Fairness and Dynamic Consistency

Daniel Schoch

1. Fairness and Expected Utility

It has been argued that valuing fairness, in contrast to valuing equality, is incompatible with Expected Utility (Diamond 1967, Broome 1982). Consider Diamond’s famous example of the following two social lotteries

\[
L_1 = \begin{cases} (0, 0) & (0, 1) \\ 1/2 & 1/2 \end{cases}, \quad L_2 = (1, 0).
\]

The society consists of two equally deserving persons. In alternative \(L_2\), one person is assigned a (utility or income) level of one for sure, while the other person has level zero. In lottery \(L_1\), the situation is completely symmetrical such that both persons have an equal chance to obtain the higher level. A common interpretation of the zero and one states are “being dead” and “being alive”. Alternatively, the state “one” could symbolise an undividable good to be distributed among the two persons by a well-meaning decider, for example a mother giving a single toy to their children (Machina 1989). Fairness seems to demand to prefer \(L_1\) over \(L_2\), but vNM-rationality evaluates both alternatives equal, even if the underlying social ordering respects distributional equality by satisfying the Pigou-Dalton transfer condition.\(^1\)

Preferring \(L_1\) over \(L_2\) implies a violation of the sure-thing principle,\(^2\) which is a necessary condition for vNM expected utility. Since we are indifferent between \((1, 0)\) and \((0, 1)\), and \(L_2\) could equivalently be written as

\[
L_2 = \begin{cases} (1, 0) & (1, 0) \\ 1/2 & 1/2 \end{cases},
\]

the sure-thing principle allows us to replace the rightmost outcome by \((0, 1)\), which leads us to \(L_1\). Thus \(L_2\) and \(L_1\) are indifferent.

---

\(^1\) For the sake of simplicity, we are dealing with lotteries only, but the argumentation below might be translated to the formalism of Savage actions.

\(^2\) The formulation of the sure-thing principle we will need here says that if one alternative with a sure outcome of \(x\) is valued indifferent to a lottery \(L\), then any mixture of \(x\) with any lottery \(L'\) is indifferent to the same mixture of \(L\) with \(L'\). Formally,

\[
x \sim L \Rightarrow p \cdot x + (1-p) \cdot L' \sim p \cdot L + (1-p) \cdot L'
\]

In short: sure things can be replaced by indifferent lotteries. We might presume that \(x\) and \(L\) on one hand, and \(L'\) on the other hand refer to disjoint events.
Some basic notation first. Let $\mathcal{P}(\mathbb{R}^n)$ denote the set of lotteries on $\mathbb{R}^n$. For $L \in \mathcal{P}(\mathbb{R}^n)$, $E(L) \in \mathbb{R}^n$ denotes the ($n$-dimensional) expectation value of $L$, i.e. the vector containing the expectation values for each person. Let $\mathbf{1} = (1, 1, \ldots) \in \mathbb{R}^n$ be the equi-distributed vector assigning 1 to every person. On $\mathbb{R}^n$, $x \cdot y = \Sigma_{i=1}^n x_i \cdot y_i$ denotes the euclidean scalar product.

We define fairness as distributional equality of the expectation values of a lottery. In an absolute sense, we call a lottery $L$ fair, if and only if it assigns equal expectations to each persons. As a partial order of distributional equality, the Lorentz half-ordering has been established (Sen 1971). We say that $x \geq y$ hold for $x, y \in \mathbb{R}^n$, if and only if $x$ results from $y$ by a sequence of Pigou-Dalton transfers and Pareto improvements.

**Definition 1.** A lottery $L \in \mathcal{P}(\mathbb{R}^n)$ is **fair** iff there exists an $\alpha \in \mathbb{R}$ with

$$E(L) = \alpha \cdot \mathbf{1}.$$ 

An ordering $\succeq$ on $\mathcal{P}(\mathbb{R}^n)$ is said to respect fairness iff for each lotteries $L, L' \in \mathcal{P}(\mathbb{R}^n)$

$$E(L) \succeq E(L') \Rightarrow L \succeq L' \text{ and }$$

$$E(L) \succeq E(L') \Rightarrow L \preceq L'.$$

It is easy to construct a social welfare ordering respecting fairness. Let $W$ be any continuous, convex\(^3\) and symmetrical function $\mathbb{R}^n \rightarrow \mathbb{R}$, which is monotonously increasing in each argument. Then the function

$$(1.1) \quad U(L) = W(E(L))$$

induces an ordering $\succeq$ on $\mathcal{P}(\mathbb{R}^n)$ satisfying the Principle of Personal Good\(^4\), the Pigou-Dalton condition, anonymity, Lehrer’s theorem (Sen 1971) and fairness\(^5\). The function $W$ can be chosen either as an egalitarianist’s or as a prioritarianist welfare function.

\(^3\) This means

\[ W(\lambda x + (1 - \lambda) y) > \lambda W(x) + (1 - \lambda) W(y) \]

for each $x, y \in \mathbb{R}^n$ and $0 < \lambda < 1$.

\(^4\) The Principle of Personal Good says that if a prospect $L$ is at least as good for each individual as a prospect $L'$, then it is overall at least as good as it, and if it is even better for some individuals, then it is overall better. It implies the Pareto principle on sure prospects. In presence of the sure-thing principle, it is equivalent to the Pareto principle and first-order stochastic dominance. For details see Broome 1991, Rabinowicz, P3 and the literature cited there.

\(^5\) It has to be noticed that the following condition for compromise fairness, which can be found in Karni & Safra (2002) is in general not satisfied: For non-certain lotteries $L \neq L'$ and $L \sim L'$

$$\frac{1}{2} L + \frac{1}{2} L' \not\succ L.$$
(Fleurbaey 2001, Rabinowicz). The ordering does not, however, differentiate between the two lotteries
\[
L_3 = \left\{ \begin{array}{ll}
(1,1) & (0,0) \\
1/2 & 1/2
\end{array} \right\}, \quad L_1 = \left\{ \begin{array}{ll}
(1,0) & (0,1) \\
1/2 & 1/2
\end{array} \right\}.
\]

It has been argued that preferring $L_3$ over $L_1$ is a benchmark to decide between egalitarianism, where $L_3$ is preferred over $L_1$ on the ground that $L_3$ results a totally equal outcome for sure, and prioritarianism, which accounts both lotteries as equal (Broome 1991, 2001). This presumes that the expectation value is formed \textit{ex post} as expected welfare. In our approach above, the two lotteries are always indifferent even for an egalitarian welfare function $W$. It might, however, be possible to linearly combine expected welfare with (1.1) in order to achieve $L_3 \succ L_1 \succ L_2$.

2. The Dynamic Inconsistency Objection

It has been proposed to discard vNM rationality for social decisions (Diamond 1967, Broome 1982), but this proposal was subject to criticism (e.g. Broome 1984). The most prominent is that giving up the sure-thing principle leads to a violation of \textit{dynamic consistency}. Dynamic consistency requires that the evaluation of a lottery does not depend on the way it is being decomposed into a decision tree and evaluated piecewise. To understand this, one first has to keep in mind that under very general assumptions each lottery has at least one \textit{certainty equivalent}, i.e. a single-outcome lottery to which it is indifferent.\footnote{Let $L$ be any lottery, $\alpha$ be the minimal outcome any person can achieve, and $\beta$ the maximal possible outcome. For any $x \in \mathbb{R}$, let $C(x)$ denote the lottery consisting of the only sure outcome $x$. By the Pareto principle and first-order dominance, $C(\beta) \succeq L \succeq C(\alpha)$. Then, by continuity (in the vNM sense), there exists a $\gamma \in [\alpha, \beta]$ with $C(\gamma) \sim L$, which is a certainty equivalent for $L$.}

Assume now, a lottery $L$ could be written $L = p \cdot L_A + (1 - p) \cdot L_B$ with lotteries $L_A$ and $L_B$ referring to disjoint events (see fig. 1). If $L_A$ and $L_B$ are optimal choices in their respective subtrees, so should be $L$ as a mixture of the \textit{values} of the two subtrees. In other words, if $L_A$ and $L_B$ are substituted by certainty equivalents in $L$, this lottery should be indifferent to $L$.\footnote{Observe, that dynamic consistency does not only require the decomposability of lotteries into sub-lotteries, which is an analytical truth, but the possibility to distribute the very rationality principle of maximisation onto subtrees without changing the order.} This is nothing but the sure-thing principle.
Broome (1984) stated several more arguments against dropping expected utility maximisation. The iteration objection is the most important of them. It states that since a fair lottery has always to be preferred over each of its determinate outcomes, a fair random selection procedure could never be finally established, since whichever alternative turned out has to be disregarded and randomisation has to be repeated. For Machina (1989), the iteration objection just presumes dynamic consistency, which does not hold. Other objections of Broome (loc. cit.) can be undermined by noting that Broome’s proposed solution of integrating fairness into the outcome states does not allow for a gradational judgement of fairness.

Defenders of intrinsic representation of fairness might point out the holistic character of a consequentialist decision making (Machina 1989, McClennan 1990). Only the best lottery over all possible outcomes is morally relevant, restrictions to subsets would presume prejudices. Non-terminal sublotteries in a decision tree are essentially lotteries over lotteries and have to be treated as such. This reflects that from the standpoint of consequentialism all intermediate steps have to be valued with respect to the final outcome. From a holistic point of view, the iteration objection loses its base: If we actually replace one of the outcomes in lottery $L_1$ from above, say $(1, 0)$, by $L_1$ itself, we obtain a suboptimal lottery

$$\begin{cases}
(1, 0) \quad (0, 1) \\
1/4 \quad 3/4 \end{cases}.$$ 

If both alternatives are being replaced by $L_1$, then the original lottery is reproduced.

Dynamic Inconsistency is often described as a change of preference. From the standpoint of a consequentialist agent looking backward from the final outcomes, it could better be described as a change in his or her aggregation principle. That means to admit that the very principle under which she or he acts, is not universally applicable. There seem to be two possible answers to this challenge. Firstly, one could try to establish certain probabilities as a matter of nature. This obviously does not apply to probabilities stemming from randomisation procedures the agent has set up in order to establish fairness. Secondly, one has to restrict man-made probabilities to such cases, in which dynamic consistency nevertheless holds. We follow the last strategy.
3. Taurek’s Example

The following example from John Taurek cited in Broome (1984) shows that in some cases dynamic consistency can nevertheless be established alongside fairness. A volcanic island is about to explode and only one (for simplicity) out of five inhabitants can be saved by the only boat around. Before approaching the island, a decision has to be taken to go either to the north (A), where three people are waiting to be rescued, or to go to the south (B), where two people stay. It seems to be preferable to go to the north where chances are more equally spread, but either determinate choice is suboptimal. The optimal randomisation procedure is shown in figure 2.

The striking point is now that in this special case dynamic consistency is preserved for an ordering of type (1.1). As above, we assume that the saved person will reach level “one”, while a “zero” is assigned to the rest. The optimisation problem for sublottery A and B are \((p_1, p_2, 1 - p_1 - p_2, 0, 0)\) and \((0, 0, 0, p_3, 1 - p_3)\), which by the Pigou-Dalton condition (convexity of W) results to \(p_1 = p_2 = 1/3\) and \(p_3 = 1/2\). The certainty equivalents of A and B are \((1/3, 1/3, 1/3, 0, 0)\) and \((0, 0, 0, 1/2, 1/2)\), respectively. The combined lottery of both has a certainty equivalent of

\[(3.1) \quad (1/3 \cdot p, 1/3 \cdot p, 1/3 \cdot p, 1/2 \cdot (1 - p), 1/2 \cdot (1 - p))\,.

Since the total sum of the components of the vector (3.1) is independent of \(p\), the optimal solution can be found by equating its components. We obtain only one equation, \(1/3 \cdot p = 1/2 \cdot (1 - p)\), which gives us \(p = 3/5\). This corresponds to an overall chance of 1/5 for each inhabitant, which is the optimal choice for a fair global lottery. Thus we

---

8 Broome seems to miss this point by taking the direction of the boat as a yes-or-no decision, which could be motivated by his general intuition against intrinsic representation of fairness.
have shown that the global optimisation problem leads to the same result, as the piece-
wise optimisation procedure belonging to the decision tree in figure 2. We are justified
in claiming that the partition given by this tree is dynamically consistent.

4. A Sufficient Condition for Dynamic Consistency

The previous example showed that we can establish dynamic consistency in those cases,
where fairness could be achieved without 'weighing goods', but utilizing general prin-
ciples which hold for any ordering of type (1.1) regardless of the special choice of the
social welfare function $W$. We consider the general case of a lottery, which can be de-
composed into $k$ lotteries $L_1, \ldots, L_k$. We presume that randomisation is always possible,
i.e. the set of lotteries, from which the agent chooses, is convexly closed. Only the last
optimisation procedure is subject of the following theorem, the optimality of the
$L_i$ is tacitly assumed. The result could be iterated to apply to more general trees.

**Theorem 1.** Let $L_1, \ldots, L_k$ be lotteries and $\preceq$ be an ordering of type (1.1), and let $L$
be a fair lottery in

$$\Delta(L_1, \ldots, L_k) = \left\{ \begin{array}{l}
p_1 \cdot L_1 + \ldots + p_k \cdot L_k | \sum_{i=1}^k p_i = 1 \end{array} \right\}.$$

If the expectation values of the lotteries $L_i$ have equal sum,

$$E(L_i) \cdot 1 = E(L) \cdot 1,$$

then $L$ is maximal in $\Delta(L_1, \ldots, L_k)$.

**Proof.** We set $c^i = E(L_i)$ and define $L'(p') = p'_1 \cdot L_1 + \ldots + p'_k \cdot L_k$ for each $p' \in \mathbb{R}^n$ with

$$\Sigma_{i=1}^n p'_i = 1.$$ Since $L \in \Delta_0 := \Delta(L_1, \ldots, L_k)$, there exists a $p \in \mathbb{R}^n$ with $\Sigma_{i=1}^n p_i = 1$ and $L = L'(p)$. We have to show that $p' = p$ is optimal. We find for the sum of components of $E(L'(p'))$,

$$E(L'(p')) = \sum_{j=1}^n \sum_{i=1}^k p'_i \cdot c^j = \sum_{i=1}^k p'_i \cdot \sum_{j=1}^n c^j.$$ By assumption, the last sum $\Sigma_{j=1}^n c^j = E(L_0) \cdot 1$ is a constant, say $n \cdot \gamma$ for convenience.
Thus the sum of components of $E(L'(p'))$ is independent of $p'$. Since $W$ is symmetrical
and convex, any optimal value for $p'$ maximising $W(E(L'(p'))) \text{ would result to a fair}
lotterie, $E(L'(p'^{\text{opt}})) = \gamma \cdot 1$. But such a solution exists, namely $p'^{\text{opt}} = p$, since $L$ is a fair
lottery in $\Delta_0$.

We have just shown that each maximal lottery in $\Delta_0$ is fair. It remains to show that
each two fair lotteries from $\Delta_0$ are equivalent. Assume $L'(p^1)$ and $L'(p^2)$ are both fair,
$E(L'(p^1)) = \alpha \cdot 1$ and $E(L'(p^2)) = \beta \cdot 1$, and assume that $L'(p^1) \preceq L'(p^2)$. Then $\alpha \neq \beta$. On
the other hand, as we have see above, $E(L'(p')) \cdot 1 = n \cdot \gamma$ for all $p'$, which contradicts

759
\[ E(L'(p^1)) \cdot 1 = n \cdot \alpha \neq n \cdot \beta = E(L'(p^2)) \cdot 1. \]

Thus \( L'(p^{opt}) \sim L \), which shows that \( L \) is maximal in \( \Delta_0 \).

**References**


