ARIEL RUBINSTEIN

THE REASONABLE MAN – A SOCIAL CHOICE APPROACH

ABSTRACT. Most legal systems rely heavily on the notion of the reasonable man. Here an attempt is made to analyze the reasonable man in a social choice model. The main argument is that if the reasonable man satisfies a certain set of axioms, he essentially coincides with one of the individuals, both in terms of his preferences and in terms of his expectations.

1. INTRODUCTION

In recent years central concepts in the social sciences have undergone formal criticism, but for some reason, the basic concepts of legal thought have been overlooked.

This article attempts to deal with one of the most fundamental concepts of legal thought, that of the reasonable man. I feel it necessary to apologise for this pretension. The connection between the reasonable man as a formal concept, as presented below, and the reasonable man as a legal concept requires much broader argument than that provided in this article. When I began dealing with the subject I was convinced that I would be able to refer to existing legal analyses. However, the only abstract conceptual analysis that I came across was Raphael Powell's article on 'The Unreasonableness of the Reasonable Man' [5]. In view of this deficiency, I have restricted this article to indicating some of the problematics of the concept of the reasonable man. I hope that this article will motivate legal scholars to deal with the conceptual aspects of the subject, after which there will be room for a deeper formal discussion.

The concept of 'reasonableness' in its various forms is central to many branches of the law; the test for conviction of an accused in criminal law is that his guilt is to be proved beyond reasonable doubt; the test for determining breach of contract as a 'fundamental breach' (which allows for automatic recession of the contract) is, that it may be assumed that a reasonable person

Theory and Decision **15** (1983) 151–159. 0040–5833/83/0152–0151$00.90

would not have entered into the contract had he foreseen the breach and its consequences; and the criterion as to negligence in criminal law has been defined as "failure to exercise such care, skill and foresight as a reasonable man in the same situation would exercise".

The most important use of the reasonable man is made in the law of torts. The difficulty in defining a standard of careful conduct caused legal systems to define a person as negligent if he failed to conform to a standard of care, reasonable in the circumstances of the case. Atiyah [2] criticizes this use of the concept:

> ... It is clear that a finding that a defendant was negligent involves making a judgement on his conduct; and it is therefore necessary to know what criteria are employed in the process of making this judgement. It is not possible to make a judgement without criteria. At one time the conventional answer to this question was to make the somewhat mystical and certainly much misunderstood figure, the 'reasonable man'.

The reasonable man has been described in numerous ways. For instance: 'average man', 'average juryman', 'man who comes home from work, rolls up his shirt sleeves and mows the lawn', 'never a woman', 'the anthropomorphic concept of justice'. However, as already mentioned, there has been no comprehensive attempt to characterize the concept, and certainly not in an axiomatic manner.

In the majority of its uses, the reasonable man functions as a measure of human conduct. Is this 'measure of human conduct' a measure of actual conduct or does it constitute a criterion for desirable conduct? Powell [5] distinguishes between the 'reasonable probable man' and the 'reasonable usual man'. The reasonable ***probable*** man provides a criterion as to how people *should* behave. From this aspect the concept of the reasonable man expresses the desire to regulate the specific legal problem in the best possible fashion by determining a standard of conduct which does not necessarily comply with the existing standard. The reasonable *usual* man (or the average man) is a criterion for *existing* human behavior. In this sense the reasonable man is a product of the individual's actual behavior and expectations, without regard to the manner according to which people *should* behave.

The various legal systems tend to use the term 'reasonable man' interchangeably, either in the sense of 'reasonable probable man' or in the sense of 'reasonable usual man', and sometimes for both. A doctor who has taken all precautionary measures ordinarily taken by his professional colleagues, will be judged in terms of the reasonable usual man. On the other hand, a driver who drives his car at an excessive speed like all the drivers on that road, will be judged according to the standard of the reasonable probable man.

In this article I will concentrate on the analysis of the reasonable usual man. Discussion of the reasonable probable man requires, to my mind, a different analysis and I hope to deal with this question in a separate article.

The formal presentation of the problem is based on a distinction between the set of actions, A, and the set of outcomes B. The set A can be interpreted as the set of all actions possible in a specific situation while the set B is the set of their possible consequences. Some other interpretations are also possible. For example, let A be the set of possible texts (like contracts or decisions made by a committee) and B be the set of interpretations of those texts.

A society consists of n individuals. Each individual is characterized by two elements: (i) a preference order on the set of outcomes, and (ii) a function from A into B which is called a realization function. A realization function expresses the manner in which an individual expects that actions in A will result in outcomes in B. Notice, that a realization function assigns to an action a single outcome. In Section 4, it will be made clear that our framework also enables us to consider probability expectations by considering B to be the set of probability measures on a basic set of outcomes.

The individuals can differ in their expectations and preferences, and the reasonable man is presented as a function that assigns to each society a preference order on B and a realization function. Our three assumptions as to the reasonable man are:

(1) The preference relation of the reasonable man is a function of preference relations alone, and the realization function of the reasonable man is a function of realization functions alone.

(2) If all individuals anticipate that a certain action will cause a particular outcome, the reasonable man also adheres to this anticipation.

(3) If every individual prefers what he believes to be the realization of one action to what he believes to be the realization of another action then it is true as well about the reasonable man.

Now we come to the main question of this study: is it possible to find an acceptable way to aggregate the individuals' expectations and preferences?

The main theorem states that assuming the above assumptions any such aggregation coincides with one of the individuals in the society in the sense that if he prefers what he believes to be the realization of $a_1 \in A$ to what he believes to be the realization of $a_2 \in A$ then it is true as well about the reasonable man. Thus, if our formal description of the reasonable man is adequate and the above assumptions are acceptable, a consistent judge should identify the choice made by a reasonable man with the choice that would be made by a certain individual in his society. This conclusion is in the spirit of the following quotation from Allen (see Atiyah [2]): "Nobody is deceived by the fiction that the judge is stating not what he himself thinks, but what he thinks an average reasonable man might think". On the other hand, the above conclusion is not in accordance with the more common view. According to this view, the judge is supposed to refer to his personal experience when determining the reasonable man, so as to find the *average* of all the individuals' outlooks.

It should be noted that this article could be regarded within the context of Social Choice Theory. The difference between the model which will be presented hereinafter and traditional problems of Social Choice Theory lies in the lack of identity between the set of actions which constitute alternatives for choice and the set of outcomes on which the individuals' preferences are defined. The reasonable man aggregates orderings of outcomes and functions mapping actions to outcomes. Essentially, a dictatorship result is proved. The result is achieved without imposing the condition of Independence of Irrelevant Alternatives. In Proposition 3 this condition is derived from the assumptions about the reasonable man. This proposition may be considered as a justification for the Independence of Irrelevant Alternatives.

Finally, after completing this research, I came across an impossibility result of Bayesian group decision-making with separate aggregation of beliefs and values due to Hylland and Zeckhauser [4]. Although there are differences in the models, their results is in the same spirit of the current paper.

2. THE MODEL

Let $N = \{1, \ldots, n\}$ be the set of individuals in society, A be the set of actions, and B be the set of outcomes which may follow from actions in A. Assume $|A| \geqslant 2$ and $|B| \geqslant 3$.

Let Ω be the set of all transitive, reflexive and connected binary relations on B.

Given $\succsim$, R and S are relations in Ω, notations like '$r \succ s$' and '$r \sim s$' are used instead of '$r \succsim s$ and not $s \succsim r$' and '$r \succsim s$ and $s \succsim r$' respectively, and the notation '$R[r, s] \Leftrightarrow S[t, u]$' is used instead of '$r\,R\,s \Leftrightarrow t\,S\,u$ and $s\,R\,r \Leftrightarrow u\,S\,t$'.

Let $\theta \subseteq \Omega$ be a set of admissible preferences. We assume θ satisfies the following three assumptions:

(A1) For every function H: $\theta^N \to \theta$ which satisfies the conditions Independence of Irrelevant Alternatives (I) and Strong Pareto (P^*) there exists $j \in N$ such that $r \succ^j s \Rightarrow r \succ s$, where $H(\succsim^1, \ldots, \succsim^n) = \succsim$, $H(\bar{\succsim}^1, \ldots, \bar{\succsim}^n) = \bar{\succsim}$ and I and P^* are the following two conditions:
Condition I. For all $r, s \in B$, $(\succsim^1, \ldots, \succsim^n)$ and $(\bar{\succsim}^1, \ldots, \bar{\succsim}^n)$ in θ^N if for any $i \in N$, $\succsim^i[r, s] \Leftrightarrow \bar{\succsim}^i[r, s]$ then $\succsim[r, s] \Leftrightarrow \bar{\succsim}[r, s]$.
Condition P^*. If for all $i \in N$, $r \succsim^i s$ then $r \succsim s$, and if there is also a j such that $r \succ^j s$ then $r \succ s$.

(A2) For any $r, s \in B$ there exists $R \in \theta$ such that $r\,R\,s$.

(A3) For any $r, s, t, u \in B$ $(r \neq s)$, if for every $R \in \theta$ $t\,R\,u \Rightarrow r\,R\,s$ then for every $R \in \theta$ $R[t, u] \Leftrightarrow R[r, s]$.

(A1) ensures that Arrow's impossibility theorem is valid in relation to θ. (A2) demands that θ be sufficiently 'rich' so that for any two outcomes r, s in B there is a relation in θ according to which r is preferable to s. (A3) demands that the fact that t is preferred to u can indicate that r is preferred to s, only if for all $R \in \theta$, there is a complete correlation between the order of R on $\{r, s\}$ and on $\{t, u\}$.

REMARK. Notice that Ω satisfies (A1), (A2) and (A3). From Arrow's impossibility theorem [1], Ω satisfies (A1). It is clear Ω satisfies (A2). For the proof that Ω satisfies (A3), let $r, s, t, u \in B$ $(r \neq s)$. If for all $R \in \Omega$ $t\,R\,u \Rightarrow r\,R\,s$, then $t = r$ and $u = s$, and therefore for all $R \in \Omega$, $R[t, u] \Leftrightarrow R[r, s]$.

A function f: $A \to B$ is called a realization function. Let F be a set of the admissible realization functions. Assume that F satisfies the following condition:

(B1) For any two different outcomes $r, s \in B$, there are two actions a, $b \in A$ and f, $h \in F$, such that $f(a) = r$, $f(b) = s$, $h(a) = s$ and $h(b) = r$.

An individual i is characterized by an ordered pair $(\succsim^i, f^i) \in \theta \times F$. $f^i(a)$ is interpreted as the outcome that i associates with action a, and $\succsim^i$ is the preference relation of i on B. Society is an n-tuple $\{(\succsim^i, f^i)\}_{i=1}^n$.

The reasonable man is defined as a function which for every society S assigns a pair $(\succsim(S), f(S)) \in \theta \times F$. The reasonable man is assumed to satisfy the following three assumptions (denote $S_j = \{(\succsim_j^i, f_j^i)\}_{i=1}^n$):

(C1) If S_1 and S_2 are societies such that for every i $\succsim_1^i = \succsim_2^i$ then $\succsim(S_1) = \succsim(S_2)$ and if for all $i, f_1^i = f_2^i$ then $f(S_1) = f(S_2)$.

Due to (C1) we can write $f(f^1, \ldots, f^n)$ and $\succsim(\succsim^1, \ldots, \succsim^n)$ instead of $f(S)$ and $\succsim(S)$. We shall abbreviate and simply write f and $\succsim$.

(C2) Let S be a society in which for all $a \in A$ there is $r \in B$ such that for all $i, f^i(a) = r$. Then $f(a) = r$.

(C3) If for all i, $f^i(a) \succsim^i f^i(b)$ then $f(a) \succsim f(b)$, and if there is also j such that $f^j(a) \succ^j f^j(b)$ then $f(a) \succ f(b)$.

Condition (C1) ensures that the reasonable man can be separated in two functions; the preference relation of the reasonable man depends on the preference relations of the individuals only, and the reasonable realization function is a function of the realization functions only.

(C2) is a natural assumption; if all the individuals expect r to follow action a, then the reasonable man shares the expectations of the individuals.

(C3) links the preferences and the expectations of the reasonable man. If all individuals expect action a to cause an outcome which is preferable to the outcome expected to follow action b, then the reasonable man also, expects the outcome expected by him to follow a, to be preferable to the outcome expected by him to follow b.

3. THE THEOREM

We will say that (r, s) and (t, u) are θ-equivalent if for every $R \in \theta$ $R[r, s] \Leftrightarrow R[t, u]$.

THEOREM. There is an individual j such that

(1) if $r \succ^j s$, then $r \succ s$

(2) for given $a, b \in A$ and $f^1, \ldots, f^n \in F$ either $f(a) = f(b)$ or $(f(a), f(b))$ and $(f^j(a), f^j(b))$ are θ-equivalent.

(3) if $f^j(a) \succ^j f^j(b)$ then $f(a) \succ f(b)$.

The proof of the theorem is split into a series of simple propositions:

PROPOSITION 1. $\succsim$ satisfies condition P^*.

Proof. Assume that for all i, $r \succsim^i s$. From (B1) there are a and b in A and $g \in F$ such that $g(a) = r$ and $g(b) = s$. Let us look at a society where for each i, $f^i = g$. For all i, $f^i(a) \succsim^i f^i(b)$ and therefore from (C3), $f(a) \succsim f(b)$. But for all i, $f^i(a) = r$ and $f^i(b) = s$ and therefore from (C2), $f(a) = r$ and $f(b) = s$, and so $r \succsim s$. In a similar way, utilizing the last part of (C3), the rest of the proposition follows.

PROPOSITION 2. Let $a, b \in A$ and $f^1, \ldots, f^n \in F$ satisfy for all i, $f^i(a) \neq f^i(b)$. Then there is j (possibly depending on a, b and $f^1, \ldots, f^n$) such that for all $R \in \theta$ $R[f(a), f(b)] \Leftrightarrow R[f^j(a), f^j(b)]$.

Proof. For all i, $f^i(a) \neq f^i(b)$. Therefore, from (A3), it is sufficient to show that there exists j such that for all $R \in \theta$ $f(a)\, R\, f(b) \Rightarrow f^j(a)\, R\, f^j(b)$. If not, then for all i, there exists $R^i \in \theta$ such that $f(a)\, R^i\, f(b)$ and $f^i(b)\, P^i\, f^i(a)$. Let us look at the society $\{(f^i, R^i)\}_{i=1}^n$. From (C3) it follows that $f(b)\, P\, f(a)$, and from Proposition 1 it follows that $f(a)\, R\, f(b)$. Contradiction!

PROPOSITION 3. $\succsim$ satisfies condition I.

Proof. Let $(\succsim^1, \ldots, \succsim^n)$ and $(\bar{\succsim}^1, \ldots, \bar{\succsim}^n)$ be in θ^N and r, s two outcomes such that for all i, $\succsim^i[r, s] \Leftrightarrow \bar{\succsim}^i[r, s]$. Assume $r \succ s$. We shall prove that $r \mathrel{\bar{\succ}} s$ $(\succsim(\bar{\succsim}^1, \ldots, \bar{\succsim}^n) = \bar{\succsim})$. From (B1) there are realization functions $g, h \in F$ and $a, b \in A$ such that $g(a) = r$, $g(b) = s$, $h(a) = s$ and $h(b) = r$. Define

$$f^i = \begin{cases} g & r \succ^i s \\ h & s \succsim^i r \end{cases}.$$

In the society $\{(\succsim^i, f^i)\}_{i=1}^n$, for all i, $f^i(a) \succsim^i f^i(b)$, and therefore $f(a) \succsim f(b)$.

There exists at least one i such that $f^i(a) \succ^i f^i(b)$; otherwise for all i $s = f^i(a) \sim^i f^i(b) = r$ and from (C2) $r \sim s$ in contradiction to the assumption that $r \succ s$.

From (C3) it follows that $f(a) \succ f(b)$ and for the same reason $f(a) \bar{\succ} f(b)$. For all i, $f^i(a) \neq f^i(b)$ and therefore from Proposition 2 there exists j such that for all $R \in \theta$ $R[f(a), f(b)] \Leftrightarrow R[f^j(a), f^j(b)]$. Therefore, either for all $R \in \theta$ $R[f(a), f(b)] \Leftrightarrow R[r, s]$, or, for all $R \in \theta$, $R[f(a), f(b)] \Leftrightarrow R[s, r]$. But $f(a) \succ f(b)$ and $r \succ s$, therefore $R[f(a), f(b)] \Leftrightarrow R[r, s]$, for every $R \in \theta$. In particular, $f(a) \bar{\succ} f(b)$ implies $r \bar{\succ} s$.

Proof of the Theorem. According to Proposition 1, $\succsim$ satisfies condition P^* and by Proposition 3, $\succsim$ satisfies Condition I. Therefore by (A1) there exists j such that if $r \succ^j s$ then $r \succ s$.

Let $f^1, \ldots, f^n \in F$. Assume that there exist $a, b \in A$ and $R \in \theta$ such that $f(a) \neq f(b)$, $f^j(a) R f^j(b)$ and not $f(a) R f(b)$. Let $\succsim^j = R$, and for all $i \neq j$ choose $\succsim^i \in \theta$ which satisfies $f^i(a) \succsim^i f^i(b)$ (the existence is guaranteed by (A2)).

Let us look at the society $\{(\succsim^i, f^i)\}_{i=1}^n$. For all i, $f^i(a) \succsim^i f^i(b)$ and therefore from (C3) $f(a) \succsim f(b)$, but from Proposition 4 it follows that $f(b) \succ f(a)$.

4. A MODEL WITH UNCERTAINTY

Let $\bar{B}$ be a finite set, $|\bar{B}| = m \geqslant 4$. Given M is a set containing m elements, denote by $P(M)$ the set of all probability measures on M and identify an element in $P(M)$ with a non-negative vector in R^m. Let $B = P(\bar{B})$.

Let θ be the set of all preference relations $\succsim$ on B which satisfy that there exists $v \in R^n$ such that

$$p \succsim q \Leftrightarrow v \cdot p \geqslant v \cdot q.$$

Denote v and v^j the vectors which represent $\succsim$ and $\succsim^j$. The elements of θ can be identified by vectors in R^n normalized $0 - 1$ (with the identification of the vectors $\mathbf{0}$ and $\mathbf{1}$). Define a topology on θ, $\succsim^n \to \succsim$ if $v^n \to v$.

In place of (A1) assume a similar condition

(A1)* For all continuous H: $\theta^n \to \theta$ which satisfy conditions I and P^* there exists j such that $r \succ^j s \Rightarrow r \succ s$.

Let F be the set of all functions from A to B. In this model the realization function assigns to each action a probability measure which represents the belief regarding how the action leads to outcomes in $\bar{B}$.

PROPOSITION. In this model if $(\succsim, f)$ is a reasonable man (satisfying (C1), (C2) and (C3)) and $\succsim$ is a continuous function, then there exist j, $\alpha \geqslant 0$ and β such that $v = \alpha v^j + \beta 1_m$ and for all $a, b \in A$ there exists $\gamma \geqslant 0$ such that $f(a) - f(b) = \gamma[f^j(a) - f^j(b)]$.

Proof. From Kalai and Schmeidler [3] θ satisfies (A1)*. It is clear that θ satisfies (A2). Let us show that θ satisfies (A3). For all $r \neq s, t, u$ in B, if for all $R \in \theta$ $t\,R\,u \Rightarrow r\,R\,s$, then for all $x \in R^n$ $(t-u)x \geqslant 0 \Rightarrow (r-s)x \geqslant 0$. From Farkash's Lemma there exists $\lambda \geqslant 0$ such that $(r-s) = \lambda(t-u)$, but $r \neq s$ and therefore $\lambda > 0$ and $(t-u) = 1/\lambda(r-s)$. Therefore, $\succsim[r,s] \Leftrightarrow \succsim[t,u]$.

From the proof of the theorem it follows that there exists j such that:

(1) If $p \succ^j q$ then $p \succ q$, i.e., for all p, q $p \cdot v^j > q \cdot v^j \Rightarrow p \cdot v > q \cdot v$. Thus $v = \alpha v^j + \beta 1_m$ when $\alpha \geqslant 0$ and $\beta \in R$.

(2) For all $a, b \in A$, if $f(a) \neq f(b)$, then for all $w \in R^n$ $[f(b) - f(a)] \cdot w \geqslant 0$ iff $[f^j(b) - f^j(a)] \cdot w \geqslant 0$, and therefore there exists $\gamma > 0$ such that $f(b) - f(a) = \gamma[f^j(b) - f^j(a)]$.

REFERENCES

[1] Arrow, K. J., *Social Choice and Individual Values*, Yale University Press, New Haven, 1963.

[2] Atiyah, P. C., *Accidents Compensation and the Law*, Weidenfeld and Nicolson, London, 1970.

[3] Kalai, E. and Schmeidler, D., 'Aggregation Procedure for Cardinal Preferences', *Econometrica* **45** (1977), 1431–1437.

[4] Hylland, A. and Zeckhauser, R., 'The Impossibility of Bayesian Group Decision Making with Separate Aggregation of Beliefs and Values', *Econometrica* **47** (1979), 1321–1336.

[5] Powell, R., 'The Unreasonableness of the Reasonable Man', *Current Legal Problems* **157** (1957), 104–126.

Department of Economics,
The Hebrew University,
Jerusalem,
Israel.