The Generality of Neutral Principles: A Game-Theoretic Perspective

Robert L. Birmingham

Follow this and additional works at: https://scholarship.law.berkeley.edu/californialawreview

Recommended Citation

Link to publisher version (DOI)
https://doi.org/10.15779/Z38JV1F

This Article is brought to you for free and open access by the California Law Review at Berkeley Law Scholarship Repository. It has been accepted for inclusion in California Law Review by an authorized administrator of Berkeley Law Scholarship Repository. For more information, please contact jcera@law.berkeley.edu.
The Generality of Neutral Principles:
A Game-Theoretic Perspective

Robert L. Birmingham*

Eleven years ago Professor Wechsler first argued that constitutional adjudication should be based on neutral principles of general applicability that transcend the factual contexts of the cases in which they are announced. Since that time, legal scholars have been engaged in a continuing debate over both the meaning and the validity of Wechsler's thesis. Professor Birmingham adds a new perspective to this debate by analyzing neutrality and generality in terms of game theory.

[W]e always pay for generality by sacrificing content, and all we can say about practically everything is almost nothing.

Kenneth Boulding

Professor Wechsler has argued that constitutional adjudication should be evaluated primarily in terms of the neutrality of the principles on which decisions are based. This standard, proposed over a decade ago, has been repeatedly attacked as unintelligible. Professor Wright, finding it the product of theories more misleading than enlightening, posits confusion and incorrectness. Professor Countryman attributes his inability to understand Wechsler's theory to what he describes as an intense lack of definition. Professor Pollak appears willing to proclaim either innocuity or incomprehensibility, but seems uncertain as to which pronouncement is more appropriate. Continuing discussion concerning the concept of neutrality and frequent approving references to it nevertheless demonstrate the existence of a core of meaning offering useful insights into the judicial process. One feels,

perhaps intuitively, that the standard introduced by Professor Wechsler has significance; the problem is to isolate and explicate those elements embedded in his analysis responsible for this significance. This Article presents a tentative step toward solution based on an often noted interrelationship between generality and neutrality. It applies distinctions generated in the theory of games, a branch of mathematics of increasing importance in economics and other social sciences.

I

GENERALITY AND NEUTRALITY

Professor Wechsler in his initial exposition conjoined generality with neutrality:

I put it to you that the main constituent of the judicial process is precisely that it must be genuinely principled, resting with respect to every step that is involved in reaching judgment on analysis and reasons quite transcending the immediate result that is achieved. To be sure, the courts decide, or should decide, only the case they have before them. But must they not decide on grounds of adequate neutrality and generality, tested not only by the instant application but by others that the principles imply? . . .

... A principled decision, in the sense I have in mind, is one that rests on reasons with respect to all the issues in the case, reasons that in their generality and neutrality transcend any immediate result that is involved.6

Juxtaposition at the dawn of debate has stimulated efforts to characterize neutrality through interpretation of the companion concept: a principle has been thought neutral if it is a rule of general application or if it is to be applied across the board either without compromise or regardless of what values may be promoted or impeded.7 Understood merely to connote comprehensiveness, the proposed criterion permits a partial ordering of principles: if one of two principles can be applied in all cases where the other can be applied, and this principle is applicable in other cases as well, it can, because of its broader applicability, be proclaimed more neutral.

That such a measure has not found acceptance demonstrates reluctance to equate neutrality with generality in the sense indicated.

6. Wechsler, supra note 2, at 15, 19. See also id. at 16, 26.
Professor Deutsch, arguing consequent maximal satisfaction of public expectations, is able to reason from a definition of neutrality as that degree of generality widely perceived as adequate—a definition which he does not employ uncritically—to a rule requiring decision premised on the most general principle available.8 His repudiation of the simple quantitative standard which this approach might suggest is nevertheless evident in his discussion of Brown v. Board of Education,9 a decision which disturbed Professor Wechsler10 because he believed it to have sought a desirable result through sacrifice of that reasoned exposition of the law indispensably associated with judicial resolution of disputes:

In what sense is a constitutional principle inadequately general if it states that all claims put forward by Negroes or Communists will be approved, or that all claims advanced by segregationists or corporations will be rejected?

. . . Wechsler's test for adequacy is general applicability; his objection to the principle that all Negro claims should be approved is that it represents "ad hoc evaluation" of the case before the Court. This objection would be met, however, if the Court did in fact approve all Negro claims coming before it—a course of action that would neither change the general agreement as to the principle's illegitimacy nor be approved by Wechsler.

. . . The general agreement as to the illegitimacy of the principle that Negro claims should always be favored does not, however, rest on the prediction that that principle might often be held inapplicable; indeed, the more often it was held applicable, the more illegitimate the Court's course of action would be. That agreement rests, rather, directly on disapproval of the content of the principle: disapproval based on the belief that other values are equally or more important than those embodied in Negro claims.11

Professor Deutsch concludes that the demand that constitutional principles be generally applicable must be taken to express pervasive aversion to attribution of legal consequences to membership in groups.12 Arguably less a resolution than a restatement of the doctrinal difficulty he identifies, his position parallels that adopted by Professor Wechsler, who asked whether evaluation of a suggested principle through consideration of its applicability to other cases, preferably cases involving opposed interests, is not the essence of judicial method.13 This concern

10. Wechsler, supra note 2, at 31-34.
11. Deutsch, supra note 8, at 191-93.
12. Id. at 195.
13. Wechsler, supra note 2, at 15.
that a rule of law offer protection in appropriate circumstances to inconsistent values asserted by competing segments of society has been reaffirmed by Professor Wechsler through direct although not entirely unambiguous restatement.14

Professor Fuchs, interpreting the position taken by Professor Wechsler, acknowledges a goal of greater and more accurate generalization, but contends that generality is a characteristic common to all principles and thus hardly the definitive measure of their sufficiency; he finds an unquantified but independent aptness simultaneously required.15 Professor Deutsch through reference to the legal consequences of membership in groups appears to introduce a parallel requirement of aptness as a component of generality itself.16 Although Professor Wechsler of course accepts the significance of decisional criteria other than comprehensiveness as previously construed,17 his semantic stance is more equivocal: if, as some interpreters suggest,18 neutrality and generality are thought to be essentially undifferentiated concepts, aptness may be considered an element of either; alternatively, aptness and comprehensiveness may be taken to inhere in neutrality and generality respectively. These linguistic complexities are symptomatic of confusion concerning underlying substantive issues, which can be most conveniently examined in a context unprejudiced by dispute over definitions. The language appropriate to inquiry is supplied by the theory of games.

II

THE THEORY OF GAMES

A game is a set of rules delineating possible actions by individuals or groups and the consequences of such actions. The rules normally designate a set of players, define the behavioral choices open to the players and the information on which these choices must be based, and associate with each combination of choices an outcome comprising a payoff to each player.19 The prototypical dispute at law may be described in terms of the strictly competitive two-person game. In this game the players have diametrically opposed preferences: if one player

17. See, e.g., Wechsler, supra note 2, at 16-17.
18. See, e.g., Brown, supra note 7, at 387; Henkin, supra note 7, at 653-54.
prefers one outcome to another, the other player will prefer the other outcome; if one player is indifferent between two outcomes, the other will also be indifferent. It is likewise in the law: if, for example, a claim for damages is litigated, the extent of the plaintiff’s victory usually measures the loss to the defendant.

$$\begin{array}{c|cc}
R & 1 & 2 \\
\hline
1 & 2 & 4 \\
2 & 0 & \\
\end{array}$$

Figure 1

Data concerning a game may be summarized diagramatically, as in Figure 1. Here one player, designated $R$, must choose row 1, while the other player, designated $C$, may unless otherwise constrained choose either column 1 or column 2. The squares, representing the two possible combinations of choices, are marked with the payoffs to the players: the numbers in the lower left and upper right corners of a square indicate gain derived by players $R$ and $C$ respectively from the combination of choices associated with that square. Gain may be measured ordinally or cardinally, although interpersonal comparison of levels of satisfaction, at least implicit in some cardinal measures, would entail assumptions severely restricting the operational significance of any conclusions which might be reached. For present purposes gain is most conveniently expressed in monetary units, dollars or perhaps thousands of dollars.

As illustrated in Figure 1, choice by player $C$ will determine which of the two sets of payoffs will be realized: if he selects column 1 each player will gain two units; if he selects column 2 he will gain four units but player $R$ will gain nothing. Let $y$ be the probability that player $C$ will select column 1. Since $C$ must select either column 1 or column 2, the probability that he will select column 2 is necessarily $1-y$. The expected payoffs to players $R$ and $C$ can therefore be written as functions of the probability $y$:

$$
E_R(y) = 2y,
E_C(y) = -2y + 4.
$$

The strictly competitive nature of the relationship between the players is evident on calculation of the rates of change of the expected payoffs to them as $y$ changes:

Because these rates of change differ in sign one player must lose if the other gains. Assuming indifference on the part of player $C$ to the welfare of player $R$, player $C$, if his choice is unconstrained, can be expected to maximize his gain by choosing column 2 or, equivalently, setting $y$ equal to zero.

Legal rules may be considered operators directly affecting either the probability $y$ or the payoffs to the players. The law may be interpreted to prevent player $C$ from choosing column 2, so that $y$ is necessarily one. Alternatively, where for example the law permits player $R$ to recover damages if player $C$ chooses column 2, it may be thought to alter the consequences of this choice and thus to change $y$ by indirection. The difficulty here is that the law in either granting or denying relief must choose to aid one player at the expense of the other. To raise $y$ is to benefit $R$ but injure $C$; to refuse to raise $y$ is to benefit $C$ but injure $R$. No rule applied to the game as presented can offer mutual advantage.

III

THE PRISONER’S DILEMMA

Assume players $R$ and $C$, having committed a serious offense, have been apprehended by the police and isolated in separate cells. Their interrogator informs each that although he is certain of their guilt he has insufficient evidence to convict them of this offense. He asserts, however, that even without further information he will be able to secure their conviction on a lesser charge, permitting the state to impose a penalty of brief imprisonment. He urges each of the players to confess, promising that if one cooperates and the other remains silent, the cooperating player will escape punishment, while his recalcitrant partner will be incarcerated for an extended period. If both confess each will receive an intermediate sentence. The players accept the accuracy of this outline of possible consequences and must decide independently whether to confess. Abstraction from their disadvantageous situation has produced the prisoner’s dilemma game.\footnote{See A. Rapoport & A. Chammah, Prisoner’s Dilemma (1965); Rapoport, Escape from Paradox, Scientific Am., July 1967, at 50. See also J. Boot, Mathematical Reasoning in Economics and Management Science 90-92 (1967).}
In Figure 2 choice of row 1 by player $R$ or column 1 by player $C$ may be taken to indicate silence by the choosing player, while choice of row 2 by player $R$ or column 2 by player $C$ may be taken to indicate confession by the choosing player. If the desirability of an outcome to a player is a decreasing function of his term of incarceration alone, he will list the possible combinations of choices in the following order, from least to most satisfactory: (1) his own silence coupled with confession by his partner; (2) mutual confession; (3) mutual silence; (4) his own confession coupled with silence by his partner. The payoffs to each player associated with alternative combinations of choices may be ranked on an ordinal scale, that is, a scale indicating the order of the elements of a set but not showing the magnitudes of the disparities between them. Thus in Figure 2 the payoffs resulting from mutual silence and mutual confession are assigned ranks of one and two respectively; confession coupled with silence offers a payoff of rank four to the confessing player and a payoff of rank zero to his partner.\footnote{22}

Protection of his individual interest requires confession by each prisoner. If player $C$ remains silent, confession by player $R$ provides player $R$ a payoff of rank four, greater than the payoff of rank two he could obtain through silence. If player $C$ confesses, confession by player $R$ similarly offers player $R$ a payoff of rank one, exceeding the

\footnote{22. Other indices of rank can be substituted for those selected so long as the order of the outcomes is not varied. The numbers chosen are appropriate to subsequent analysis, where gain is measured not ordinally but in monetary units.}
payoff of rank zero he would otherwise procure. Player C faces an identical pattern of payoffs. Mutual confession, however, offers each player a payoff of only rank one, which is inferior to that payoff of rank two which the players can each obtain if both remain silent. Thus decisions rational in the individual instance paradoxically combine to determine a collectively undesirable outcome.

The prisoner's dilemma game generalizes a pattern of choice occurring not only with respect to silence and confession in the situation described but also in other situations where cooperation can prove mutually beneficial although no external constraints prevent profit through injury to a player who has behaved cooperatively. Disadvantage is more precisely demonstrated if the payoffs to the players shown in Figure 2 are taken to measure gain in monetary units. Let player R select row 1 with probability $x$ and player C select column 1 with probability $y$. The expected payoffs to the players can now be presented as functions of the probabilities $x$ and $y$:

$$E_R(x, y) = -xy - x + 3y + 1,$$
$$E_C(x, y) = -xy + 3x - y + 1.$$

The rate of change of the expected payoff to each player as that probability governing his conduct changes may also be calculated:

$$\frac{\partial E_R}{\partial x} = -y - 1,$$
$$\frac{\partial E_C}{\partial y} = -x - 1.$$ 

Both $x$ and $y$, being probabilities, must be nonnegative. Because the expression $-y - 1$ is less than zero, the gain to player R varies inversely with $x$. He can maximize his gain by setting $x$ equal to zero without regard to the behavior of the other player. Therefore he is likely to select row 2. Because the expression $-x - 1$ is less than zero, the gain to player C varies inversely with $y$. He can maximize his gain by setting $y$ equal to zero without regard to the behavior of the other player. Therefore he is likely to select column 2.

But benefit to player R depends in part on the choice of player C, and benefit to player C depends in part on the choice of player R. This interdependence is apparent from the rate of change of the expected payoff to each player as that probability governing the conduct of the other changes:

$$\frac{\partial E_R}{\partial y} = -x + 3,$$
$$\frac{\partial E_C}{\partial x} = -y + 3.$$
Neither \( x \) nor \( y \) can exceed unity. Therefore the expressions \(-x + 3\) and \(-y + 3\) must be positive: increases in \( y \) and \( x \) will benefit players \( R \) and \( C \) respectively. An increase in either variable alone will nevertheless cause loss to that player who does not benefit. Although this loss will normally be less than the concomitant gain to the other player, \( x \) and \( y \) are determined by that player who would be injured by their increase. If the players act to protect their individual interests each will obtain a smaller gain than is otherwise available because their non-cooperation will cause reciprocal injury.

IV

THE PARETO STANDARD

One may differentiate, at least conceptually, a set of states of society from which movement benefiting any one individual may be made only at the cost of injury to another. Such states may be characterized as Pareto optimal or efficient.\(^{23}\) Although choice among Pareto optimal states entails evaluation of competing individual claims, the superiority of at least one such state over any designated state outside the set of Pareto optimal states may be defended as almost tautological. If social welfare is obtained by aggregating individual utilities, affirmance of such superiority follows directly from acceptance of the proposition that scarce means should be allocated so as to maximize the satisfaction of ends; if this is an ethical premise it is at least one which generally enjoys unqualified assent.\(^{24}\)

Intervention in a game to secure a particular outcome may or may not reduce the expected payoff to one of the players. Where such reduction occurs, intervention can be justified only through claims ultimately founded on personal preference: loss to one player must be balanced against gain to another. Where no player is injured and at least one player is benefited, however, intervention can be justified by appeal to the Pareto standard. The strictly competitive relationship shown in Figure 1 offers no opportunity for intervention without choice among competing values because gain to one player is invariably associated with loss to the other. In the prisoner's dilemma game shown in Figure 2 all outcomes other than that resulting from mutual confession are optimal in the Pareto sense, since substitution of another outcome

---


for any of them cannot occur without injury to at least one player. Mutual confession does not yield an optimum in this sense because the accompanying payoff to each player is less preferred than the payoff he can obtain if both players remain silent. Each player if unconstrained will nevertheless choose to confess, since by confessing each can maximize his own advantage irrespective of the choice of the other; their combined choices will determine an outcome less satisfactory to either than the outcome that they could have achieved if both had chosen differently. In this situation limitation of choice by external compulsion offers opportunity for benefit to both players.

\[
\begin{bmatrix}
C & 1 \\
1 & 2 \\
1 & 2 \\
2 & 0 \\
2 & 4
\end{bmatrix}
\]

Figure 3

The two patterns of interaction are however not independent: the strictly competitive relationship previously presented can be derived from the prisoner's dilemma game as generalized to provide monetary payoffs if player R in this game is assumed to have committed himself to selection of row 1. As indicated by Figure 3, a strictly competitive relationship paralleling that of Figure 1 but composed of the first column rather than the first row of Figure 2 can be constructed by positing that player C rather than player R has undertaken the initial cooperative commitment. These interdependencies are summarized in Figure 4.

\[
\begin{bmatrix}
C & 1 & 2 \\
1 & 2 & 4 \\
1 & 2 & 0 \\
2 & 0 & 1 \\
2 & 4 & 1
\end{bmatrix}
\]

Figure 4
Enforcement of a decision to compel an outcome optimal to the players in the prisoner's dilemma game by restricting player R to choice of row 1 and player C to choice of column 1 would normally occur in the context of one or the other of the strictly competitive relationships illustrated in Figures 1 and 3: a player whose conduct conforms to the norm thus established should have standing to complain that prospective violation of that norm by the other player will, unless prevented, cause him loss. If deviant behavior is viewed from the narrow perspective of the strictly competitive game, resolve to uphold the prohibition which such behavior offends must rest on an objectively unverifiable assertion that one player rather than the other should benefit. Analysis within the broader framework provided by the prisoner's dilemma game, however, supports the opposite conclusion: intervention to compel cooperation may be justified with reference to at most a minimal value statement because the standard of conduct which such intervention sustains is ultimately advantageous to both players.

V

NEUTRALITY AND GENERALITY RE-EXAMINED

Neutrality and generality and the relationship between these concepts can be given concrete content in terms of the games described. Reference to them identifies at least one source of the significance of the standard introduced by Professor Wechsler. In the context they provide, a principle may be regarded as neutral if its application will not cause loss to any individual. A neutral principle is thus ideally without distributive consequences, serving solely to promote movement toward a position of Pareto optimality.

If disputes at law must without exception be characterized as strictly competitive games, no legal principle can be neutral in the primitive sense posited; by definition, any judicial decision granting or refusing relief must aid one player at the expense of another. The generality of a principle may be taken to be a measure of the conviction with which one can reformulate a dispute to which it is applied as a product of the prisoner's dilemma. A principle is therefore general to the extent that a player whose conduct is challenged as inconsistent with it can be considered likely to invoke the same principle to challenge the conduct of other players in other situations. To this extent judicial action may arguably be justified as universally advantageous.

Interpreted in this way generality is one source of primitive neutrality. As such it is the basis of the social contract and therefore of law itself:

If a covenant be made, wherein neither of the parties perform
presently, but trust one another; in the condition of mere nature, which
is a condition of war of every man against every man, upon any rea-
sonable suspicion, it is void: but if there be a common power set over
them both, with right and force sufficient to compel performance, it
is not void. For he that performeth first, has no assurance the other
will perform after; because the bonds of words are too weak to bridle
men's ambition, avarice, anger, and other passions, without the fear of
some coercive power; which in the condition of mere nature, where
all men are equal, and judges of the justness of their own fears, cannot
possibly be supposed. And therefore he which performeth first, does
but betray himself to his enemy . . . .

But in a civil estate, where there is a power set up to constrain
those that would otherwise violate their faith, that fear is no more
reasonable; and for that cause, he which by the covenant is to perform
first, is obliged so to do.26

The surrogate standard thus introduced retains a substantial compo-
nent of subjectivity; it nevertheless structures an otherwise essentially
formless distinction.

CONCLUSION

In his discussion of Brown v. Board of Education,28 Professor
Wechsler summarized the contentions of the disputants in terms appro-
priate to a strictly competitive game:

But if the freedom of association is denied by segregation, inte-
gration forces an association upon those for whom it is unpleasant or
repugnant. Is this not the heart of the issue involved, a conflict in
human claims of high dimension, not unlike many others that involve
the highest freedoms . . . . Given a situation where the state must
practically choose between denying the association to those individuals
who wish it or imposing it on those who would avoid it, is there a basis
in neutral principles for holding that the Constitution demands that the
claims for association should prevail? I should like to think there is,
but I confess that I have not yet written the opinion. To write it is for
me the challenge of the school-segregation cases.27

Neutrality in this instance requires resort to a principle which on re-
peated application would offer satisfaction to both players. A prin-
ciple of requisite breadth would appear vacuous. Although all individu-
als arguably have an ultimate interest in justice, more precise specifica-
tion of the grounds supporting the decision here seems to compel reli-
ance on a nonneutral rule. That the result is nevertheless desirable in-
dicates that the standard of neutrality does not provide a comprehen-
sive measure of judicial performance.

27. Wechsler, supra note 2, at 34.