of the courts and legislature to ascribe weight to them, reduces the parity principle from a concrete principle to an aspirational ideal” (Bagaric and Pathinayake 2013, 400).

### **Safeguards.**

While we argue that the second horn is preferable to the first horn in the realm of research ethics, we also acknowledge that embracing a discretionary process that inevitably involves inconsistency requires safeguards in order to prevent too many morally irrelevant factors from entering the decision process, as well as to ensure that those factors that are morally relevant are given consideration. These safeguards should be aimed both at reducing unfairness and perceptions of unfairness.

Unfairness can be reduced by improving content consistency, which can be achieved by way of improvements in procedural consistency. Many of these are discussed in detail elsewhere, so here we merely flag some of the ways in which procedural consistency can be improved (Emanuel et al. 2004). One important aspect of the decision-making procedure that ought to be consistent is the make-up of the committee evaluating research protocols. This includes the members who sit on the committee and their backgrounds and expertise, as well as the training they received before beginning their membership. Although some have pointed out how consistent expertise and shared training will not be sufficient to ensure consistencies between committees, it is nonetheless an important step towards it (Edwards, Ashcroft, and Kirchin 2004, Sayers 2007). Similarly, guidelines that require the discussion of IRBs to cover several questions or categories in relation to each research protocol can contribute to procedural consistency. While in the US, IRB members take guidance from the Belmont Report’s emphasis on respect for persons, beneficence, and justice (Ryan et al. 1979), and ten ethical “review domains” have been identified to guide IRB discussions in the UK (Trace and Kolstoe 2017), evidence suggests that the vast majority of IRB meetings are spent discussing consent forms, and very

little time is spent on other important topics such as the impact of the research on the community involved (Abbott and Grady 2011, Angell et al. 2007). This suggests that more might be need to be done to ensure that IRB members truly consider the importance of several ethical aspects of a protocol during discussion. Such a process would resemble the current use of Sentencing Guidelines as recommendations, rather than as a mandatory framework for sentencing decisions. In hopes of better understanding the decision-making processes of IRBs, the National Research Ethics Service (NRES) in the UK has been conducting an explicit exercise, the Shared Ethical Debate (ShED), in order to collect details about their decision, rationale and the process of deliberation (Heasman, Gregoire, and Davies 2011, Trace and Kolstoe 2017, Davies 2008). The aim is that the findings from this exercise will then be used to improve standard operating procedures for IRBs and enable each IRB to reflect on how they deliberate and make decisions.

As with trust and trustworthiness, where trust should only be place in processes, individuals, or institutions that are in fact trustworthy, perceptions of IRBs as fair ought only to be improved if the process they engage in is in fact fair. Given this, two interrelated changes that might help to improve both fairness and perceptions of fairness in research oversight are for IRBs to increase the transparency of their processes and to give reasons for their decisions (Klitzman 2012, Lynch 2018). While some argue that there are good reasons to refrain from making IRB decision-making processes available to the public (e.g. confidentiality, freedom of discussion) (Sheehan 2008), allowing investigators to peek within the black box of IRB reasoning may also help them to better understand and learn about the common concerns discussed within research ethics. Lynch proposes several ways that IRB transparency could be increased, including releasing redacted IRB minutes, sharing information about internal processes (e.g., timelines, mechanisms for self-quality improvement), or opening up IRB meetings to various stakeholders (e.g., investigators, the public, participants) (Lynch 2018). While it is common in some countries (e.g., the UK) and uncommon in others (e.g., the US), many have suggested that inviting investigators to discuss their protocols at IRB meetings can have a significant reduction in tensions and frustrations between IRBs and investigators (Stark 2011, Klitzman 2012). Furthermore, increases in IRB transparency may not only lead to improvements in

investigators' perceptions of fairness in IRB functioning, but may also increase public and participant trust in the process of research ethics review, improve the efficiency of IRB processes, and decrease inconsistencies within and across IRBs (Lynch 2018).

A particular form of transparency that is likely to have a significant impact on perceptions of fairness involves transparency with regards to the reasons behind IRB decisions and requests <sup>11</sup>. As Lynch has pointed out, federal regulations in the United States require IRBs to provide reasons for their decisions when they disapprove a protocol, but not when they request modifications for one, which is much more common. As she suggests, providing reasons for modifications can help investigators to learn and encourage self-reflection amongst IRB members, as “the process of articulating reasons may cause IRBs to reassess their own views” (Lynch 2018, 153). The connection between fairness and giving reasons for one's decisions also falls out of the research of Keith-Spiegel and colleagues discussed above, which measures the significance that researchers place on the importance of fairness in research ethics processes, and the link between retaliatory behavior and perceptions of unfairness. The authors suggest that requiring IRB members to explain their requests for modifications of protocols in ethical terms to researchers may well reduce perceptions of injustice, and thereby reduce noncompliant behaviors (Keith-Spiegel, Koocher, and Tabachnick 2006). As discussed above, when only the content of decisions is available, perceptions of the fairness of a process are filled in by whatever one is able to imagine, and when the decision is not the desired one, the imagination is not always generous. Such a requirement may also have the effect of reducing the number of unsubstantiated requests made by IRBs (e.g. changing the font), given that members will be required to reflect on each request and offer an ethical reason in support of it. Furthermore, it is also possible that increased transparency of IRB deliberations may give researchers insight into ethical issues

---

<sup>11</sup> Drawing on Daniels and Sabin's concept of 'accountability for reasonableness', McGuinness also emphasizes the importance of reason giving in IRB decision making (McGuinness 2008, Daniels and Sabin 1997).

related to their research that could aid them in the development of future protocols, making everyone's job easier<sup>12</sup>.

## **Conclusion**

In response to persisting concerns related to inconsistency in the ethical review of research, we have offered a partial defense of such inconsistency. This defense rests on the recognition of a dilemma faced by all who aspire to consistently approach complex and normative decisions. This dilemma results in either the achievement of consistency, but the exclusion of morally relevant factors, or inconsistency, but the inclusion of both morally relevant and irrelevant factors. As a result of similarities between systems of research governance and systems of justice, we argue that the second horn is preferable to the first when approaching the decisions faced by IRBs, but warn that certain safeguards should be taken in order to prevent too many morally irrelevant factors from entering the decision-making process.

---

<sup>12</sup> It is worth noting that another aspect of IRB decision-making that can improve perceptions of fairness is the availability of an appeals process. Currently, such a process is available in several countries (e.g. the UK, Australia, Canada), but not in the US.

## References

- Abbott, Lura, and Christine Grady. 2011. "A systematic review of the empirical literature evaluating IRBs: what we know and what we still need to learn." *Journal of Empirical Research on Human Research Ethics* 6 (1):3-19.
- Angell, Emma L, Alan Bryman, Richard E Ashcroft, and Mary Dixon-Woods. 2008. "An analysis of decision letters by research ethics committees: the ethics/scientific quality boundary examined." *BMJ Quality & Safety* 17 (2):131-136.
- Angell, Emma L, CJ Jackson, RE Ashcroft, A Bryman, K Windridge, and Mary Dixon-Woods. 2007. "Is 'inconsistency' in research ethics committee decision-making really a problem? An empirical investigation and reflection." *Clinical Ethics* 2 (2):92-99.
- Angell, EMMA, Alex J Sutton, KATE Windridge, and Mary Dixon-Woods. 2006. "Consistency in decision making by research ethics committees: a controlled comparison." *Journal of Medical Ethics* 32 (11):662-664.
- Bagaric, Mirko, and Athula Pathinayake. 2013. "The Paradox of Parity in Sentencing in Australia: The Pursuit of Equal Justice That Highlights the Futility of Consistency in Sentencing." *The Journal of Criminal Law* 77 (5):399-416. doi: 10.1350/jcla.2013.77.5.866.
- Baumgartner, Frank R, Emma Johnson, Colin Wilson, and Clarke Whitehead. 2015. "These Lives Matter, Those Ones Don't: Comparing Execution Rates by the Race and Gender of the Victim in the US and in the Top Death Penalty States." *Alb. L. Rev.* 79:797.
- Byrne, Margaret M, Jeanne Speckman, Ken Getz, and Jeremy Sugarman. 2006. "Variability in the costs of institutional review board oversight." *Academic Medicine* 81 (8):708-712.
- Caulfield, Timothy, Nola Ries, and Graham Barr. 2011. "Variation in ethics review of multi-site research initiatives." *Amsterdam LF* 3:85.
- Daniels, Norman, and James Sabin. 1997. "Limits to health care: fair procedures, democratic deliberation, and the legitimacy problem for insurers." *Philosophy & public affairs* 26 (4):303-350.
- Davies, H. 2008. "Standards for research ethics committees: purpose, problems and possibilities."
- Dawson, Angus J, and Steve M Yentis. 2007. "Contesting the science/ethics distinction in the review of clinical research." *Journal of medical ethics* 33 (3):165-167.
- Department of Health. 1997. Multi-Centre Research Committees (HSG(97)23). Richmond House, 79 Whitehall, London, SW1A 2NS.
- Department of Justice. 2006. FACT SHEET: THE IMPACT OF UNITED STATES V. BOOKER ON FEDERAL SENTENCING
- Driscoll, Andrea, Judy Currey, Linda Worrall-Carter, and Simon Stewart. 2008. "Ethical dilemmas of a large national multi-centre study in Australia: time for some consistency." *Journal of clinical nursing* 17 (16):2212-2220.
- DuBois, James M. 2004. "Is Compliance a Professional Virtue of Researchers? Reflections on Promoting the Responsible Conduct of Research." *Ethics & Behavior* 14 (4):383-395. doi: 10.1207/s15327019eb1404\_8.
- Dziak, Kathleen, Roger Anderson, Mary Ann Sevick, Carol S Weisman, Douglas W Levine, and Sarah Hudson Scholle. 2005. "Variations among Institutional Review Board reviews in a multisite health services research study." *Health services research* 40 (1):279-290.
- Edwards, S. J., R. Ashcroft, and S. Kirchin. 2004. "Research ethics committees: differences and moral judgement." *Bioethics* 18 (5):408-27.
- Edwards, S. J., T. Stone, and T. Swift. 2007. "Differences between research ethics committees." *Int J Technol Assess Health Care* 23 (1):17-23. doi: 10.1017/s0266462307051525.
- Emanuel, Ezekiel J, Anne Wood, Alan Fleischman, Angela Bowen, Kenneth A Getz, Christine Grady, Carol Levine, Dale E Hammerschmidt, Ruth Faden, and Lisa Eckenwiler. 2004. "Oversight of

- human participants research: identifying problems to evaluate reform proposals." *Ann Intern Med* 141 (4):282-91.
- Faden, Ruth R., Nancy E. Kass, Steven N. Goodman, Peter Pronovost, Sean Tunis, and Tom L. Beauchamp. 2013. "An Ethics Framework for a Learning Health Care System: A Departure from Traditional Research Ethics and Clinical Ethics." *Hastings Center Report* 43 (s1):S16-S27. doi: 10.1002/hast.134.
- Freedman, Benjamin. 1987. "Scientific value and validity as ethical requirements for research: a proposed explication." *IRB: Ethics & Human Research* 9 (6):7-10.
- Friesen, Phoebe, Lisa Kearns, Barbara Redman, and Arthur L. Caplan. 2017. "Rethinking the Belmont Report?" *The American Journal of Bioethics* 17 (7):15-21. doi: 10.1080/15265161.2017.1329482.
- Friesen, Phoebe, Barbara Redman, and Arthur Caplan. 2018. "Of Straws, Camels, Research Regulation, and IRBs." *Therapeutic innovation & regulatory science*:2168479018783740.
- Goldman, Jerry, and Martin D Katz. 1982. "Inconsistency and institutional review boards." *Jama* 248 (2):197-202.
- Green, Lee A., Julie C. Lowery, Christine P. Kowalski, and Leon Wyszewianski. 2006. "Impact of Institutional Review Board Practice Variation on Observational Health Services Research." *Health Services Research* 41 (1):214-230. doi: 10.1111/j.1475-6773.2005.00458.x.
- Grinnell, Frederick, John Z Sadler, Victoria McNamara, Kristen Senetar, and Joan Reisch. 2017. "Confidence of IRB/REC Members in Their Assessments of Human Research Risk: A Study of IRB/REC Decision Making in Action." *Journal of Empirical Research on Human Research Ethics* 12 (3):140-149.
- Heasman, Peter, Alain Gregoire, and Hugh Davies. 2011. "Helping Research Ethics Committees share their experience, learn from review and develop consensus: an observational study of the UK Shared Ethical Debate." *Research Ethics* 7 (1):13-18.
- Hirshon, Jon Mark, Scott D. Krugman, Michael D. Witting, Jon P. Furuno, M. Rhona Limcangco, Andre R. Perisse, and Elizabeth K. Rasch. 2002. "Variability in Institutional Review Board Assessment of Minimal-risk Research." *Academic Emergency Medicine* 9 (12):1417-1420. doi: 10.1197/aemj.9.12.1417.
- Keith-Spiegel, P., G. Koocher, and B. Tabachnick. 2006. "What scientists want from their research ethics committee." *J Empir Res Hum Res Ethics* 1 (1):67-82. doi: 10.1525/jer.2006.1.1.67.
- Keith-Spiegel, Patricia, and Gerald P Koocher. 2005. "The IRB paradox: could the protectors also encourage deceit?" *Ethics & Behavior* 15 (4):339-349.
- Khan, Muhammad A., Michelle S. Barratt, Scott D. Krugman, Janet R. Serwint, and Marilyn Dumont-Driscoll. 2014. "Variability of the Institutional Review Board Process Within a National Research Network." *Clinical Pediatrics* 53 (6):556-560. doi: 10.1177/0009922814527504.
- Klitzman, Robert. 2011. "The myth of community differences as the cause of variations among IRBs." *AJOB primary research* 2 (2):24-33.
- Klitzman, Robert. 2012. "From anonymity to "open doors": IRB responses to tensions with researchers." *BMC research notes* 5 (1):347.
- Klitzman, Robert. 2015. *The ethics police?: the struggle to make human research safe*: Oxford University Press.
- Lynch, Holly Fernandez. 2018. "Opening Closed Doors: Promoting IRB Transparency." *The Journal of Law, Medicine & Ethics* 46 (1):145-158.
- Maslen, Hannah. 2015. "Penitence and Persistence: How Should Sentencing Factors Interact?" In *Exploring Sentencing Practice in England and Wales*, 173-193. Springer.
- McGuinness, Sheila. 2008. "Research ethics committees: the role of ethics in a regulatory authority." *Journal of Medical Ethics* 34 (9):695-700.
- McWilliams, Rita, Julie Hoover-Fong, Ada Hamosh, Suzanne Beck, Terri Beaty, and Garry Cutting. 2003. "Problematic variation in local institutional review of a multicenter genetic epidemiology study." *Jama* 290 (3):360-366.

- Mhaskar, R., E. B. Pathak, S. Wieten, T. M. Guterbock, A. Kumar, and B. Djulbegovic. 2015. "Those Responsible for Approving Research Studies Have Poor Knowledge of Research Study Design: a Knowledge Assessment of Institutional Review Board Members." *Acta Inform Med* 23 (4):196-201. doi: 10.5455/aim.2015.23.196-201.
- Mustard, David B. 2001. "Racial, ethnic, and gender disparities in sentencing: Evidence from the US federal courts." *The Journal of Law and Economics* 44 (1):285-314.
- Needham, Allison C., Mufiza Z. Kapadia, and Martin Offringa. 2015. "Ethics Review of Pediatric Multi-Center Drug Trials." *Pediatric Drugs* 17 (1):23-30. doi: 10.1007/s40272-014-0098-9.
- Page, Stacey A., and Jeffrey Nyeboer. 2017. "Improving the process of research ethics review." *Research Integrity and Peer Review* 2 (1):14. doi: 10.1186/s41073-017-0038-7.
- Petersen, L. A., K. Simpson, R. SoRelle, T. Urech, and S. Chitwood. 2012. "How variability in the institutional review board review process affects minimal-risk multisite health services research." *Annals of Internal Medicine* 156 (10):728-735. doi: 10.7326/0003-4819-156-10-201205150-00011.
- Puschmann, Cornelius, and Engin Bozdog. 2014. "Staking out the unclear ethical terrain of online social experiments." *Internet Policy Review*.
- Quigley, Dianne. 2001. "Compilation on Environmental Health Research Ethics Issues with Native Communities." *Syracuse, NY: Syracuse Initiative for Research Ethics in Environmental Health*.
- Rawls, John. 1971. *A Theory of Justice*: Harvard University Press.
- Ryan, KJ, JV Brady, RE Cooke, DI Height, AR Jonsen, P King, K Lebacqz, DW Louisell, DW Seldin, E Stellar, and RH Turtle. 1979. *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research*. edited by United States National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Washington, D.C.
- Sayers, Gwen M. 2007. "Should research ethics committees be told how to think?" *Journal of medical ethics* 33 (1):39-42.
- Schneider, Carl. 2015. *The Censor's Hand: The Misregulation of Human-Subject Research*: MIT Press.
- Shah, Seema, Amy Whittle, Benjamin Wilfond, Gary Gensler, and David Wendler. 2004. "How do institutional review boards apply the federal risk and benefit standards for pediatric research?" *Jama* 291 (4):476-482.
- Sheehan, Mark. 2008. "Should research ethics committees meet in public?" *Journal of medical ethics* 34 (8):631-635.
- Silverman, Henry, Sara Chandros Hull, and Jeremy Sugarman. 2001. "Variability among institutional review boards' decisions within the context of a multicenter trial." *Critical care medicine* 29 (2):235.
- Skarlicki, Daniel P., and Robert Folger. 1997. "Retaliation in the Workplace: The Roles of Distributive, Procedural, and Interactional Justice." *Journal of Applied Psychology* 82 (3):434-443.
- Stair, Thomas O, Caitlin R Reed, Michael S Radeos, Greg Koski, Carlos A Camargo, and THE MARC\*. 2001. "Variation in institutional review board responses to a standard protocol for a multicenter clinical trial." *Academic emergency medicine* 8 (6):636-641.
- Stark, AR, JE Tyson, and PL Hibberd. 2010. "Variation among institutional review boards in evaluating the design of a multicenter randomized trial." *Journal of Perinatology* 30 (3):163.
- Stark, L., and J. A. Greene. 2016. "Clinical Trials, Healthy Controls, and the Birth of the IRB." *N Engl J Med* 375 (11):1013-5. doi: 10.1056/NEJMp1607653.
- Stark, Laura. 2011. *Behind closed doors: IRBs and the making of ethical research*: University of Chicago Press.
- The National Institutes of Health. 2016. *Final NIH Policy on the Use of a Single Institutional Review Board for Multi-Site Research*.
- Trace, Samantha, and Simon Erik Kolstoe. 2017. "Measuring inconsistency in research ethics committee review." *BMC medical ethics* 18 (1):65.

- Vadeboncoeur, Claudia, Nick Townsend, Charlie Foster, and Mark Sheehan. 2016. "Variation in university research ethics review: reflections following an inter-university study in England." *Research Ethics* 12 (4):217-233.
- Vick, Catherine C, Kelly R Finan, Catarina Kiefe, Leigh Neumayer, and Mary T Hawn. 2005. "Variation in Institutional Review processes for a multisite observational study." *The American Journal of Surgery* 190 (5):805-809.
- Whitney, Simon N, Kirsten Alcer, Carl E Schneider, Laurence B McCullough, Amy L McGuire, and Robert J Volk. 2008. "Principal investigator views of the IRB system." *International journal of medical sciences* 5 (2):68.