We discuss here the connection between the game-theoretic representation of the principal-agent problem and its representation in terms of standard production and consumption sets. In addition we consider the issue of a-rationality as it applies in this framework.

The simplest game-theoretic representation of a principal-agent problem consists of a game with two players – the principal \( P \) and agent \( A \). Assume that the principal selects an element \( d_P \) from its decision set \( D_P \) in advance of the agent. The agent, after having perfectly observed which \( d_P \) has been chosen, then selects an element from its own decision set \( D_A \). The following highly simplified example captures the essential features of the problem:

\[
D_P = \{d'_P, d''_P\}, \quad D_A = \{d'_A, d''_A\} .
\]

An outcome of the game emerges when both players have selected decisions. The set \( X \) of outcomes has as its elements

\[
x_1 = (d'_P, d'_A), \quad x_2 = (d'_P, d''_A), \quad x_3 = (d''_P, d'_A), \quad x_4 = (d''_P, d''_A).
\]

Call \( X \) a game form; fig. 1 depicts \( X \).

Let \( P \) have preferences

\[
x_1 \succ_P x_3 \succ_P x_4 \succ_P x_2 ,
\]

whereas \( A \) exhibits

\[
x_3 \succ_A x_4 \succ_A x_2 \succ_A x_1 .
\]

* The authors have benefited greatly from discussions with Tom Russell on this topic.
Since this game is finite and is characterized by perfect information it possesses a perfect Nash equilibrium (P.N.E.). In the above case there are two P.N.E.'s, $x_3$ and $x_4$. The reason for this is easy to see.

If $P$ chooses $d'_P$, then $A$ responds with $d''_A$ so that the resulting outcome is $x_2 = (d'_P, d''_A)$ which is judged worst under $P$'s preferences. So $P$ chooses $d''_P$. Notice that if $A$'s preferences were such that

$$x_1 >_{A} x_2,$$

or if the game form is given by fig. 2, then $x_4$ would emerge as the P.N.E. This point is extremely important and we return to it again below.

Fig. 1.

Fig. 2.
Consider the case of a firm which faces a principal-agent problem with the above described features. For instance, the principal might correspond to the owner of the firm and the agent its manager. The difference in preferences might then reflect different attitudes to risk [see Russell and Thaler (1980)] or a difference in incentives [see Jensen and Meckling (1976)]. It turns out though that the specific reason for this difference is not important for the questions discussed below. Let the firm produce a single product whose output level $q$ is determined by the decisions of the two players,

$$q = f(x),$$

where $x \in X$ and $f$ is a ‘production function’,

$$f(x_1) > f(x_3) = f(x_4) > f(x_2).$$

Suppose that the firm’s principal is able to choose among various game forms. Thus the feasible game forms really constitute sub-games of a larger game. Each game form can be thought of as describing an alternative hierarchical management structure for the firm. Given the preferences of the agent then, the P.N.E. stipulates the output level $q$ which emerges when the game form in question is employed. Since agent preferences are a variable in this setting we distinguish two types of agents in the example under discussion. Agent $A_1$ has preferences which satisfy (4) while $A_2$’s preferences satisfy (5).

From the principal’s point of view the selection of a game form together with a specification of the agent(s) employed gives rise to a point in the production set of the firm. Let $y$ denote such a point. Then $y$ will have as its components $-y_1$, $-y_2$, and $q$ where

$$y_1 = 1 \quad \text{if agent is } A_1,$$

$$= 0 \quad \text{otherwise},$$

$$y_2 = 1 \quad \text{if agent is } A_2,$$

$$= 0 \quad \text{otherwise},$$

$q =$ output level and satisfies $q = f(x)$.

Let $Y$ denote the set of feasible production points. Then $(-1, 0, f(x_3)) \in Y$ when the game form $\{x_1, x_2, x_3, x_4\}$ is available to $P$, $(-1, 0, f(x_1)) \in Y$ when the game form $\{x_1, x_3, x_4\}$ is available, and $(0, -1, f(x_1)) \in Y$ if either $\{x_1, x_2, x_3, x_4\}$ or $\{x_1, x_3, x_4\}$ is available. Those game forms which generate the boundary of $Y$ are especially important in the discussion that follows.

It should be clear that the procedure outlined above is quite general and can be used to generate a standard production set $Y$ for a firm which faces a potential principal-agent conflict.

It has been noted elsewhere [see Hammond (1976)] that the changing tastes problem violates the condition of $\alpha$-rationality. This also applies to the class of
hierarchical games under consideration. For an $A_1$ agent, the game form \{x_1, x_2, x_3, x_4\} has \{x_3, x_4\} as the P.N.E. set. Yet the smaller game form \{x_1, x_3, x_4\} has \{x_1\} as the P.N.E. set. $\alpha$-rationality is violated because \{x_1, x_3, x_4\} \{x_1, x_2, x_3, x_4\} but $\sim$\{x_1\} $\subset$ \{x_3, x_4\}. In other words, there is choice reversal: $x_3$ is chosen over $x_1$ in \{x_1, x_2, x_3, x_4\} while $x_1$ is chosen over $x_3$ in \{x_1, x_3, x_4\}. A standard conclusion in welfare economics is that when $\alpha$-rationality is violated, the choice mechanism is not consistent with an underlying preference ordering. One is led to ask, therefore, whether it is meaningful to describe the firm above as choosing its production plan in accordance with the maximization of some objective function?

In order to answer the above question, let $p = (p_1, p_2, p_q)$ be a price system: $p_1$ and $p_2$ are input prices while $p_q$ is the price of output. If $y \in Y$, then $p \cdot y$ denotes the value of production plan $y$. Clearly the function $\pi_p(y) = p \cdot y$ ranks the elements of $y$ for each $p$. Indeed it seems entirely reasonable that the principal would choose $y \in Y$ to maximize $\pi_p(y)$. Consequently the violation of $\alpha$-rationality does not preclude the ranking of alternative production plans by their profitability.

The condition which leads $\alpha$-rationality to be violated is easily interpreted in the production set framework. Let $y' = (-1, 0, f(x_3))$ and $y'' = (-1, 0, f(x_1))$. Clearly $\pi_p(y'') > \pi_p(y')$ as long as $p_q$ is strictly positive. Then the two properties: (i) $\pi_p(y'') > \pi_p(y')$, and (ii) $y'$ is chosen from $Y$, just imply that $y'' \notin Y$. It is the statement $y'' \notin Y$ which is really critical in reconciling the violation of $\alpha$-rationality with choice based upon the $\pi_p(\cdot)$ ranking of $Y$. Remember that $Y$ only admits production plans which can be generated as P.N.E.'s. With this view, the production set embodies not only the technical knowledge of how inputs can be converted into outputs but the associated managerial skills and preferences as well. Thus two requirements must be met if a production plan $y$ is to be admitted to $Y$. First, $y$ must be technically feasible: if $y$ is associated with outcome $x$, then the output component $y_q$ of $y$ must satisfy $y_q \leq f(x)$. Second, $y$ must be managerially feasible: the outcome $x$ associated with $y$ must emerge as a P.N.E. from an available game form. Notice that if only game form \{x_1, x_2, x_3, x_4\} is available, then $y'' = (-1, 0, f(x_1))$ satisfies technical feasibility but not managerial feasibility. When viewed in this way, there is no choice reversal. 1

This same argument carries over to the consumer choice problem with changing tastes. The principal and agent represent a consumer at different stages of his life, and changing tastes can be depicted by the preference conflict between $P$ and $A$. 2

---

1 In the exercise used above to establish that $\alpha$-rationality is violated there are two implied production sets $Y^1$ and $Y^2$. $y'$ and $y''$ belong to $Y^1$, but only $y' \in Y^2$ (because $\{x_1, x_3, x_4\}$ is not available). When choosing from $Y^1$, $y''$ would be selected over $y'$. When choosing from $Y^2$, $y''$ cannot be selected over $y'$ because $y'' \notin Y^2$.

2 Alternatively, the individual can be modeled as having a conflict between aspects of his personality at a point in time. This is the approach taken in Thaler and Shefrin (1980) and Shefrin and Thaler (1980).
A consumption plan is defined by a function on outcomes (like $f$) and the consumption set $C$ is obtained from the P.N.E.'s. $\alpha$-rationality is never violated provided the consumption plans being compared are chosen from $C$.

Hammond (1976) concludes that an individual with changing tastes does violate $\alpha$-rationality. He reaches this conclusion because he allows $C$ to vary. We think it makes more sense to restrict the analysis to the choices that are both technologically and managerially feasible. [For an alternative view, see Kreps (1978).] In addition, we believe that any welfare judgments should be made in the same domain, that is wherein $C$ is held fixed. Moreover, implications about welfare are in large part independent of whether or not the choice mechanism satisfies $\alpha$-rationality. Notice that $x_3$ and $x_4$ are Pareto-efficient outcomes when preferences are given by (3) and (4), but Pareto-inefficient if $A$ prefers $x_2$ over $x_3$ and $x_4$.

References

Kreps, D., 1978, On sophisticated choice of opportunity sets, Graduate School of Business mimeo. (Stanford University, Stanford, CA).