Necessity and proportionality: towards a balanced approach?

Book or Report Section

Accepted Version

Bilchitz, D. (2014) Necessity and proportionality: towards a balanced approach? In: Lazarus, L., McCrudden, C. and Bowles, N. (eds.) Reasoning Rights: Comparative Judicial Engagement. Hart Publishing, Oxford, pp. 41-62. ISBN 9781849462525 Available at http://centaur.reading.ac.uk/93175/

NECESSITY AND PROPORTIONALITY: TOWARDS A BALANCED APPROACH?

David Bilchitz[1]

Domestic courts and international tribunals are often faced by government measures that place limits upon fundamental rights. The circumstances under which such a limitation may be justified are generally laid out in very broad terms in limitation clauses which differ in a variety of ways. [2] Despite these differences, what is notable, in recent years, is that the bodies responsible for making decisions in relation to fundamental rights have exhibited a remarkable convergence upon a general approach towards the limitation of rights that centers on the notion of proportionality.

The proportionality enquiry ultimately seeks to evaluate the benefits to be achieved by the infringing measure against the harms caused through violating fundamental rights. Judges have developed a particular reasoning process to give structure to such an analysis. The first part of this process involves considering the purpose of the measure that limits a fundamental right. Jurisdictions vary on how they characterize this stage: in Germany, for instance, the purpose

[1] Professor, Faculty of Law, University of Johannesburg; Director, South African Institute for Advanced Constitutional, Public, Human Rights and International Law; Secretary-General, International Association of Constitutional Law

[2] A Barak 'Proportional Effect: The Israeli Experience' (2007) 57 *University of Toronto Law Journal* 369-370 states that 'Limitation clauses are central to the understanding of constitutional democracy and of judicial review of the constitutionality of statutes. They are an expression of the democratic values of society'. In some countries, such as Canada or South Africa, there are general limitation clauses; in others like Germany there are specific clauses authorizing the violation of particular rights under particular conditions.

must simply be a 'legitimate purpose';[3] in Canada, the objective must be of 'sufficient importance to warrant overriding a constitutionally protected right or freedom'.[4]

The second part of this process is the proportionality enquiry proper which 'examines the relationship between the object and the means of realizing it. Both the object and the means must be proper. The relationship between them is an integral part of proportionality'.[5] There are three key components to the proportionality test in this regard. The first stage requires that the infringing measures be 'rationally connected to the objective'.[6] I shall refer to this as the 'suitability requirement' which essentially holds that a measure that infringes a right can only be justified if it is suitable for realizing the purpose it is designed to achieve. The second stage requires that the means 'impair "as little as possible" the right or freedom in question'.[7] I shall refer to this as the 'necessity' requirement. Finally, the third stage requires that the benefits of the infringing measure must be proportional to the violation of fundamental rights caused. This I shall term the 'proportionality stricto sensu requirement' which ultimately involves weighing up the harms caused to fundamental rights against the benefits of the infringing measure. At this stage, for instance, '[t]he more deleterious effects of a measure, the more important the

[3] See D Grimm 'Proportionality in Canadian and German Constitutional Jurisprudence' (2007) 57 *The University of Toronto Law Journal* 383, 388.

[4] *R v Big M Drug Mart Ltd* [1985] 1 SCR 295 at 352.

[5] Barak above n 1 at 371.

[6] *R v Oakes* [1986] 1 SCR 103 at [70].

[7] Ibid.

objective must be if the measure is to be reasonable and demonstrably justified in a free and democratic society'.[8]

Each stage of the proportionality enquiry is deserving of detailed analysis. This paper will focus upon the 'necessity' component which has often been of great importance in the jurisprudence of courts around the world in determining whether a particular measure unacceptably violates fundamental rights. In the first section, I outline what I term the 'strict interpretation of necessity'. This understanding has been defended seminally by Robert Alexy who provides a logical derivation for it from a theoretical view of fundamental rights as principles or 'optimisation requirements'. This interpretation, however, gives rise to two particular problems when it has been applied by courts. The first problem relates to its being formulated in a manner that is too strong: this would reduce the chances that any limitation would pass constitutional muster. Courts have thus found various methods of circumventing the test which then reduces the protection it offers to fundamental rights. The second problem relates to the necessity enquiry being too weak and thus having little value in the judicial review of measures that infringe fundamental rights. The arguments in this part of the paper demonstrate the need for a more detailed engagement with the various sub-components of the necessity enquiry, which I seek to accomplish in the second part of this paper.

Four components of the necessity enquiry are distinguished and analysed: possibility, instrumentality, impact, and comparativity. Each is shown to involve both normative and qualitative judgments that prevent the enquiry from neatly being reduced to a notion such as

[8] Ibid. at [71].

'optimisation'. Moreover, when the qualitative dimensions of necessity are understood, the enquiry also can be seen to require an element of balancing within it that cannot be eliminated. This analysis provides support for the approach of Courts – such as in Canada – where the necessity stage of the proportionality enquiry has assumed greater importance. I seek to show, however, that, despite necessity including an element of balancing within it, the second and third stages of the proportionality enquiry should not be collapsed. This is important for courts and academic writers who are concerned about the broad discretion and ad hoc nature of the reasoning often present at the third stage of the enquiry. This article allows for an alternative approach to the third stage to be developed (such as is evident in the work of Jochen von Bernstorff in this volume) yet suggests that a more restrained form of balancing in the necessity enquiry cannot be dispensed with. I conclude by outlining a revised moderate interpretation of necessity that provides a much better conceptual framework and guided process of reasoning which courts can employ to test the constitutionality of laws and executive action that infringe fundamental rights.

PART I: TWO PROBLEMS WITH STRICT NECESSITY

1.1 The strict formulation of necessity

In examining the question of necessity, it is important to start from how it has been formulated by courts in various jurisdictions. The German Constitutional Court - which is the first court to have made extensive use of the proportionality enquiry in relation to fundamental rights – holds that a statute which limits a fundamental right 'is necessary if the legislator could not have chosen a different means which would have been equally effective but which would have

infringed on fundamental rights to a lesser extent or not at all'.[9] Similarly, the Canadian Constitutional Court in the celebrated case of *R v Oakes*, formulates the requirement as entailing that the means 'should impair "as little as possible" the right or freedom in question'.[10] The Israeli Supreme Court has similarly stated that the means used 'must injure the individual to the least extent possible. In the spectrum of means which can be used to achieve the objective, the least injurious means must be used'.[11]

These formulations share a high degree of similarity and express, what I shall term, the 'strict interpretation of necessity' ('SN'). Analytical justification for this formulation of the test of necessity is given in the impressive work of Robert Alexy. The principle of necessity flows logically, according to Alexy from a particular conception of fundamental rights. For Alexy, rights are principles rather than rules. Rules are norms that are always either fulfilled or not; whereas principles are 'norms which require that something be realized to the greatest extent possible given the legal and factual possibilities'.[12] This characterization of principles has implications for how to deal with conflicts between them: it means that where they conflict, one principle has to be weighed against the other and a determination has to be made as to which has greater weight in this context.[13] Alexy contends that this process is governed by the

[9] Cannabis decision BVerfGE 90 (9 March 1994) at [172].

[10] *Oakes* above n 5 at [70].

[11] *Beit Sourik Village Council v The Government of Israel* HJC 2056/04 at [41] ('Beit Sourik case').

[12] R Alexy *A Theory of Constitutional Rights* (Oxford, Oxford University Press, 2002) 47.

[13] Ibid. 50.

principle of proportionality: indeed, 'the nature of principles implies the principle of proportionality'.[14]

The strong interpretation of necessity, according to Alexy, is entailed logically by an understanding of rights as optimization requirements.[15] If rights are norms that must be optimized to the greatest extent possible, then when engaged in proportionality analysis, courts should adopt a stringent approach ensuring that any limitation of a right that is being proposed should strictly speaking be *necessary* to realize the purpose sought to be achieved. This means that no other alternative must be available that can equally realize the purpose and be less invasive of the right in question.

To see why this is so, it is important to recognize that the notion of principles as optimization requirements means that the outcome must be adopted that allows for the greatest possible realization of each principle so far as this is compatible (both factually and legally) with the greatest realization of the others. [16] Thus, any degree of intrusion into Principle 1 (a right) may only be allowable to the extent necessary to realize another vital competing Principle 2 (an important social objective). If another means is available that equally realizes Principle 2, and has a lesser impact on Principle 1, then Principle 1 is not being optimized to the greatest extent possible compatible with the realization of Principle 2. This strict interpretation of necessity, thus, can be seen to derive from the characterization of rights as 'optimisation requirements'.

[14] Ibid. 66.

[15] Ibid. 67-68 and 399.

[16] K Möller 'Balancing and the Structure of Constitutional Rights' (2007) 3 *International Journal of Constitutional Law* 453, 459.

Part of the problem with Alexy's argument relates to making sense of the notion of 'optimisation' in this context. It is important to recognize that Alexy's language is metaphorical. The notion of 'optimization' in a field such as economics, where clear quantification is possible, has a clear meaning: we clearly understand what it means, for instance, to optimize profit. However, the exact meaning of this term in the legal or moral context is less clear as the principles and values to be evaluated and balanced are qualitative and not quantifiable in the same manner.[17] I shall, in the second part of the paper, consider the extent to which the language of optimization is apposite when engaging with the various sub-components of the necessity enquiry.

2.2 *Is Strict Necessity Too Strong?*

Two significant problems flow from the strict interpretation of necessity, both of which have a significant impact on the protection it accords to fundamental rights.[18] The first problem relates to the fact that the strict interpretation places a considerable burden upon other branches of government to justify any measure that limits a right. Requiring there to be no other measure that equally realizes the purpose and has a less restrictive impact on the right could allow for very few limitations on fundamental rights to pass constitutional muster. As Blackmun J famously wrote '[a] Judge would be unimaginative indeed if he could not come up with something a little less 'drastic' or a little less 'restrictive' in almost any situation, and thereby

[17] K Möller ibid. 462.

[18] An analysis of the various sub-components of necessity in the second part of this paper will also show why the language of 'optimization' is not apposite in many ways.

enable himself to vote to strike the legislation down'.[19] If the strict interpretation were to be applied scrupulously, judges would be required to consider all possible alternatives that could realize the government objective and be less restrictive on the right. The identification of any alternative that was even slightly less restrictive and equally effective would be sufficient to defeat a legislative measure. As a result, if the strict interpretation were to be applied rigorously, it would be very difficult for any law – however well-motivated - to pass the proportionality enquiry with the result that courts would often be striking down legislation as unconstitutional.

This would be problematic for a variety of reasons. The first problem is that the test may prevent limitations from passing constitutional muster that are indeed normatively justified when the balance of reasons is considered. The strict interpretation means that any slightly less drastic means would be sufficient to defeat a legislative measure, however well-considered it is. To illustrate this point, consider the case before the German Constitutional Court challenging the regulatory requirement that tobacco manufacturers had to put warnings on cigarette packets to protect people from the attendant health risks. [20] Alexy, in his discussion of the case, argues that it involves weighing what he terms the 'minor' infringement on the freedom of profession of tobacco manufacturers against the sizable impact of tobacco smoke on people's health. He thus concludes that the outcome of the balancing process here is 'obvious'[21] and

[19] *Illinois State Board of Elections v Socialist Workers Party* et al 440 US 173 (1979) at 188-9

[20] The case is BVerfGE 95, 173. Alexy n 11 above discusses the case at 402-3.

[21] Alexy ibid. quoting the Court ibid. at 187.

the government measure is justified. However, if we consider the matter from the perspective of the necessity enquiry, Alexy's conclusion, in my view, is too quick.

Alternative measures could easily be thought of to achieve the purpose which do not require tobacco producers to place health warnings on their packets (and thus be less restrictive of the right in question). Consider, for instance, requiring manufacturers to contribute towards broader public educational measures, or a major advertising campaign which would in all likelihood be a more effective means to achieve the public health objectives set by the government and be less restrictive of the rights of manufacturers. A strict interpretation of necessity might require a court to strike down this measure on these grounds. Furthermore, it will always be difficult to determine whether the particular measure adopted by the government is constructed in the manner least restrictive of rights. In the context of the health warnings, there would always be a question as to whether the size of the warning required by the legislature was the least intrusive possible of the manufacturers' right to place commercial information on the cigarette packet?[22]

Apart from the problem that limitations of rights will rarely be capable of being justified on a strict application of this standard, this example also illustrates how the strict interpretation of necessity would lead to a separation of powers problem. Courts would be required to strike down any law that did not meet this strict test and to substitute their judgment concerning which measure is 'least restrictive' for that of the legislature. Doubts may be raised concerning

[22] For a discussion of this example , see D Bilchitz 'Does Balancing Adequately Capture the Nature of Rights?' (2010) 25 *Southern African Public Law* 423, 436.

the competence of the court to do so. A legitimacy problem may also arise if the courts are seen ultimately to narrow democratic, legislative discretion too significantly.[23]

Three types of responses can be identified to address this problem, none of which I shall argue go to the root of the problem which lies in the strict interpretation of necessity itself. The first response involves essentially denying that necessity requires categorically that the legislature adopt the less intensively interfering means. Alexy writes that '[t]he point is simply that if the legislature wishes to pursue its goal further, it may only use the less intensive of the means, or an equally mild means, or a still milder one. That is no optimization to the highest point, but simply a ban on unnecessary sacrifices of constitutional rights'.[24] It is hard to see the force of Alexy's argument here. Indeed, it is true that, when considering a particular limitation of a right, a court could strike it down if there are indeed *any* alternative means that is equally effective and less restrictive of the right. This in fact reinforces the strictness of the test. Moreover, it is not true that the legislature then has the discretion to adopt any alternative means it likes. In order for a revised measure to pass constitutional muster, the alternative must be the 'least intensive' or mildest one possible: if the legislature fails to adopt such a measure, then it will fail to optimize both the objective sought to be achieved and the fundamental right in question and will be acting unconstitutionally. The strict conception of necessity does reduce the discretion of the legislature significantly and, in a system of judicial review, requires intrusive intervention on the part of the judiciary.

[23] This is indeed a worry that several German academics have raised and is the subject of a reply by Alexy n 12 at 389.

[24] Alexy n 12 above 399.

The second response seeks to address the problem of the strict interpretation being too strong by reducing the epistemic burden on the party seeking to justify the limitation of a right. It is argued that determining whether the legislature has adopted the least restrictive means is not a matter that admits of a certain judgment. As Choudhry points out, '[p]ublic policy is often based on approximations and extrapolations from the available evidence, inferences and comparative data, and, on occasion, even educated guesses. Absent a large-scale policy experiment, this is all the evidence that is likely to be available'.[25] The Canadian Supreme Court has recognized that decisions on matters relating to the limitation of rights 'must inevitably be the product of a mix of conjecture, fragmentary knowledge, general experience and knowledge of the needs, aspirations and resources of society'.[26] If we take these points seriously, then it means that a government, when seeking to justify a measure according to the strict interpretation of necessity, will usually be unable to do so to any high degree of certainty. This opens the door to reducing the strength of the necessity standard through lessening the epistemic conditions needed to satisfy the standard.

It is quite clear that if no evidence is produced that the test of necessity is met, then the protection of rights will be weakened significantly; on the other hand, too strong a requirement seems unreasonable in light of prevailing evidence.[27] Different possibilities have been suggested to address this problem. Some judges in the Israeli Supreme Court have sought to

[25] S Choudhry 'So What is the Real Legacy of Oakes? Two Decades of Proportionality Analysis under the Canadian Charter's Section 1' (2006) 32 *Supreme Court Law Review* 512, 524.

[26] *Mckinney v University of Guelph* [1990] 3 SCR 229 at 304.

[27] See Alexy n 11 above 417-8.

impose a presumption that a law of parliament is constitutional, claiming that those challenging the law have the task of proving its unconstitutionality. Goldberg J for instance writes: '[t]he party defending the law need not show that there are other alternatives that more severely infringe the right and that the less-infringing alternative was chosen, but rather the party arguing against the validity of a law must show that there exists a specific, clear alternative that fulfills the proper purpose, while infringing the protected right in a manner that is significantly less than the infringement of the law'.[28] Such a presumption in favour of the necessity test having been met means that a court can avoid going through this reasoning process unless convincing evidence is presented to establish that alternative means exist. Part of the point of the necessity enquiry is to force courts to engage closely with the effect of the government's measure and possible alternatives in relation to the stated objective and the fundamental right in question. A presumption of constitutionality would usually involve an abdication of the court's responsibility to reason through whether the requirement of necessity has been met or not, thus weakening any protection it offers to fundamental rights significantly. Moreover, such a generalized deferential presumption offers very weak protection for fundamental rights and appears to go against the very normative importance they are to have in a system with judicial review. Fundamental rights are not just any normative consideration: they are particularly strong protections that can only be limited on the basis of a strong justification. This requires there to be a presumption against limiting rights rather than a presumption in favour thereof.

[28]See the judgments of Cheshin, Goldberg and Bach JJ in *United Mizrachi Bank v Migdal Cooperative Village* CA 6821/ 93, PD 49(4) 221 (1994) at 421.

In the face of the evidentiary problems that exist in dealing with necessity, the Canadian court has developed a test requiring a 'reasonable basis' for the judgment that a particular measure is necessary.[29] Alexy has sought to address this problem by developing what he terms a second law of balancing. That law states '[t]he more heavily an interference in a constitutional right weighs, the greater must be the certainty of its underlying premises'.[30]

All these approaches recognise that complete epistemic certainty is not possible in judging whether the necessity test has been met or not. However, it is of vital importance to distinguish between the various tests involved in determining proportionality and the evidentiary basis necessary to make a judgment that these tests are satisfied. It is this distinction which points to a more generalized difficulty with an approach that addresses too strict a test of necessity through reducing the evidence required to meet the test. To ostensibly retain the strict interpretation of necessity yet effectively weaken it through fiddling with the epistemic conditions for its realization is to reduce the clarity or guidance as to what is required to justify the limitation of a fundamental right. We thus need to develop a test that is sufficiently protective of rights and can be met by prevailing epistemic conditions rather than utilizing an unrealistic test and then undermining it through reducing the evidence necessary to establish that it has been satisfied. Otherwise, it is unclear in fact what the evidence we have must be directed at proving and we are left with judgments that will be ad hoc and arbitrary. If the requirements of the test are to be understood differently, then this should be addressed

[29] *Irwin Toy v Quebec (Attorney General)* [1989] 1 SCR 927 at 994.

[30] Alexy n 11 above 418. I cannot within the scope of this paper evaluate these responses to the problem of epistemic uncertainty.

directly. Fiddling with the evidence in order to reduce the requirements of the test will cause significant confusion and thus reduce any possible protection afforded to fundamental rights by the necessity test significantly.

The third and final response to addressing the problem that the formulation of the strict interpretation of necessity renders the test too strong is to focus on *who* should determine whether the test is met or not. Courts have often adopted a strategy of 'deference' to the legislature concerning what constitutes the 'least restrictive means'. The German Constitutional Court, for instance, in its judgment on the constitutionality of criminalizing the possession of and dealing in cannabis, immediately qualifies the requirement of necessity as follows: '[i]n forming a judgment as to whether the chosen means is suitable and necessary for achieving the desired goals the legislator has a certain degree of discretion. The same applies to the estimation and prediction of the dangers which threaten other individuals or the public good which must be undertaken in this context. The Federal Constitutional Court can only review the exercise of this discretion to a limited extent, the precise extent depending on the nature of the subject in question, the feasibility of forming a sufficiently clear view, and the nature of the legal interests which are at stake'.[31]

Where the strict interpretation of necessity thus appears to lead to an unreasonable outcome or to constrain the legislature too much, the court can claim that it must defer to the other branches of government in judging whether the standard has in fact been met. The German Constitutional Court is not alone in its expressions of the need for deference in relation to the

[31] Cannabis decision n 9 above 173.

necessity enquiry. The Canadian Courts[32], Israeli courts[33] and South African courts[34] have all expressed similar sentiments in some of their jurisprudence.

However, the approach - of adopting a strict interpretation of the requirement and then being highly deferential to the other branches as to whether it has been met- suffers from several flaws. First, it seems to attack the effect rather than the cause: the separation of powers problem appears to arise from an unduly strict interpretation of the necessity requirement. The court attempts to remedy the problem by deferring to the legislature rather than addressing the standard itself. Secondly, by shifting the focus in this way, the court in fact may abdicate its proper responsibility in a society that accepts judicial review, to determine whether a violation of a right is justifiable or not. That enquiry requires determining whether such a measure is proportionate, which in turn requires that the court be satisfied that the principle of necessity is met. For a court to defer to other branches of government is this regard, is for it to avoid actually determining whether a particular violation of a right meets the requirements of this test or not. Either a limitation of a right must be necessary or not; it makes no sense to defer the judgment concerning whether this requirement has been met to the very body whose measure in this regard is being impugned. The deference strategy thus allows us to pretend that a measure is in fact necessary given that parliament has deemed it to be so, when it is not in fact so.

---

[32] *Her Majesty the Queen in Right of the Province of Alberta v Hutterian Brethren of Wilson Colony* [2009] 2 S.C.R. 567 at para 53 ('Hutterian Brethren')

[33] *Beit Sourik Village Council v the Government of Israel* n 11 above paras 40-48 and 58

[34] *S v Manamela* 2000 (3) SA 1 (CC) at [95].

Finally, there is, ironically, a significant danger that the supposedly strict protection offered by necessity for fundamental rights will be significantly weakened by a deferential approach. When, for instance, should courts look deeply into the question of necessity and when should they defer? Without a clear principled basis for deference in some cases rather than others, the necessity enquiry could be effectively avoided in cases where it is most important. For instance, in a case such as the cannabis judgment – where courts face protecting rights in the face of a high degree of controversy – how can we be assured that deference will not be accorded too readily? In other cases, the legislature may claim that too little deference is shown. Without clear criteria for determining when this should occur, deference seems to lead to an undue subjectivism on the part of the court, and rendering the protection of rights through a stringent test such as necessity precarious.

Unfortunately, the factors outlined by the German Constitutional Court concerning when deference should be accorded do not appear to be helpful. The nature of the subject suggests that some subjects require less deference than others: without more specification, it is unclear why this is so and to which subjects a more deferential strategy should apply. A similar point applies to the nature of the legal interests at stake. The criterion relating to the 'feasibility of forming a sufficiently clear view' is also extremely vague and once again raises important questions concerning the evidentiary basis upon which necessity is determined.[35] According to Choudhry, the Canadian court too has struggled to articulate an adequate basis for according deference in some cases and not doing so in others. It has also not been consistent in when it

[35] See the discussion of the second response above: see S Choudhry n 25 above 512-521.

accords deference.[36] The ability and consistency of courts to provide clear criteria as to when to defer is thus to be doubted; without doing so, the necessity test loses much of its value.

Having a strict standard with a high level of deference thus grants rights little protection. Indeed, the strictness of the standard can be seen to be part of the problem in that motivates courts to reduce what they expect of the legislature and thus impacts upon the effectiveness of the standard. A better approach, it seems, would be to have a more adequate understanding of the standard which places realistic requirements on the parties before the court and institutional capacity plays a role only insofar as this renders a party better placed to make a judgment on a particular issue. In Part II, I shall elaborate upon what such a realistic approach to necessity entails.

2.3 *Is Strict Necessity Too Weak?*

The strict interpretation of necessity may also lead to the converse problem that the actual test itself is construed in a manner that renders it too weak. This worry arises from the element of the test which requires that any alternative measure considered should be equally effective in realizing the purpose sought to be achieved by the other branches of the government. The question arises as to what is meant by 'equal effectiveness'?

Aharon Barak is a proponent of a strict approach and writes that '[t]he first element of the necessity test examines the question of whether alternative means can fulfill the law's purpose

[36] See who outlines how the Canadian court attempted to outline categories of deference and lack of consistency with which this was applied.

at the same level of intensity and efficiency as the means determined by the limiting law. If such an alternative does not exist, the law is necessary and the necessity test is met' (my emphasis).[37] He goes further and holds that 'the necessity test is based on the assumption that the only change that should be brought about by the alternative means is that the limitation on the constitutional right would be of a lesser extent'.[38] Thus, any increase in any cost associated with an alternative measure would lead it to fail to meet the requirement that it be equally effective in realizing the purpose. Barak claims that a consideration of alternatives which are not equally effective in this way should take place at the final balancing stage of the proportionality enquiry. The strict approach is exemplified by the judgment of the Israeli Supreme Court in the *Beit Sourik Village Council v Government of Israel*[39].

This extremely strict interpretation of the requirement that the measure realize the purpose in an equally effective manner strongly reduces the importance of the necessity test.[40] Barak recognizes that this interpretation creates significant weakness in the test and that, perhaps, for this reason, he no longer views it as being at the "heart" of the proportionality enquiry.[41] The balancing stage of proportionality thus becomes 'the most important of proportionality's

[37] A Barak *Proportionality: Constitutional Rights and Their Limitations* (Cambridge, Cambridge University Press, 2012) 323.

[38] Ibid. 325.

[39] HCJ 2056/04.

[40] See also P Yowell 'Proportionality in United States Constitutional Law' (this volume) 9 who describes this as a 'technical approach' to necessity and agrees that this renders the necessity enquiry of minimal importance.

[41] Barak n 30 339.

tests'.[42] Indeed, this interpretation can render the necessity enquiry virtually useless: if equal effectiveness is interpreted strictly, it may be difficult to show that any alternative measure could ever realize the purpose 'equally'. A lot of importance will also be placed on the level of abstraction and manner in which the purpose is construed which will determine how the level of effectiveness is to be judged. [43]

Barak's interpretation of equal effectiveness unsurprisingly also seems to be at odds with how courts have employed the test. Take, for instance, the case in Germany relating to a regulation that prohibited the selling of confectionery that was not a genuine chocolate product but rather consisted mainly in puffed rice with cocoa powder. The aim of the regulation was to protect consumers from mistaken purchases. Yet, the Federal Constitutional Court found that the trade prohibition was not necessary: labeling could interfere with the rights of manufacturers less whilst 'equally effectively' protecting individuals against deception and confusion.[44] Yet, if we adopt the strict interpretation of equal effectiveness advocated by Barak, it appears that the Court was unjustified in reaching its conclusion. A ban on such products will no doubt be more effective than a labeling requirement and entirely prevent confusion: labeling can be more or less effective and must take into account social scientific facts such as that consumers often do not read labels and some consumers are more literate than others. The court thus inevitably

[42] Ibid. 340.

[43] This is a problem that affects both the suitability and necessity parts of the test: for a discussion, see Ibid. 331-333.

[44] BVerfGE 53, 135: see the discussion of the case in Alexy n 12 398-9.

had to weaken its understanding of the degree to which the purpose was realized in order to reach its conclusion that the government measure was not necessary.

Alexy, in his analysis of this case, appears to approve of the court's reasoning, thus rejecting Barak's strict approach to equal effectiveness. He writes that '[t]he principle of consumer protection (P2) is broadly equally well satisfied by a duty to label (M1) as by a trade prohibition (M2). So for P2 it is [sic] irrelevant whether M1 or M2 is adopted'.[45] Alexy's language - 'broadly equally well satisfied' - suggests that he must realize that a ban realizes the purpose to a greater extent than labeling: yet, as his language suggests, the requirement must be interpreted to include some flexibility in order to render the necessity enquiry meaningful.

This discussion has shown that, if the requirement of equal effectiveness is understood too strictly, it can effectively render the necessity test otiose. This essentially leads to confusion and opens up the space for courts to vary the standard when expedient: courts will adopt a strict approach when they wish to save a government measure but adopt a more flexible approach when they wish to strike such a measure down. The failure adequately to specify what is involved in the test can thus lead it to be applied inconsistently and for it to offer little protection for fundamental rights in controversial cases. This raises the question once again as to how best to understand the content of this component of the necessity enquiry and whether the strict interpretation correctly captures what is at stake in the enquiry. The next section of this article provides a detailed consideration of the various sub-components of necessity and,

[45] Alexy n 12 above 399.

on this basis, then seeks to evaluate its role within the broader proportionality enquiry as a whole.

PART II: TOWARDS A MODERATE INTERPRETATION OF NECESSITY

In Part I, I sought to outline what I termed the strict interpretation of necessity (SN) and several difficulties that it causes. In order to address these problems, it is useful to break down the elements of this interpretation into four sub-components that are central to this enquiry:

(SN1) a range of possible alternatives to the measure the government wishes to employ must be identified ('the possibility component');

(SN2) the relationship between the government measure under consideration, the alternatives identified in SN1 and the objective sought to be achieved must be determined. Only those alternatives that are 'equally effective' in realizing the objective must remain for consideration in the following parts of the test ('the instrumentality component');

(SN3) the differing impact upon fundamental rights of the measure and the alternatives identified in SN2 must be determined ('the impact component'); and

(SN4) given the findings in SN2 and SN3, an overall comparison must be undertaken between the government measure and the possible alternatives and a judgment made concerning whether the measure adopted by the government is the least restrictive of the rights in question that can achieve the government objective in comparison with all other possible and equally effective alternatives ('the comparative component').

I shall now seek to understand in more depth what is involved in each of these sub-components.

2.1 Possibility

On the strict interpretation of necessity, a measure will only be necessary if there is no possible alternative that will equally realize the objective whilst having a lesser impact on the right. This raises the question of what in fact is to be included within the realm of 'all possible alternatives'? Indeed, this notion plays a crucial role in the idea, defended by Alexy, that rights are optimization principles; as we saw, the necessity principle flows logically from this conception in cases where two principles clash. If each principle must be realized to the greatest extent possible, how far does this extend? The field of possibilities that are considered will to a large extent determine how strict the necessity enquiry will be. The wider the notion of possibility applied, the more alternatives there will be to a government measure and thus the harder it will be to show that measure as being necessary. I shall now consider some interpretations of possibility in this context and the difficulty of rendering this parameter entirely fixed.

When referring to what is legally possible (in the context of the balancing requirement), Alexy seems to be employing the idea of logical possibility.[46] The outer limits of logical possibility, however, do not seem to be apposite in the context of the necessity enquiry: first, such an understanding of possibility would render it extremely difficult for the legislature to establish

[46] Alexy n 11 above 51-53.

that an infringing measure met the test of necessity given that alternative logically possible measures could usually be imagined. The second problem is that these imagined logically possibilities may not be physically possible to achieve: an alternative to a measure, for instance, that involved the logical possibility that humans can fly without the aid of an aircraft could surely not pass any meaningful test of necessity.

What then about employing the notion of that which lies within the range of what is physically possible for human beings to achieve? Again, this notion is very broad and seems to place too severe restrictions on the legislature in attempting to justify a particular measure. It also fails to take account of the fact that what is physically possible may not be practically feasible: it may be physically possible to create overhead cable cars to transport people above cities but so expensive and complex that it is not practically feasible.

Why not then use the notion of what is 'practically feasible?' This notion seems more promising yet it is important to recognize that it lacks the specificity of the other notions of possibility mentioned above. Feasibility itself can admit of differing interpretations: does this notion, for instance, include such constraints as economic scarcity (a measure is unaffordable for a country given current lending policies, although it could be affordable if these policies changed) or political sensitivity (a measure cannot be passed as it will cause massive protests and civil disobedience if it is)?[47] The way in which this requirement is interpreted will affect how the necessity test is applied. Feasibility also has certain dangers. If, for instance, political sensitivity is considered, then it can prevent the adoption of measures to accommodate minorities, for

[47] Bilchitz n 21 above 434.

instance, where the majority finds these to be controversial. That would run counter to a rights-based society that is meant to accommodate individual freedom and difference. Economic scarcity could also always be used as a defence against alternatives to a government measure that are perhaps slightly more resource intensive. The standard of practical feasibility, if interpreted too weakly, runs the risk of offering too little protection for rights.

This discussion has sought to demonstrate that the notion of possibility that informs the consideration of alternatives is central to the necessity enquiry and will determine the stringency of the test that is applied. The strict interpretation of necessity assumes this idea is clear and admits of definite application: as I have attempted to show it does not and admits of significant variability. How then should courts proceed in dealing with this factor? What is clear is that, if the necessity test is to have some bite, court cannot simply defer to the government in determining the range of feasible alternatives. They also should not interpret the notion of 'feasibility' too restrictively and in a way that runs counter to their role to ensure strong protection for fundamental rights. Ultimately, both parties must place before the court the leading alternatives they wish to be considered. If a claim is made that one of these is not feasible, then the court must interrogate these arguments and provide reasons for its finding.

An example of a court explicitly considering such a matter occurred in South Africa in *S v Williams.* [48] In this case, the court found the imposition of corporal punishment as a sentence for juvenile offenders to be unconstitutional. The state argued that 'sentencing alternatives for juveniles were limited and that this country did not have a sufficiently well-established physical

[48] *S v Williams* 1995 (3) SA 632 (CC)

and human resource base which was capable of supporting the imposition of alternative punishments'.[49] Langa J, writing for the majority, recognized that the court needed 'to examine available resources to determine whether there are indeed appropriate sentencing options. It has to be borne in mind that the presence of various options in a number of legislative provisions may not always reflect practical realities'.[50] This was not determinative, however, and the court believed it to be 'important that resources should be made available and that they should be utilised properly, so that the values expressed in the Constitution may be upheld and maintained'.[51] The court then proceeded to discuss some of the alternative possibilities that exist for punishing juveniles including correctional supervision and community service.[52] The court here clearly recognizes the need not to be side-tracked by considering 'unfeasible' alternatives; the constraints of the society had to be considered but courts also could not simply defer consideration of what was feasible to the legislature and accept its view of what resources were available. Ultimately, the values of the Constitution had to play a central role in determining what was to be considered a feasible alternative. The case thus shows that, far from being a technical formal enquiry, the very notion of possibility that lies at the heart of necessity requires courts to engage in substantive and normative reasoning relating to the alternatives that are under consideration. Thus, it is evident that - in relation to this very first

[49] Williams para 63.

[50] Ibid at [64]

[51] Ibid.

[52] Ibid. at para 73 and 75.

sub-component of necessity - any simple quantitative notion of 'optimisation' is not apposite and a range of qualitative factors come into the enquiry.

2.2 Instrumentality

Once a court has identified the range of possible measures it will consider, it must turn to considering the relationship between these differing means under consideration and the objective sought to be achieved by the original governmental measure. Again, the notion of optimization suggests that any alternative must realize the objective in question to the same *maximum* degree. Courts have often formulated the requirement as requiring each alternative worthy of consideration to be 'equally effective' in realizing the objective in question. I have considered in the first part of this paper the manner in which too strict an interpretation of this requirement effectively can render the necessity enquiry meaningless.

It is important to recognise at this point that, once again, an understanding of this requirement has been bedeviled by the language of quantification and optimisation which is inapposite in this context. Courts are required to consider legal and policy measures and the extent to which they realize a particular purpose. It is important to chart the various qualitative judgments involved in this process.

First, of course, as the SA Constitutional Court points out, this requirement requires a careful analysis of the purpose of the provision in question.[53] If too strict an interpretation of the

[53] Manamela n 27 above at [96].

purpose is given, then the measure can effectively be guaranteed to be the only way to realize the purpose; if too weak or broad an interpretation is given, then alternatives can be considered that will very weakly achieve the purpose the government wishes to achieve. The manner in which a court characterizes the purpose is thus significant for the necessity enquiry. As Barak puts it, '[t]he more the Court lowers the level of abstraction, the more difficult it is to find less drastic means for the realization of the object'.[54] Barak suggests that courts should determine a law's purpose in accordance with the 'actual purpose designated by the law'.[55] That is not a complete solution, however, since determining the actual purpose usually involves an act of construction on the part of the Court.[56]

Secondly, it is important to recognize that a judgment is required that two (or more) measures realize the purpose 'equally effectively'. The very nature of law and policy means that this judgment does not admit of exact quantification; we are in the realm of a qualitative judgment that the realization of the purpose is 'more or less' equal. That immediately leads to some flexibility in the test itself. Yet, it may be argued that the very notion that an alternative must be equally effective is flawed and imposes too strict a requirement upon alternatives that are under consideration. McLachlin J in the *Hutterian Brethren* case formulates the problem as

[54] Barak n 1 above at 373.

[55] Barak n 30 above 333.

[56] At the necessity stage, it appears to me that the goal of the court is to achieve a bona fide construction of the legislative purpose rather than any detailed evaluation of this purpose. Such an evaluation should be considered when the court considers whether the purpose is legitimate and in relation to whether its realization justifies the infringement of the right at the final 'proportionality stricto sensu' stage. For a contrary view, see P Yowell (note 39 above) 7.

follows: 'the court need not be satisfied that the alternative would satisfy the objective to exactly the same extent or degree as the impugned measure. In other words, the court should not accept an unrealistically exacting or precise formulation of the government's objective which would effectively immunize the law from scrutiny at the minimal impairment stage. The requirement for an 'equally effective' alternative measure…should not be taken to an impractical extreme'. [57] This dictum recognizes that to insist upon an equally effective alternative will often involve too weak a test for the government measure to meet and thus embody too deferential an approach. The Canadian Supreme Court here also recognizes that the purpose of the government can often be realized to differing degrees.

McLachlin J suggests an alternative formulation of the instrumentality sub-component that appears to be more consonant with the legal and policy context and can render the necessity enquiry a meaningful element of the proportionality enquiry. She holds that alternatives should be considered that 'give sufficient protection, in all circumstances to the government's goal…While the government is entitled to deference in formulating its objective, that deference is not blind or absolute. The test at the minimum impairment stage is whether there is an alternative, less drastic means of achieving the objective in a *real and substantial manner*.' (my emphasis)[58].

This re-formulation suggests that courts should not understand the test to require that a measure realize a purpose to exactly the same degree. That strict formulation generally only

[57] *Alberta v Hutterian Brethren of Wilson Company* 2009 SCC 38 at [55].

[58] Ibid. at [55].

provides an excuse for a high degree of deference that offers little protection for fundamental rights. Instead, a more realistic standard is developed that allows courts to grapple with the question whether there are alternatives that realize the government's objective in a 'real and substantial manner'. The qualitative, normative judgment is here made explicit but allows the courts to deal in a more transparent manner with the *degree* to which alternative measures succeed in realizing the objective and whether they cross a more realistic threshold of 'acceptability'.

2.3 Impact

Once the possible measures have been identified that can realize the objective in a real and substantial manner, it becomes necessary to determine the impact of each of those measures on the fundamental rights of individuals. This enquiry requires ascertaining the degree to which a right is affected by the differing measures under consideration. The assumption here is that rights can be affected to differing degrees: some invasions, for instance, may go to the heart of the right whilst others affect it more peripherally.

In order to understand the degree of invasion, it is necessary to have a clear understanding of the underlying justificatory basis for the right as well as some understanding of its content. This is a matter that should often addressed when determining whether there is indeed a prima facie infringement of a right prior to an application of the proportionality test.[59] If the court has

[59] Indeed, in my view, it is crucial for the correct application of proportionality that the courts seek to understand, at the first stage of the enquiry, the substantive content of the right in the particular context and thus the nature of the violation that has taken place. Here, I disagree with P Yowell (note 39 above) 4 this volume who suggests that all human rights adjudication can simply be collapsed into the proportionality enquiry.

not, at this first stage of analysis, engaged with the content of the right and specified the exact nature of the violation, it will need to do so as part of the necessity enquiry; only by doing so, will it be able to estimate the degree of invasion of the right. An assessment at this stage will also have to lack place concerning the *degree* of impact that alternative measures would have on the right. Some 'vagueness' will of course attach to understanding the degree to which a right has been infringed by a measure and the various alternatives under consideration. Again, this is a qualitative normative judgment to be made by a court that does not admit of exact precision. I accept though that it is broadly possible to adopt a justifiable methodology to determine the degree to which a right is infringed.[60] The problem that often arises at this stage is that courts often do not demonstrate a willingness to analyse the content of a right and thus fail to articulate a transparent normative basis upon which to justify the degree of impairment that differing alternatives may cause. Without doing so, however, in the context of the necessity enquiry, courts are left without an adequate basis to determine whether the government has indeed adopted means that are necessary.

2.4 Comparativity

Already at the two previous stages, it could be argued we are required to engage in a form of comparative reasoning. At the instrumentality stage, we are required to consider the relationship between means and ends of the government measure and various alternatives, and, at the impact stage, to consider the the degree of the impact on fundamental rights of the

[60] Alexy n 12 above, for instance, suggests how this may be done with a three-stage scale of 'minor', 'moderate', and 'serious' elements at 402.

government measure and various alternatives. SN4 requires us to bring these two stages together and to make a judgment that involves both of these criteria and determine whether the government measure is the least restrictive means of the feasible alternatives under consideration. This involves evaluating whether an alternative exists that realizes the objective in a real and substantial manner whilst having a lesser impact on the right. A few important points need to be made here.

The first is that the judgment is ultimately comparative in nature. I have already outlined the difficulty of determining the range of alternatives that are considered in relation to what is considered 'possible'. Once the range is determined, courts are required to make a judgment that the measure adopted by the government is 'better than' any others that are available. If the notion of optimization is to be of any use here, it means that the government measure is the best against the background of feasible alternatives. The judgment as to whether it is the 'best' in these circumstances also involves two dimensions: the manner in which it realizes the objective and the impact on fundamental rights. Given the existence of these two axes in question, and the lack of precision in relation to each, the judgment ultimately becomes a more complex one than may initially be expected. Indeed, understanding the sub-components in the manner indicated above entails that the necessity enquiry will, to a degree, include within it an element of balancing.

To see why this is so, consider the following scenario. Let us imagine that a government measure (M1) realizes an Objective (01) to a certain degree and infringes a right to a certain degree (R1). Now, a court considers an alternative (M2) that realizes the objective (O1) to a

lesser (yet real and substantial) degree yet also infringes the right to a lesser degree (R1). Thus, M1 is better than M2 in the realization of the objective (O1); yet M1 is worse than M2 in relation to its impact on the right in question (R1). The objective is still realized by M2 but not as well as in relation to M1; yet there is a clear difference in relation to their respective impact upon rights. The difficult question that arises is whether M1 is in fact necessary and whether less restrictive means exist in these circumstances?

This problem, for instance, arose for the Israeli Supreme Court in a case relating to the erection of a separation barrier between Israelis and Palestinians.[61] The court determined the issue on the basis of the notion of the proportionality principle in international humanitarian law which is similar to its application in the context of limiting rights. The court found that the purpose of building the wall was to provide security for the citizens of Israel who faced terror attacks emanating from the occupied Palestinian territories. The existing route of the wall was challenged as having a drastic impact upon the human rights of Palestinians. The court was presented by the applicants with alternative routes for the barrier which would have a lesser impact upon the lives of Palestinians. The court accepted, however, the perspective of the military that these alternative routes provided less security for Israelis. The Supreme Court thus considered the necessity test to have been met given that the suggested alternative routes would realize the security objective to a lesser extent. Instead, it decided the case in terms of the balancing component of proportionality (requiring the harms caused by a measure to be in proportion to its benefits). It found that 'the military commander's choice of the route of the

[61] Beit Sourik n 11 above.

separation fence is disproportionate…[w]here the construction of the separation fence demands that inhabitants be separated from their lands, access to these lands must be ensured, in order to minimize the damage to the extent possible'.[62]

Is the court here correct to recognize that the test of necessity has been met by the military's measure and to focus its objection as being made under the third stage or balancing component of proportionality? It reaches this conclusion given a strict formulation of the instrumentality requirement of necessity: '[t]he question is whether the former route satisfies the security objective of the security fence *to the same extent* as the route set out by the military commander' (my emphasis).[63] Yet, as we have already seen, this formulation of the requirement is not apposite in the legal and policy context and is also too strict to render the test useful or meaningful. If we adopt the more moderate view of the Canadian Supreme Court, a means must be such as to realize the objective in a real and substantial manner. Once we weaken the test in this way, it becomes possible to recognize that two measures may both realize a government's objective in a substantive manner though one may be better at doing so than another. If the alternative to the government measure impacts upon rights to a lesser degree (even if worse from the perspective of realizing the objective), then it has to be determined whether the gain for fundamental rights can off-set the loss in respect to the government's objective. Here we see that a balancing component becomes part of the necessity enquiry itself. This is a clear consequence of recognizing the qualitative dimensions

[62] Ibid. [61].

[63] Ibid. [58].

involved in the necessity enquiry and the inadequacy of the strict understanding of instrumentality.

The reasoning in the *Beit Sourik* case could thus have declared the route of the wall unacceptable because of a failure to meet the necessity test. Indeed, the core of the reasoning that led to a finding of a lack of proportionality involved recognizing that there was in fact an alternative route for the barrier which would have provided a little less security for Israelis yet significantly reduced the harms to Palestinians. Since an alternative existed that protected the security interests of Israelis in a real and substantial manner whilst having a lesser impact upon Palestinian rights, the measures adopted by the army were not necessary on the revised conception I have articulated above.

2.5 The Relationship between Necessity and Balancing

Le Bel J in a dissenting opinion in the *Hutterian Brethren* case claims that 'proportionality analysis depends on a close connection between the final two stages in the *Oakes* test'.[64] The above analysis provides reasons for this close connection between the last two stages. Yet, Dieter Grimm, for instance, has criticized the Canadian Supreme court for incorrectly employing

[64] Hutterian Brethren n 25 above [191]. Le Bel goes on to say at [192] that 'it may be tempting to draw sharp analytical distinctions between the minimal impairment and balancing of effects parts of the Oakes test. But determining whether a measure limiting a right successfully meets the justification test should lead to some questioning of the purpose in the course of the proportionality analysis, to determine not only whether an alternative solution could reach the goal, but also to what extent the goal itself ought to be realized. This part of the analysis may confirm the validity of alternative, less intrusive means'. In a similar vein, he states at [195] that 'alternative solutions should not be evaluated on a standard of maximal consistency with the stated objective. An alternative measure might be legitimate even if the objective could no longer be obtained in its complete integrity'.

in the necessity enquiry a 'kind of language that is typical of the balancing process reserved for the third step in Germany'.[65] He contends that the reason for this approach may be to avoid too wide a discretion at the third stage of the proportionality enquiry which suggests that judges are involved too heavily in making value judgments that should be the preserve of elected branches. However, Courts in Canada, Grimm argues, risk 'self-deception when all the value-oriented considerations have been made under the guise of a seemingly value-neutral category'.[66] He suggests that it is important that different stages of the proportionality analysis be separated: '[a] confusion of the steps creates the danger that elements enter the operation in an uncontrolled manner and render the result more arbitrary and less predictable'.[67]

The argument I have provided through an analysis of the qualitative sub-components of necessity provide some support for the Canadian approach and demonstrates that an element of balancing is contained within the necessity enquiry when understood in a plausible manner. The importance of the second stage in Canada can be explained not only by an attempt to avoid value judgments on the part of the court but rather as a result of the inherent content of the enquiry. The discussion of these sub-components also highlights the fact that necessity is not simply a value-neutral or factual enquiry. It rather involves several substantive and qualitative elements that cannot be avoided by courts. I thus do not agree with Grimm that such judgments are called for only at the third stage of the proportionality enquiry; we are in

[65] Grimm n 2 above 395.

[66] Ibid. 397.

[67] Ibid.

agreement, however, that courts should not deceive themselves that they can avoid value judgments *even* in the second stage of the proportionality enquiry.

If balancing is included within the necessity enquiry, however, the question then becomes how are we to distinguish the second and third stages of the proportionality enquiry. Have we not essentially collapsed these two elements of the proportionality test and thus confirmed Grimm's fears of rendering the enquiry more amorphous and unstructured? It is important to point out that a similar but converse problem arises on the approach advocated by Barak, Grimm and the Israeli Supreme Court in *Beit Sourik*. By adopting a strict interpretation of equal effectiveness, they land up depriving the necessity enquiry of utility and essentially requiring all the work to occur at the third stage of the proportionality enquiry. Barak clearly writes that where any additional expense or burden is placed upon the state by an alternative measure, it should not be considered at the necessity stage but within the 'framework of proportionality strict sensu'.[68]

The problem with collapsing the two stages and placing all the emphasis on the third stage is that the balancing that takes place there is itself highly controversial and has attracted much criticism from the academic community. It has been contended, for instance, that the third stage weakens the protection afforded to human rights[69] and provides too much discretion to

[68] Barak n 30 above at 324-326.

[69] O Gross '"Once More Unto the Breach": The Systemic Failure of Applying the European Convention on Human Rights to Entrenched Emergencies' (1998) 23 *Yale Journal of International Law* 437ff.

courts, creating a serious separation of powers problem.[70] In this volume, Jochen von Bernstorff argues that the main problem with balancing is its *ad hoc* character. For him, balancing involves weighing 'the seriousness of the infringement against the importance and urgency of the factors that justify it'.[71] 'Ad hoc balancing turns human rights adjudication into an exercise of political decision-making, which fails to create and stabilize legal expectations within the legal system'.[72] This exacerbates the problems of providing too little protection to victims of fundamental rights violations as well as providing too much scope for judicial intervention with legislation. Von Bernstorff proposes a form of proportionality analysis that largely eschews ad hoc balancing. His alternative nevertheless requires courts to test measures that infringe rights for both suitability and necessity. Instead of the third stage, courts are invited to develop various sub-tests and bright-line rules for specific groups of cases that determine when the limitation of a right is justified or not.[73]

Much of von Bernstorff's analysis is illuminating and deserves more detailed attention than I am able to provide here. It nevertheless remains important to point out that the implications of

[70] T Alexander Aleinikoff 'Constitutional Law in the Age of Balancing' (1987) 96 *Yale Law Journal* 943, 984-986. .

[71] J von Bernstorff 'Proportionality Without Balancing: Why Judicial ad hoc balancing is unnecessary and potentially detrimental to the realization of individual and collective self-determination' ??? this volume (ed. C Mcrudden and L Lazarus) 11.

[72] Ibid. 4.

[73] The scope of this paper does not allow a detailed examination of von Bernstorff's alternatives. I am broadly sympathetic to his attempt to strengthen human rights protection and to try and achieve a greater principled approach towards the limitation of rights that is not purely based upon ad hoc, case specific analysis. I, however, doubt it will be possible to develop the categorical rules von Bernstorff wants without engaging in some form of balancing reasoning in order to establish these rules. It nevertheless would be preferable to have certain general provisional rules that apply in specific cases rather than the much wider unstructured judgment often made by courts at the third stage of the proportionality enquiry.

my argument in this chapter are that adopting von Bernstorff's 'bright-line' rules will not eliminate some form of balancing from the alternative he suggests since a degree of balancing is included within the necessity enquiry itself, which von Bernstorff does not wish to dispense with. It is thus much harder to eliminate balancing from the proportionality enquiry than may have been thought.

Nevertheless, I do not believe my argument completely destroys von Bernstorff's attempt to introduce more categorical reasoning into the proportionality enquiry. The reason for this is that it remains possible to distinguish the balancing that takes place within necessity from the balancing at the third stage of the enquiry.[74] This has implications both for the traditional proportionality enquiry as well as for von Bernstorff's argument.

The balancing enquiry that takes place within the necessity enquiry is limited in nature. Ultimately, at the necessity stage, courts are concerned with evaluating the feasible alternatives to the measure desired by the government. Those alternatives must be evaluated against the measure to determine whether there is at least one that realizes the government's objective in a real and substantial manner but has a lesser impact on the right. As has been discussed above, this can involve the court in balancing a reduction in the effectiveness (in realizing the government's objective) of an alternative measure against a lesser impact upon fundamental rights. The need to evaluate alternatives in this manner provides an important

[74] Here I disagree with Yowell (note 39 above) 11-13, 48 who argues that the existence of some *degree* of balancing at the necessity enquiry ultimately leads to a collapse of the second and third stages of the proportionality enquiry into a single test of means-end proportionality. As I seek to indicate below, in my view, it is possible to recognise that a much more limited form of balancing occurs at the necessity stage than at the 'proportionality strict sensu' stage.

process of reasoning that is protective of fundamental rights since it requires an engagement with whether the government's very objective could have been achieved adequately with a lesser impact on fundamental rights. Importantly though, in the necessity enquiry, what remains of central importance are two axes: the relation between any means and the objective in question as well as the impact on fundamental rights. This places a more restrained focus on the balancing process that occurs in this enquiry. It is hard to see how this more limited form of balancing can be removed from the proportionality enquiry without losing its essential character.

At the final stage of the proportionality enquiry, however, balancing involves a much wider enquiry. Let us say, for instance, it is concluded that a governmental measure is necessary. There still remains the question whether, in the context of the matter under review, the benefits of that measure (in relation to the purpose the government wishes to achieve) outweigh the costs for fundamental rights. As Grimm puts it, the two previous steps of the proportionality enquiry 'cannot evaluate the relative weight of the objective of the law, on the one hand, and the fundamental right, on the other, in the context of the legislation under review'.[75] Consider a law that allows the police to shoot a person to death if this is the only means to protect a property. Here, there is a lawful purpose and the law ensures that the means adopted are both suitable and necessary. Yet, '[i]f one had to stop here, the balance

[75] Grimm n 2 above 396.

between life and property could not be made. The law would be regarded as constitutional and life would not get the protection it deserves'.[76]

The example illustrates correctly the kind of non-comparative normative enquiry required by the third stage of proportionality. Grimm, in this example, assumes that courts would reach a conclusion protective of life. Yet, it is equally possible that – in a context of widespread theft and robbery – courts would reach the opposite conclusion. The broad discretion involved in such value judgments at the third stage of the enquiry as well as the utilitarian cost-benefit calculus in particular cases with limited principled guidelines appears to be what rightly worry theorists such as von Bernstorff. Whether it is possible or desirable to remove the broader normative evaluation required by the third stage when considering the limitation of rights is a matter I cannot take further here. What I have sought to demonstrate, however, is that a more limited form of balancing cannot be eliminated from the necessity enquiry. This more restrained form of balancing can be distinguished from that which is traditionally undertaken at the third stage of the proportionality enquiry, thus leaving room for alternatives to the third stage such as those proposed by von Bernstorff.

CONCLUSION

Proportionality has become the lingua franca of today's conversation across borders concerning the circumstances under which it is appropriate to limit fundamental rights. This paper has sought to engage with the necessity component of the proportionality enquiry. In the first

[76] Ibid.

segment of this paper I sought to identify what I termed the strict interpretation of necessity. I sought to demonstrate that this understanding of necessity leads can lead to two opposing problems: either it is considered to be too strong triggering substantial deference on the part of courts to other branches or it is too weak as a result of a strict construal of the equal effectiveness component. Either way, the test fails to offer adequate protection for fundamental rights by inviting courts either to circumvent its requirements or simply rendering it meaningless. The second part of the paper then sought to break down the enquiry into four parts and to consider what was entailed in each. Each component was seen to involve both qualitative and normative judgments that meant that the strict interpretation could not adequately be justified.

This analysis led to the conclusion that what I term a moderate interpretation of necessity (MN) needed to be adopted. The various components of MN can be summarized as follows:

(MN1) All feasible alternatives need to be identified, with courts being explicit as to criteria of feasibility;

(MN2) The relationship between the government measure under consideration, the alternatives identified in MN1 and the objective sought to be achieved must be determined. An attempt must be made to retain only those alternatives to the measure that realize the objective in a real and substantial manner;

(MN3) The differing impact of the measure and the alternatives (identified in MN2) upon fundamental rights must be determined with it being recognised that this requires a recognition of approximate impact; and

(MN4) Given the findings in MN2 and MN3, an overall comparison (and balancing exercise) must be undertaken between the measure and the alternatives. A judgment must be made whether the government measure is the best of all feasible alternatives, considering both the degree to which it realizes the government objective and the degree of impact upon fundamental rights ('the comparative component').

The moderate interpretation of necessity makes it clear that courts no longer are required to evaluate governmental measures against an unreasonably high standard. As Justice Kriegler stated in the SA Constitutional Court, '[w]here section 36(1)(3) speaks of less restrictive means it does not postulate an unattainable norm of perfection.'[77] As I have sought to show, however, the enquiry is not some overarching amorphous judgment. Necessity involves a process of reasoning designed to ensure that only measures with a strong relationship to the objective they seek to achieve can justify an invasion of fundamental rights. That process thus requires courts to reason through the various stages of the moderate interpretation of necessity.

The more moderate interpretation of necessity allows courts also to move away from any general strategy of deference; rather, courts must sensitively evaluate the evidence put before them by both parties in relation to each stage of the necessity enquiry. Where other branches

[77] *S v Mamabolo* 2001 (3) SA 409 (CC) at [49].

of government have institutional advantages and expertise, courts may, in cases where the evidence is finely balanced, find in favour of the government as they are better placed to make the qualitative judgments involved. This should not involve any general presumption in favour of the government but rather allows for its evidence and arguments to be given some weight where they are presented on matters in which the government has more institutional capacity and expertise. The converse will also apply in that its evidence and arguments will be given less weight in determining the impact on rights, for instance, where the applicant may have greater knowledge and understanding.

The key purpose of the necessity enquiry is to offer an explicit consideration of the relationship between means, objectives and rights. The discipline of this reasoning process helps identify flaws in any attempted justification for the limitation of fundamental rights. The process, as I have argued, is not mechanical nor is it factual. It is essentially normative and qualitative in nature. Failure to conduct the necessity enquiry with diligence, however, means that a government measure can escape close scrutiny in relation to both the realization of the objective and impact upon fundamental rights.

Fundamental rights are not absolute yet they deserve strong protection. The way in which the necessity enquiry is conducted can tend towards either rendering these rights absolute or offering them little protection. I have sought in this piece to highlight certain ways in which courts and academic writers have approached the matter which tend towards these two extremes. Instead, I have sought to grapple with the complexity of the enquiry itself, seeking to demonstrate that a more moderate interpretation is possible which affords significant

protection for rights but allows for their limitation in suitable circumstances. Ultimately, good judgment is ineliminable from a determination of necessity; a clear reasoning process helps to guide the judgments that must be made through clarifying the nature of the enquiry and normative considerations in question.