X-Games*

Kfir Eliaz[†] and Ran Spiegler[‡]

December 23, 2013

Abstract

What is common to the following situations: incentivizing collective action in the presence of social preferences, monopoly pricing when consumers are loss averse, arms races when players are privately informed of their armament costs? We present a simple formalism, called X-games, which unifies these situations as well as others, and use it to unify and extend the separate analyses that they received in the literature.

1 Introduction

Consider the following problems.

1. A central planner designs an incentive scheme in order to encourage organ donations in a large population, when agents' preferences exhibit conformism: their preference for an action increases with the fraction of agents in the population who take it. What is the cheapest way to implement donation by everyone as the unique Nash equilibrium, assuming that the planner can use personalized transfers conditional on donation?

*This paper supersedes an earlier version, titled "Incentivizing Participation under Reference-Dependent Preferences: A Note", which in turn was directly inspired by Kőszegi and Heidhues (2012). Spiegler acknowledges financial support from the European Research Council, Grant no. 230251. We thank Botond Kőszegi, Benny Moldovanu, David Parkes, Neil Thakral and an associate editor for helpful comments. We owe special thanks to Yair Antler, Alex Frug and Eeva Mauring for excellent research assistance.

[†]Tel-Aviv University and University of Michigan. Email: kfire@post.tau.ac.il.

[‡]Tel Aviv University and University College London. URL: http://www.tau.ac.il/~rani. E-mail: rani@post.tau.ac.il.

2. Two countries are engaged in an arms race, where armament costs are *i.i.d.* and privately known. For which cost distributions is mutual armament the unique Bayesian-Nash equilibrium?

3. A monopolist faces consumers who exhibit an "attachment effect": when they encounter an unexpectedly high price, they are more willing to pay it if they initially thought that the purchase is likely to take place. What is an optimal sales (i.e., random price) policy for the monopolist?

4. A set of risk-averse agents with privately observed wealth shocks decide whether to invest in a risky project. The probability of success increases with the proportion of agents who invest. For which distributions of wealth will mutual investment be the unique Bayesian-Nash equilibrium?

5. Firms in a large industry contemplate whether to switch to a new, superior technology, when there are positive network externalities. Under which distributions of individual benefits from the superior technology will firms remain "locked" in the old technology as a unique Nash equilibrium outcome?

These five problems are not only different in terms of economic content, but also seem to involve different classes of models: complete-information population games in problems 1 and 5, individual consumer choice with reference-dependent preferences in 3, and two-player Bayesian game in 2 (between a consumer and himself) and 4. Furthermore, items 1 and 3 are mechanism-design problems in which the planner's instrument is a distribution of monetary transfers, while problems 2, 4 and 5 are descriptive game-theoretic models in which the questions are formulated in terms of the distribution of player types.

Despite these differences, the stories do seem to share two common features. First, agents' preferences exhibit *positive externalities* (although in situation 3, the consumer's externality is "internal", in the sense that it is defined w.r.t. the expectations he had about his own behavior). Second, all problems are concerned with uniqueness of equilibrium outcomes (however defined). It is therefore natural to suspect that game-theoretic analysis of these problems will have shared features, despite their different interpretations and formal set-ups.

In this short paper we offer a simple formalism, called *X-games*, which unifies the above problems as well as others. Some of these problems have been analyzed in the literature, while others are novel. Our contribution is fourfold: first, we believe that we are the first to point out the connection among all these problems; second, we provide a characterization result that extends previous treatments and we apply it to a number of new problems; third, the X-game formalism covers the case of negative externalities as well; and finally, the formalism may be of interest in its own right and suggest future applications.

This paper emphatically builds on prior work. Rather than giving an overview of the relevant literature now, we believe it is more effective to discuss the precedents in detail as we go along. After stating our main results, we present a number of applications; some of them are new, others literally replicate existing results; and yet others are very close to existing works in terms of interpretation and underlying logic, but provide new results or variations on existing results, simply because the questions addressed by our main results were not posed in the original context.

2 The Model and the Main Result

An *X-game* is a pair $<\pi, \mu>$, where $\pi : [0,1] \times \mathbb{R} \to \mathbb{R}$ and μ is an atomless *cdf* over $\mathbb{R}$ with support T_μ.[1] We use x and t to denote the first and second arguments of π, respectively. We say that the X-game is *linear* if $\pi(x,t) \equiv a + bx - t$ for some $a, b \in \mathbb{R}$, $b \neq 0$.

Let S be the set of integrable functions $s : \mathbb{R} \to \{0,1\}$. For every $s \in S$, define

$$x(s,\mu) = \int_t s(t) d\mu(t)$$

We say that $s^* \in S$ is an *equilibrium* in the X-game $<\pi,\mu>$ if for every $t \in T_\mu$, $s^*(t) = 1$ (0) whenever $\pi(x(s^*,\mu),t) > 0$ (< 0).

Because the value of this formalism lies in its abstraction and variety of potential applications, we refrain from giving the elements of the model a specific economic interpretation. Instead, we illustrate it with a concrete example of a linear X-game (which echoes the first problem mentioned in the Introduction). A measure one of agents face a decision whether to participate in a government-provided program. Interpret t as the (possibly negative) price for participation, such that $s(t) = 1$ means that the agent participates when he faces the price t. Let x represent the fraction of agents in

[1] Allowing μ to have atoms would make the exposition more cumbersome without changing any of the results.

the population who participate, and let $\pi(x,t) = a + bx - t$ be the agent's net utility gain from participation. The *cdf* μ represents a distribution of individualized prices across the population. Equilibrium in the X-game corresponds to pure-strategy Nash equilibrium in the population game.

The main problem we address in this paper is the following: *For which X-games* $< \pi, \mu >$ *is it the case that* $x(s^*, \mu) = 1$ *in every equilibrium* s^**?* Note that in a linear X-game an equilibrium $s^*(t)$ is necessarily a *threshold function*, i.e., there is some t^* such that $s^*(t) = 1$ for $t < t^*$ and $s^*(t) = 0$ for $t > t^*$. In Section 4 we introduce the broader class of quasi-linear X-game which also have the feature that an equilibrium is a threshold function. Therefore, in the class of X-games we study, if $x(s^*, \mu) = 1$ in every equilibrium, then all equilibria are identical except for a zero μ-measure set of values of t. Hence, instead of stating that $x(s^*, \mu)$ is equal to some constant x^* (say, 1) in *every* equilibrium, we will say from now on that there is an *essentially unique* equilibrium s^* in which $x(s^*, \mu) = x^*$.

The following result provides a complete characterization for the case of linear X-games. The result makes use of the notion of first-order stochastic dominance (FOSD). For any pair of *cdf*s F and G defined over $\mathbb{R}$, we say that F FOSD G if $F(t) \leq G(t)$ for every t. If $F(t) < G(t)$ for all t with $F(t) < 1$ and $G(t) > 0$, then F is said to *strictly* FOSD G. We use $\mathcal{U}[c, d]$ to denote the uniform distribution over an interval $[c, d]$.

Proposition 1 *Let* $< \pi, \mu >$ *be a linear X-game. If* $\mathcal{U}[\min\{a, a+b\}, a+b]$ *strictly FOSD* μ*, then there is an essentially unique equilibrium* s^* *and this equilibrium satisfies* $x(s^*, \mu) = 1$*. Conversely, if there exists an essentially unique equilibrium* s^* *and in this equilibrium* $x(s^*, \mu) = 1$*, then* $\mathcal{U}[\min\{a, a+b\}, a+b]$ *FOSD* μ*.*

The proof of Proposition 1 will appear as a corollary of a more general result we provide in Section 4. We discuss it now in terms of the above government-program example. Suppose that the government wishes to implement full participation in its program with the maximal possible revenue. When $b < 0$, it cannot do better than a uniform price $t = a + b - \varepsilon$, where $\varepsilon > 0$ is arbitrarily small. In contrast, when $b > 0$, the uniform distribution $\mathcal{U}[a, a+b]$ attains the infimal cost over all price distributions that implement $x(s^*, \mu) = 1$ in every equilibrium s^*, hence price discrimination is necessary for optimality.

To see the intuition for the case of $b > 0$, suppose first that the government cannot price-discriminate, i.e., it is restricted to degenerate μ. In this case, it would have to set a price $t < a$ in order to ensure that all agents participate in any equilibrium because otherwise, there would be an equilibrium s with $x(s,\mu) = 0$. Now suppose that the government can price-discriminate. Then, it can assign a low price $t < a$ to a small group of agents, thus turning participating into a dominant strategy for them. Having secured a positive mass of participating agents, the government can turn to another small group and offer them to participate at a slightly higher price. Thanks to the positive externality, this second group will accept the offer because the first group is already known to participate. The government can proceed in this manner and leverage the positive externality, targeting new groups of agents at ever higher prices.

This argument is reminiscent of "infection" arguments introduced by Rubinstein (1989) and developed in the literature on global games (see Morris and Shin (2003)). Morris and Shin (2005) extended this type of argument to a somewhat broader class of "interaction games". In several of the papers we cite in the sequel (Winter (2004), Baliga and Sjöström (2004), Heidhues and Kőszegi (2012) and Sákovic and Steiner (2012)), essentially the same quasi-infection logic plays a central role in the proof of results, despite the fact that these papers employ distinct classes of games and solution concepts. Finally, Blonski (1999, 2000) characterizes Nash equilibria in anonymous binary-action games with a continuum of players.

3 Applications

In this section we apply Proposition 1 to a variety of settings, thereby showing its links to the existing literature.

Application 1: Overcoming a habit
Consider the following principal-agent situation. In every time period, the agent makes a *static* choice between two actions, 1 and 0. The agent has reference-dependent preferences: he is basically indifferent between the two actions, but he suffers a “mental switching cost” of 1 if he takes an action he is not used to. Specifically, *the agent’s utility from each action is equal to the long-run frequency that he takes it.* In the absence of monetary incentives, both actions are stable: given that the agent always plays action i, it is (strictly) optimal for him to stick to this habit. Kőszegi (2010) and Kőszegi and Rabin (2006) introduced this notion of stability and called it “*personal equilibrium*” (PE henceforth).

The principal would like the agent to choose action 1 at all times. Suppose that he can give the agent a (possibly negative) transfer conditional on taking this action.

The principal wants the agent to choose 1 in *every* PE, thus overcoming any habit that the agent might have. If the principal is restricted to deterministic transfers, he must commit to a transfer above 1 (the agent's mental switching cost) in order to knock out the PE in which the agent always plays 0.

Can the principal do better with a transfer that *fluctuates* over time? Imagine that the principal adheres to a long-run distribution μ over transfers, such that the transfer the agent faces at each period is independently drawn from μ.[2] Kőszegi (2006) extended the definition of PE to such stochastic environments. A strategy for the agent - namely, a function that assigns an action to every realized transfer - is a PE if for every realized transfer, the agent's action maximizes his expected reference-dependent payoff, where the expectation is taken w.r.t the long-run frequency over his actions, induced by μ and his own strategy.

This model can be mapped into the linear X-game formalism, where x represents the long-run frequency that the agent chooses the action 1, $-t$ is the transfer he receives conditional on playing this action, μ is the long-run distribution over t, and $\pi(x,t) = -1 + 2x - t$. An equilibrium in our model corresponds to a PE. Proposition 1 then implies that randomization indeed benefits the principal. In particular, a uniformly distributed transfer with support $[-1+\varepsilon, 1+\varepsilon]$ implements action 1 as a unique PE outcome, where $\varepsilon > 0$ can be arbitrarily small. Thus, the principal can attain his objective at virtually no cost, as if the mental switching cost did not exist.

Application 2: Arms races with private armament costs

This example is literally taken from Baliga and Sjöström (2004), except for a necessary change in notation. Two countries play the following 2×2 symmetric Bayesian armament game. We present player i's payoff only:

$a_i \backslash a_j$	B	N
B	$-t_i$	$g - t_i$
N	$-l$	0

where the actions B and N represent building new weapons and refraining from building new weapons, respectively; $l > 0$ is the loss of a country that chooses N while its rival chooses B; $g > 0$ is the gross gain of a country that chooses B while its rival chooses N; and t_i is the cost of armament, which is independently drawn from the *cdf* μ, and constitutes country i's private information.

[2] One interpretation is that the principal commits ex-ante to μ. A less literal, more interesting interpretation is that μ is a reduced form of a long-run price distribution, and that a patient principal will have an incentive to develop a repulation for playing it.

This model can be mapped into our X-game formalism, such that our notion of equilibrium corresponds to symmetric Nash equilibrium in the Bayesian game. Let x stand for the ex-ante probability that the opponent plays B. The function π is the net gain from playing B and is given by $\pi(x,t) = g + (l-g)x - t$. Proposition 1 thus implies that the unique symmetric Nash equilibrium strategy is "always B" if armament costs are low in the sense that $\mathcal{U}[\min\{g,l\},l]$ strictly FOSD μ. This result is similar to Theorem 1 in Baliga and Sjöström (2004).

Application 3: Technology adoption in industries with network externalities
Many technological innovations exhibit network externalities, in the sense that a user's benefit from switching to a new technology depends (because of compatibility issues) on the proportion of other users who also switch. Because these externalities have important implications for market competition between firms that offer such innovations, they have received much attention in the I.O. literature. The following is a reduced form of Farrell and Saloner (1985). There is a unit mass of firms in the industry. Each firm chooses whether to switch to a new technology (action 0). The firm derives an intrinsic benefit t from switching, which is independent of the other firms' behavior. This benefit is the firm's private information and is *i.i.d.* across firms according to a *cdf* μ over $\mathbb{R}$. The gain from retaining the old technology for a firm of type t when a fraction x of firms also retain it is $\pi(x,t) = bx - t$ with $b > 0$.

Network externalities naturally give rise to a coordination problem that can result in multiple equilibria. The above situation can be written as an X-game, where our notion of equilibrium corresponds to Nash equilibrium in the population game. Using Proposition 1, we can derive tight conditions for technology lock-in to emerge as the unique equilibrium outcome. Specifically, if $\mathcal{U}[0,b]$ strictly FOSD μ, then all firms remain locked in the old technology in Nash equilibrium. If $\mathcal{U}[0,b]$ does not FOSD μ, there exist Nash equilibria in which some firms switch to the new technology.

4 Quasi-Linear X-Games

In this section we extend our analysis to a larger class of X-games. We say that an X-game $\langle \pi, \mu \rangle$ is *quasi-linear* if $\pi(x,t)$ satisfies the following properties:

(P1) $\pi(x,t)$ is continuous in t and linear in x.

(P2) For every x there is a unique $t(x)$ that solves $\pi(x, t(x)) = 0$.

(P3) $\pi(x,t) \cdot [t - t(x)] < 0$ for every $t \neq t(x)$.

Clearly, every linear X-game is also a quasi-linear X-game with $t(x) = a + bx$. Note that the above properties imply that for any equilibrium s, there exists a unique cutoff t^* such that $s(t) = 1$ (0) if $t < t^*$ ($t > t^*$). In addition, these properties also have the following implication. Note that since π is linear in x, it is either strictly increasing, strictly decreasing or constant in x.

Lemma 1 *If π is strictly increasing / strictly decreasing / constant in x, then $t(x)$ is strictly increasing / strictly decreasing / constant (respectively).*

Proof. Assume first that $\pi(x,t)$ is strictly increasing in x. Then $x' > x$ implies $\pi(x',t) > \pi(x,t)$ for any t. In particular, $\pi(x', t(x)) > 0$. Hence by (P3), $t(x) < t(x')$. Since by (P2), $t(\cdot)$ is a function, we conclude that it must be strictly increasing. Similarly, if $\pi(x,t)$ is strictly decreasing in x then $t(\cdot)$ is a strictly decreasing function. Finally, if π is constant in x, then by definition $t(x)$ is constant. ■

Define the following auxiliary function

$$\tilde{F}(t) = \frac{\pi(0,t)}{\pi(0,t) - \pi(1,t)}$$

By (P1), $\tilde{F}$ is continuous. By (P2)-(P3), $\tilde{F}(t) \in [0,1]$ for every $t \in [\min\{t(0), t(1)\}, t(1)]$, with $\tilde{F}(t(0)) = 0$ and $\tilde{F}(t(1)) = 1$. However, $\tilde{F}$ is not necessarily an increasing function. Define the following continuous *cdf* over $[\min\{t(0), t(1)\}, t(1)]$:

$$F(t) = \max_{\tau \in [\min\{t(0),t(1)\},t]} \tilde{F}(\tau)$$

Thus, F is the lowest non-decreasing function that lies weakly above $\tilde{F}$.

Proposition 2 *Let $\langle \pi, \mu \rangle$ be a quasi-linear X-game. If F strictly FOSD μ, then there exists an essentially unique equilibrium s^* and this equilibrium satisfies $x(s^*, \mu) = 1$. Conversely, if there exists an essentially unique equilibrium s^* and this equilibrium satisfies $x(s^*, \mu) = 1$, then F FOSD μ.*

Proof. We consider two cases.

Case 1. π is strictly increasing in x (i.e., $t(0) < t(1)$).

(Sufficiency). Assume F strictly FOSD μ. Let $s^*(t) = 1$ for all $t \in T_\mu$. Then $\pi(x(s^*, \mu), t) = \pi(1, t)$, and by (P3), $\pi(1,t) \geq 0$ for all $t \leq t(1)$. By strict FOSD, any

$t > t(1)$ is not in T_μ. It follows that $\pi(1,t) \geq 0$ for all $t \in T_\mu$, implying that s^* is an equilibrium.

Suppose there exists an equilibrium s' with $x(s',\mu) < 1$. Therefore, there must exist some $t' < t(1)$ such that $t' \in T_\mu$ and $s'(t) = 0$ for all $t \geq t'$. By (P3) and the definition of equilibrium, this implies that $\pi(\mu(t'),t) < 0$ for all $t > t'$. By (P1), $\pi(x,t)$ is linear in x and so,

$$\pi(\mu(t'),t') = \pi(0,t') + \mu(t')(\pi(1,t') - \pi(0,t')) \tag{1}$$

By the definition of $\tilde{F}$, $\pi(\tilde{F}(t'),t') = 0$, i.e.,

$$\pi(0,t') + \frac{\pi(0,t')}{\pi(0,t') - \pi(1,t')} \cdot (\pi(1,t') - \pi(0,t')) = 0$$

Since F strictly FOSD μ, we have $\mu(t') > F(t')$. By definition, $F(t') \geq \tilde{F}(t')$. Because $\pi(1,t') > \pi(0,t')$, it follows that $\pi(\mu(t'),t') > 0$. By continuity, $\pi(\mu(t'),t'+\varepsilon) > 0$ for any small enough ε, a contradiction.

(Necessity). Assume that there exists an essentially unique equilibrium s^* and that this equilibrium satisfies $x(s^*,\mu) = 1$. This means that μ must satisfy the following two properties: $\pi(1,t) \geq 0$ for all $t \in T_\mu$, and $\mu(t(1)) = 1$. In order *not* to have an equilibrium s in which $s(t) = 0$ for all $t \in T_\mu$, it must be that $\pi(0,t) > 0$ for some $t \in T_\mu$. This means that T_μ must include values strictly below $t(0)$. Hence, $\mu(t(0)) > 0$ and μ is strictly above F at $t(0)$ and weakly above it at $t(1)$. Since μ is a *cdf*, it is upper-semi-continuous. Hence, either μ is strictly above F for all $t \in (t(0),t(1))$ (in which case our proof is complete) or there exists $t^* \in (t(0),t(1))$ such that $F(t^*) = \mu(t^*)$. Let us show that this implies the existence of some $t^{**} \in (t(0),t^*]$ such that $\tilde{F}(t^{**}) = \mu(t^{**})$. If $F(t^*) = \tilde{F}(t^*)$, this is immediate. Now suppose $F(t^*) > \tilde{F}(t^*)$. Since μ is an increasing function, by the definition of F, there must be some $t' \in (t(0),t^*)$ such that $\tilde{F}(t') = F(t^*) \geq \mu(t')$. By the intermediate value theorem, there must be some $t^{**} \in [t',t^*]$ such that $\tilde{F}(t^{**}) = \mu(t^{**})$. Plug the definition of $\tilde{F}(t^{**})$ into (1), and obtain $\pi(\tilde{F}(t^{**}),t^{**}) = \pi(\mu(t^{**}),t^{**}) = 0$. But this means that in contradiction to our initial assumption, there exists an equilibrium s' in which $s'(t) = 1$ if and only if $t \leq t^{**}$.

Case 2. π is not strictly increasing in x (i.e., $t(0) \geq t(1)$).

(Sufficiency). In this case F is a degenerate distribution that puts a mass of one on $t(1)$. If F strictly FOSD μ, then $t \leq t(1) < t(0)$ for all $t \in T_\mu$, with a strict inequality for some $t \in T_\mu$. Hence, $\pi(1,t) \geq 0$ for all $t \in T_\mu$, with a strict inequality for some

$t \in T_\mu$. Therefore, a function s^* that satisfies $s^*(t) = 1$ for all $t \in T_\mu$ is clearly an equilibrium.

Suppose there exists an equilibrium s' with $x(s', \mu) < 1$. Consider equation (1). Since $t' < t(1) < t(0)$, both $\pi(0, t')$ and $\pi(1, t')$ are positive, and therefore $\pi(\mu(t'), t') > 0$. The continuity of π in t then implies that $\pi(\mu(t'), t' + \varepsilon) > 0$ for any small enough ε, a contradiction.

(*Necessity).* Since the support of F is $\{t(1)\}$, we have that F FOSD μ. ■

When t is interpreted as a price for participation, Proposition 2 has a simple implication for the random price scheme that maximizes revenues subject to full participation.

Corollary 1 *F is a solution to the problem*

$$\sup_{\mu} \int_t ts^*(t)d\mu(t)$$

subject to the constraint that s^* is an essentially unique equilibrium of (π, μ) and $x(s^*, \mu) = 1$.

The following applications illustrate the value of extending our framework from linear to quasi-linear X-games.

Application 4: Moral hazard in teams
Consider a project which is carried out by a team of two agents, who need to choose between exerting effort and shirking. Exerting effort entails a cost of 1, while shirking is costless. The probability that the project succeeds depends on the agents' effort decisions: if both exert effort the project succeeds with probability one; if only one exerts effort the probability of success is $\beta < \frac{1}{2}$; and if both shirk the project fails for sure. An agent's payoff matrix is as follows:

$a_i \backslash a_j$	E	S
E	$-t-1$	$-\beta t - 1$
S	$-\beta t$	0

Now suppose that a principal employing the agents pays a transfer of $-t_i$ to agent i conditional on a successful project, where t_i is *i.i.d* according to some *cdf* μ that the principal commits to ex-ante. A stochastic transfer captures occasional bonuses, and the identity of the transfer distribution for the two agents may represent an ex-ante fairness norm.

This situation can be described as a quasi-linear X-game, where x is the probability that the opponent exerts effort, and an agent's gain from exerting effort is

$$\pi(x,t) = xt(2\beta - 1) - (\beta t + 1)$$

It is straightforward to verify that $\pi(x,t)$ satisfies properties (P1)-(P3). In addition, the function $\tilde{F}$ is defined over the interval $[-1/\beta, -1/(1-\beta)]$, and given by

$$\tilde{F}(t) = \frac{\beta t + 1}{2\beta t - t}$$

Note that $F = \tilde{F}$. By Proposition 2, if F strictly FOSD μ, then μ induces a unique Bayesian Nash equilibrium in which both agents always exert effort. If F does not FOSD μ, there exist Nash equilibria under μ in which agents shirk with positive probability.

The expected transfer in absolute terms according to F is

$$\frac{1}{1-2\beta}\ln(\frac{1-\beta}{\beta})$$

By comparison, if the principal were restricted to deterministic transfer, he would have to commit to a transfer above $1/\beta$ (in absolute terms) in order to induce effort in every equilibrium. For illustration, when $\beta = \frac{1}{4}$, randomization reduces the expected transfer by roughly 45%.

This example is similar to the moral hazard problem studied in Winter (2004), where a group of agents independently decide whether to make a costly investment in a project, where the probability of the project's success increases with the number of agents who invest. Instead of a personalized random transfer (conditional on success) which is ex-ante identical, Winter (2004) considered a profile of personalized deterministic transfers (representing ex-post discrimination among agents). He showed that the profile of transfers that induces a unique Nash equilibrium in which all agents invest is discriminatory if the success probability is supermodular in the number of agents who invest. (For related models, see Spiegler (2000) and Sákovic and Steiner (2012).)

Application 5: Investment by risk-averse agents

There is a unit mass of agents. Each agent i has initial wealth w_i, which is only known to him. The distribution of initial wealth is common knowledge. The agent's utility from a prospect that gives a payoff of R is $\ln(w_i + R)$. Each agent has to decide whether to invest in a risky project. If he invests, then with probability $\alpha(1+x)$ the project results in a bonus $B > 0$ (i.e., $R = B$), where $\alpha < \frac{1}{2}$ and x is the total fraction of agents who invest. With probability $1-\alpha(1+x)$, the project results in a loss of $L < \alpha B/(1-\alpha)$ (i.e., $R = -L$). Assume the distribution of wealth satisfies $w_i \geq L$ for each i and that B/L is large enough such that an agent with initial wealth B will strictly prefer to invest even if no one else does.[3] If an agent does not invest, his total wealth remains equal to his initial wealth. We are interested in the following question: for which distributions of initial wealth does this game have a unique symmetric Bayesian Nash equilibrium in which all agents invest?

To address this question, we map this investment game into a quasi-linear X-game. Let $t = -w$ (minus the agent's initial wealth), and use the above definition of x. Then, $\pi(x, t)$ is the gain from investing for an agent with initial wealth $-t$ when a total fraction x of the agents invest:

$$\pi(x, t) = \alpha(1 + x) \ln(\frac{t - B}{t + L}) + \ln(\frac{t + L}{t})$$

Clearly, $\pi(x, t)$ is continuous in t. Since $t = -w$ and $w \geq L$ it follows that the function is linear and increasing in x in the relevant domain. To verify (P2) and (P3) note that $\pi(x, t)$ asymptotes to zero as $t \to -\infty$, is increasing up to $t = -LB/(\alpha B-(1-\alpha)L)$ and then decreases thereafter approaching $-\infty$ as $t \to -L$. This implies that if $\pi(x, t) > 0$ for some t, then for any $x > 0$ the function $\pi(x, t)$ will cross the horizontal axis only once and from above. Properties (P2) and (P3) then follow from our assumption that $\pi(0, -B) > 0$ and from the fact that $\pi(x, t)$ increases in x.

By Proposition 2, if the wealth distribution strictly FOSD (recall that $w = -t$)

$$F(t) = \frac{1}{\alpha} \cdot \frac{\ln(t + L) - \ln(t)}{\ln(t + L) - \ln(t - B)} - 1,$$

then all agents invest in equilibrium, whereas if the wealth distribution does not FOSD F, there are equilibria in which some agents do not invest. It can be verified that the function F is increasing over the interval $[t(0), t(1)]$.

5 Revenue Maximization

In this section we interpret t as a monetary transfer from an agent to a "planner" conditional on the agent "participating" ($s(t) = 1$). The planner's payoff depends on the participation rate and on the expected transfer he receives. The planner's

[3] I.e., B/L is large enough such that $\alpha \ln(2) + (1 - \alpha) \ln(1 - \frac{L}{B}) > 0$.

objective is to design a transfer scheme that maximizes his expected payoff subject to the constraint that the transfer induces a unique equilibrium participation rate.

To solve this design problem the planner first needs to derive for each possible participation rate x the optimal distribution of transfers μ that induces an essentially unique equilibrium in which a fraction x of the agents choose 1. The planner can then choose the pair (x, μ) with the highest expected payoff. The previous sections focused on the case in which the planner wished to implement $x = 1$, but there are situations in which $x < 1$ may be optimal. Recall the government-program example of Section 2. If the government faces capacity constraints, it may wish to implement only partial participation.

We focus on quasi-linear X-games. We say that μ implements $x^* \in [0, 1]$ in (π, μ) if there exists an essentially unique equilibrium s^*, and this equilibrium has the property that $x(s^*, \mu) = x^*$. Let $M(x^*)$ denote the set of *cdf*s μ that implement x^*. For a given x^*, the planner's problem is

$$\sup_{\mu \in M(x^*)} \int_t ts^*(t) d\mu(t)$$

subject to the constraint that s^* is an equilibrium in (π, μ). (By the definition of $M(x^*)$, s^* is thus the essentially unique equilibrium.) Thus, a *cdf* that solves the planner's problem implements the participation rate x^* at the highest possible revenue.

Fix $x^* \in (0, 1)$, and define

$$\tilde{F}_{x^*}(t) = \frac{\pi(0, t)}{\pi(0, t) - \pi(x^*, t)}$$

For every $t \in [\min\{t(0), t(x^*)\}, t(x^*)]$, define

$$F_{x^*}(t) = \max_{\tau \in [\min\{t(0), t(x^*)\}, t]} \tilde{F}_{x^*}(\tau)$$

As in the previous section, it can be verified that F_{x^*} is a well-defined *cdf*.

Proposition 3 *Fix $x^* \in (0, 1)$. The following distribution solves the planner's problem. With probability x^*, it chooses a value t from $[\min\{t(0), t(x^*)\}, t(x^*)]$ according to the cdf $F_{x^*}(t)$, and with probability $1 - x^*$, it chooses a value t above $\max\{t(0), t(1)\}$.*

Proof. Consider a solution $\mu \in M(x^*)$, and fix an equilibrium s^* such that $x(s^*, \mu) = x^*$. Thus, there are values of t for which $s^*(t) = 0$. Specifically,

$$\int_{t | s^*(t) = 0} d\mu(t) = 1 - x^*$$

Construct a *cdf* μ' that shifts all the weight on $\{t|s^*(t)=0\}$ into values of t above $\max\{t(0),t(1)\}$. Define s' as follows:

$$s'(t)=\begin{cases} 1 \quad if & s^*(t)=1 \\ 0 \quad if & t>\max\{t(0),t(1)\} \end{cases}$$

We need not pin down s' for other values of t because they lie outside the support of μ'. Note that $x(s',\mu')=x^*$.

Let us show that s' is an equilibrium. First, by the monotonicity of $t(x)$ and the definition of $t(0)$ and $t(1)$, we have $s'(t)=0$ for every $t>\max\{t(0),t(1)\}$. Second, since s^* is an equilibrium in $\langle\pi,\mu\rangle$ and since $x(s',\mu')=x(s^*,\mu)=x^*$ and the supports of μ and μ' coincide over the set $\{t\leq\max\{t(0),t(1)\}|s^*(t)=1\}$, it must be the case that for every $t\leq\max\{t(0),t(1)\}$ in $T_{\mu'}$, $\pi(x^*,t)\geq 0$. It follows that w.l.o.g we can restrict attention to μ's that assign probability $1-x^*$ to $t>\max\{t(0),t(1)\}$. Our task now is to derive the structure of such a solution μ conditional on $t\leq\max\{t(0),t(1)\}$.

Given any μ, define the *cdf* $\mu_{x^*}(t)=\min\{1,\mu(t)/x^*\}$ over $(-\infty,\max\{t(0),t(1)\}]$. Let $z=x/x^*$, and define $\tilde{\pi}(z,t)\equiv\pi(x,t)$. Define $\tilde{t}(z)$ by $\tilde{\pi}(z,\tilde{t}(z))=0$. Note that since z is proportional to x, if (π,μ) is a quasi-linear X-game, then so is $(\tilde{\pi},\mu_{x^*})$. By construction, $x(s,\mu_{x^*})=1$ for every equilibrium s in $(\tilde{\pi},\mu_{x^*})$ if and only if μ implements x^* in (π,μ). By Corollary 1, the *cdf*

$$\mu_{x^*}(t)=\max_{\tau\in[\min\{\tilde{t}(0),\tilde{t}(1)\},t]}\frac{\tilde{\pi}(0,\tau)}{\tilde{\pi}(0,\tau)-\tilde{\pi}(1,\tau)}$$

defined over $[\min\{\tilde{t}(0),\tilde{t}(1)\},\tilde{t}(1)]$ solves the optimization problem

$$\sup_{\tilde{\mu}}\int_t ts^*(t)d\tilde{\mu}(t)$$

subject to the constraint that s^* is an essentially unique equilibrium in $(\tilde{\pi},\tilde{\mu})$ and

$$\int_t s^*(t)d\tilde{\mu}(t)=1$$

Using the definitions of z, $\tilde{\pi}$ and $\tilde{t}$, we obtain $\mu_{x^*}(t)=F_{x^*}(t)$ over the interval $[\min\{t(0),t(x^*)\},t(x^*)]$. ■

Application 1 revisited

Recall the linear X-game described in Application 1, where $\pi(x,t)=-1+2x-t$. Unlike the original application, assume that the principal derives direct utility from the agent's behavior. Specifically, the principal's net payoff under (s, μ) is

$$v(x(s,\mu)) + \int_t ts(t)d\mu(t)$$

where $v(x)$ is the principal's gross payoff when the agent's long-run frequency of action 1 is x. We assume that v is concave, twice differentiable and attains a unique interior maximum. This captures a trade-off between the benefit from inducing a "good" habit and the cost of accommodating such a habit (for instance, the socially optimal level of recycling may be interior due to processing costs).

By Proposition 3, the least costly random transfer scheme that implements a given interior x^* has the following structure: with probability $1 - x^*$, $t > 1$ (the agent will choose action 0 when faced with such realizations of t, independently of x); and with probability x^*, t is uniformly distributed over the interval $[-1, 2x^* - 1]$. It follows that the optimal x^* will be chosen to maximize

$$v(x^*) + x^*(x^* - 1)$$

If v is sufficiently concave (say, $v'' < -2$), x^* is given by the FOC $v'(x^*) = 1 - 2x^*$.

Application 7: Selling to a consumer with reference-dependent preferences
The following is a variation on Heidhues and Kőszegi (2012). A monopolist interacts with a single consumer with reference-dependent preferences, who chooses according to the concept of PE (as in Application 1). The monopolist faces a constant marginal cost $c \geq 0$. It commits ex-ante to a random price strategy. The consumer's utility from not buying is 0, and his utility from buying is $1 + \lambda \cdot \Pr(\text{buying})$, where $\Pr(\text{buying})$ is the long-run frequency of buying induced by the monopolist's random price strategy and the consumer's own purchase strategy (namely, his decision whether to buy as a function of the price realization he faces).

This model can be written as a linear X-game, where t is the product price, μ is the random price strategy, $s(t) = 1$ means that the consumer buys at the price t, $x = \Pr(\text{buying})$, and $\pi(x,t) = 1 + \lambda x - t$. Thus, $t(x) = 1 + \lambda x$. An equilibrium in the X-game corresponds to a PE defined w.r.t the consumer's reference-dependent preferences. The monopolist's profit given (s, μ) is $\int_t (t - c)s(t)d\mu(t)$.

Let us first consider the case of $\lambda > 0$ (studied by Heidhues and Kőszegi (2012)). This captures an "attachment effect", i.e., an increase in the consumer's willingness to pay for the product as a result of being accustomed to buying it. By Proposition 3, the most profitable random price strategy that implements $\Pr(\text{buying}) = x^*$ has the following structure: with probability $1 - x^*$, $t > 1 + \lambda$ (the consumer will not buy when faced with such a price realization, independently of x); with probability x^*, t is uniformly distributed over the interval $[1, 1 + \lambda x^*]$. The induced profit for the monopolist is

$$x^* \left[\frac{1 + (1 + \lambda x^*)}{2} - c \right]$$

Therefore, if $c < 1+\lambda/2$, the monopolist will charge a price that is uniformly distributed over the interval $[1, 1 + \lambda]$, and the consumer will always buy. This strategy extracts (in expectation) the consumer's bare willingness to pay plus half the attachment-effect term. If $c > 1 + \lambda/2$, the monopolist will opt out.[4]

Now suppose that $\lambda < 0$. Here the consumer exhibits an "anti-attachment" or "boredom" effect: his willingness to pay for the product is a decreasing linear function of the long-run frequency of buying. Because $t(0) > t(1)$ in this case, an optimal random price strategy that implements $\Pr(\text{buying}) = x^*$ assigns probability $1 - x^*$ to $t > 1$ and probability x^* to $t(x^*) = 1+\lambda x^*$. The monopolist will choose x^* to maximize

$$x^* \left[1 + \lambda x^* - c\right]$$

Hence, as long as $c < 1$, $x^* = (c - 1)/2\lambda$ and $t(x^*) = (1 + c)/2$, whereas if $c \geq 1$ the monopolist will opt out. Thus, the product price will fluctuate between two levels: a high price for which not buying is a dominant action, and a lower price for which buying is the only optimal action given the upper bound on the long-run buying frequency implied by the high price.

References

[1] Baliga, S. and T. Sjöström (2004), "Arms Races and Negotiations," *Review of Economic Studies* 71, 351-369.

[2] Blonski, M. (1999), "Anonymous Games with Binary Actions," Games and Economic Behavior 28, 171-180.

[4] The original Heidhues-Kőszegi model does not fall into the X-game formalism, and differs from Application 7 in the following dimensions. First, while we insist on a unique PE buying probability, Heidhues and Kőszegi allow for multiple PEs and select the one that gives the consumer the highest ex-ante payoff. Second, they depart from quasi-linear utility and assume that reference dependence carries over to the money dimension, such that if the consumer pays a price above his expectation, he experiences the difference as an extra cost. These two considerations imply that the monopolist's optimal pricing strategy consists of an atom on a "regular price" as well as a smooth distribution over an interval of "sale prices" that are bounded below the regular price.

[3] Blonski, M. (2000), “Characterization of Pure Strategy Equilibria in Finite Anonymous Games,” Journal of Mathematical Economics 34, 225-233.

[4] Carlsson, H. and E. van Damme (1993), “Global Games and Equilibrium Selection,” *Econometrica* 61, 989-1018.

[5] Farrell, J. and G. Saloner (1985), “Standardization, Compatibility and Innovation,” *Rand Journal of Economics* 16, 70-83.

[6] Kőszegi, B. (2010), “Utility from Anticipation and Personal Equilibrium,” *Economic Theory* 44, 415-444.

[7] Kőszegi, B. and P. Heidhues (2010), “Regular Prices and Sales,” *Theoretical Economics*, forthcoming.

[8] Kőszegi, B. and M. Rabin (2006), “A Model of Reference-Dependent Preferences,” *Quarterly Journal of Economics* 121, 1133-1166.

[9] Morris, S. and H.S. Shin (2003). “Global Games: Theory and Applications,” in *Advances in Economics and Econometrics (Proceedings of the Eighth World Congress of the Econometric Society)*, edited by M. Dewatripont, L. Hansen and S. Turnovsky; Cambridge University Press.

[10] Morris, S. and H.S. Shin (2005). “Heterogeneity and Uniqueness in Interaction Games,” in *The Economy as an Evolving Complex System III*, edited by L. Blume and S. Durlauf; Oxford University Press, Santa Fe Institute Studies in the Sciences of Complexity.

[11] Rubinstein, A. (1989), “The Electronic Mail Game: A Game with Almost Common Knowledge,” *American Economic Review* 79, 385-391.

[12] Sákovics, J. and J. Steiner (2012). “Who matters in coordination problems?”, *American Economic Review* 102, 3439-61.

[13] Spiegler, R. (2000), “Extracting Interaction-Created Surplus,” *Games and Economic Behavior* 30, 142-162.

[14] Winter, E. (2004), “Incentives and Discrimination,” *American Economic Review* 94, 764-773.