On the Structure of Legal Principles*

ROBERT ALEXY

Abstract. The author offers a sketch of his thesis that legal principles are optimization commands. He presents this thesis as an effort to capture the structure of weighing or balancing and to provide a basis for the principle of proportionality as it is applied in constitutional law. With this much in place, he then takes up some of the problems that have come to be associated with the optimization thesis. First, he examines the objection that there are no such things as principles, but only different modes of the application of norms. Second, he discusses problems concerning the concept of an optimization command and the character of the “ought” contained in principles. He concludes that the distinction between commands to optimize and commands to be optimized is the best method for capturing the nature of principles.

The distinction between rules and principles had already been thoroughly considered in Germany by Josef Esser during the 1950s, albeit with a slightly different terminology (Esser 1974). In Austria, Walter Wilburg in the 1940s had anticipated major developments in his theory of flexible systems (Wilburg 1941; 1951; 1963). Still, it was Ronald Dworkin’s major challenge to H. L. A. Hart’s version of legal positivism, initially in “The Model of Rules,” that marked the beginnings of a broad discussion (Dworkin 1967). During the past three decades the distinction between rules and principles, including its implications for legal methodology, the concept of the legal system, the relation between law and morality, and legal dogmatics, especially that of basic rights, has been subject to a great number of in part very detailed studies. Two main positions have emerged. One is that principles express the idea of optimization. This can be expressed in the short formula that principles are optimization commands, and it is this feature that represents the main distinction between principles and rules. This understanding may be termed “principle theory.” The other position is less

* I should like to thank Bonnie Litschewski Paulson and Stanley L. Paulson for help and advice on matters of English style.

uniform. But there is a consensus of opinion to the effect that the optimization thesis is either wrong or at any rate that its explanatory power is greatly exaggerated.

The number and variety of contrary positions are such that I cannot discuss all of them. Instead, I will consider just a few objections concerning the structure of principles as norms. First, however, I shall provide a sketch of the three main theses of principle theory; they delineate the core of the controversy.

I. Three Theses

1. The Optimization Thesis

According to the standard definition of principle theory (Alexy 1996, 75ff.), principles are norms commanding that something be realized to the highest degree that is actually and legally possible. Principles are therefore optimization commands. They can be fulfilled in different degrees. The mandatory degree of fulfilment depends not only on actual facts but also on legal possibilities. The field of legal possibilities is determined by counter-vailing principles and rules. Contrariwise, rules are norms that can only be either complied with or not. If a rule is valid, it requires that one do exactly what it demands, nothing more and nothing less. Rules therefore comprise a decision in the fields of actual and legal possibilities. They are definitive commands. This means that the difference between rules and principles is a difference in quality and not only one of degree. Every norm is either a rule or a principle (Alexy 1996, 77ff.; 1995, 203).

2. The Collision Law

The difference between rules and principles emerges most clearly when one turns to collisions of principles and conflicts of rules. Collisions of principles and conflicts of rules share the feature that two norms, when applied separately, lead to incompatible results, namely to two contradictory specific or concrete legal “ought”-judgments. But they differ most fundamentally in their respective solutions to the conflict.

a) Conflicts of Rules

A conflict between two rules can only be solved by either introducing an exception clause into one of the two rules or declaring at least one of them invalid. An example of the first is a school regulation that prohibits one’s leaving the classroom before the signal but requires that one do just that in the event of a fire alarm. This conflict is easily solved, namely by introducing, for the case of the fire alarm, an exception into the prohibition to leave the classroom before the signal. If such a solution is out of reach, the
only remaining possibility is to declare at least one of the rules invalid. This is the main purview of collision rules such as “lex posterior derogat legi priori,” “lex superior derogat legi inferiori,” and “Federal law shall override Land law” (art. 31 of the German Basic Law, GG).

b) Collisions of Principles
A collision of principles is solved in an altogether different way. An example is a decision of the German Federal Constitutional Court concerning the inability to attend sessions of a court proceeding (Decisions of the Federal Constitutional Court, BVerfGE vol. 51, 324). This decision takes up the question whether a trial may be held in the case of an accused who would be in danger of suffering from a stroke or heart attack owing to the stress of the trial. The colliding norms are, on the one hand, art. 2, par. 2 s. 1 GG, guaranteeing to everyone the right to life and the inviolability of one’s body, and on the other, the rule-of-law principle (Rechtsstaatsprinzip) in so far as it obligates the state to provide for a functioning system of criminal justice. If only the basic right existed, the conduct of the trial could easily be qualified as prohibited owing to the danger posed to the life and health of the accused that is connected therewith. If, vice-versa, the only obligation that existed were that of the state to provide for a functioning system of criminal justice, the performance of the trial would be classified without difficulty either as obligatory or at least as permitted. Now the court could have solved the case by either declaring the basic right or the rule-of-law principle as invalid. In this case the court would have treated the collision as a conflict of rules and solved it in terms of validity. It is obvious, however, that neither the invalidation of the basic right to life and inviolability of the body nor of the principle of a functioning criminal justice system as a sub-principle of the rule-of-law principle is a live option here. The second possibility for solving a conflict of rules, namely introducing an exception, also fails to comprehend what is to be done in this case. The basic right to life and inviolability of the body does not count as an exception to the principle of a functioning system of criminal justice, nor does this principle count as an exception to the right to life and inviolability of the body. Rather, the court solves the problem by determining a conditional priority of one of the colliding principles over the other with respect to the circumstances of the case. The basic right to life and to the inviolability of the body shall have priority over the principle of a functioning system of criminal justice as a sub-principle of the rule-of-law principle where “there is a clear and specific danger that the accused will forfeit his life or suffer serious bodily harm in case the trial is held” (BVerfGE vol. 51, 234, 346). Under these conditions the basic right has greater weight and therefore takes priority; under different conditions precisely the opposite may well be the case.
The priority of the basic right implies that its legal effects are mandatory. The fulfilment of the conditions of priority brings about the legal effects of the preceding principle. This can be formulated as a general collision law. It runs:

The conditions under which one principle takes priority over another constitute the operative facts of a rule giving legal effect to the principle deemed prior.2

A more technical version is:

If principle \( P_1 \) takes priority over principle \( P_2 \) under conditions \( C \): \( (P_1 \mathbf{P} P_2) \) \( C \), and if \( P_1 \) under conditions \( C \) implies legal effect \( R \), then a rule is valid that comprises \( C \) as the operative facts and \( R \) as legal effect: \( C \rightarrow R \). (Alexy 1996, 83)

The collision law expresses the fact that the priority relations between the principles of a system are not absolute but only conditional or relative. The task of optimizing is to determine correct conditional priority relations. The fact that a determination of a conditional priority relation in accordance with the collision law is always the determination of a rule formed on the occasion of the case demonstrates that the respective levels of principles and rules are by no means unconnected. To solve a case by weighing is to decide by means of a rule that is substantiated by giving priority to the preceding principle. In this respect, principles are necessarily reasons for rules.

3. The Balancing Law

The practical significance of principle theory in the form of the optimization thesis is found above all in its equivalence to the principle of proportionality (Verhältnismäßigkeitgrundsatz). Principle theory implies the principle of proportionality and the principle of proportionality implies principle theory (Alexy 1996, 100ff.). The fact that principle theory implies the principle of proportionality means that the three sub-principles it contains, the principle of appropriateness, of necessity, and of proportionality in a narrow sense follow logically from it; hence, they are deducible from it in a strict sense. The same is true if we proceed from the other side of the equivalence relation, namely, that the principle of proportionality implies principle theory. Thus, one who rejects principle theory must reject the principle of proportionality, too. The dispute over principle theory can therefore be seen as a reflection of the dispute over the proportionality principle.

The implications of the proportionality principle turn on the definition of the concept of principle. Principles \textit{qua} optimization commands demand

---

2 This phrasing of the collision law relates to the case in which the legal effect of the preceding principle comes fully into force. If it does not come fully into force, modifications are necessary; see Alexy 1996, 83 fn. 42.
realization as far as is possible relative to the actual and legal possibilities. A relativization in the direction of the actual possibilities leads to the principles of appropriateness and necessity. Let us assume a measure $M$ that encroaches on the freedom of trade, occupation, or profession ($P_1$) in order to promote consumer protection ($P_2$) but which is not appropriate to promoting $P_2$ in any way whatever. It is possible to abandon $M$ at no cost to $P_2$, consumer protection. The optimization of $P_1$ and $P_2$ demands, then, that $M$ not be used. This is exactly the content of the principle of appropriateness. The principle of necessity says that a measure $M_1$ is prohibited in respect of $P_1$ and $P_2$ if there is an alternative measure $M_2$ that promotes $P_2$ approximately as well as $M_1$ but encroaches less intensively on $P_1$. Let us assume that $P_2$ stands, again, for consumer protection, in particular, for the consumers’ protection against buying products that they do not in fact want. Let us also assume that $M_1$ is an absolute prohibition of goods that look like chocolate but are not chocolate. $M_2$ stands in this case for the obligation clearly to designate the nature of the goods. This obligation, namely ($M_2$), obviously encroaches less intensively on the freedom of trade, occupation, or profession ($P_1$) than would an absolute prohibition ($M_1$), and it serves consumer protection more or less equally well; therefore, the absolute prohibition ($M_1$) is prohibited in relation to $P_1$ and $P_2$ as an unnecessary means (BVerfGE vol. 53, 135, 145ff.).

The principles of appropriateness and necessity stem from the obligation of a realization as great as possible relative to the actual possibilities. They express the idea of Pareto-optimality. The principle of proportionality in a narrow sense stems from the obligation of a realization as far as possible relative to the legal possibilities, that is, relative most of all to the countervailing principles. Here we are concerned with balancing or weighing in a narrow and true sense. This is necessary whenever the fulfilment of one principle leads to the non-fulfilment of another, hence, whenever one principle is only realizable at the cost of another. For this kind of case the following balancing law can be formulated:

The more intensive the interference in one principle, the more important the realization of the other principle. (Alexy 1996, 146)

The problems connected with this formula are the main topics of the discussion concerning the question whether balancing is a rational procedure.

II. Two Objections

Balancing and collision law are attempts to describe the core idea of principle theory, namely the optimization thesis, more precisely. This thesis claims that principles are norms that, owing to their structure, are fundamentally distinct from rules. Numerous arguments have been adduced against this position. Two are discussed here.
1. Principle Structure and Norm Application

A radical “norm theoretic” criticism would have us believe that principles do not exist at all; rather, only norms exist, albeit norms that are used in different ways. Klaus Günther thus asserts that the difference between rules and principles is not a difference in structure but merely a “different kind of use.” Whether “we treat a norm as a rule by using it without regard to the unequal characteristics of the situation or whether we treat a norm as a principle by considering all (actual and legal) facts” (Günther 1988, 270), is a question of the “conditions of action” (Günther 1988, 273) or of the “conditions of conversation” (Günther 1988, 270). In legal systems, the conditions of conversation and of action are institutional in nature. The separation of powers and the rule-of-law requires that the decisions of the legislator be treated as rules and that exceptions be admitted only in special cases.

The heart of Günther’s criticism is the thesis that, independently of the institutional frame, one can “direct the claim that a norm be used relative to the actual and normative (legal) possibilities present in a situation, [...] to any norm whatever” (Günther 1988, 272). This is true, but it fails to address the decisive point. By a “use relative to the actual and normative (legal) possibilities present in a situation,” Günther understands the “consideration of all circumstances” (Günther 1988, 272). To consider something, however, is different from optimizing. This is demonstrated by the fact that the consideration of all circumstances is also possible using norms that can either be fulfilled or not, whereas optimization requires that one be able to adhere to a norm to a greater or lesser extent. While optimization implies the consideration of all circumstances, the consideration of all circumstances does not imply optimization. It is this point that makes clear why Günther’s criticism fails to meet what is essential—optimization itself.

Gradual fulfilment is not the only reason for constructing principles in terms of the structure of norms. Only principle theory can explain why a norm, set aside in a balancing decision, is neither violated nor partially or totally rendered invalid. The idea of optimization is necessary if we are to comprehend the dimension of weight in the case of a norm, in contrast to its validity. This has a great many consequences in legal dogmatics. An adequate theory of the limits of a right, for example, is not possible without principle theory (Alexy 1996, 249ff.).

One can grant that Günther is right when he says that there are cases in which it is not easy to decide whether a norm should be treated as a rule or as a principle (Günther 1988, 272). This is a question of interpretation, and, as is usual with interpretation, there are no criteria providing for simple and

---

3 Stelzer (1991, 215) argues along the same lines, contending “that ‘rule’ and ‘principle’ characterize, ultimately, the argumentative use of norms.”

4 Habermas (1998, 429) has raised against the optimization thesis the objection that it cannot take account of the case in which “one right can ‘yield’ to another right, without loss of validity, when the two happen to conflict.” Precisely the opposite is the case.
clear answers in all cases. But this is not an objection to the point that “principle” captures a characteristic of norm structure. The very question whether a norm is a rule or a principle presupposes that norms qua principles be possible entities.

2. Optimization Commands

a) Commands to Optimize and Commands to Be Optimized

The discussion of Günther’s “norm theoretic” objection has shown that the distinction of rules and principles generally is dependent on the character of principles as commands to optimize. Aulis Aarnio and Jan-Reinard Sieckmann have raised the objection that the concept of a command to optimize is ill-suited for distinguishing between rules and principles. According to the standard definition presented above, principles are optimization commands, that is, commands to optimize, because they impose the obligation that something be realized to the highest degree that is actually and legally possible. This obligation has, indeed, a definitive character. It can only either be fulfilled or not fulfilled, and its complete fulfilment is always obligatory (Sieckmann 1990, 65). Aarnio puts this point as follows: “Either one does or one does not optimize. For example, in the case of conflict between two value principles, the principles must be brought together in the optimum manner, and only in the optimum manner” (Aarnio 1990, 187). Optimization commands therefore have the structure of rules.

This in no way says that principle theory in the guise of the optimization thesis collapses; it simply gives it a sharper focus. A distinction is to be made between commands to be optimized and commands to optimize. Commands to be optimized are the objects of balancing or weighing. They can be termed “the ideal ‘ought’” or “ideals” (Alexy 1995, 203ff.). An ideal “ought” is something that is to be optimized and thereby transformed into a real “ought” (Alexy 1995, 204). As the object of optimization, it is placed on the object level. Contrariwise, the commands to optimize, that is, the optimization commands, are placed on a meta-level. On this level they prescribe what is to be done with that which is found on the object level. They impose the obligation that their subject matter, the commands to be optimized, be realized to the greatest extent possible. As optimization commands they are not to be optimized but to be fulfilled by optimization.

Principles, therefore, as the subject matter of balancing are not optimization commands but rather commands to be optimized. As such they comprehend an ideal “ought” that is not yet relativized to the actual and legal possibilities. In spite of this, it is useful to talk about principles as optimization commands or obligations. It expresses in an altogether straightforward way the nature of principles. In saying what is to be done with principles, one says all that matters from the point of view of legal practice. This practical aspect is lent support by a theoretical consideration.
There is a necessary connection between the ideal “ought,” that is, the principle as such, and the optimization command as a rule. The ideal “ought” implies the optimization command and *vice-versa*. These are two sides of the same coin. The question “whether the command to weigh, necessarily accompanying a principle, is ‘inside’ or ‘outside’ of the meaning of this principle,” thus formulated by Peczenik (1989, 78), can be answered by pointing out that the optimization command is comprised in the concept of principle. To abolish it would mean that the principle had lost its character as a principle. Peczenik adds that the question of whether or not principles should be called “optimization commands” has no substantial consequences for moral or legal philosophy (Peczenik 1989, 78). All this seems to show that it is advisable, for reasons of simplicity, to designate principles as “optimization commands” and to employ more precise distinctions only where necessary.

b) Reiterated Validity Obligations

One might well think that none of this leads to clarity where the structure of principles is concerned. Certainly one knows that they are to be optimized and, therefore, that they are commands or obligations to be optimized, but one knows nothing about the nature of whatever is to be optimized. Sieckmann tries to answer this question with his theory of reiterated validity obligations. His starting point is a distinction among three kinds of sentences: (1) norm formulations, (2) validity sentences, and (3) validity obligations (Sieckmann 1997, 352). Norm formulations or norm sentences (Alexy 1996, 42) express norms semantically, that is, norms as simple meaning-contents, without saying anything about their validity. Here the basic form can be represented with the help of the obligation operator “$O$” (It is obligatory ...) and the letter “$p$” standing for what is obligatory. For example, “$Op$” can express: “It is obligatory to return the thing.” Instead of “$Op$” one can insert a simple “$n$” for norm formulations or norm sentences. Validity sentences or norm-validity sentences (Alexy 1996, 51) have a more far-reaching content than norm formulations. They represent norms not only semantically but also express that these norms are valid. Using the validity predicate “$V$”, they can be represented as “$VOp$” or, more simply, as “$Vn$”.$^5$ Finally, validity obligations are obligations expressing that a specific norm (n) ought (O) to be valid (V), which can be captured by “$OVn$”. Once these distinctions have been made, the question arises: How, employing this machinery, are principles best represented? It is easy to see that the first alternative has to be eliminated. A principle is more than simply a norm in the semantic sense, that is, a mere norm content (n), if its role as a reason in balancing or weighing is to be captured. For this, it must, in one way or another, have validity (Sieckmann 1994, 209). Now one might assume that

---

$^5$ In the following the simpler form is used. For “$n$” one can always substitute “$Op$".
this could be accomplished by attaching to the norm sentence “n” the validity predicate “V”, which would imply the classification of the principles as “Vn”, the second of the alternatives stated above. According to Sieckmann, however, precisely this is not permissible. Validity sentences, that is, sentences of the form “Vn”, express what Sieckmann calls “normative statements.” One who utters such a sentence gives expression to an assumption about the definitive validity of a norm (Sieckmann 1997, 351). If principles had definitive validity, they would not lend themselves to and would not be in need of balancing, and they would therefore not be principles at all.

Now one might ask whether it is really cogent to reserve validity predicates like “V” solely for definitive validity. There is the possibility of introducing different validity predicates, for example, one for definitive and one for prima facie validity. Nothing in our normative language precludes such a possibility; the language is sufficiently flexible to give expression to the differing strengths of different kinds of validity. The point of Sieckmann’s effort, however, is to obstruct this very path. If one chose it, the problem concerning the argumentative power of principles would be concealed without being analysed in a specific validity predicate. Seen from an analytical point of view, the interpretation of principles as reiterated validity obligations is said to be superior to this.

If the validity predicate “V” is understood solely in the sense of definitive validity, a dilemma arises between the first and second alternatives representing the point of departure of the thesis of reiterated validity obligations. Simple norm sentences of the sort “n” or “Op” are not sufficient to represent the argumentative power of principles. They merely express norms as the content of thought sans validity. Anyone could come up with a norm, thus understood. This alternative is, therefore, too weak. On the other hand, validity sentences of the sort “Vn” or “VOp” are too strong. This is indeed the case where the validity predicate “V”, as Sieckmann stipulates, is to be reserved solely for definitive validity. Principles, after all, are the opposite of definitive obligations.

According to Sieckmann, the solution of the dilemma lies in the third alternative, that is, in the interpretation of principles as validity obligations. Problems, however, arise at once. According to Sieckmann, validity obligations have the form of “OVn”. That which begins with “O”, not “V”, is nothing other than a norm in the semantic sense, thus, a mere content of thought without validity, and this will be true even if the thought refers to validity. It is of no help, here, to insert another “V” before the “O”, for this would lend to the thought, initially too weak, the strength of definitive validity. Sieckmann, therefore, has to find a path between the too weak “O” and the too powerful “V”—he does, at any rate, if he wants to stay within the co-ordinates of his elements “n”, “V”, and “O”. It is said that this middle course can be found in the connection of “O” and “V”, namely, in an
infinitely reiterating connection. The true logical form of the principle is therefore said to be “… OVOVn” (Sieckmann 1997, 352).

Now it may be asked: Is this an answer to the question of what principles come to that goes beyond what we already know? To be sure, it is clearly said what principles are not. They cannot be either definitely valid or non-obligatory norms. But this is nothing new. Only a positive answer would have real interest. The thesis of reiterated validity obligations offers us a positive answer, however, in only a very limited way. Reiteration illustrates the problem in an interesting manner but does not solve it. It is interesting, for the two things that principles cannot be are connected by infinite reiteration in such a way that a continuous oscillation takes place between them. “Vn” says that “n”, for example, a norm sentence of the form “Op”, is definitely valid. This is too much. “OVn” weakens this radically. It is only a norm without validity that turns up in the world as a mere thought of a norm. This is too little. “OVn” overcomes this insufficiency, but goes beyond the target, for definitive validity—and this according to Sieckmann is what “V” stands for—exceeds that which can be claimed of a principle. Thus, “O” must immediately be poured as water on the fire once again, and the game continues endlessly, for the quenched fire is to be rekindled straightaway, ad infinitum. It seems as though the reiteration thesis describes in a relatively complicated way the notion that principles stand somewhere between definitive validity and non-obligation.

A different position would be close at hand if it were possible to explain, by appeal to reiteration, distinctive features of principles that could not be explained without it. This is, however, not the case. Sieckmann regards principles as validity obligations, that is, as obligations expressing that specific norms ought to be valid, a state of affairs that can be captured by “OVn”. In so doing, he gives the infinite reiteration of the form “… OVOVn” a certain priority over the form “… VOVOVn”. This, however, is not at all evident and does not follow from the assumption of infinity, for the assumption speaks on behalf of a balance between validity (V) and “ought” (O). How can an obligation operator (O) that stands ahead of the validity predicate (V), and thus does not include validity, substantiate validity? Sieckmann’s answer is plain. Thanks to infinite reiteration, there is no obligation operator (O) standing alone ahead of a validity predicate (V), for ahead of every obligation operator there is always a validity predicate that also does not stand alone—and this for the reason that it is immediately connected to an obligation operator. Perhaps one has to imagine that all this takes place at one time. None of this, however, serves as an explanation of the character of the principle. That question is, indeed, encapsulated neither by a validity predicate nor by some deontic operator, but is left undecided between “V” and “O”. In the end, nothing is claimed other than that the argumentative power of principles is to be found somewhere between definitive validity and non-obligation.
Still, Sieckmann’s attempt to capture the character of principles with the help of the reiteration thesis is not without value. It deserves attention as a painstaking effort to understand the nature of principles by means of classical deontic logic and a binary validity predicate. However, nothing said in this connection alters the fact that there is, presently, no explanation of what principles come to that is more promising than a construction of them as obligations to be optimized, where this corresponds to obligations to optimize. This construction seems to be the best expression of the idea of an ideal “ought” and of ideal validity.

Christian Albrechts University
Faculty of Law
Olshausenstraße 40
D-24118 Kiel
Germany
E-mail: alexy@law.uni-kiel.de

References
