Hiding or Revealing – Their Indirect Evolution in the Acquiring-a-Company game*

Werner Güth
Max Planck Institute
for Research on Collective Goods
Kurt-Schumacher-Str. 10
53113 Bonn, Germany
gueth@coll.mpg.de

Stefan Napel[1]
Dept. of Economics,
University of Bayreuth
Universitätsstr. 30
95447 Bayreuth, Germany
stefan.napel@uni-bayreuth.de

Version: January 4, 2022

Abstract

The Acquiring-a-Company game of Bazerman and Samuelson (1983) is modified by letting the privately informed seller send a numerical value message to the prospective buyer. A population of sellers reveal or hide their private information according to their categorical type rather than by engaging in consequentialistic decision making. Population shares of the types evolve according to expected profits (fitness). The analysis illustrates how specific institutional and behavioral aspects shape the creation of surplus in the market and possibilities for maintaining a positive share of revealing sellers.

Keywords: Indirect evolutionary approach; categorical types; asymmetric information; gains from trade

JEL codes: D01, C70, C73, C91

*We gratefully acknowledge very constructive comments made by an anonymous referee.

[1]Corresponding author.

1 Introduction

Asymmetric information of traders on markets can imply no-trade results (see Akerlof 1970; Bazerman and Samuelson 1983; Samuelson and Bazerman 1985). Policy-makers may try to limit private information (see, for our setting, the Takeover Bid Directive 2004/25/EC, the Transparency Directive 2004/109/EC, and the Unfair Commercial Practices Directive 2005/29/EC of the European Union) or encourage its disclosure, for instance by rewarding "whistle blowing" or by public records about previous behavior. Regarding the latter, market participants may themselves decide whether to hide or to reveal what they privately know in each and every case, i.e., by hiding or revealing acts given the specific situation. Instead we will focus on situations where they rely on a general rule and are committed to categorical hiding or revealing (see Harsanyi 1979, who distinguishes between act and rule utilitarianism).

We focus on rule-guidedness but not, like in rule utilitarianism, by engaging in a consequentialist decision between categorically hiding or revealing. Instead, like in evolutionary biology and game theory, categorical inclinations are viewed as inherited rules of behavior whose population shares evolve according to their relative reproductive success. As usual in evolutionary analysis of markets, average or expected profit is assumed to measure the fitness of the different categorical types.

Our indirect evolutionary approach (see Berninghaus, Güth, and Kliemt 2012, as well as Alger and Weibull 2013 for reviews) allows some market behavior, namely the price choices, to be rationally decided, based on how traders perceive and evaluate their market environment. As often in indirect evolution we distinguish between decision-driving *utility*, which traders maximize given their market perception, and evolution-driving *fitness*. Market traders need not be aware of the latter; they may be completely ignorant of evolutionary selection and its determinants.[2]

The Acquiring-a-Company game (Bazerman and Samuelson 1983; Samuelson and Bazerman 1985) features a privately informed seller and an uninformed potential buyer of a company. It was investigated by Samuelson and Bazerman by running – to the best of our knowledge – the first stochastic ultimatum experiments which, later on and rather independently of ultimatum experiments (see Güth and Kocher 2014), were repeated by other authors with more or less modifications. Di Cagno et al. (2016) modified the original Acquiring-a-Company game by letting the privately informed seller send a numerical value message to the prospective buyer, who then proposes a price for the seller's company. Rather than studying the respective (game-theoretically innocent) cheap-talk communication, we use Di Cagno et al.'s setup to investigate the evolution of categorical hiding or revealing. When sellers categorically reveal,[3] welfare enhancing trade is guaranteed. But revealing sellers may be exploited, i.e., make zero profit, due to the ultimatum power of the buyer. On the other hand when sellers hide, one obtains a no-trade result for a generic parameter region. Trade takes place only outside this region with both traders, the seller and the buyer, gaining from trade in expectation (how much depends on the specific parameter level).

Our analysis confirms the intuition that revealing can hardly ever survive without buyers rewarding sellers who are revealing. In line with the robust evidence of ultimatum experiments it will be assumed that sellers do not accept exploitation

[2]Rule utilitarianism would require awareness of survival prospects in case of long-lived traders.

[3]The Acquiring-a-Company game would then become a usual ultimatum game. Only in case of hiding, the game remains stochastic and yields striking experimental findings (see the discussion in Güth, Marazzi, and Panaccione 2020).

what, in turn, induces buyers to reward categorically revealing sellers. Will such rewarding increase the population share of revealing individuals and how would this depend on institutional and behavioral aspects? Regarding behavioral aspects we consider loss aversion of buyers who, when confronting a hiding seller, experience gains and losses with positive probabilities. Institutional aspects, on which we focus, are the average potential gains from trade, competition between potential buyers and the technology through which buyers may learn about the categorical type of their seller.

Section 2 describes the modified Acquiring-a-Company (AaC) game, which allows for hiding and revealing value messages and which, in case of hiding, does not question the pooling solution of the original AaC game. A first evolutionary analysis justifies that revealing types may not survive and confirms this pooling solution (Section 3). The possibility that decision utility and fitness in the sense of expected profits differ is highlighted by allowing buyers to be loss averse. Section 4 lets the buyer invest in costly detection of the seller's type. Section 5 assumes, in line with the abundant evidence of ultimatum experiments, that buyers reward revealing. We focus on the implications of imperfect type signals in Section 6. Section 7 concludes.

2 The modified AaC game

Seller S owns a company and is aware of its own evaluation, $q \cdot v$, and that this is linearly linked with the one, v, of the potential buyer B according to a commonly known parameter q satisfying $0 < q < 1$. Buyer B cannot observe v but expects it to be randomly drawn from the unit interval $[0, 1]$ with uniform density, what is known to S.

In the original Acquiring-a-Company (AaC) game, studied by Bazerman and Samuelson (1983) and Samuelson and Bazerman (1985), buyer B directly proposes a price p to seller S which S can accept, $\delta(p) = 1$, or reject, $\delta(p) = 0$. Seller S then earns $\delta(p) \cdot (p - qv)$ and buyer B earns $\delta(p) \cdot (v - p)$. Our analysis adopts the modification of the original AaC game by Di Cagno et al. (2016): the seller S, who is aware of v, sends a so-called "cheap-talk" value message $\hat{v} = \hat{v}(v)$ to buyer B before standard AaC play unfolds.

Applying backward induction yields $\delta^*(p) = 1$ only if $p \geq q \cdot v$, i.e., acceptance is restricted to $p/q \geq v$. Anticipating this and realizing that message $\hat{v}$ is cheap talk, the risk and loss neutral buyer B expects to earn

$$\frac{p}{q} \cdot \int_0^{\frac{p}{q}} (v - p) \cdot \frac{q}{p} dv + \left[1 - \frac{p}{q}\right] \cdot 0 = \int_0^{p/q} (v - p)\, dv = \frac{p^2}{q} \left(\frac{1}{2q} - 1\right) \tag{1}$$

from proposing $p \in [0, q]$. So B's payoff increases (decreases) in p when $q < \frac{1}{2}$ (resp. $q > \frac{1}{2}$). Hence the optimal price proposal is $p^* = q$ for $q \leq \frac{1}{2}$, the lowest price guaranteeing acceptance $\delta^*(p^*) = 1$ for all v. It is $p^* = 0$ for $q > \frac{1}{2}$, what precludes trade that would be welfare enhancing due to $(1 - q) \cdot v > 0$ for all $v > 0$.

Although the timing structure of the modified AaC game allows for signaling, it does not question the above pooling equilibrium[4] which, like the lemon market analyzed by Akerlof (1970), involves no trade and inefficiency in case $\frac{1}{2} < q < 1$ and $v > 0$ meaning that the positive surplus $(1 - q)v$ from trade is not realized. Value

[4]Like in deterministic ultimatum games there exist other equilibria in weakly dominated responder strategies (if for $q < \frac{1}{2}$ seller S rejects all price proposals p with $q < p < \underline{p} < \frac{1}{2}$ it would be optimal for B to propose $p = \underline{p}$ what, in turn, renders the acceptance behavior a best response).

messages $\hat{v}(v)$ of the seller qualify as cheap talk and do do not affect prices or payoffs. If, for instance, higher $\hat{v}$ would trigger higher prices p, this would be anticipated by the seller and render overreporting, $\hat{v} > v$, profitable for the seller.[5]

Trade for $q \leq \frac{1}{2}$ at price $p^* = q$ – though it yields positive profits in expectation – implies a loss for buyer B whenever the value of the company satisfies $v < q$, i.e., with positive probability.[6] It is, in our view, interesting how (anticipated) loss aversion of the buyer affects the (parameter) range of no-trade and, respectively, of always-trade predictions.

A *loss averse* (rather than risk and loss neutral as assumed above) buyer weighs losses higher – by a factor $k > 1$, say – than the respective gains for values $v > q$ when evaluating expected payoffs. Such a loss averse buyer B's expected utility from proposing price p is illustrated in Figure 1 for $q = 0.25$ and $q = 0.5$, together with three possible realizations of v. B's expected utility evaluates to

$$\int_p^{p/q} (v - p)\, dv + k \int_0^p (v - p)\, dv = \frac{p^2}{q}\left(\frac{1}{2q} - 1\right) - \frac{k-1}{2}p^2 \tag{2}$$

and increases in price p only when $q < q^*(k)$ with

$$q^*(k) = \frac{\sqrt{k} - 1}{k - 1}. \tag{3}$$

Hence the buyer proposes price $p^* = q$ if $q \leq q^*(k)$ and otherwise $p^* = 0$. L'Hôpital's

[5]The experimental evidence of Di Cagno et al. (2016) on the modified AaC game confirms dominance of over-reporting, i.e. $\hat{v} > v$, although a significant and, over multiple rounds, rather stable share of participants satisfy $\hat{v}(v) \leq v$.

[6]This has to be distinguished from the winner's curse that arises for inexperienced buyers or bidders in auctions. The latter entails losses not just for some realizations of v but even in expectation (see, e.g., Thaler 1988).

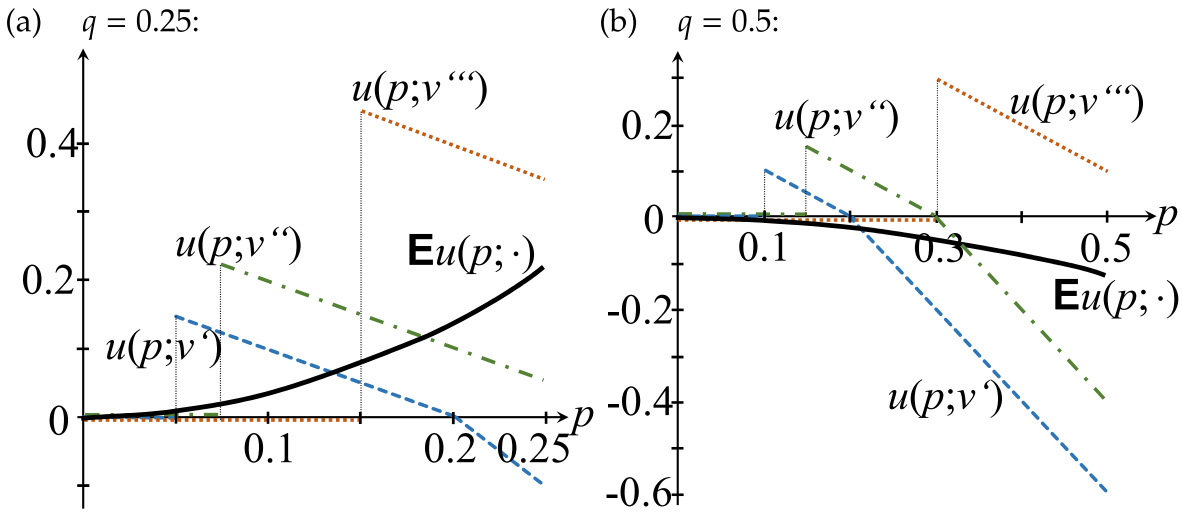

Figure 1: (Expected) Utility of a loss-averse buyer with $k = 2$ (vertical axis) when proposing positive prices $p \leq q$ (horizontal axis) for $[0, 1]$-uniformly distributed v as well as for realizations $v' = 0.2$, $v'' = 0.3$, $v''' = 0.6$ in case of $q = 0.25$ (left panel) and $q = 0.5$ (right panel)

rule, i.e., taking derivatives with respect to k separately for numerator and denominator, yields

$$\lim_{k \searrow 1} q^*(k) = \lim_{k \searrow 1} \frac{1}{2\sqrt{k}} = \frac{1}{2} \tag{4}$$

whereas $k > 1$ implies $q^*(k) < \frac{1}{2}$. So idiosyncratic loss aversion of the buyer further hampers trade because it induces price offers $p^* = 0$ also in the range $q^*(k) < q < 1/2$, independently of whether seller S is aware of the buyer's loss aversion or not.[7]

3 Evolution of recognizable seller types

We put the modified AaC-setup described above into a larger context by letting company owners and potential acquirers come from a large population of agents that fall into different moral categories. The latter are presumed to affect the handling

[7]Backward induction only requires that seller S is opportunistic (maximizing own profit) and that buyer B is aware of this and the own opportunism, based on loss aversion.

of private information as a seller. Namely, for a straightforward illustration of the indirect evolutionary approach assume that seller S can be of only two categorical types; namely

- a truthful type t who is revealing the privately known value of the company, i.e., $\hat{v}(v) = v$ for all v, or
- a lying type l who is hiding it by sending no message, i.e., $\hat{v}(v) = \emptyset$ for all v; or via excessive overreporting, $\hat{v}(v) = 1$ for all v. Only when later allowing an l-type to be falsely seen as a t-type it will matter whether the l-type sends no message or message $\hat{v}(v) = \bar{v} = 1$ for all v (and we will then assume the latter).

Distinguishing the two types, t and l, means to complement material motives, e.g. profits, in economic interaction by moral motives, e.g., in case of type t by intrinsic motivation in addition to material opportunism. Indirect evolutionary analysis lets traits for intrinsic motivations, which may create behavioral constraints or imply conflicting goals, endogenously evolve: the trait for categorical truthfulness, for instance, can survive only if its carriers end up at least as well off in material terms as their opportunistic peers.

So let x with $0 \leq x \leq 1$ denote the population share of t-types in an infinite population of agents who are randomly paired and assigned to role S or B to play the modified AaC-game. We assume x to be commonly known and generating the probability x of B-players to encounter a t-type seller in case of no type recognition.

This first illustration of indirect evolution assumes type recognition to be perfectly reliable and costless. Since a t-seller S truthfully reports v, i.e., $\hat{v}(v) = v$ for all v, an opportunistic, i.e., own profit maximizing buyer B encountering a t-seller will propose prices

$$p^*(\hat{v}) = q \cdot \hat{v} = q \cdot v \text{ for all } \hat{v}. \tag{5}$$

Whereas buyer B expects to earn $v - qv = (1-q)v$ for all v and in expectation $(1-q)/2$, the t-type seller S earns nil as in other deterministic ultimatum games.

By contrast an l-type seller S would earn 0 only for $q > \frac{1}{2}$ and for $q \leq \frac{1}{2}$ accept the price $p^* = q$, i.e., S earns in expectation $p^* - q\int_0^1 v\,dv = q - q/2 = q/2$. So for $q < \frac{1}{2}$ the share x of t-type sellers will decrease, i.e., for $q < \frac{1}{2}$ only the monomorphic l-population, $x^* = 0$, is evolutionarily stable. Hence for $q > \frac{1}{2}$ both seller types have the same fitness although for different reasons: a t-type seller would be exploited via ultimatum price proposals $p^*(\hat{v}) = q\hat{v} = qv$ whereas an l-type seller would induce the buyer to refrain from trade via $p^* = 0$. This renders the x-dynamics ambiguous – although one could rely on rare tremble driven evolution with rare price offers $p^* = q$, predicting a slow x-decline towards $x^* = 0$ for $q > 1/2$ (see Selten 1983 who – unlike in Selten 1975 – claims such small trembles to be a natural aspect of habitats; see also Berninghaus, Güth, and Kliemt 2012 for illustrations).

When buyers are loss averse, recognized t-types would still earn nil as seller S whereas an l-type seller expects to earn nil only for $q \geq q^*(k)$ but $p^* - q\int_0^1 v\,dv = q - q\int_0^1 v\,dv = \frac{q}{2}$ for $q < q^*(k)$. Compared to the analogous result for $k = 1$ all what has changed is that the (fast) decrease of x now applies to the smaller range $0 < q < q^*(k)$ of q-parameters.

The analysis demonstrates how the range of q-parameters triggering trade depends on k only when assessing the relative fitness of the l-type seller. Assuming B

to be also fitness-guided would require to neglect k (> 1),[8] i.e., that B's fitness is B's payoff for $k = 1$. Due to our focus on hiding versus revealing we restrict the (indirect) evolution analysis to seller types, specifically the evolution of categorical type t- or l-reporting. Other behavior, namely that of B and the acceptance decisions by sellers, is presumed to be rationally determined.

We readily admit that commercial take-overs often involve at least multiple buyer candidates and then are better modeled by auctions than by bilateral bargaining. However, the synergy effects of such take-overs can be very specific for different buyer candidates. This can render it likely that in the end only the most promising buyer candidate engages with the seller in actual price negotiations. Field examples could be car producers trying to complement their own expertise by taking over specialized firms, e.g., in battery and electric engine production, or pharmaceutical producers acquiring research firms with expertise in developing vaccines.

4 Costly type recognition

Assume now that buyer B must invest the monetary (fitness) cost $C > 0$ to recognize the seller's type t or l without doubt (see Güth and Kliemt 2000 for a previous indirect evolutionary analysis of more or less costly type recognition). As before let B be loss averse. Thus B, when confronting an l-type seller or when not investing in type

[8] Namely, the loss-weighted expectation for $v < q$

$$\int_p^{p/q} (v - p)\, dv + k \int_0^p (v - p)\, dv = \frac{p^2}{q}\left(\frac{1}{2q} - 1\right) - \frac{k-1}{2}p^2$$

– which is B's decision utility – would not represent B's fitness if B-behavior co-evolved with x rather than being rationally decided.

recognition, expects to earn (see fn. 8)

$$q\left(\frac{1}{2q}-1\right)-\frac{k-1}{2}q^2=\frac{1}{2}-q-\frac{k-1}{2}q^2 \text{ for } q \le q^*(k) \tag{6}$$

due to $p^* = q$ for $q \le q^*(k)$ whereas B earns nil due to $p^* = 0$ for $q > q^*(k)$.

If for $q < q^*(k)$ instead B invests in type recognition, B is matched with probability x with a t-type seller and avoids any losses by proposing $p^*(\hat{v}) = q\hat{v}$. So the expected utility from investing in type recognition is

$$(1-x)\left(\frac{1}{2}-q-\frac{k-1}{2}q^2\right)+x\frac{1-q}{2}-C \text{ for } q \le q^*(k). \tag{7}$$

Investing C is strictly better than not investing if

$$C < C^*(k) = [1+(k-1)q]\frac{xq}{2} \text{ for } q \le q^*(k). \tag{8}$$

So for $C < C^*(k)$ buyer B will invest in type recognition and, when recognizing the t-type, "follow its signals" via believing in $\hat{v}(v) = v$ for all values v and offering $p^*(v) = q \cdot v$ for all $\hat{v}$, yielding nil for the seller. Otherwise B reacts by $p^* = q$ in the range $q \le q^*(k)$.

Analogously if $q > q^*(k)$ the expected utility from investing in type detection is

$$(1-x)\cdot 0 + x\frac{1-q}{2} - C \text{ for } q > q^*(k). \tag{9}$$

In this case investing is strictly better for B than not investing if $C < C^{**}(k) = \frac{x}{2}(1-q)$. However, evolutionary dynamics among sellers are not affected by whether $C < C^{**}(k)$

or $C \geq C^{**}(k)$: when B invests, a detected t-type of the seller earns zero profit due to exploitation and just as much as the l-type of the seller who is offered price $p^* = 0$.

For $C < C^*(k)$ and $q < q^*(k)$ the l-type seller S earns more than the t-type when there is trade, due to buyer B proposing $p^* = q$ to the l-type seller. Hence x, the population share of t-types, will decline due to buyer B appropriating all the surplus from welfare enhancing trade when confronting a t-type seller S.

When seller S instead confronts at least two potential risk and loss neutral buyers B who both invest in type recognition, this triggers the extremely opposite result: the competing buyer candidates would choose competitive prices $p^*(\hat{v}) = v$ for all values v when confronting the t-type seller, irrespective of q. So now the seller acquires all the surplus when trade is predicted: for $q > \frac{1}{2}$ the t-type expects to earn $(1-q)/2$ and thus more than an l-type, who would earn nil. For $q \leq 1/2$ the t-type still earns $(1-q)/2$ in expectation, which coincides with $\int_0^1 (p^* - qv)dv = 1/2 - q/2$, what the l-type expects due to the competitive prices $p^* = 1/2$.

However anticipating that all surplus will accrue to sellers (the t-type for $q > 1/2$ and either type for $q \leq 1/2$) and that investment costs $C > 0$ are sunk, not all competing buyers would invest $C > 0$ in type recognition. Unfortunately, assuming that just one of the buyers invests results in pricing equilibria involving mixing (using multiple prices with positive probability). These are behaviorally unconvincing and could be avoided, for example, by relying on sequential pricing, e.g., by assuming that the, in equilibrium, unique buyer B with type recognition states the price after learning about the proposals of all uninformed buyers.

Rather than analyzing this in more detail in the following we will return to assuming only one risk and loss neutral potential buyer and to moderating its monopolistic (ultimatum) power not via buyer competition but via anticipation of altruistic sanctioning, which is robustly confirmed by the abundant evidence of deterministic ultimatum experiments (see Güth and Kocher 2014).

5 Honoring revealing

We capture that exploitative ultimatum price offers will almost surely be rejected by assuming that the single buyer B, who is risk and loss neutral, honors value revelation of the t-type seller by proposing

$$p(\hat{v}) = (q + \varepsilon) \cdot \hat{v} \text{ with } 0 < \varepsilon < 1 - q. \tag{10}$$

When instead this buyer confronts the l-type seller, no rewarding and moderation of exploitation is needed:[9] for $q \leq 1/2$ the expected payoffs $q/2$ of the l-type seller S and $1/2 - q$ of the (loss neutral) buyer B in case of trade at the price $p^* = q$ are more or less (un)equal depending on q (for $q = 1/3$ expected payoffs would be equal).

As a consequence the recognized t-type earns $\varepsilon\hat{v}$ and in expectation $\varepsilon/2$ for all $0 < q < 1$, whereas the l-type of S triggers trade only for $q \leq \frac{1}{2}$ and expects to earn $q/2$ in that case. The evolutionarily stable composition x^* is therefore

$$x^* = \begin{cases} 1 & \text{for } \varepsilon > q \text{ or } q > \frac{1}{2} \\ 0 & \text{for } \varepsilon < q < \frac{1}{2}. \end{cases} \tag{11}$$

[9]We restrict ε-generosity to the case when B encounters a t-type seller. The latter renders the market interaction a deterministic ultimatum game for which such generosity is robustly confirmed experimentally. No such generosity can be expected in case of B confronting the l-type seller (see Bazerman and Samuelson 1983, as well as Güth, Marazzi, and Panaccione 2020).

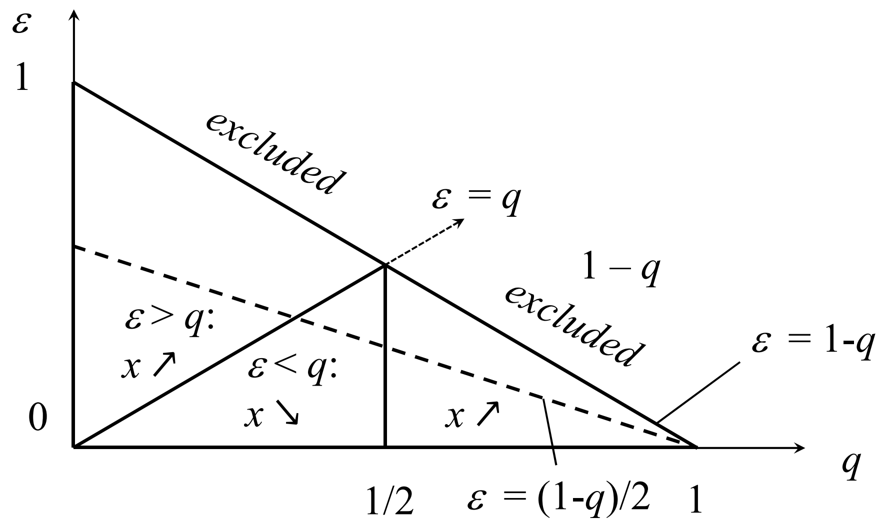

Figure 2: The x-dynamics in the ε, q-diagram

Figure 2 illustrates that, in the admissible (q, ε)-space, $\varepsilon = q$ and $q = \frac{1}{2}$ function as "watersheds" between three parameter regions with distinct evolutionary dynamics. In the rightmost triangle ($q > 1/2$), possible gains from trade are small and will only be realized if buyers obtain reliable value information. This gives the t-type an evolutionary advantage and x increases. In the middle region ($\varepsilon < q < 1/2$), trade is beneficial for buyers even when paying the maximal price q guaranteeing trade for all v. Here an honest seller, who signals acceptance of all prices $p \leq q$ to the buyer by revealing $\hat{v}(v) = v < 1$ for all v, is too little rewarded. So the l-type thrives. Finally, in the leftmost triangle ($q < \varepsilon < 1 - q$) revelation is rewarded by a larger than q share of the gains from trade. An honest seller, who signals $\hat{v}(v) = v$ for all v, receives a substantial fair share of the surplus. This lets the t-type seller earn more than what the l-type receives according to the pooling reservation price q in the range $q < 1/2$. Since the t-type has greater fitness than the l-type, x increases.

Now q is a structural parameter while the "honoring t-type sellers"-parameter ε is a behavioral aspect allowing t-type sellers to gain from trade. This behavioral parameter ε might be estimated based on data for example from an AaC-experiment that lets seller participants send value messages $\hat{v}(v)$ only after deciding between categorical t-type revealing and l-type hiding. The often (in deterministic ultimatum experiments) modal equal split of surplus would require $(q + \varepsilon - q) = \varepsilon = 1 - q - \varepsilon$ or $\varepsilon = (1 - q)/2$. This is indicated by the broken line in Figure 2 and excludes none of the three generic parameter regions $\varepsilon > q$, $\varepsilon < q < 1/2$, and $q > 1/2$, what allows for increasing as well as for decreasing x.

6 Imperfect type recognition

We now assume investment costs $C = 0$ for type recognition in order to focus on another important institutional aspect of nearly all field situations, namely that type recognition is imperfect in the sense that

- when S is of type t buyer B receives the correct signal $\hat{t}(t)$ with probability $\mu_t \in (\frac{1}{2}, 1]$ but also the false signal $\hat{l}(t)$ with possibly positive probability $1 - \mu_t$ and
- when S is of type l the correct signal $\hat{l}(l)$ has probability $\mu_l \in (\frac{1}{2}, 1]$ and the false signal $\hat{t}(l)$ has probability $1 - \mu_l$.

So perfect recognition as assumed above means that $\mu_t = 1 = \mu_l$. For any commonly known population composition x buyer B, after receiving the signal $\hat{t}$, actually expects the t-type seller with conditional probability (see Figure 3)

$$P(t|\hat{t}) = \frac{x \cdot \mu_t}{x \cdot \mu_t + (1 - x) \cdot (1 - \mu_l)}. \tag{12}$$

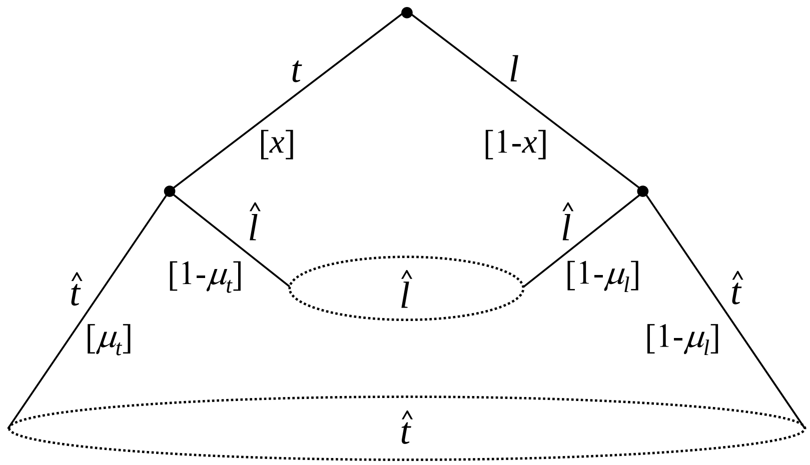

Figure 3: Random seller type signals according to population composition x, respectively $1-x$, of t- and l-type sellers and type signals $\hat{t}$ and $\hat{l}$ according to their reliability parameters μ_t and μ_l. So $P(t|\hat{t})$ is the probability of the left node after t and $\hat{t}$, conditional on reaching the $\hat{t}$ information set.

Due to $\mu_t, \mu_l > \frac{1}{2}$ the type signals are informative but imperfect for $\mu_t, \mu_l < 1$. Obviously the reliability parameters μ_t and μ_l are less relevant when x is either very small or large (see Güth, Kliemt, and Napel 2009 for an analysis of population-dependent reliabilities meaning that – in the present context – μ_t and μ_l could depend on x).

Sufficiently informative type signals should induce buyer B to "follow" them by proposing prices

- $p(\hat{v}) = (q + \varepsilon)\hat{v}$ in case of signal $\hat{t}$ and
- $p^* = \left\{ \begin{array}{ll} 0 & \text{for } q > \frac{1}{2} \\ q & \text{for } q \leq \frac{1}{2} \end{array} \right\}$ in case of signal $\hat{l}$.

"Following" an $\hat{l}$-signal lets B behave as in case of no signals. It therefore suffices to compare the expected payoffs of "following" and "non-following" the $\hat{t}$-signal.

For $q > \frac{1}{2}$ "following" the $\hat{t}$-signal yields

$$\begin{aligned} &P(t|\hat{t}) \cdot \int_0^1 [v - (q+\varepsilon)v]\, dv + \left[1 - P(t|\hat{t})\right] \cdot \int_0^1 [v - (q+\varepsilon) \cdot 1] dv \\ =&P(t|\hat{t}) \cdot \underbrace{\frac{1-q-\varepsilon}{2}}_{>0} + \left[1 - P(t|\hat{t})\right] \cdot \underbrace{(1/2 - q - \varepsilon)}_{<0} \end{aligned} \tag{13}$$

in expectation for B, whereas "non-following" lets B propose $p^* = 0$ and earn nil. So for costless type recognition and $q > \frac{1}{2}$, "following" is better than "non-following" the $\hat{t}$-signal if

$$P(t|\hat{t}) \cdot \frac{1-q-\varepsilon}{2} + \left[1 - P(t|\hat{t})\right] \cdot \frac{1-2q-2\varepsilon}{2} > 0 \tag{14}$$

or

$$P(t|\hat{t}) = \frac{x \cdot \mu_t}{x \cdot \mu_t + (1-x) \cdot (1-\mu_l)} > \frac{2\varepsilon}{q+\varepsilon} + \frac{2q-1}{q+\varepsilon}. \tag{15}$$

Thus, if accuracies μ_t and μ_l of the type signals as well as the population share x of t-types are sufficiently high to render $P(t|\hat{t})$ larger than the right-hand side of inequality (15), following the $\hat{t}$-signal is better for B.

Now B's decision to "follow" is a commitment to believe in all value messages $\hat{v}$ in case of the $\hat{t}$-signal, even when $\hat{v} = \bar{v} = 1$ (and an uncommitted buyer would be rather suspicious). Conversely, "non-following" lets B ignore $\hat{v}$ after receiving the signal $\hat{t}$: as in case of signal $\hat{l}$, buyer B views the value message $\hat{v}$ as uninformative and relies entirely on a priori beliefs. The analysis above lets B suffer from misclassifying l-types as t-types and offering them the price $p^* = p^*(\hat{v}) = (q+\varepsilon)\hat{v} = q + \varepsilon$ due to $\hat{v} = 1$. Even when acknowledging the cost of such misclassification, high enough μ_t, μ_l and

x would induce B to "follow" the $\hat{t}$-signal.

For $q < 1/2$ "following" a $\hat{t}$-signal also yields

$$P(t|\hat{t}) \cdot \frac{1 - q - \varepsilon}{2} + \left[1 - P(t|\hat{t})\right] \cdot (1/2 - q - \varepsilon) \tag{16}$$

as for $q > 1/2$ but now the second summand may be positive and always exceeds $-\varepsilon$. Now $q < 1/2$ lets B offer the smallest price $p^* = q$ guaranteeing acceptance for all values v. Instead by "non-following" B expects to earn $\frac{1}{2} - q > 0$. Thus "following" is better if

$$P(t|\hat{t}) \cdot \frac{1 - q - \varepsilon}{2} + \left[1 - P(t|\hat{t})\right] \cdot (1/2 - q - \varepsilon) > \frac{1}{2} - q \tag{17}$$

or

$$P(t|\hat{t}) = \frac{x \cdot \mu_t}{x \cdot \mu_t + (1 - x) \cdot (1 - \mu_l)} > \frac{2\varepsilon}{q + \varepsilon}. \tag{18}$$

On the one hand, condition (18) for $q < 1/2$ is less demanding than condition (15) for $q > 1/2$ due to $(2q - 1)/(q + \varepsilon)$ being positive for $q > 1/2$. On the other hand, $q < 1/2$ makes satisfying inequality $\varepsilon > q$ easier which renders condition (18) impossible.

Altogether "following" and proposing $p(\hat{v}) = (q + \varepsilon)\hat{v}$ after receiving a $\hat{t}$-signal is better for B if

(I) $q > 1/2$ and $P(t|\hat{t}) > \frac{2\varepsilon}{q+\varepsilon} + \frac{2q-1}{q+\varepsilon}$, or

(II) $q < 1/2$ and $P(t|\hat{t}) > \frac{2\varepsilon}{q+\varepsilon}$.

Otherwise an optimizing buyer B will ignore the seller's value message, $\hat{v}$, and always propose $p^* = q$ for $q \leq 1/2$ and $p^* = 0$ for $q > 1/2$. For large ε, the high cost of rewarding senders of a $\hat{t}$-signal for seemingly truthful value messages, may – and for $\varepsilon > q$ always does – induce "non-following the $\hat{t}$-signal".

In case (I), the t-type of seller S earns $\mu_t\varepsilon/2$ in expectation. Instead a clever l-type seller S, who is falsely recognized as a t-type with probability $(1 - \mu_l)$ and always sends the value message $\hat{v} = 1$, would induce the price proposal $p = q + \varepsilon$, what yields $(1 - \mu_l) \cdot (q + \varepsilon - q/2) = (1 - \mu_l) \cdot (q/2 + \varepsilon)$ in expectation. The expected payoff of t-types exceeds that of l-types and lets the population share x of t-types increase iff

$$\mu_t \frac{\varepsilon}{2} > (1 - \mu_l) \cdot \left(\frac{q}{2} + \varepsilon\right) \tag{19}$$

or

$$\frac{\mu_t}{1 - \mu_l} - 2 > \frac{q}{\varepsilon}. \tag{20}$$

This requires that accuracies μ_t and, especially, μ_l are sufficiently close to 1. Otherwise, l-types earn more and x decreases until it is no longer optimal for B to "follow".

In case (II), B "following" for $q < 1/2$ lets the t-type expect to earn

$$\mu_t \frac{\varepsilon}{2} + (1 - \mu_t)\left(q - q \int_0^1 v \, dv\right) = \mu_t \frac{\varepsilon}{2} + (1 - \mu_t)\frac{q}{2} \tag{21}$$

whereas the l-type expects

$$(1 - \mu_l)\left(q + \varepsilon - \frac{q}{2}\right) + \mu_l\left(q - \frac{q}{2}\right) = (1 - \mu_l)\left(\frac{q}{2} + \varepsilon\right) + \mu_l \frac{q}{2} = \frac{q}{2} + (1 - \mu_l)\varepsilon \tag{22}$$

from always claiming $\hat{v}(v) = \bar{v} = 1$ what is, however, effective only with probability

$1-\mu_l$, i.e., when being misclassified as t-type. The l-type of S is more successful iff

$$\mu_t \frac{\varepsilon}{2} + (1-\mu_t)\frac{q}{2} < \frac{q}{2} + (1-\mu_l)\varepsilon \tag{23}$$

or

$$\frac{q}{\varepsilon} > 1 - \underbrace{\frac{2(1-\mu_l)}{\mu_t}}_{>0}. \tag{24}$$

A sufficient condition for this inequality is $\varepsilon < q$. So *if* buyers are "following" for $q < 1/2$, case (II), the population share x of t-types will decline – similar to the dynamics in the central parameter region in Figure 2. However, when x has decreased sufficiently, buyers stop "following" because this requires $P(t|\hat{t}) > \frac{2\varepsilon}{q+\varepsilon}$ in case (II). Once this happens, type signals are ignored by B. Without appealing to "trembles" (see Section 3 above) there would be no evolutionary pressure on x to decline below the level x^* characterized by

$$\begin{aligned}
& \frac{x^* \cdot \mu_t}{x^* \cdot \mu_t + (1-x^*) \cdot (1-\mu_l)} = \frac{2\varepsilon}{q+\varepsilon} \\
\Leftrightarrow \quad & x^* \cdot (\mu_t q + \mu_t \varepsilon) = x^* \cdot 2\varepsilon(\mu_t + \mu_l - 1) + 2\varepsilon(1-\mu_l) \\
\Leftrightarrow \quad & x^* \cdot \Big[\mu_t(q-\varepsilon) + 2\varepsilon(1-\mu_l)\Big] = 2\varepsilon(1-\mu_l) \\
\Leftrightarrow \quad & x^* = \frac{1}{1 + (\frac{q}{\varepsilon} - 1) \cdot \frac{\mu_t}{2(1-\mu_l)}}.
\end{aligned} \tag{25}$$

Overall, the case of imperfect type detection with $\mu_t, \mu_l < 1$, considered in this section, has different implications than the case $\mu_t = \mu_l = 1$, considered in Section 5. Imperfect type detection involves other case distinctions than perfect detection. First, the x-increase in the right-most parameter region of Figure 2 prevails only if the ‘signal-to-noise’ ratio of $\hat{t}$-signals, $\mu_t/(1-\mu_l)$, is high enough. Otherwise, dynamics vary with the current share x of t-types: if x is sufficiently large, buyers are still “following” but the gains of l-type sellers due to being misclassified as t-types grant them a fitness advantage over honest t-types, what lets x decline. Once x has declined enough to enter the region $x \leq x^{**}$, determined by

$$\frac{x^{**} \cdot \mu_t}{x^{**} \cdot \mu_t + (1 - x^{**}) \cdot (1 - \mu_l)} = \frac{2\varepsilon}{q + \varepsilon} + \frac{2q - 1}{q + \varepsilon}, \tag{26}$$

buyers prefer “non-following”. Since the buyers do not anymore condition on type signals and only condition price proposals on q via $p^* = q$ for $q < 1/2$ and $p^* = 0$ for $q > 1/2$, there is no evolutionary pressure on x, neither upward nor downward (unless when appealing to rare trembles).

Second, the decrease of x in the middle parameter region of Figure 2 survives imperfections in type detection as long as $x > x^*$ with buyers “following”. Once $x \leq x^*$ is reached (or reflects the initial share of t-types), type signals and value messages are ignored and x stops being driven down for lack of a fitness difference.

For the left-most parameter region of Figure 2 with $\varepsilon > q$, we already mentioned that inequality (18) is violated. This highlights an important assumption in Section 5, namely that of *automatically* following t-type signals, believing in the truth of $\hat{v}$-messages and honoring them by price offers $p(\hat{v}) = (q + \varepsilon) \cdot \hat{v}$ for all $\hat{v}$. So Section 5 has neglected that B may *not want to know* whether seller S is truthful, i.e. of type t, and rather not offer $p(\hat{v}) = (q + \varepsilon) \cdot \hat{v} > qv$ which leaves expected rents of $\varepsilon/2$ to S.

Instead in this section buyer B rationally decides between “following” vs. “non-following”. Only the latter commitment means to ignore type signals and value messages, and forgoes welfare-enhancing trade if $q > 1/2$. In contrast, trade takes place in case of either commitment for $q < 1/2$, what allows for better average terms of trade for buyer B when B remains uninformed. Specifically, costless type detection combined with a behavioral norm to reward t-types in case of $q < 1/2$ and $\varepsilon > q$ is less profitable than "non-following".

So our theoretical analysis indicates that buyer B would rather ignore information, what lets both buyer types obtain identical profits and fitness, even if $\mu_t = 1 = \mu_l$. We are unaware whether such wish to react to behavioral norms by active ignorance, here meaning to avoid receiving costless and reliable information because it might trigger a fairness concern towards the prospective trade partner, has already been explored experimentally. In any case the left-most parameter region of Figure 2, identified in Section 5 as implying x-increases, actually yields a constant population share x of t-types if buyers receive perfectly reliable type information but can choose between "following" and "non-following".

Remember that the analysis above of "following the truthful type signal $\hat{t}$" lets buyer B offer the prices $p(\hat{v}) = (q + \varepsilon)\hat{v}$ what, in turn, induces a clever l-type seller to send the value message $\hat{v} = 1$, irrespective of the actual value v. This means for $x < 1$ and $\mu_l < 1$ that value messages $\hat{v} = 1$ are sent to the buyer with positive probability $(1 - x)(1 - \mu_l)$ although the actual value $v = 1$ has 0-probability. One may argue in view of this that value message $\hat{v} = 1$ should render B suspicious and S unbelievable. But should the rare t-type seller with $v = 1$ wrongly be regarded as a liar and receive price offer $p = 0$ in case of $q > 1/2$, respectively $p = q$ in case of $q < 1/2$? We have abstained from speculating about how value messages $\hat{v}$ per se may cause suspicion. One could for instance assume $\hat{v}$-dependent reliabilities μ_t and μ_l (possibly coupled also with x-dependencies of μ_t and μ_l similar to Güth et al. 2009).

One may also question that, although only the seller is aware of value v and the surplus $(1 - q)v$ from trade, the uninformed buyer B is the (ultimatum) price proposer. If instead seller S would propose the price $p = p(v)$ together with sending a value message $\hat{v}(v)$, the setup seems more suitable to speculate about how special value messages like $\hat{v} = 1$ may render the uninformed buyer B suspicious (see Di Cagno et al. 2016 for a related experiment which, instead of possibly imperfect type recognition, assumes leaking value information, i.e., buyer B may learn about v before proposing a price).

Instead of questioning the AaC-setup, which has received considerable attention in the literature, we wanted to explore how categorical hiding and revealing interacts with the crucial institutional parameters q, μ_t, μ_l, and C as well as with the behavioral ones, like k and ε, in shaping the evolution of population share x.

Our evolutionary analysis has been purely theoretical but might contribute to the motivation of future AaC experiments. On the buyer side, the issue of active ignorance or information avoidance in the context of fairness norms seems to deserve attention. It could also be interesting to let seller participants rationally decide whether to commit to categorically revealing or not before sending value messages. Moreover, future studies should check whether standard experimental findings on fairness concerns and lying aversion, by which we justified some of our theoretical assumptions, can indeed be confirmed also in more complex setups like ours. More basically, one could analyze whether and when behavioral adaptation, as observed in experiments allowing for extensive learning, is selecting among behavioral types in ways which resemble fitness-driven evolutionary selection with fitness measured by profits; in other words: when learning selects in ways similar to evolutionary selection.

7 Conclusions

Rather than doing a full-fledged evolutionary analysis of the parameter-rich environment, which would result in excessively complex case distinctions, we have illustrated how various institutional aspects affect the evolutionary dynamics of categorical "hiding" versus "revealing" in the seller role. Specifically, for the – via value messages – modified Acquiring-a-Company game we have discussed how loss aversion and competition of buyers with costly and more or less imperfect type recognition affects the evolution of such seller types.

A crucial aspect is that ultimatum power of a single buyer, who confronts a "revealing" seller, results in exploitation of the seller whereas competition of at least two potential buyers grants all the surplus from trade to the seller. After illustrating these border cases our main focus has been on markets with a monopolistic buyer who, however, "honors the revealing seller" by offering a positive share of surplus from trade. Although this may let the population share of revealing increase, this increase is endangered by unreliable type detection and its interaction with structural parameters, namely the underevaluation ratio q of seller and buyer valuations and the behavioral sharing parameter ε, as analyzed in Section 6.

The behavioral parameter ε has been justified by the robust evidence of ultimatum bargaining experiments. This evidence, however, predominantly rests on non-stochastic ultimatum games. For stochastic ultimatum experiments with less informed proposers[10], like those based on the Acquiring-a-Company game, the evidence of other-regarding concerns (see Cooper and Kagel 2016 for a review) is much more scarce (see Bazerman and Samuelson 1983 and for a related setup Güth et al. 2020) and deserves to be extended.

In future theoretical research one might want to study the simultaneous co-evolution of other-regarding concerns in the buyer role and of hiding versus revealing in the seller role. One could also question the AaC-setup more fundamentally by assigning (ultimatum) price power to the seller.

Declaration on Compliance with Ethical Standards

Conflicts of interest

No funding was received for conducting this study. The authors have no conflicts of interest to declare that are relevant to the content of this article.

Research involving Human Participants and/or Animals

Not applicable (theoretical research w/o human participants and/or animals)

Informed consent

All authors have given explicit consent to submit this work to the *Evolutionary and Institutional Economics Review* and confirm they abide with the authorship principles

[10] Asymmetry of information has mainly been studied by assuming a better informed proposer rather than a better informed responder (see, for instance, Mitzkewitz and Nagel 1993, and Harstad and Nagel 2004).

and ethical rules mentioned in the journal's submission guidelines (https://www.springer.com/journal/40844/submission-guidelines#Instructions for Authors_Ethical Responsibilities of Authors).

References

Akerlof, G. A. (1970). The market for 'lemons': Quality uncertainty and the market mechanism. *Quarterly Journal of Economics 84*(3), 488–500.

Alger, I. and J. W. Weibull (2013). Homo moralis – preference evolution under incomplete information and assortative matching. *Econometrica 81*(6), 2269–2302.

Bazerman, M. H. and W. F. Samuelson (1983). I won the auction but don't want the prize. *Journal of Conflict Resolution 27*(4), 618–634.

Berninghaus, S., W. Güth, and H. Kliemt (2012). Pull, push or both: indirect evolution in economics and beyond. In S. Okasha and K. Binmore (Eds.), *Evolution and Rationality: Decisions, Co-operation and Strategic Behavior*, Chapter 6, pp. 110–132. Cambridge: Cambridge University Press.

Cooper, D. J. and J. H. Kagel (2016). Other-regarding preferences: A selective survey of experimental results. In J. H. Kagel and A. E. Roth (Eds.), *The Handbook of Experimental Economics*, Volume 2, Chapter 4, pp. 217–289. Princeton University Press.

Di Cagno, D., A. Galliera, W. Güth, N. Pace, and L. Panaccione (2016). Make-up and suspicion in bargaining with cheap talk: An experiment controlling for gender and gender constellation. *Theory and Decision 80*(3), 463–471.

European Commission (2004a). Directive on takeover bids. Directive 2004/25/EC.

European Commission (2004b). Directive on the harmonisation of transparency requirements in relation to information about issuers whose securities are admitted to trading on a regulated market. Directive 2004/109/EC.

European Commission (2005). Directive concerning unfair business-to-consumer commercial practices in the internal market. Directive 2005/29/EC.

Güth, W. and H. Kliemt (2000). Evolutionarily stable co-operative commitments. *Theory and Decision 49*(3), 197–222.

Güth, W., H. Kliemt, and S. Napel (2009). Population-dependent costs of detecting trustworthiness: An indirect evolutionary analysis. In *Preference Change: Approaches from Philosophy, Economics and Psychology, Theory and Decision Library A 42*, pp. 243–259. Berlin: Springer.

Güth, W. and M. G. Kocher (2014). More than thirty years of ultimatum bargaining experiments: Motives, variations, and a survey of the recent literature. *Journal of Economic Behavior & Organization 108*, 396–409.

Güth, W., F. Marazzi, and L. Panaccione (2020). Exploiting ultimatum power when responders are better informed – Theoretical and experimental analysis of conflict resolution. Working Paper, Max Planck Institute for Research on Collective Goods.

Harsanyi, J. C. (1979). Bayesian decision theory, rule utilitarianism, and Arrow's impossibility theorem. *Theory and Decision 11*(3), 289–317.

Harstad, R. and R. Nagel (2004). Ultimatum games with incomplete information on the side of the proposer: An experimental study. *Cuadernos de Economía 27*, 37–74.

Mitzkewitz, M. and R. Nagel (1993). Experimental results on ultimatum games with incomplete information. *International Journal of Game Theory 22*(2), 171–198.

Samuelson, W. F. and M. H. Bazerman (1985). Negotiation under the winner's curse. In V. Smith (Ed.), *Research in Experimental Economics*, Volume 3, pp. 105–38. Greenwich, CT: JAI Press.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory 4*(1), 25–55.

Selten, R. (1983). Evolutionary stability in extensive two-person games. *Mathematical Social Sciences 5*(3), 269–363.

Thaler, R. H. (1988). Anomalies: The winner's curse. *Journal of Economic Perspectives 2*(1), 191–202.