TOO COOL FOR SCHOOL? SIGNALING AND COUNTERSIGNALING

NICK FELTOVICH, RICK HARBAUGH, AND TED TO

Abstract. In signaling environments ranging from consumption to education, high quality senders often shun the standard signals that should separate themselves from lower quality senders. We find that allowing for additional, noisy information on sender type can radically alter the predictions of signaling models, potentially explaining why those with the greatest ability to signal might choose not to. We examine equilibria where medium types separate themselves from low types by signaling, but high types then differentiate themselves from medium types by not signaling, or *countersignaling.* High types not only save the cost of signaling by relying on the additional information to stochastically separate them from low types, but in doing so they separate themselves from the signaling medium types. Hence they may countersignal even when signaling is a productive activity. To evaluate this theory we report on a two–cell experiment in which the unique Nash equilibrium of one cell involves countersignaling by high types. Experimental results confirm that subjects can learn to countersignal.

Journal of Economic Literature Classification Categories: C72, D82, D83.

Date: November 1999.
We thank Masaki Aoyagi, Chris Avery, John Bishop, Bruno Broseta, Martin Cripps, Richard Disney, William Hamilton, Robin Hanson, Bill Harbaugh, John Hardman-Moore, Morten Hviid, John Kagel, Ajit Mishra, Stephen Morris, Barry Nalebuff, Paul Newbold, Al Roth, Karl Schlag, Richard Zeckhauser and seminar participants at various conferences and departments.

> "For Nash to deviate from convention is not as shocking as you might think. They were all prima donnas. If a mathematician was mediocre he had to toe the line and be conventional. If he was good, anything went."
>
> – Z. Levinson from *A Beautiful Mind*
>
> (Nasar, 1998, p. 144)

1. Introduction

Following in the tradition of Veblen's (1899) analysis of conspicuous consumption and Akerlof's (1970) model of adverse selection, Spence's (1973a; 1974b) signaling model of overeducation showed how seemingly wasteful actions can be valued as evidence of unobservable quality. Signaling models have since been applied to economic phenomena from advertising (Nelson, 1974) to financial structure (Ross, 1977), social phenomena from courtship (Spence, 1973b) to gift exchange (Camerer, 1988), and biological phenomena from a peacock's plumage (Zahavi, 1975) to a tree's autumn foliage (Brown and Hamilton, 1996). These models conclude that in a separating equilibrium "high" types (high in productivity, wealth, fecundity, or some other valued attribute) send a costly signal to differentiate themselves from lower types.

Contrary to this standard implication, high types sometimes avoid the signals that should separate them from lower types, while intermediate types often appear the most anxious to send the "right" signals. The nouveau riche flaunt their wealth, but the old rich scorn such gauche displays. Moderate quality goods are advertised heavily, while high-quality goods rely on their reputation. Minor officials prove their status with petty displays of authority, while the truly powerful show their strength through gestures of magnanimity. The middle classes are bastions of mainstream culture, while privileged youth are drawn to countercultural lifestyles. People of average education show off the studied regularity of their script, but the well–educated often scribble illegibly. Mediocre students answer a teacher's easy questions, but the best students are embarrassed to prove their knowledge of trivial points. Acquaintances show their good intentions by politely ignoring one's flaws, while close friends show intimacy by teasingly highlighting them. People of moderate ability seek formal credentials to impress employers and society, but the talented often downplay their credentials even if they have bothered to obtain them. A person of average reputation defensively refutes accusations against his character, while a highly-respected person finds it demeaning to dignify accusations with a response.

How can high types be so understated in their signals without diminishing their perceived quality? Most signaling models assume that the only information available on types is the signal, implying that high types will be confused with lower types if they do not signal. But in many cases other information is also available to the receiver. For instance, wealth is inferred not just from conspicuous consumption, but also from information about occupation and family background. This extra information is likely to be only partially informative, meaning that types of medium quality may still feel compelled to signal so as to separate themselves from low types. But even noisy information will often be sufficient to adequately separate high types from low types, leaving high types more concerned with separating themselves from medium types. Since medium types are signaling to differentiate themselves from low types, high types may choose to not signal, or "countersignal," to differentiate themselves from medium types.

We investigate such countersignaling behavior formally with a model that incorporates extra, noisy information on type into a signaling game. We find that countersignaling can emerge as part of a standard sequential equilibrium in which all players are forming rational beliefs and are acting rationally given these beliefs. Countersignaling is naturally interpreted as a sign of confidence. While signaling proves the sender is not a low type, it can also reveal the sender's insecurity. Since medium types have good reason to fear that the extra information on type will not differentiate them from lows, they must signal to clearly separate themselves. In contrast, high types can demonstrate by countersignaling that they are confident of not being confused with low types.

The extra information on type in our model can be seen as a second signal following the literature on multidimensional signals (Quinzii and Rochet, 1985; Engers, 1987). This literature is primarily concerned with whether such signals can ensure complete separation when sender type varies in multiple dimensions. We assume that sender type varies in only one dimension and concentrate instead on the opposite problem of how the extra information can encourage partial pooling rather than complete separation. Given the noisy nature of the extra information, it might seem that high types should signal to further emphasize their quality. Instead, we find that the information asymmetry arising from the noisy extra information can give perverse incentives. Pooling with low types can become a signal in itself—a way for high types to show their confidence that the extra information is favorable to them by taking an action that is too risky for medium types.[1]

While countersignaling can be interpreted as a sign of confidence, it is not simply the absence of signaling by types whose high quality is already evident. We show that countersignaling is more interesting than this for several reasons. First, when the extra information sufficiently differentiates high types from low types, signaling can actually lower a high sender's payoff. Countersignaling can therefore arise even when signaling is a desirable activity that high types would pursue in a perfect information environment. Second, countersignaling reduces the efficiency of receiver estimates of sender quality. Since countersignaling depends on the existence of additional information on sender quality, eliminating this information can actually increase estimate efficiency. Third, when there is a range of possible signals, high types not only choose a cheaper signal than medium types, but choose the same cheap signal being sent by low types. Only by pooling with low types can high types successfully discourage medium types from mimicking their behavior. Fourth, low signaling costs can paradoxically reduce signaling behavior by encouraging high types to countersignal. In an educational context, an increase in the difficulty of an assignment can "challenge" high-ability students to stop countersignaling and to send the signal of completing it.

The idea that signaling-like behavior need not be monotonically increasing in quality has appeared in different formulations in several areas. Teoh and Hwang (1991) develop a model where firms decide whether to immediately disclose favorable earnings information or wait for the information to be revealed by other sources. Waiting makes higher quality firms look bad at first but eventually separates them from lower quality firms which face more immediate pressure to prove themselves. Hvide (1999) examines a labor market model in which education serves partly to inform workers of their true abilities. Under certain conditions, the value of this information is lower for students who perceive themselves to be well above or well below average than for those who perceive themselves to be near average quality, implying only these last types will choose to become educated. Bhattacharyya (1998) looks at the decision of how large a dividend firms should declare. While a standard signaling model predicts higher quality managers should offer higher

[1] Hertzendorf (1993) allow for two endogenous signals, one of which is noisy, but consider only two types of senders, precluding the possibility of countersignaling.

dividends, he finds a screening model predicts that, conditioned on earnings, higher quality managers will declare lower dividends since they can use the funds more efficiently than lower quality managers. Our model differs from these models in following a standard signaling model exactly with the sole exception of allowing for the presence of additional information on sender type. This added realism is sufficient to significantly expand the set of equilibria from a standard signaling game. In particular, there are many types of non-monotonic equilibria[2] which are robust to the standard refinements. The model therefore yields a rich set of results that are applicable to any signaling environment where such information is available. A number of applications are discussed in Section 4.

Countersignaling theory takes the intuition of signaling and shows how it can lead to quite different behavior than normally supposed, offering insight into phenomena which appear inconsistent with the standard signaling model. Of course, countersignaling is somewhat complex and there remains the issue of whether economic agents are capable of such behavior. To help answer this question we report results of an experimental test conducted in the fall of 1995. The experiments involved two games with three types of senders, high, medium and low quality, and a binary signal. The first game is isomorphic to a standard signaling game and has a unique equilibrium in which high and medium types signal. The second game is identical to the first game, except noisy exogenous information on types leads to the unique equilibrium involving countersignaling by high types. Experimental results tend to support the theory's predictions. From almost identical initial play in the two games, subject behavior diverged to a large amount of countersignaling by high types in the latter game and almost none in the former game.

2. A simple example

Continuing the signaling literature's traditional emphasis on education, consider the following stylized example. A prospective employee who had good grades in high school is considering whether to mention her grades in a job interview. Assume that grading standards are weak so that medium and high productivity employees are known to have good grades and low productivity employees are

[2] For expositional purposes, we focus on the simplest equilibria where high types countersignal by pooling with low types. However, there also exist equilibria where some high types signal while others countersignal and "counter-countersignaling" equilibria where very high types differentiate themselves from countersignaling high types by signaling (see Section 3.2).

known to have poor grades. Lying about grades involves the chance of being caught, so the signal of mentioning good grades is costly to low types but free to medium and high types. In addition to this signal, the interviewer will receive from a former boss a recommendation regarding the prospective employee's abilities. Assume low productivity employees expect to receive bad recommendations from their old boss and high productivity employees expect to receive good recommendations, while medium productivity employees receive good or bad recommendations with equal probability.

What should an interviewee do? Without the recommendation, medium and high types should clearly mention their good grades since it costs them nothing and since the grades differentiate them from low types. With the addition of the extra information as embodied by the recommendation, the situation is less obvious. Consider if the interviewer believes that only medium types mention their grades. Then if mediums don't mention their grades they take the chance of either receiving a good recommendation and being thought of as a high or receiving a bad recommendation and being thought of as a low. If lows are sufficiently unproductive relative to mediums and highs, not mentioning grades is too risky. High types face a different situation because they expect to receive a good recommendation. Since they need not worry about being perceived as a low type, they face a clear choice between being perceived as a medium if they mention their grades and a high if they do not. Since receiver beliefs are consistent with sender strategies and sender strategies make sense given receiver beliefs, a countersignaling equilibrium exists in which highs show off their confidence by not mentioning their grades.[3]

A numerical example may help illuminate this case. Assume that productivity is 400, 700, and 900 for lows, mediums, and highs respectively, and that lows and highs are equally prevalent in the population. Given the interviewer's beliefs, medium types can choose to receive either 700 by mentioning their grades or $(400 + 900)/2 = 650$ by deviating from the equilibrium and mimicking the lows and highs. Meanwhile, highs are perfectly separated from low types so they receive 900 by countersignaling versus 700 by deviating and mimicking the mediums. As long as lying about grades costs lows at least 300, they do not gain from mimicking the mediums and a countersignaling equilibrium exists.

[3] A signaling equilibrium is still possible in which both high and medium types are believed to signal, but if lows are sufficiently unproductive relative to mediums and highs and if the recommendations completely separate lows and highs, the equilibrium does not survive the intuitive criterion (Cho and Kreps, 1987). If, as assumed in the theory section of the paper, the recommendations only stochastically separate lows and highs, signaling and countersignaling equilibria may coexist. This issue is explored further in the theory section.

For simplicity this example assumes the signal of “bragging” about one’s grades is free for both medium and high types, but the results do not depend on this assumption. Hence this model would still apply if we were looking not at the decision to report grades, but at the potentially costly decision of whether to get good grades in the first place. Countersignaling would still be an equilibrium even if signaling cost medium types as much as 50 and cost high types as little as *negative* 200. Note that countersignaling can break down not just if signaling is too attractive for high types, but also if signaling is too expensive for medium types, e.g. the grading standard makes it difficult for mediums to get good grades. When signaling by medium types becomes too expensive and they stop signaling in equilibrium, high types no longer benefit from separating themselves from medium types by not signaling. Hence an increase in signaling costs can actually induce high types to start signaling.

Regarding the extra information embodied by the former boss’s recommendation, the extremely dichotomous information structure simplifies the problem, but less informative information can still support a countersignaling equilibrium. In this example even if low types receive a good recommendation 25% of the time and high types receive a bad recommendation 25% of the time, a countersignaling equilibrium still exists. First note that the expected quality of an interviewee who mentions grades is still 700 since only medium types are expected to signal. For an interviewee who doesn’t mention grades, if a bad recommendation is observed the expected quality of the interviewee is $(3/4)400 + (1/4)900 = 525$, while if a good recommendation is observed the expected quality is $(1/4)400 + (3/4)900 = 775$. Since medium types expect good and bad recommendations with equal probability, they still expect to receive 650 if they countersignal versus 700 if they signal. Regarding low types, they expect 3/4 of the time to receive a bad recommendation and 1/4 of the time to receive a good recommendation, so they expect to receive $(3/4)525 + (1/4)775 = 587.5$ by not signaling, giving them even less incentive to deviate than in the previous case. Regarding high types, their quality will be estimated at $(1/4)525 + (3/4)775 = 712.5$ if they countersignal so deviating is unprofitable and the countersignaling equilibrium still stands. Depending on the exact model parameters, even a little bit of extra information can disrupt the standard result that signals are non-decreasing in type.

In this example interviewees faced a simple binary choice of mentioning their grades or not. While signaling decisions are often binary, in many cases a wider range of signals is available, *e.g.*, how expensive a car one buys. In such cases it is less obvious that high types will be willing to pool with low types since they have the extra option of breaking off and sending a higher signal that is not worthwhile for mediums to mimic. The following section develops the theory for this case, showing that highs can still choose to countersignal by pooling with lows. We also consider an extension where there is both a continuum of types and a continuous signal. Readers who feel comfortable with the basic intuition of the above example and are not interested in the theory details may want to skip to the subsequent section on various applications of countersignaling. In the final section we return to the simplest case of a binary signal with three types to report on an experimental test of countersignaling. This test is closely related to the above example.

3. A theory of countersignaling

In this sender–receiver game, we allow for three sources of receiver information. First, there is common knowledge about the distribution of types which incorporates all the background information which both senders and receivers know. For instance, if it is common knowledge that all senders are in a certain age group, then the distribution of types is conditioned on this knowledge. Letting the set of types be $Q \subset \mathbb{R}_+$, this information is summarized by the probability distribution $f(q)$. For simplicity, we define expectations for the case when Q is discrete so that $f(q)$ is the proportion of the population that is of type q.

Second, the sender sends a signal s in the set $S \subset \mathbb{R}_+$ which is observed by the receiver. The receiver observes this signal noiselessly, but does not know which type sent the signal. This signal costs the sender $c(s,q)$ where c is increasing in s and decreasing in q. To ensure there is some signal that everyone would be willing to send, assume $0 \in S$ and c satisfies $c(0,q) = 0$. Also assume that when S is an interval c is convex in s. Since c is also increasing, this implies that c is continuous and almost everywhere twice differentiable in s with $c_{ss}(s,q) \geq 0$ so that the marginal cost of a signal is increasing in the signal. Further, assume the standard "single-crossing property" so that if $q < q'$ then $c_s(s,q) > c_s(s,q')$, *i.e.*, not only is it less costly for "high" types to send any given signal than it is for "low" types but the marginal cost of that signal is also lower. Note that we have assumed (as is standard) that signaling costs are increasing in the signal. This will be relaxed in Section 3.2.

Finally, and this is the unique aspect of the model, the receiver has noisy information about the sender's type. This information is sent at no cost to the sender and is exogenous in the sense that sender actions cannot at this stage affect it. The sender knows that the receiver has this information, but is unaware of exactly what the receiver knows. We model this information as a noisy exogenous signal, $x \in X$, distributed according to the conditional probability distribution $g(x|q)$. Assume that for any q, g has full support over X.[4] The conditional distribution of the exogenous signal is common knowledge but the actual value of x is not known to the sender at the time of sending the *endogenous* signal. In general, the exogenous signal can be thought of as a summary measure of all the other noisy information that the receiver will have about the sender at the time of making the signaling choice. To reduce confusion with the signal, s, we will refer to the noisy exogenous signal, x, as just "exogenous information." We will want x to be an "informative" signal of q as will be defined after the equilibrium concept definition below.

The structure of the game is as follows. First, a sender is drawn randomly from the distribution of types. The sender then sends the endogenous signal without knowing what was or will be the realized value of the exogenous information. Finally, the receiver observes both the exogenous information and the sender's signal. Given this information and her beliefs about sender signaling strategies, the receiver rewards the sender with the sender's expected quality.[5] This can be thought of as a reduced form of a game where senders are workers and receivers are firms which simultaneously make wage offers.[6]

Except for a brief discussion of mixed strategy equilibria in Section 3.2, we consider only pure strategy Nash equilibria, so a strategy is a mapping between types and signals. Let s_q represent the pure strategy of a sender of type q and let the function $\mu(q|s,x)$ be a probability distribution representing receiver beliefs about which types q send observed signal s and information x. Receiver

[4] The assumption of full support simplifies the discussion of out-of-equilibrium beliefs. When the support of g is less than full (*i.e.*, $X(q) \subsetneq X$ is g's support when the sender is of type q), interesting examples with extreme information structures can yield unique Intuitive Criterion countersignaling equilibria (such as the simple example given in Section 2 and used in the experimental test of countersignaling (Section 5)).

[5] The signals are thereby assumed to play a purely informational role, having no effect on the sender's productivity or other valued attributes.

[6] Alternatively, we could assume as in the example from Cho and Sobel (1990) that receivers have utility function $-(q-a)^2$ where a is the receiver's payment to the sender.

expectations of sender quality, given receiver beliefs and the observed signals are

$$\sum_{q' \in Q} q' \mu(q'|s,x).$$

Assuming sender risk-neutrality for simplicity, the gross of costs return to type q of sending signal s is the sender's expected perceived quality

$$E_\mu[q'|s,q] = \int_{x \in X} \left(\sum_{q' \in Q} q' \mu(q'|s,x) \right) g(x|q) dx. \tag{1}$$

Definition 1. A pure strategy *Perfect Bayesian Equilibrium* is given by a type contingent strategy profile s_q and receiver beliefs $\mu(q|s,x)$ where

(i) $E_\mu[q'|s_q,q] - c(s_q,q) \geq E_\mu[q'|s',q] - c(s',q)$ for any $s' \in S$ and

(ii) for any $s \in S$, $\mu(q|s,x)$ is such that if $\{q' \mid s_{q'} = s\} \neq \emptyset$ then

$$\mu(q|s,x) = \frac{g(x|q)f(q)}{\sum_{\{q'|s_{q'}=s\}} g(x|q')f(q')}. \tag{2}$$

Condition (i) requires that agents choose signals as a best response to the receiver's beliefs. Condition (ii) requires that for any information set that can be reached on the equilibrium path, the receiver's beliefs are consistent with Bayes rule and the equilibrium sender strategy.[7]

As it turns out, there will be many pure strategy Perfect Bayesian Equilibria. Furthermore, various standard refinements do not yield a unique equilibrium (*i.e.*, Intuitive Criterion and Divinity-like refinements (Cho and Kreps, 1987; Banks and Sobel, 1987)). However, with the exception of Proposition 4 and a brief discussion, we defer these details to Appendix A.

We adopt the convention here of calling a perfect Bayesian equilibrium a *signaling equilibrium* if s_q is weakly monotonic in the sender's type and strictly monotonic at at least one point. If s_q is strictly monotonic, we call it a *strictly monotonic signaling equilibrium.* Our definition of signaling includes some forms of partial pooling equilibria and deviates from standard terminology in the interest of maintaining consistency with the initial motivating example (Section 2) and the later experimental test (Section 5). For now, we concentrate on the strictly monotonic signaling equilibria which will sometimes be referred to as simply signaling equilibria. Finally, any equilibrium in which

[7] Note that if g has less than full support, (ii) would need to be modified to read " ... such that if there exists a $q \in \{q' \mid s_{q'} = s\}$ such that $g(x|q) > 0$ then ... "

s_q is non-monotonic, will be called a *countersignaling equilibrium*. We will concentrate on the case where the signal is rising and then falling in type.

As mentioned earlier, we require that the exogenous information, x, should be in some sense informative. Informally, first note that in order for the exogenous signal to have any information content in equilibrium, at least two types, $\Lambda \subset Q$, must send the same signal. Otherwise, with perfect separation, the exogenous information plays no role. A sender of type q must believe that if she pools with agents in Λ, in expectation she will be rewarded more than some other type $q' < q$. That is, the sender may do worse than lower types *ex post* once the receiver has observed the available information, but the information is correct on average so that *ex ante* the sender does better in expectation.

To define this notion more precisely, we need to first provide some additional notation. Since, we are only interested in pure strategy equilibria, this assumption will be defined in terms of sets of agents, $\Lambda \subset Q$, who "pool" together.[8] For any nonempty Λ, let $\bar{q}_\Lambda(q)$ be a sender of type q's gross expected payoff, given that the receiver uses Bayes rule and believes her to be of some type belonging to Λ. That is,

$$\bar{q}_\Lambda(q) = \int_{x \in X} \left(\sum_{q' \in \Lambda} q' \frac{g(x|q')f(q')}{\sum_{q'' \in \Lambda} g(x|q'')f(q'')} \right) g(x|q)dx. \tag{3}$$

The term within the parentheses is the receiver's Bayesian estimate of the sender's quality, having observed x. Integrating over all $x \in X$ yields a type-q sender's *ex ante* expected payoff from "pooling" with the agents in Λ. It is easy to see that if $\Lambda = \{q'\}$ for $q' \in Q$ (*i.e.*, it is a singleton) then $\bar{q}_\Lambda(q) = q'$ for any $q \in Q$. That is, if q' is the only type sending some signal s then upon observing s the receiver must believe that the sender is of type q'.

We will consider the conditional distribution, g, to be *informative* if for any $|\Lambda| \geq 2$ and for any $q, q' \in Q$, whenever $q < q'$ then $\bar{q}_\Lambda(q) < \bar{q}_\Lambda(q')$. A sufficient condition for this to hold is that x and q are affiliated or, equivalently, that $g(x|q)$ satisfies the monotone likelihood ratio property. Note that types sending the same endogenous signal are imperfectly separated by the exogenous information since g has full support. This implies that $\inf \Lambda < \bar{q}_\Lambda(q) < \sup \Lambda$ for all $q \in Q$ and $|\Lambda| \geq 2$.

[8] That is, suppose the receiver only knows that the agent belongs to the set Λ with priors based on f and may subsequently adjust those priors based on new information (*i.e.*, x).

3.1. **Countersignaling with three types.** For this subsection, we will consider the following special case. We will take $S = \mathbb{R}_+$ and $Q = \{L, M, H\}$ where $L < M < H$. Throughout the text, we will refer to these types as Lows, Mediums and Highs.

It is clear that in any strictly monotonic signaling equilibrium, each type's expected payoff is equal to her quality and Lows always send signal $s_L = 0$. Thus, in a strictly monotonic signaling equilibrium, $E_\mu[q'|s_q, q] = q$ for $q = L, M, H$ and $c(s_L, L) = 0$.

Before we characterize the signaling equilibria, we first need to define several critical values. Let $\tilde{s}_M$ and $\hat{s}_M$ solve $L = M - c(\tilde{s}_M, L)$ and $M - c(\hat{s}_M, M) = L$. For $s_M \in [\tilde{s}_M, \hat{s}_M]$, let $\tilde{s}_H(s_M)$ and $\hat{s}_H(s_M)$ solve $M - c(s_M, M) = H - c(\tilde{s}_H, M)$ and $H - c(\hat{s}_H, H) = M - c(s_M, H)$. The single-crossing property ensures that $\tilde{s}_M$ and $\hat{s}_M$ are unique and that for any given s_M, $\tilde{s}_H(s_M)$ and $\hat{s}_H(s_M)$ are unique. The signal $\tilde{s}_M$ is the minimum which is required to deter Lows from imitating the Mediums and the signal $\hat{s}_M$ is the maximum that a Medium is willing to send before they would prefer to imitate the Lows. Similarly, $\tilde{s}_H(s_M)$ is the minimum signal which deters Mediums from imitating Highs and $\hat{s}_H(s_M)$ is the maximum signal that Highs are willing to send before they would prefer to imitate Mediums. Given our assumptions it is easy to see that $\tilde{s}_M < \hat{s}_M$, $\tilde{s}_H(s_M) < \hat{s}_H(s_M)$ and $s_M < \tilde{s}_H(s_M)$.

Under our assumptions, the standard signaling result holds and can be written as:

Proposition 1. *The strategy profile (s_L, s_M, s_H) can be supported as a strictly monotonic signaling equilibrium if and only if $s_L = 0$, $s_M \in [\tilde{s}_M, \hat{s}_M]$ and $s_H \in [\tilde{s}_H(s_M), \hat{s}_H(s_M)]$.*

Notice that the characterization of strictly monotonic signaling equilibria does not depend at all on either the distribution of types $f(q)$ or the distribution of the exogenous signals $g(x|q)$. As we will now see, the distribution functions play a significant role in the partial pooling equilibria that can arise. Under the standard signaling framework, such equilibria do not survive commonly used refinements. We will particularly be interested in equilibria where, as discussed in the introduction, high types may choose not to signal or to send a low signal (*i.e.*, countersignal). With only three types, a countersignaling equilibrium must have the Lows and the Highs pooling so there are two candidate classes of pure strategy countersignaling equilibria: u-shaped equilibria and hump-shaped equilibria. Since the former class can be ruled out by our informational assumptions,[9] all countersignaling equilibria must be in the latter class.

Suppose that senders play strategy $\mathbf{s}^* = (s^*, s^*_M, s^*)$ where $s^* < s^*_M$. Let μ, describe beliefs that are Bayes consistent with playing $\mathbf{s}^*$. Then the expected gross payoff to sender q from signal s^*_M is

$$E_\mu[q'|s^*_M, q] = M$$

and the expected gross payoff to sender q from signal s^* is

$$E_\mu[q'|s^*, q] = \int_{x\in X} (\mu(L|s^*, x)L + \mu(H|s^*, x)H)g(x|q)dx.$$

Since the equilibrium has Highs and Lows sending the same signal s^* and since μ is Bayes consistent, $E_\mu[q'|s^*, q] = \bar{q}_{\{L,H\}}(q)$. Therefore, by assumption, $E_\mu[q'|s^*, L] < E_\mu[q'|s^*, M] < E_\mu[q'|s^*, H]$. Due to the noisy exogenous information, the difference in the gross returns from sending signals s^*_M and s^*, $M - E_\mu[q'|s^*, q]$, is declining in quality. Since signaling costs are also decreasing in quality, this is crucial to the existence of countersignaling. For example, if there were no exogenous information on type then $M - E_\mu[q'|s^*, q]$ would be flat. Since the cost of signaling is declining in quality, if Mediums found it advantageous to send signal s^*_M then so would Highs.

Recall the definition of $\bar{q}_{\{L,H\}}(q)$—the expected gross payoff of a type-q individual when the receiver believes them to belong to $\{L, H\}$. We can use the fact that $\bar{q}_{\{L,H\}}(q) = E_\mu[q'|s^*, q]$ to define some necessary critical values which we will then use to characterize the set of countersignaling equilibria. Let $\hat{s}^*_L$ solve $\bar{q}_{\{L,H\}}(L) - c(\hat{s}^*_L, L) = L$. Since $\bar{q}_{\{L,H\}}(L) > L$, $\hat{s}^*_L > 0$. Now for $s^* \geq 0$ let $\tilde{s}^*_M(s^*)$ and $\hat{s}^*_M(s^*)$ solve $\bar{q}_{\{L,H\}}(L) - c(s^*, L) = M - c(\tilde{s}^*_M, L)$ and $M - c(\hat{s}^*_M, M) = \bar{q}_{\{L,H\}}(M) - c(s^*, M)$ respectively. These values have similar roles to those defined for the signaling equilibria. In particular, $\tilde{s}^*_M(s^*)$ is the minimum signal required to deter Lows from imitating the Mediums and $\hat{s}^*_L$ and $\hat{s}^*_M(s^*)$ are the maximum signals that Lows and Mediums are willing to send before they would prefer to send some alternative signal.

[9] Suppose $\mathbf{s}^* = (s^*, s^*_M, s^*)$ is a countersignaling equilibrium where $s^* > s^*_M$. Since $\mathbf{s}^*$ is an equilibrium then both $\bar{q}_{\{L,H\}}(L) - M \geq c(s^*, L) - c(s^*_M, L)$ and $\bar{q}_{\{L,H\}}(M) - M \leq c(s^*, M) - c(s^*_M, M)$. Furthermore, since the cost function for Mediums is flatter than that for the Lows (the single-crossing property), $s^* > s^*_M$ implies that $c(s^*, L) - c(s^*_M, L) > c(s^*, M) - c(s^*_M, M)$. However, this means that $\bar{q}_{\{L,H\}}(L) - M > \bar{q}_{\{L,H\}}(M) - M$, a contradiction since $\bar{q}_{\{L,H\}}(L) < \bar{q}_{\{L,H\}}(M)$.

Proposition 2. *The strategy profile* $\mathbf{s}^* = (s^*, s^*_M, s^*)$ *can be supported as a countersignaling equilibrium if and only if* $s^* \in [0, \hat{s}^*_L]$, $s^*_M \in [\tilde{s}^*_M(s^*), \hat{s}^*_M(s^*)]$ *and* $\bar{q}_{\{L,H\}}(H) - c(s^*, H) \geq M - c(s^*_M, H)$.

Proof. ($\Leftarrow$) For any $x \in X$, let beliefs be given by $\mu(L|s,x) = 1$ whenever $s \notin \{s^*, s^*_M\}$. Let $\mu(L|s^*,x) = g(x|L)f(L)/(g(x|L)f(L)+g(x|H)f(H))$, $\mu(H|s^*,x) = 1-\mu(L|s^*,x)$ and $\mu(M|s^*_M,x) = 1$. These beliefs obviously satisfy (ii) of the definition of a PBE. Thus we need to show that each agent's best response is to follow their prescribed strategy. Obviously no type has an incentive to choose $s \notin \{s^*, s^*_M\}$ since they would receive a payoff of $L - c(s,q)$ which is no greater than the payoff L they would get by choosing $s = 0$.

Now, since $s^* \leq \hat{s}^*_L$, $E_\mu[q'|s^*, L] - c(s^*, L) \geq \bar{q}_{\{L,H\}}(L) - c(\hat{s}^*_L, q) = L$ so that it is individually rational for a type-L player to choose s^*. Since, $s^*_M \geq \tilde{s}^*_M(s^*)$, it follows that $E_\mu[q'|s^*, L] - c(s^*, L) = M - c(\tilde{s}^*_M(s^*), L) \geq M - c(s^*_M, L)$ so that no L-type individual has an incentive to choose $s = s^*_M$.

Since $s^*_M \leq \hat{s}^*_M(s^*)$, $M - c(s^*_M, M) \geq M - c(\hat{s}^*_M, M) = \bar{q}_{\{L,H\}}(M) - c(s^*, M)$ so that no type-M individual has an incentive to choose $s = s^*$.

Finally, in order for Highs to be willing to send signal s^*, they must get at least as much as they would if they imitated the Mediums (*i.e.*, $\bar{q}_{\{L,H\}}(H) - c(s^*, H) \geq M - c(s^*_M, H)$).

($\Rightarrow$) Follows by reversing the previous arguments. ■

Figure 1 provides an illustration of the most efficient countersignaling equilibrium in which the smallest possible signals, $s^* = 0$ and $s^*_M = \tilde{s}^*_M(0)$, are sent. Level sets represent sender indifference between being various payoff/signal combinations, where utility increases in a Northwesterly direction. Following the single-crossing property, the sets are flattest for high types and steepest for low types, ensuring that the indifference curves of different types cross only once. The intercepts represent the utility payoffs of each indifference curve. Thus in this equilibrium, Highs get the greatest payoff ($\bar{q}_{\{L,H\}}(H)$) while the Lows get the least ($\bar{q}_{\{L,H\}}(L)$). According to level set l, Lows are just indifferent between sending signal s^*_M (pretending to be a Medium) and sending the equilibrium signal of zero (pooling with the Highs). That is, s^*_M is the minimum signal that Mediums can send and deter the Lows from mimicking them (*i.e.*, $s^*_M = \tilde{s}^*_M(0)$). The level set h represents the High utility from playing according to equilibrium and not signaling. Highs are willing to pool with the Lows as long as the they get a greater payoff ($\bar{q}_{\{L,H\}}(H)$) than from pretending to be a Medium and sending signal s^*_M (getting payoff $M - c(s^*_M, H)$)). This is true in our case since indifference

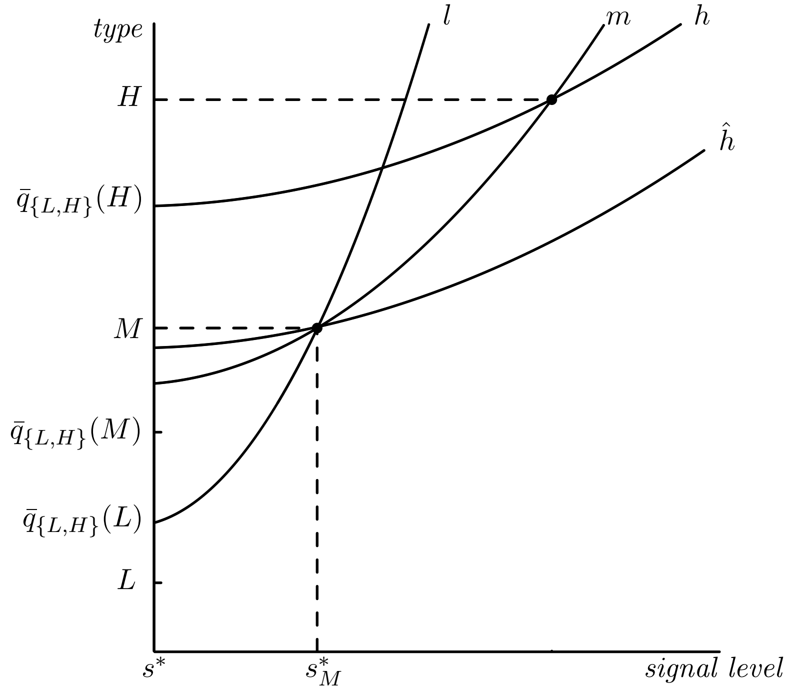


Figure 1. Countersignaling equilibrium with a continuous signal

curve h is higher than $\hat{h}$. Finally, Mediums must prefer to send signal s^*_M than to send s^* and pretending to belong to $\Lambda = \{L, H\}$. This is true since the intercept of m (her equilibrium payoff) is greater than $\bar{q}_{\{L,H\}}(M)$.

From Proposition 2 it is clear that countersignaling equilibria do not always exist. The question is what conditions are required to ensure existence. If the exogenous information is such that the x's for Mediums and Lows tend to be different, then Mediums have little incentive to signal since they can be easily distinguished from Lows. On the other hand, Highs and Lows must tend to have relatively different x's in order for Highs to be willing to pool with the Lows. That is, the exogenous information must sufficiently separate Highs from Lows, but insufficiently separate Mediums from Lows. Formally, these constraints are conditions on the relative positions of $\bar{q}_{\{L,H\}}(L)$, $\bar{q}_{\{L,H\}}(M)$ and $\bar{q}_{\{L,H\}}(H)$. This intuition is made concrete in the following proposition.

Proposition 3. *For $\bar{q}_{\{L,H\}}(M)$ sufficiently small and $\bar{q}_{\{L,H\}}(H)$ sufficiently large, a countersignaling equilibrium exists.*

Proof. Suppose that (i) $\bar{q}_{\{L,H\}}(M) \leq M - c(\tilde{s}^*_M(0), M)$ and (ii) $\bar{q}_{\{L,H\}}(H) \geq M - c(\tilde{s}^*_M(0), H)$. By the definition of $\tilde{s}^*_M(s^*)$, $M - c(\tilde{s}^*_M(0), L) > L$. Hence there exists $\bar{q}_{\{L,H\}}(M)$ sufficiently small that, (i) is satisfied.

By definition $\bar{q}_{\{L,H\}}(M) = M - c(\hat{s}^*_M(0), M)$ and by supposition, this is no greater than $M - c(\tilde{s}^*_M(0), M)$ so that $M - c(\hat{s}^*_M(0), M) \leq M - c(\tilde{s}^*_M(0), M)$. But this implies that $\hat{s}^*_M(0) \geq \tilde{s}^*_M(0)$ and thus $[\tilde{s}^*_M(0), \hat{s}^*_M(0)]$ is non-empty. Finally, take the strategy profile $\mathbf{s}^* = (0, \tilde{s}^*_M(0), 0)$. Since $[\tilde{s}^*_M(0), \hat{s}^*_M(0)]$ is non-empty and (ii) holds then given appropriately constructed beliefs, $\mathbf{s}^*$ is a countersignaling equilibrium. ■

This highlights the intuition behind countersignaling: Mediums signal in order to differentiate themselves from Lows, while Highs do not signal to both save signaling costs and to differentiate themselves from Mediums. They can afford to send the same signal as Lows because they are confident the exogenous information will differentiate them. Looking at Figure 1, if $\bar{q}_{\{L,H\}}(M)$ is too large, Mediums will prefer to send signal s^* rather than their own signal s^*_M. Similarly, if $\bar{q}_{\{L,H\}}(H)$ is too small (less than the intercept of $\hat{h}$), Highs would prefer to pretend to be Medium rather than pool with Lows.[10]

As with the standard signaling model, ours is subject to multiple equilibria. A substantial literature has developed in an effort to "refine" away "undesirable" equilibria in signaling models (Banks and Sobel, 1987; Cho and Kreps, 1987; Cho and Sobel, 1990). As we show in Appendix A, contrary to the standard signaling framework, refinements such as the Intuitive Criterion, D1 and D2 are unable to yield a unique equilibrium. In particular, under conditions qualitatively identical to those given in Proposition 3, countersignaling equilibria continue to exist under the Intuitive Criterion, D1 and D2. Furthermore, countersignaling equilibria which survive such refinements might be, in terms of welfare, more appealing. This is demonstrated formally for the Intuitive Criterion as follows.

Proposition 4. *The Pareto dominant countersignaling equilibrium surviving the Intuitive Criterion Pareto dominates all strictly monotonic signaling equilibria. In particular, every type is strictly better off under the Pareto dominant countersignaling equilibrium.*

[10] The value of $\bar{q}_{\{L,H\}}(H)$ as shown in Figure 1 also turns out to be the smallest such value for which the countersignaling equilibrium depicted survives the Intuitive Criterion, D1 and D2 (Appendix A).

Proof. To prove this we need only make a comparison with the Pareto dominant signaling equilibrium. Let $(0, s_M, s_H)$ be that equilibrium. Since the equilibrium is Pareto dominant, s_M solves $L = M - c(s_M, L)$ and s_H solves $M - c(s_M, M) = H - c(s_H, M)$. Now take the Pareto dominant countersignaling equilibrium, $(0, s_M^*, 0)$ where s_M^* solves $\bar{q}_{\{L,H\}}(L) = M - c(s_M^*, L)$.

First, since $\bar{q}_{\{L,H\}}(L) > L$, it is obvious that the Lows are strictly better off. It also follows that $s_M^* < s_M$ and therefore the Mediums are also strictly better off. Suppose that the Pareto dominant countersignaling equilibrium does not strongly Pareto dominate the Pareto dominant signaling equilibrium *i.e.*, $\bar{q}_{\{L,H\}}(H) \leq H - c(s_H, H)$. By definition, $M - c(s_M, M) = H - c(s_H, M)$. Since $s_M^* < s_M$, $M - c(s_M^*, M) > H - c(s_H, M)$. This implies that beliefs which satisfy the Intuitive Criterion must put probability zero on the event that a signal $s \geq s_M$ was sent by an M. Similarly, Lows would never send such a signal. But since $\bar{q}_{\{L,H\}}(H) \leq H - c(s_H, H)$, type H agents would have an incentive to deviate from the equilibrium with any signal $s \in [s_M, s_H)$. Therefore $(0, s_M^*, 0)$ fails the Intuitive Criterion. ∎

The argument is roughly as follows. Suppose that the Pareto dominant countersignaling equilibrium does not Pareto dominate the best signaling equilibrium. This implies that it must be the Highs who are worse off. However, suppose that the Highs deviate and send their equilibrium signal from the signaling equilibrium. With probability 1 they would be thought to be Highs since Mediums would never be willing to send this signal. Since their signaling payoff is greater than their countersignaling payoff, Highs have an incentive to deviate from the countersignaling equilibrium by playing according to the signaling equilibrium. Thus the countersignaling equilibrium fails the Intuitive Criterion.

Finally, in practice Proposition 4 might overstate the efficiency of countersignaling. We have assumed the receiver is risk neutral, but if the receiver is risk averse or benefits from matching senders to particular jobs based on their quality, the loss in information to the sender in the countersignaling equilibrium might exceed any cost savings to the senders. As discussed below, inefficiencies can also result if signaling is to some extent a desirable activity.

3.2. **Extensions.**

Productive signaling. While the wasteful nature of signaling is often emphasized in the literature, many forms of signaling are, in moderation, productive or otherwise desirable. For instance, while education may be excessive in a signaling equilibrium, it is often a preferred activity in moderation. When signaling is to some extent desirable, countersignaling by high types might be inefficient because of insufficient signaling.

To illustrate the point in a simple manner, suppose we relax the condition that signaling costs are strictly increasing in the signal. In particular, assume that costs are initially decreasing but eventually increasing.[11] With a few minor modifications, although somewhat more complicated to analyze, equilibria like those we have already looked at exist under similar conditions. Let $s_L = \arg\min_s c(s, L)$. All countersignaling equilibria will now have $\hat{s}_L^*$ solving $\bar{q}_{\{L,H\}}(L) - c(\hat{s}_L^*, L) = L - c(s_L, L)$ while the remaining critical values are defined as before. Since signaling costs are negative, signaling might be too little rather than excessive as in a standard signaling model.

Bounded signals. So far we have assumed that the signaling range has no upper bound. While in many cases this might be quite reasonable, it might be more realistic in other cases to include an upper bound on the highest signal that can be sent, *e.g.*, the best signal that a high school student can send is to get straight A's. Consider if there is some maximum signal $\bar{s}$ so that $S = [0, \bar{s}]$. If $\bar{s} < \tilde{s}_H(s_M)$ then from Proposition 1 a strictly monotonic signaling equilibrium does not exist. If $\bar{s} \geq \tilde{s}_M^*(0)$ then provided the conditions from Proposition 3 hold, we know that a countersignaling equilibrium still exists. Thus by eliminating the possibility for a strictly monotonic signaling equilibrium, putting an upper limit on the signal can be considered conducive to countersignaling. High types cannot signal their capabilities relative to mediums by sending a higher signal, so the only alternatives are countersignaling equilibria and signaling equilibria that are not strictly monotonic. That is, the only remaining equilibria are partial pooling equilibria which imply reduced levels of signaling and a resulting loss of information to the receiver.

Counter-countersignaling and mixed strategy equilibria. Real world signaling behavior may be more complicated than the pure strategy equilibria we have derived in our simple three type example. For example, even in situations conducive to countersignaling, some high types may be observed to be countersignaling while others may be observed to be signaling. Two means of getting more

[11] With the education story, one could think of this as being a "joy of learning" benefit to education as opposed to productivity enhancing education. While productivity enhancing education might be more realistic, declining costs will be sufficient to illustrate our point.

complicated signaling behavior are to add more types or to look at mixed strategy equilibria. We briefly consider each of these possibilities in turn.

Suppose there is a fourth type, $H^+ \geq H$ for which signaling is completely costless. This modification can yield a "counter-countersignaling" equilibrium where L, M and H types play according to $\mathbf{s}^*$ and type H^+ agents send an arbitrarily large signal. The presence of H^+ types has no effect on equilibrium beliefs over L, M and H types but a sufficiently large signal, $s^*_{H^+}$, will deter imitation by any type even under beliefs which survive the Intuitive Criterion, D1 and D2. Obviously, less extreme cost structures will yield yet other types of counter-countersignaling behavior. For example, it may be that rather than sending a higher signal, there may be equilibria where H^+ types pool with M's.

Returning to the three-type example, now consider the possibility that senders play mixed strategies or that some proportion of each type plays different strategies. In particular, consider the mixed strategy profile where L and M types and $1-\Delta$ of the H types play according to $\mathbf{s}^* = (0, \tilde{s}^*_M(0), 0)$ and the remaining Δ of the H types send a signal, $\dot{s}^*_H$, at which they get H and are indifferent between sending 0 and $\dot{s}^*_H$. Since fewer high types are now pooling with the lows, this will have the effect of reducing $\bar{q}_{\{L,H\}}(q)$ for all $q \in Q$. Provided that $\bar{q}_{\{L,H\}}(H) \geq M - c(\tilde{s}^*_M(0), H)$ and $\bar{q}_{\{L,H\}}(M) \leq M - c(\tilde{s}^*_M(0), M)$, this strategy profile is clearly an equilibrium. Furthermore, if $\bar{q}_{\{L,H\}}(H)$ is sufficiently large, $\mathbf{s}^*$ survives the Intuitive Criterion, D1 and D2. Notice that when pure strategy countersignaling equilibria exist, there is in general a continuum of Δ's with equilibria where some high types signal and some high types countersignal.

Countersignaling with a continuum of types. If senders come in only three types, a countersignaling equilibrium must involve a positive signal by medium types and a lower or zero signal by low and high types. When types form a continuum and signals are continuous, countersignaling equilibria could take a wide variety of forms. For example, it might increase to some point, fall immediately to zero, and begin increasing again in a manner analogous to counter-countersignaling. Rather than attempting any general statements about such varied possibilities, this section presents an example to demonstrate that countersignaling is possible.[12] We present a particularly conservative example in which the countersignaling equilibrium tracks a standard separating equilibrium, except types

[12] This example is based on a suggestion by Barry Nalebuff.

outside of a middle range send a zero signal. We can then make a noncontroversial assumption about receiver beliefs. If a signal is observed which would not be sent in the countersignaling equilibrium but would be observed in the signaling equilibrium, the receiver believes it was sent by the type that would have sent it in the signaling equilibrium. In other words, the receiver believes the sender is playing the countersignaling equilibrium, but if any information is unexpectedly received that is only consistent with the signaling equilibrium, the receiver responds accordingly.

Following the standard Spence model and earlier assumptions, signaling costs $c(s, q)$ are increasing in the signal s for a given type q, and decreasing in type q for a given signal s (Spence, 1974a). In particular, let the cost function be $c(q, s) = s/q^3$.[13] Types are distributed uniformly over the unit interval, $Q = [0, 1]$. Assume that $X = \mathbb{R}$ and $g(x|q) = q + \epsilon$ where ϵ is a random variable with normal distribution with zero mean and standard deviation $1/4$. Let $\phi(\cdot)$ represent the probability density function of a standard normal distribution.

In the Spence separating equilibrium each type is believed to send a unique signal and the exogenous information has no impact. Representing this mapping from signals to receiver inferences of type by $\hat{q}(s)$, the return to a signal is then $\hat{q}(s) - s/q^3$. Maximizing with respect to s gives $d\hat{q}(s)/ds = 1/q^3$. In equilibrium, beliefs are consistent with actions so $\hat{q}(s) = q$, implying $q^3 dq = ds$. Integrating gives the family of solutions $s = q^4/4 + K$, each of which is a separating equilibrium. Of these the Riley outcome $s = q^4/4$ is the only reasonable solution since type $q = 0$ never benefits from sending a positive signal.

We are interested in a countersignaling equilibrium where there are two types $0 < q_a < q_b < 1$ such that $s^*_q = q^4/4$ for $q \in [q_a, q_b]$ and $s^*_q = 0$ for $q < q_a$ and $q > q_b$. For $s = 0$ there will be pooling so the exogenous information and distribution of types will affect receiver beliefs. In particular, upon observing $s = 0$ and x, Bayes consistent receiver beliefs are

$$\mu(q|0, x) = \frac{\phi(4(q - x))}{\int_0^{q_a} \phi(4(q' - x))dq' + \int_{q_b}^1 \phi(4(q' - x))dq'}.$$

[13] For $c(q, s) = s/q^n$, larger values of n imply signaling costs decrease more rapidly in type and increase more rapidly in the signal. Senders are therefore more readily discouraged from sending higher signals and the amount of "waste" from signaling is less, giving each sender a higher payoff in equilibrium. For countersignaling equilibria of the form that we will be examining, it appears to be easier to construct equilibria for larger n. For smaller values medium types must send such high signals to separate themselves from each other that they do not benefit sufficiently from signaling.

On the other hand, for $s \in (0, 1/4]$, neither distribution plays any role and we assume that beliefs are given by $\mu((4s)^{1/4}|s, x) = 1$. For $s > 1/4$, let $\mu(1|s, x) = 1$.

Given these beliefs, the expected gross payoff to sender q from signal $s = 0$ is

$$E_\mu[q'|0, q] = \int_{-\infty}^{\infty} \left(\int_0^{q_a} q'\mu(q'|0, x)dq' + \int_{q_b}^{1} q'\mu(q'|0, x)dq' \right) \phi(4(x - q))dx.$$

The conditions for the marginal types to be indifferent are

$$E_\mu[q'|0, q] - c(0, q) = q - c(q^4/4, q) \ \text{ for } q = q_a, q_b.$$

Solving numerically, one solution is $q_a \approx 0.521$ and $q_b \approx 0.961$. Further calculations confirm types $q < q_a$ and $q > q_b$ prefer $s = 0$ to $s = q^4/4$ and types $q_a < q < q_b$ prefer $s = q^4/4$ to $s = 0$.

The countersignaling equilibrium tracks the signaling equilibrium over the range $[q_a, q_b]$, but outside of this range higher and lower types "pool" with each other by sending a zero signal. Type q_b is just indifferent between signaling (and being identified as type q_b) and not signaling (and relying on the exogenous information x). Higher types not only save costs by countersignaling, but since $E_\mu(q|0, q_b) \approx 0.728 > q_a$ they are estimated to be of higher quality than many types sending a strictly positive signal.

4. Applications of Countersignaling

As suggested in the introduction, many phenomena seem inconsistent with the standard signaling model in that the signal is not always monotonically increasing in type. This section discusses a few of these phenomena in more detail. In some of these examples high types just save costs by countersignaling, while in others they appear to receive a higher type estimate by separating themselves from signaling medium types.

4.1. **Education.** Signaling models predict overeducation in equilibrium, but the opposite problem of underachievement by talented students seems to be pervasive in some countries. Applied to education, our model predicts that some high ability students may choose to underperform, relying instead on other information to separate themselves from low ability students.

The U.S. education system is unusual in its extensive reliance on tests such as the SATs which are based primarily on aptitude rather than on acquired knowledge. Interpreting these tests as the exogenous signal and grades as the endogenous signal, eliminating the SATs would force talented high school students to signal their ability through better grades. Alternatively, making the SATs evaluate knowledge rather than aptitude would force students to expend effort on studying for the exams and achieve a similar result. The U.S. system is also widely criticized for its low grading standards and the consequent ease of attaining A's. This situation can be thought of as a bounded signal where high quality students have limited ability to distinguish themselves by sending a more costly signal. Talented students may back away from the cheap maximum signal because medium types are using the same signal to separate themselves from low types.[14] If grading standards are raised sufficiently, a strictly monotonic signaling equilibrium is possible whereby the best students feel compelled to separate themselves from their less talented peers through better grades. Wider availability of more difficult courses such as Advanced Placement classes, as is the recent trend, would have the same effect.

Unlike in the U.S., acceptance to universities in the U.K. is largely dependent on performance in the "A-levels" (knowledge-based entrance exams). As a result, one would expect that countersignaling behavior would be less prevalent in the British university admissions process. However, other elements of the British education system can induce countersignaling behavior as has recently been documented. In particular, Smith and Naylor (1998) show that after controlling for socioeconomic background, gender, age, marital status, etc., the university performance of students who attended private high schools was on average poorer than those who attended public schools. Controlling in addition for university performance, Naylor *et al.* (1998) find that these students also tended to have higher post graduation earnings. That is, despite the fact that, as a group, private high school graduates have poorer than average university performance, a student who attended a private high school earned on average more than a student with the same major, from the same university, with an identical grade point average and identical background but who instead attended a public high school. Interpreting these results within our framework, one's high school represents part of the background information available to prospective employers. Knowing that graduating from a

[14] Sending other signals which interact with the cost of getting good grades is also possible. If grades are too easy, high-ability students may conspicuously engage in activities which make studying difficult. If they can still get good grades then this handicap signaling shows them to be more capable than other students with comparable grades. To the extent that these actions sometimes make it excessively difficult to study, the best students might on average have lower grades than less talented students who take the easier route, implying a signaling pattern that is similar to that of a countersignaling equilibrium.

private school is viewed favorably by many employers,[15] some of these students[16] rationally do not perform as well at the university level as their publicly educated peers.

4.2. **Advertising.** Advertising can be a signal of quality, demonstrating the advertiser's faith that consumers who try the product will come back for more and thereby allow recovery of advertising costs (Nelson, 1974). The countersignaling model indicates that this relation between advertising and quality need not always hold. While advertising may prove that a product is not of low quality, it also reveals the firm's concern that the product would be perceived as low quality without the advertising. Firms of high quality products may therefore avoid advertising so as to show their confidence in the product's reputation. This might explain why advertising seems rare by high-quality firms in reputation-intensive fields such as investment banking, law, and medicine.

Separate from signaling quality through advertising expenditures, the content of advertising may be designed to convince consumers that purchase of the product is the best signal of one's wealth, fashionability, or other attributes. Sellers of relatively inexpensive or unstylish products may try to convince consumers that, contrary to the signaling logic of conspicuous consumption (Veblen, 1899), purchasing these goods may actually increase their status. For instance, Hyundai had a controversial advertising campaign suggesting that Hyundai buyers, unlike the owners of more expensive sports cars, were not insecure about their masculinity. Relatedly, firms may try to position their products as "classy" by rejecting gaudy displays of wealth in favor of simple designs. General Motors has marketed the Buick Park Avenue under the countersignaling slogan, "Where the power of understatement is understood."

Advertising campaigns that soft sell the qualities of the product can also be strategic. Hyperbole may persuade consumers that there is at least some basis for the claims, but also leave consumers wondering why the firm is so worried that the product's strengths are unappreciated. Irony takes the soft sell further by reference to and rejection of signaling. Of course, such attempts need

[15] There are many plausible reasons which are consistent with countersignaling for the apparent favoring of private high school graduates. One of these is the notion that these schools belong to an "old boys network" where hiring is based on membership to the network. This is consistent with countersignaling if a student's membership to such a network is correlated with their contribution towards employer profits, even if such membership is uncorrelated with ability.

[16] In fact, they perform more poorly with sufficient frequency to affect, at a statistically significant level, the econometric results of Smith and Naylor (1998).

not be successful. While the “Image is nothing” campaign boosted Sprite among Generation Xers, Coca-Cola’s “OK” drink failed to capture the targeted slacker market.

4.3. **Counterculture.** Signaling theory helps explain how complex and expensive cultural phenomena can be supported in a material world. Countersignaling theory offers some insight into how countercultures can arise even as growing prosperity would seemingly favor the further expansion of mainstream culture. For instance, in the 1960s a counterculture developed in Western countries as many upper-middle and upper class students dropped out of college to pursue alternative lifestyles. The spread of tertiary education to the broad middle class in this period left privileged youth less able to distinguish themselves from their peers through a college diploma. As suggested by countersignaling theory, these youth may have reacted to the expansion of types sending the signal by choosing not to send it. Expecting their abilities and family connections to separate them from the underclass, they chose to differentiate themselves from the middle class by rebeling against mainstream culture.

4.4. **Informality.** Less extreme than the counterculture of the 1960s and its more recent manifestations is the everyday tendency towards informal dress. Ironically, but consistent with the theory, dressing down is most common in the United States, the country with the cheapest clothing in the developed world. Indeed, this trend towards informality arose during the last several decades as clothing costs dropped sharply in relative terms.

Informality in social interactions also lends itself to analysis with countersignaling theory. Politeness is usually not lavished on complete strangers nor on close friends, but rather reserved for relationships that fall somewhere in between. To be overly polite towards a close friend is not just to waste effort, but to signal that the relationship is of a more “medium” nature. The friend will wonder if one is deliberately putting distance in the relationship, or whether one is afraid that some information is available that would put the friendship in an even worse light than that which politeness signals. The suspicion behind the question, “Why are you being so nice?” is well-founded in a countersignaling equilibrium.

4.5. **Incomplete contracts.** Countersignaling offers some new insight into why mutually beneficial contracts are not always adopted. Consider a proposed marriage where one partner is significantly richer than the other and fearful of an expensive divorce settlement. A prenuptial agreement as part of a signaling equilibrium would seem to be the logical outcome. If the poor partner did not suggest or agree to such a contract the rich partner would conclude that the motive for marriage was money, not love. When there are more than two types, motivated to varying degrees by both love and money, the situation is more complicated. Offering a prenuptial agreement proves that one is driven by more than just money, but also reveals concern that rumors or other information might suggest otherwise. Hence a countersignaling equilibrium is possible where a poor partner with unimpeachable intentions does not propose a prenuptial agreement, or may even indignantly refuse to sign one, in order to show confidence.[17] Of course, similar situations can arise whenever business partners are making commitments to each other but have some doubts regarding the reliability of the other party. To offer a contract which helps protect against losses does not necessarily assuage the fears of the other side, and may even aggravate them. This may help explain the popularity of "gentleman's agreements" in which reputable parties proceed without a written contract. A legalistic culture in which both sides believe quality firms always use contracts can help avoid such an equilibrium.

4.6. **Fashion Cycles.** Many fashions behave cyclically. Ties move from fat to thin and back, hem lengths move up and down, fabric colors range between somber and outrageous. Signaling models can help explain how new fashions are continually introduced to allow consumers to signal their wealth or fashion-consciousness (Pesendorfer, 1995), but have difficulty explaining how the new fashion can be the same as the old fashion, especially when the old fashion still has some currency. While the countersignaling model is static, it offers some insight into this situation. Assume consumers vary in their fashionability and that the costliness of acquiring the current fashion (signaling) is decreasing in type. This could be for informational reasons—knowing where to shop for instance. There is some additional information on consumer fashionability that is noisy and exogenous. Assume that the current fashion is for thin ties, and that the lower half of the fashionability distribution is comprised of fashion laggards who find it too difficult to purchase new thin ties and continue to wear fat ties. Among the upper half of the distribution most types are fashion followers who wear thin ties, except for a fringe of fashion leaders who countersignal by wearing old fat ties. Some fashion followers might think of countersignaling, but given the other

[17] Spier (1992) shows that contracting may also fail when there is uncertainty over the rich rather than the poor partner's type.

information about their fashionability they are worried that such a bold move will be misunderstood and they will be labeled as fashion laggards. This fear by fashion followers is exactly what makes cycling back to fat ties so attractive to fashion leaders. By pooling with fashion laggards they are able to clearly separate themselves from the fashion followers who lack the confidence to make such a risky fashion statement.

4.7. **Stubbornness and flexibility.** The economist's usual admonition to ignore sunk costs can be inappropriate when there are information asymmetries. Substantially revising an initial plan suggests that the basis for the original decision might not have been so solid, reflecting unfavorably on the decision maker (Prendergast and Stole, 1996). Even though constantly adjusting the plan to accommodate new information might be efficient, decision makers might benefit from stubbornly sticking to their original plans in order to signal that their initial assessment of the situation was well-founded.

Despite the signaling logic of stubbornness, a willingness to admit mistakes is not always seen as lack of gravitas. People who ignore the potential embarrassment and publicly change their minds are sometimes credited with being "big enough to admit their mistakes" rather than labeled as flighty. This could reflect a countersignaling equilibrium in which the highest quality types demonstrate their confidence by a flexibility that middle-range types are afraid to show.

4.8. **Display behavior.** Displays signal an animal's unobservable qualities, *e.g.*, a loud roar signals a lion's strength. While such signals are common, "cool" behavior is also observed. An animal may act nonchalant, holding back its display in order to show its confidence. Excessive displays might be interpreted not as a sign of strength but as a fear of otherwise being seen as weak. Cho and Kreps (1987) illustrate signaling behavior with the example of a person proving their toughness (or hiding their weakness) by choosing beer rather than quiche for breakfast. As countersignaling shows, a truly tough person might instead choose quiche to show they are beyond having to rely on such displays.[18] A similar reluctance to use displays can arise in mating relationships. Males are often portrayed as inveterate signalers constantly trying to impress females, but popular wisdom suggests that the quiet confidence of the "strong, silent type" is not unappreciated.

[18] Indeed, the most ruthless fictional criminals are often portrayed as almost effete in their refined tastes.

4.9. **Anonymous giving.** Charitable donations are usually publicized, suggesting that donors receive some benefit from having their generosity exposed. For instance, public donations may signal one's trustworthiness or other socially desirable characteristics (Harbaugh, 1998). Yet, donations are sometimes made anonymously. In many cases this may be to avoid further solicitations, but in some cases the donors might feel uncomfortable with a public display of their magnanimity. Concern that such self-promotion might be seen as "tacky" and lead to a lower type estimate is justifiable in a countersignaling equilibrium. Trying too hard to prove one's goodness may reveal concern that other information suggests a less favorable evaluation.

4.10. **Understatement and sarcasm.** Linguists have noted that understatement can paradoxically intensify a description despite being a weaker and therefore less risky commitment by the speaker (Hubler, 1983). Interpreting statements about quality as a simple signaling game in which the speaker wants to convey a positive assessment but exaggeration is costly due to the danger of being uncovered, understatement is readily interpreted as a countersignal. It is less costly to the speaker because it is less exaggerated, and it may also convey an even stronger message than praise alone. The key, as linguists have noted, is other contextual information.

Sarcasm can also be seen as a countersignal, different from understatement in that the speaker wishes to convey a negative impression. Reporting what is in fact quite bad to be "excellent" or some other superlative can represent a strongly negative assessment if the context is sufficient to make it highly unlikely that the object of such praise is as good as claimed (Haiman, 1998).

5. A test of countersignaling

To investigate whether agents can engage in countersignaling behavior, we use a simple experiment with two cells, one corresponding to a standard signaling game and the other to a countersignaling game. We follow the three–type model outlined above with several simplifications. First, to reduce the computational burden on the subjects we use a greatly simplified information structure in the countersignaling cell of the experiment; we assume that there is no overlap between the exogenous information sent by low types and that sent by high types. Second, we assume the endogenous signal is binary.[19] Third, we automate the receiver side so that senders

[19] These simplifications are not necessary for the existence or uniqueness of a countersignaling equilibrium, but substantially reduces the computational burden on subjects.

receive their exact expected quality conditional on the signals. This models a situation where the receiver can observe the average quality of all types with a given combination of signals, but cannot observe individual quality. Since the game then becomes one of simultaneous play of the senders, rather than of sequential play of senders first and receivers second, Nash equilibrium, rather than perfect Bayesian or sequential equilibrium, is the appropriate equilibrium concept. It can easily be shown that each Nash equilibrium of our simplified version of the strategic environment corresponds to a sequential equilibrium of the original environment, as long as receiver payoff functions are appropriately modeled.

We make the test of countersignaling difficult in two ways. First, we test for countersignaling when high types have negative signaling costs, *i.e.*, they receive a bonus for signaling, implying they would always signal in a perfect information environment. (High types are willing to forgo this bonus in the countersignaling equilibrium since they are assigned a lower type estimate if they choose to signal.) Second, we give the games a strong educational context in which signaling is described as getting a "good grade" and not signaling as getting a "bad grade," terminology which might well encourage signaling by our subjects (primarily university students) of all quality types.

5.1. **Description of the game.** Both the signaling and countersignaling cells share the same basic structure. We motivate the cells to the subjects with an education model in which students signal their skill levels to firms. There are three types of student: High, Medium, and Low skill level. They signal in two ways: by grades, which they choose, and by test scores, which are exogenous and may have a random component. Test scores are uncorrelated with skill level in the signaling game but correlated with skill level in the countersignaling game. After students have signaled they are hired by competitive risk–neutral firms.

Since we assume firms to be competitive and risk neutral, their decision problem is trivial; they pay each student a wage equal to that student's expected productivity, conditional on the student's grade and test score, and the overall distribution of grades and test scores for each type (consistent with Bayes's rule). We therefore suppress firms' role in the experiment; their role is played by computer, rather than by subjects.[20]

[20] We do this in an attempt to simplify the experimental environment for the subjects, and because we are primarily interested in whether students can learn to play their part of a countersignaling equilibrium. If they can do so, it is probably reasonable to expect that firms, which tend to be long–lived and make many hiring decisions over long spans of time, can learn to play their part of such an equilibrium. The technique of automating some parts of a

The population of students consists of 25% Highs, 50% Mediums, and 25% Lows.[21] The distribution of skill levels is common knowledge but the skill level of an individual student is that student's private information, unknown to other students or to firms. Students' observable characteristics come in two forms, the signal (grades) and exogenous information (test scores). Grade is a binary choice, either G (good grade) or B (bad grade). A bad grade is costless, while the cost of a good grade varies inversely with skill level. After students have chosen their grades they are informed of their test score, which is either P (pass) or F (fail). Test scores do not depend directly on grades, but in the countersignaling game they do depend on skill level. The test score corresponds to the exogenous information in the theory section.

After grades and test scores have been determined, students are hired by firms. Firms value students according to the students' skill levels; High types are the most valuable and Low types the least. Students' skill levels are unobservable to firms, but firms know the *distribution* of skill levels for each possible combination of grade and test score (for example, firms know how many High, Medium, and Low types had good grades and passed the test, but do not know the type of a particular student who had good grades and passed the test). Since firms are competitive and risk–neutral, each student is paid the mean productivity of all students with the same observable characteristics.

We use two sets of parameter values, one for the 'S' (signaling) cell and the other for the 'C' (countersignaling) cell. The characteristics of each player type are given in Table 1. In each case the "population" of students consists of 4 High types, 8 Medium types, and 4 Low types. Note that Highs have a negative cost of signaling, *i.e.*, Highs receive material benefits by signaling.

The unique Nash equilibrium outcome of the S cell has all High and Medium types earning a good grade and none of the Low types doing so. In contrast, the unique Nash equilibrium outcome of the C cell has all Medium types earning a good grade but none of the High or Low types doing so. This difference arises from the different nature of the exogenous signal (test score) in the two games.

game in order to simplify an experimental environment is quite common, and has been done in a wide variety of settings, including other signaling games (*e.g.*, Cooper *et al.*, 1997a,b), simple markets (Roth *et al.*, 1991), Cournot–Stackelberg duopoly games (Huck *et al.*, 1999), common–value auctions (Garvin and Kagel, 1994; Kagel and Levin, 1986), asymmetric–information "lemons" markets (*e.g.*, Ball, 1991; Ball *et al.*, 1991; Cifuentes and Sunder, 1991), and bargaining under incomplete information (Archibald, 1998; Archibald and Wilcox, 1999). We will discuss the possible implications of this design component on the results in Section 5.4.

[21] Since the only players making actual decisions will be the students, we will use the terms "student" and "player" interchangeably.

Player Type	Number	Cost of	% Passing Test		Productivity
(Skill Level)	of Type	Good Grade	S cell	C cell	
High	4	−25	50	100	900
Medium	8	+25	50	50	700
Low	4	+350	50	0	400

Table 1. Characteristics of Player Types

The exogenous information in the S cell is completely uninformative; there is no difference—even probabilistically—between the exogenous information sent by the different types of student. The game thus reduces to a standard signaling model in which higher–quality types signal to distinguish themselves from lower–quality types. In equilibrium, Highs and Mediums earn a good grade so that they do not look like Lows, while Lows do not attempt to pool with Highs and Mediums by earning a good grade since the cost is too high. Highs would like to be distinguished from Mediums also, but the only way to accomplish this is to not signal, in which case they would look like Lows, which gives them a much lower payoff. Thus, we have Highs and Mediums pooling with each other, but separated from Lows.

In the C cell the extra information included in the exogenous information allows for a different equilibrium. Since Lows always fail the exam and Mediums fail the exam half the time (and don't know their test score until after they choose their grades), Mediums are still concerned with differentiating themselves from Lows. Hence they continue to earn a good grade. But Highs are no longer concerned with differentiating themselves from Lows since Highs always pass the exam and Lows always fail it. Instead, they are concerned with being mistaken for the Mediums who pass the exam. Since Mediums are earning a good grade to separate themselves from Lows, Highs react by earning a bad grade to separate themselves from Mediums. They do this at some sacrifice since earning a good grade has negative cost.[22]

5.2. **Experimental procedures.** The experiment consisted of eight sessions, four of the S cell and four of the C cell, with 16 subjects in each session. Subjects, mostly undergraduates at the University of Pittsburgh, were recruited by announcements in economics classes and by notices

[22] Recall from Section 2 that inclusion of informative exogenous information is not a sufficient condition for a countersignaling equilibrium. If the difference between pass rates for High and Medium types is sufficiently small then a signaling equilibrium is possible as in the S cell. The cost function also affects the set of equilibria. For instance, if the cost of earning a good grade was increased sufficiently for all types, the model predicts that Medium types would stop signaling due to the expense and High types would start signaling to differentiate themselves from Medium types.

		Grade	
		G	B
Test Score	P	#High: 2 #Medium: 4 #Low: 0 Salary: 734	#High: 0 #Medium: 0 #Low: 2 Salary: 400
	F	#High: 2 #Medium: 4 #Low: 0 Salary: 734	#High: 0 #Medium: 0 #Low: 2 Salary: 400

(a) Signaling Cell

		Grade	
		G	B
Test Score	P	#High: 0 #Medium: 4 #Low: 0 Salary: 700	#High: 4 #Medium: 0 #Low: 0 Salary: 900
	F	#High: 0 #Medium: 4 #Low: 0 Salary: 700	#High: 0 #Medium: 0 #Low: 4 Salary: 400

(b) Countersignaling Cell

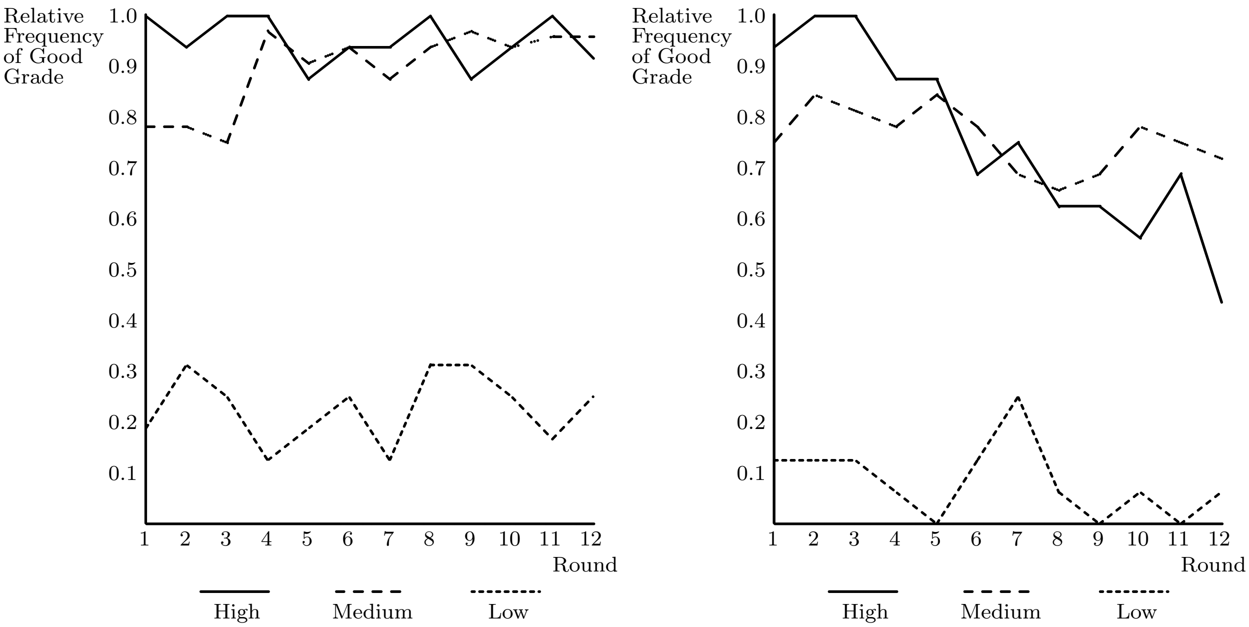
Table 2. Sample of post-round information given to subjects

placed on bulletin boards. Sessions lasted for roughly 90 minutes and consisted of one practice round and 12 rounds, except for one session which, due to time constraints, consisted of one practice round and 10 rounds. We kept the number of rounds small in order to ensure that subjects had time to think about their decisions and that payoffs per decision were relatively high.

The experiment was conducted with pen and paper. Instructions were read aloud to subjects (a copy is given in Appendix B), and a table similar to Table 1 (but with only the information concerning the cell they were in) was written at the front of the room. Subjects were given record sheets with spaces to write for each round their skill level, grade, test score, salary (called "gross payoff"), cost of earning a good grade (when applicable), and net payoff.[23] In each round, each subject drew (from a bucket) one of sixteen slips of paper, on which were printed a skill level and test score, and a space for subjects to write their choice of grade. The slips of paper were prepared in advance and were repeatedly folded and sealed so that test scores could not be seen without breaking the seal. Skill levels and test scores were block–randomly assigned to the slips of paper, so that, in the S cell for example, in each round there were 4 Medium types that passed the test, 2 Low types that failed the test, and so on.

In each round, each subject copied her skill level onto her record sheet, chose her grade, and wrote this grade on her record sheet and on the slip of paper. After this was done, a monitor came to the subject's desk and watched her break the seal, revealing her test score. The subject wrote

[23] Because High types have a negative "cost" of earning a good grade, we used the phrase "bonus or penalty" rather than "cost" in the experiment.



(a) Signaling Cell

(b) Countersignaling Cell

Figure 2. Plot of Relative Signaling Frequency

her test score on her record sheet and the monitor collected the slip of paper. When all the slips had been collected, the distribution of skill levels and the salary corresponding to each possible pair of observable characteristics were posted at the front of the room. Examples of this posted information are given in Tables 2(a) and 2(b) for the signaling equilibrium of the S cell and the countersignaling equilibrium of the C cell respectively. Subjects then wrote their salaries on their record sheets, and calculated and recorded their net payoffs. The next round then began. The posted information remained posted until it was replaced by information from the following round.

Subjects were paid a \$5.00 participation fee. In addition, one of the non–practice rounds was randomly selected from those played, and subjects were paid their net payoffs from that round at the exchange rate of 100 points/\$1.00, in addition to the \$5.00. Average earnings were approximately \$11.00 for a 90–minute session. All earnings were paid to the subjects in cash at the end of the session.

5.3. **Experimental results.** Figures 2(a) and 2(b) show the frequency with which players in the two cells chose G (good grade). Note that early–round play for the first three rounds is very similar

	C Sessions			S Sessions		
	High	Medium	Low	High	Medium	Low
Round 1	15/16 (.938)	24/32 (.750)	2/16 (.125)	16/16 (1.000)	25/32 (.781)	3/16 (.188)
Rounds 1–3	47/48 (.979)	77/96 (.802)	6/48 (.125)	47/48 (.979)	74/96 (.771)	12/48 (.250)
Rounds 10–12	27/48 (.562)	72/96 (.750)	2/48 (.042)	38/40 (.950)	76/80 (.950)	9/40 (.225)
Round 12	7/16 (.438)	23/32 (.719)	1/16 (.062)	11/12 (.917)	23/24 (.958)	3/12 (.250)
Equilibrium Prediction	.000	1.000	.000	1.000	1.000	.000

Table 3. Aggregate Early– and Late–Round Play (Frequency of Signaling)

Round(s)	Skill Level	p–value (1–tail permutation test)
1	H	$p > .20$
	M	$p > .20$
	L	$p > .20$
1–3	H	$p > .20$
	M	$p > .20$
	L	$p < .15$
10–12	H	$p < .05$
	M	$p < .02$
	L	$p < .05$
12	H	$p < .10$
	M	$p < .10$
	L	$p > .20$

Table 4. Significance of Differences Between Play in S Sessions and Play in C Sessions

in the two cells, but that play begins to diverge thereafter as Highs choose good grades less and less frequently in the countersignaling cell. Mediums in the signaling cell increase the frequency of choosing a good grade somewhat, with no such increase in the countersignaling cell. Lows also appear to choose good grades less and less frequently in the countersignaling cell, but this difference is slight. By the final three rounds, differences between the two cells seem apparent.

Table 3 reports the frequency with which players in the two cells chose G in early and in late rounds and Table 4 evaluates the significance of the observed frequency differences using the one–tail permutation test (Seigel and Castellan, 1988).[24] For each player type, we see that there are no significant differences in early–round play in the S and C cells.[25] Thus, at least in early rounds, players do not seem to behave in accordance with the equilibrium predictions. However, some aspects of play in later rounds are consistent with the theory's predictions. Recall that the Nash equilibrium predicts High types should choose G more often in the S cell than in the C cell, while play of Medium and Low types should be the same in both cells. In rounds 10–12 High types are far more likely to play G in the S cell than in the C cell, and this difference is significant. However, we see that Medium and Low types also play G significantly more often in the S cell than in the C cell.[26]

Because not only High types, but also Medium types, choose G more often in the C cell than in the S cell, differences between the play of Highs and that of Mediums in the C cell are smaller than one would hope, even in late rounds of the experiment. (Differences in the S cell are even smaller, but that is exactly the equilibrium prediction.) A chi–square test, using the individual–level data from round 12, gives a test statistic of 2.5 with 1 degree of freedom, equivalent to a p-value of about 0.12. This p-value is suggestive, but not significant at standard levels. A robust rank–order test, using the individual–level data from rounds 10–12, gives a test statistic of 1.52 and thus a p-value of 0.0643 (the robust rank–order test statistic has a standard normal distribution).[27] Using

[24] This test makes no assumptions about the underlying population distributions, unlike the Mann–Whitney test, which gives broadly similar results in this case, but is inappropriate because it assumes that second– and higher–order moments of the population distributions are the same and can therefore give rejections of the null hypothesis based solely or partly on differences in, for example, variances. The robust rank–order test, which also makes no assumptions about population distributions, gives generally less significant results in our case than the permutation test since it assumes data are only measurable on an ordinal scale. Our frequency data are measurable on an interval scale, allowing use of the more powerful permutation test.

[25] Data for the permutation test are calculated by averaging the frequency of good grades in each session for the given skill level over the periods noted. Aggregating data in this way leaves only four data points for each cell, but ensures that the data are independent.

[26] Lower choices of G by Medium and Low types in the C cell are not surprising given the differences between the cells. In the signaling cell choosing G by Medium and Low types has the advantage of pooling with High types who choose G frequently. Although the cost of G for Low types is sufficiently high that G is never a best response, the net payoff loss is small. In the countersignaling cell High types choose G less frequently so the advantage to Medium types of choosing G is smaller. And since High types and Low types are always distinguished in the countersignaling cell, there is no chance for Low types to pool with High types by choosing G, increasing the net payoff loss to Low types of choosing G.

[27] In order to implement this test, as well as to eliminate one possible source of dependence among data points, play over each player and type was averaged. Specifically, if the player with ID# 6 was a Medium type in rounds 11 and 12, and played G in 11 and B in 12, she was listed as having chosen G with relative frequency 0.5. If the player with

session–level data, rather than individual–level data, yields no significant differences between the play of Mediums and that of Highs in the C cell.

5.4. **Discussion.** The only difference between our S and C cells is that in the S cell, the exogenous signal is uninformative, while in the C cell, it is informative. Though not always significant at standard levels, our results are affected by this difference between cells in the direction predicted by countersignaling theory. In the S cell, bad grades are a dominant strategy for Low types; by choosing bad grades, they earn at least 400, while good grades earn them at most 383.33. Indeed, Low types (for the most part) quickly learn to get bad grades. Once Low types are mostly choosing bad grades, M and H types do best by earning good grades, and they learn this quickly, too. Play thus moves quickly toward the equilibrium.[28]

In the C cell, bad grades are again dominant for Low types, and the payoff differential is even larger—now, good grades earn them at most 290. Again, they quickly learn to choose bad grades. Once most of the Low types are choosing bad grades, good grades become a best response for Medium types, and they indeed tend to choose good grades. Once Medium types are choosing good grades, High types do better by choosing *bad* grades. While the experiment never reaches the point where all High types choose bad grades, by the final round, about half of them are doing so, while hardly any High types in the S cell do.

The reason for the apparent lack of convergence of play in the C cell to the countersignaling equilibrium may have to do with the incentives players face when play is not in equilibrium. In order for bad grades to be a best response for High types, *all* Medium types (or at least all of those who eventually turn out to receive a Pass score on the exogenous signal) must be earning good grades. Since the relative frequency of good grades by Medium types never actually reaches one, countersignaling by High types is not usually the best choice until later rounds."[29]

ID# 7 was a Medium type in only round 10, and played B, he was listed as having chosen G with relative frequency 0.

[28] There is a consistent small fraction of Low types choosing good grades, even in the last rounds of the experiment. This may be due to the fact that in equilibrium, the wage of students with bad grades is 400, while that of students with good grades is 766.67. Since subtracting the 350 cost of good grades from 766.67 yields 416.67—higher than 400—a myopic Low type might choose good grades in the hope of earning a higher payoff. Of course, when a Low type chooses good grades, he becomes part of this group, reducing the expected productivity of the group; the wage of this group falls to 714.29, so that the L's payoff to defecting from equilibrium in this way is only 364.29, and his hopes are dashed.

[29] Also, note that in the countersignaling equilibrium, a Medium type who passes the exogenous signal earns 675, while if he had chosen bad grades, while still passing, he would have earned 860. Thus, if he concentrates on what follows after passing the exogenous signal, *ignoring the 50% chance of failing the exogenous signal*, he might mistakenly

We conclude this section with a remark about the use of programmed play by signal *receivers*, *i.e.*, firms. Almost certainly, this affected the results relative to an alternative experimental design in which subjects could play the role of receivers also. Allowing receivers to be subjects instead of computer programs introduces an extra source of complexity into the experiment: from a design standpoint, motivating experimental subjects to act as competitive firms is a difficult problem. In such an experiment, receiver–subjects could be assigned a worker with a particular grade and test score, and would be asked to choose a wage for that worker. However, it would be rather difficult to set up reasonable incentives for receivers to choose a wage equal to the worker's expected productivity.[30] Even if receivers were properly motivated, it is likely that some will choose mistakenly, so that from the *senders'* point of view, they are not being paid the average productivity of workers with the same characteristics, but rather this average plus some random component.

Then, differences in expected payoff between senders' strategies (choosing good grades versus bad) will become less important, and senders should become less likely to choose the theoretical expected–payoff–maximizing action, and particularly less so when the theoretical difference in expected payoff is small to begin with. Consider now the situation faced by High type senders in the C cell, when all L types are choosing bad grades, and all Medium and High types are choosing good grades. (This situation corresponds to the equilibrium of the S cell, and was reached many times in the C cell also.) A High type earns 825 in this case, and would have earned 900 by choosing bad grades instead. Thus the payoff to bad grades is higher than that to good grades, and High types should choose bad grades. But the difference in expected payoff is quite small, and smaller yet when receivers make mistakes. Without automated receivers, it is likely that larger differences in expected payoffs would be necessary to achieve comparable results.

6. Conclusion

Addition of noisy information on type to a standard signaling model allows for equilibria in which medium types signal to distinguish themselves from low types but high types do not. Such countersignaling by high types can be seen as a sign of confidence. Signaling proves the sender

conclude that bad grades were the better choice. This type of reasoning may explain the relatively low frequency of good grades by Medium types.

[30] This is closely related to the problem of eliciting accurate probability estimates from subjects; a brief survey of attempts to deal with this latter problem is given by Camerer (1995).

is not a low type but also reveals the sender's insecurity that they would be perceived as such if they did not signal. In contrast, countersignaling indicates the sender's faith that whatever other information the receiver has on the sender will probably be consistent with the sender being of high quality.

Countersignaling equilibria can invert a number of the standard implications of signaling models. Whereas signaling equilibria can be inefficient because of excessive signaling, countersignaling equilibria may be inefficient because of inadequate signaling. While signaling equilibria can play an informational role in increasing the efficiency of receiver estimates of type, countersignaling equilibria may lower the efficiency of these estimates. And while higher costs tend to reduce signaling in a signaling model, a limited increase in costs can lead to more signaling in a countersignaling model.

Since countersignaling is more complicated behavior than signaling, the question of whether economic agents can countersignal was tested with a two-cell experiment in which exogenous information on sender type was available. In one cell the exogenous information was completely uninformative and signaling by medium and high types was the unique Nash equilibrium. In the other cell the exogenous information was partially informative and the unique Nash equilibrium involved countersignaling by high types even though high types had a negative cost to signaling. The experimental results confirm that adding noisy exogenous information on types to signaling games can affect behavior in directions consistent with the predictions of countersignaling theory. Countersignaling by high types was rare in the signaling cell but was the most common choice by the last period of the countersignaling cell.

Appendix A. Refinement issues

A.1. **Intuitive Criterion.**

Since in our game, the role of the receiver has been reduced to that of rewarding the sender her expected quality, we simplify the definition of the Intuitive Criterion. Our notation is from Fudenberg and Tirole (1991).

Definition 2. Let $u^*(q)$ be a vector of Perfect Bayesian Equilibrium payoffs for the sender. For any $s \in S$, define

$$J(s) = \{q \mid u^*(q) > H - c(s,q)\}. \tag{A.1}$$

A pure strategy Perfect Bayesian Equilibrium satisfies the *Intuitive Criterion* if and only if there does not exist $s \in S$ and $q \in Q$ such that

$$u^*(q) < \min\{Q \setminus J(s)\} - c(s,q) \tag{A.2}$$

for $\{Q \setminus J(s)\} \neq \emptyset$.

As defined, $J(s)$ is the set of types that could never do better by deviating from the equilibrium and sending signal s. According to the Intuitive Criterion, out of equilibrium beliefs must therefore put zero probability on the event that some $q \in J(s)$ sends signal s. As a result, a type $q \in Q$ agent who sends signal s can never expect to get less than $\min\{Q \setminus J(s)\} - c(s,q)$. If, for some agent, this is greater than her expected equilibrium payoff then the equilibrium fails to survive the Intuitive Criterion.

As is well known, when there are more than two types the Intuitive Criterion does not imply a unique signaling equilibrium. As we will now see, it cannot always eliminate countersignaling equilibria. In the following proposition, we characterize the set of countersignaling equilibria. It is then easy to show that under tighter conditions, though qualitatively similar to those set out in Proposition 3, this set is non-empty.

In order to characterize the set of countersignaling equilibria that satisfy the Intuitive Criterion, we first define the following additional notation. For some countersignaling equilibrium, (s^*, s^*_M, s^*), let $\dot{s}^*_L(s^*)$, $\dot{s}^*_M(s^*_M)$ and $\dot{s}^*_H(s^*)$ solve $\bar{q}_{\{L,H\}}(L) - c(s^*, L) = H - c(\dot{s}^*_L, L)$, $M - c(s^*_M, M) = H - c(\dot{s}^*_M, M)$ and $\bar{q}_{\{L,H\}}(H) - c(s^*, H) = H - c(\dot{s}^*_H, H)$. These critical values represent the highest signal agents of any type would ever be willing to send (*i.e.*, if they were believed to be of type H with probability 1).

Proposition 5. *A countersignaling equilibrium,* (s^*, s^*_M, s^*)*, survives the Intuitive Criterion if and only if* $\bar{q}_{\{L,H\}}(H) - c(s^*, H) \geq H - c(\dot{s}^*_M(s^*_M), H)$.

Proof. ($\Rightarrow$) Given some countersignaling equilibrium (s^*, s_M^*, s^*), $J(s)$ is the set of types that satisfy $\bar{q}_{\{L,H\}}(q) - c(s^*, q) > H - c(s,q)$ for $q = L, H$ and $M - c(s_M^*, M) > H - c(s, M)$. To consider this, note that by the single-crossing property, $\dot{s}_L^*(s^*) < \dot{s}_M^*(s_M^*)$. Since for sufficiently large $\bar{q}_{\{L,H\}}(H)$, $\dot{s}_H^*(s^*) \to 0$, there are three relevant, mutually exclusive cases: a) $\dot{s}_L^*(s^*) < \dot{s}_M^*(s_M^*) < \dot{s}_H^*(s^*)$, b) $\dot{s}_L^*(s^*) \leq \dot{s}_H^*(s^*) \leq \dot{s}_M^*(s_M^*)$ and c) $\dot{s}_H^*(s^*) < \dot{s}_L^*(s^*) < \dot{s}_M^*(s_M^*)$.

Under case a),

$$J(s) = \begin{cases} \emptyset & \text{if } s \leq \dot{s}_L^*(s^*) \\ \{L\} & \text{if } \dot{s}_L^*(s^*) < s \leq \dot{s}_M^*(s_M^*) \\ \{L, M\} & \text{if } \dot{s}_M^*(s_M^*) < s \leq \dot{s}_H^*(s^*) \\ \{L, M, H\} & \text{if } s > \dot{s}_H^*(s^*) \end{cases}$$

Only the second and third cases impose any restrictions on equilibrium beliefs.

For signals $s \in (\dot{s}_M^*(s_M^*), \dot{s}_H^*(s^*))$, the right hand side of (A.2) is $H - c(s,q)$. By definition, Highs are indifferent between sending signal $\dot{s}_H^*(s^*)$ for a payoff of $H - c(\dot{s}_H^*(s^*), H)$ and sending signal s^*. But $s < \dot{s}_H^*(s^*)$ and beliefs which satisfy the Intuitive Criterion put probability 1 on the event $q = H$. Thus, since s is less expensive than $\dot{s}_H^*(s^*)$, Highs must strictly prefer s to s^* (*i.e.*, $H - c(s, H) > H - c(\dot{s}_H^*(s^*), H) = \bar{q}_{\{L,H\}}(H) - c(s^*, H)$). Therefore *any* countersignaling equilibrium where $\dot{s}_H^*(s^*) > \dot{s}_M^*(s_M^*)$ fails the Intuitive Criterion and case a) cannot hold in any countersignaling equilibrium that survives the Intuitive Criterion.

Under case b),

$$J(s) = \begin{cases} \emptyset & \text{if } s < \dot{s}_L^*(s^*) \\ \{L\} & \text{if } \dot{s}_L^*(s^*) \leq s < \dot{s}_H^*(s^*) \\ \{L, H\} & \text{if } \dot{s}_H^*(s^*) \leq s < \dot{s}_M^*(s_M) \\ \{L, M, H\} & \text{if } s \geq \dot{s}_M^*(s_M^*) \end{cases}$$

Again, only the second and third cases impose restrictions on out of equilibrium beliefs.

In both cases (*i.e.*, either $s \in [\dot{s}_L^*(s^*), \dot{s}_H^*(s^*))$ or $s \in [\dot{s}_H^*(s^*), \dot{s}_M^*(s_M)))$, the right hand side of (A.2) becomes $M - c(s,q)$. Since $s_M^* < \dot{s}_L^*(s^*) < \dot{s}_H^*(s^*)$—the first inequality follows from the single-crossing property and the second by assumption—no type will ever deviate to s. That is, $u^*(q) \geq M - c(s_M^*, q) > M - c(s, q)$ for $q \in Q$ (the first inequality follows by the fact that $\mathbf{s}^*$ is an equilibrium and the second by the fact that $s \geq \dot{s}_L^*(s^*) > s_M^*$).

Under case c),

$$J(s) = \begin{cases} \emptyset & \text{if } s < \dot{s}_H^*(s^*) \\ \{H\} & \text{if } \dot{s}_H^*(s^*) < s \leq \dot{s}_L^*(s^*) \\ \{L, H\} & \text{if } \dot{s}_L^*(s^*) < s \leq \dot{s}_M^*(s_M^*) \\ \{L, M, H\} & \text{if } s > \dot{s}_M^*(s_M^*) \end{cases}$$

Again, only the second and third cases affect beliefs. Since for the second case, the right hand side of (A.2) becomes $L - c(s, q)$, no type will deviate to such an s. For the third case, the right hand side of (A.2) becomes $M - c(s, q)$. Again, since $s > \dot{s}_L^*(s^*) > s_M^*$, no type will ever deviate to such a signal.

In sum, a countersignaling survives the Intuitive Criterion if and only if $\dot{s}_M^*(s_M^*) \geq \dot{s}_H^*(s^*)$. Since $H - c(\dot{s}_H^*(s^*), H) = \bar{q}_{\{L,H\}}(H) - c(s^*, H)$, this is true if and only if $\bar{q}_{\{L,H\}}(H) - c(s^*, H) \geq H - c(\dot{s}_M^*(s_M^*), H)$.

($\Leftarrow$) Follows by reversing previous argument. ■

Thus the existence of countersignaling equilibria under the Intuitive Criterion requires that the exogenous information should further separate the Highs from the Lows.[31]

A.2. ***Divinity*-like refinements.** Consider the Cho and Kreps (1987) equilibrium refinements, D1 and D2. Clearly since the exogenous information is irrelevant to the signaling equilibria, they will never be eliminated by D1 or D2. On the other hand, we would like to determine whether

countersignaling equilibria can also survive the more stringent equilibrium concepts D1 and D2. The definition of D1 requires defining the sets of rationalizable gross payoffs that would give the sender a greater net payoff than candidate equilibrium payoff for sending signal s.

$$D(q, s) = \{\pi \mid L \leq \pi \leq H \text{ and } u^*(q) < \pi - c(s, q)\} \tag{A.3}$$

[31] Compare $\bar{q}_{\{L,H\}}(H) \geq H - c(\dot{s}_M^*(\tilde{s}_M^*(0)), H)$ to $\bar{q}_{\{L,H\}}(H) \geq M - c(\tilde{s}_M^*(0), H)$ from the proof of Proposition 3. Since $\dot{s}_M^*(\tilde{s}_M^*(0))$ is further up the Medium indifference curve, it must be the case that $H - c(\dot{s}_M^*(\tilde{s}_M^*(0)), H) > M - c(\tilde{s}_M^*(0), H)$.

$$D^0(q,s) = \{\pi \mid L \leq \pi \leq H \text{ and } u^*(q) = \pi - c(s,q)\} \tag{A.4}$$

where $u^*(q)$ is the equilibrium payoff of a type q sender.

Definition 3. The criterion D1 puts probability zero on any type q sending signal s if there is some other type q' such that

$$\{D(q,s) \cup D^0(q,s)\} \subset D(q',s). \tag{A.5}$$

The argument being that if there are a wider range of receiver payments rationalizable by some beliefs, then type q' should be infinitely more willing to send s than type q.[32]

Let $\mathbf{s}^* = (s^*, s^*_M, s^*)$ be some countersignaling equilibrium. Define $\underline{\pi}(q,s)$ to solve:

$$\begin{aligned} \underline{\pi}(L,s) - c(s,L) &= u^*(L) \\ \underline{\pi}(M,s) - c(s,M) &= u^*(M) \\ \underline{\pi}(H,s) - c(s,H) &= u^*(H) \end{aligned} \tag{A.6}$$

The $\underline{\pi}(q,s)$ should be interpreted as the minimum out of equilibrium payoff that would be necessary to induce an agent of type q to send signal s and are simply given by agents' equilibrium indifference curves. Given these definitions, it is easy to see that:

$$\begin{aligned} D(q,s) &= \{\pi \mid H \geq \pi > \underline{\pi}(q,s)\} \\ D(q,s) \cup D^0(q,s) &= \{\pi \mid H \geq \pi \geq \underline{\pi}(q,s)\} \end{aligned} \tag{A.7}$$

To analyze the stategy/type combinations which can be eliminated under D1, define $\ddot{s}_{qq'}$, for $q < q'$, to solve $\underline{\pi}(q,s) = \underline{\pi}(q',s)$. Given the single-crossing property, each of these values is unique and exists. Since $\mathbf{s}^*$ is a countersignaling equilibrium, the single-crossing property further implies that $\ddot{s}_{LM} \leq \ddot{s}_{LH} \leq \ddot{s}_{MH}$.

Proposition 6. *Countersignaling equilibrium (s^*, s^*_M, s^*) survives criterion D1 (and hence D2) if and only if $s^*_M = \tilde{s}_M(s^*)$ and $\bar{q}_{\{L,H\}}(H) - c(s^*,H) \geq H - c(\dot{s}^*_M(s^*_M),H)$.*

[32] For our model, D1 and D2 are identical because criterion D2 puts probability zero on any type q sending signal s if $\{D(q,s) \cup D^0(q,s)\} \subset \cup_{q' \neq q} D(q',s)$. D1 and D2 are equivalent here because for any q', $D(q',s)$ is an interval of the form $(\pi, H]$ and any finite union of such intervals is equal to one of its members.

Proof. ($\Rightarrow$) Let $s \neq s^*, s^*_M$. If $s \in [0, \ddot{s}_{LM})$ then $\underline{\pi}(L,s) < \underline{\pi}(M,s) < \underline{\pi}(H,s)$ so that $\{D(H,s) \cup D^0(H,s)\} \subset \{D(M,s) \cup D^0(M,s)\} \subset D(L,s)$. Thus, out of equilibrium beliefs must assign zero probability to the event that a Medium or a High sends such a signal so that $\mu(L|s,x) = 1$. Using similar arguments it is easy to show that if $s = \ddot{s}_{LM}$ then $\mu(H|s,x) = 0$, if $s \in (\ddot{s}_{LM}, \ddot{s}_{MH})$ then $\mu(M|s,x) = 1$, if $s = \ddot{s}_{MH}$ then $\mu(L|s,x) = 0$ and if $s > \ddot{s}_{MH}$ then $\mu(H,s,x) = 1$.

Since $\mathbf{s}^*$ is a countersignaling equilibrium, the only deviations that need to be considered are non-equilibrium deviations (*i.e.*, $s \neq s^*, s^*_M$). It is clear that no type will ever deviate to $s \in [0, \ddot{s}_{LM}]$ since the worst possible belief in this case is $\mu(L|s,x) = 1$ and the associated deviation payoff would be $L - c(s,q) \leq u^*(q)$ for every q.

Consider $s \in (\ddot{s}_{LM}, \ddot{s}_{MH}]$. The worst (from the sender's viewpoint)D1-consistent receiver beliefs for deterring deviations in are $\mu(M|s,x) = 1$. Mediums have an incentive to deviate to s if and only if $s < s^*_M$. Thus if $\mathbf{s}^*$ survives criterion D1, it must be that $\ddot{s}_{LM} \geq s^*_M$. But it can be seen that in any countersignaling equilibrium, $\ddot{s}_{LM} \leq s^*_M$, with equality only if $s^*_M = \tilde{s}^*_M(s^*)$ (*i.e.*, equilibria that survive D1 must have Mediums sending the lowest possible signal, given s^*). Henceforth, assume that $s^*_M = \tilde{s}^*_M(s^*)$. Since $s^*_M = \tilde{s}^*_M(s^*)$, by definition, neither Lows nor Highs have an incentive to deviate to s.

Now consider $s > \ddot{s}_{MH}$. In this case, $\mu(H|s,x) = 1$. Thus type q senders can profitably deviate to some signal s if $\ddot{s}_{MH} < \dot{s}^*_q(s^*_q)$. Since $\dot{s}^*_L(s^*) < \dot{s}^*_M(s^*_M)$, countersignaling equilibria that survive criterion D1 must have $\ddot{s}_{MH} \geq \max\{\dot{s}^*_M(s^*_M), \dot{s}^*_H(s^*)\}$. We now need to show that this condition is equivalent to $\bar{q}_{\{L,H\}}(H) - c(s^*, H) \geq H - c(\dot{s}^*_M(s^*_M), H)$. Consider if this latter condition holds with equality. By definition of $\dot{s}^*_H(s^*)$, $\dot{s}^*_H(s^*) = \dot{s}^*_M(s^*_M)$ so the former condition holds with equality. Now consider if the latter condition is a strict inequality. Since it is readily shown that $d\dot{s}^*_H(s^*)/d\bar{q}_{\{L,H\}}(H) < 0$, $d\dot{s}^*_M(s^*_M)/d\bar{q}_{\{L,H\}}(H) = 0$ and $d\ddot{s}_{MH}/d\bar{q}_{\{L,H\}}(H) > 0$, the former condition still holds.

($\Leftarrow$) Follows similarly. ■

Although D1 further restricts the set of countersignaling equilibria (there are fewer equilibrium signals available to Mediums), the condition for their existence is identical to that under the Intuitive Criterion. As a result, we can conclude that the Pareto dominant countersignaling equilibrium under D1 and D2 is identical to that under the Intuitive Criterion and therefore strongly Pareto dominates all signaling equilibria.

Furthermore, unlike the standard signaling model, not only does D1 not select a unique equilibrium, there are many countersignaling equilibria which survive D1. At first glance, this is somewhat puzzling because sender preferences are completely standard and clearly satisfy conditions (i), (ii) *and* (iii) of the Cho-Sobel Theorem (as defined in Fudenberg and Tirole (1991, p. 458)). Cho-Sobel fails because our assumption about the receiver's priors are different. In Cho-Sobel, D1-consistent beliefs are identical across all senders prior to observing the signal s, so within any group of pooling sender types the highest type would prefer to signal her type at a marginally higher cost. In our model after observing x the receiver has different beliefs for each sender which are "on average" correlated with the sender's actual type. In any equilibrium in which two or more types pool, the highest type need not prefer to break away since the exogenous information ensures that the highest type already expects to receive a higher type estimate than the lower types.

As a final note, the out of equilibrium beliefs required to support these equilibria suggest a weakness in some of the Divinity-like refinement concepts. D1 and stronger beliefs (stronger refinements include Universal Divinity but exclude Divinity) involve a discontinuity which does not seem entirely plausible. In particular, upon observing (s^*, x), the receiver believes that the sender is of type L with probability $g(x|L)f(L)/(g(x|L)f(L) + g(x|H)f(H))$ and of type H with probability $g(x|H)f(L)/(g(x|L)f(L) + g(x|H)f(H))$. However for signals, $s \in N(s^*, \epsilon) \setminus s^*$, these D1 beliefs jump discontinuously to probability 1 that the sender is of type L.

Appendix B. Instructions Used in the Experiment

Below is the text of the instructions read to subjects participating in the experiment. Text in brackets was read only to subjects participating in S sessions; text in parentheses was read only to subjects participating in C sessions. Text not in parentheses or brackets was read to all subjects.

> This is an experiment in the economics of decision-making. If you have a question at any time, please feel free to ask the experimenter.
>
> You will each be playing the role of a high school student. The experiment is made up of several rounds. In each round, you will have a certain skill level, you

will decide whether or not to earn a good grade, and you will receive a test score. Based on your grade and your test score, you will earn a salary.

The sequence of play in a round is as follows: First, players are assigned their skill levels. Then, players choose what grade to earn. Then, players find out their test score and salary, as well as the decisions of the entire group. After this, the next round begins. Skill levels and test scores will be chosen again.

At the beginning of each round, you will be randomly assigned a skill level – either High, H, Medium, M, or Low, L. The skill levels are different in [two] (three) ways. First, the skill levels vary in how likely they are to get a good test score. High types always get good scores, Medium types sometimes get good scores, and Low types never get good scores. (Second, they vary in the bonus or penalty you get for earning a good grade. Low types have a large penalty, Medium types have a small penalty, and High types have a bonus. Thirdly,) [Secondly,] firms value different skill levels differently. High types are the most valuable to firms and Low types the least. You will not be told the skill levels of any of the other players, but in each round, there will always be 4 High types, 8 Medium types, and 4 Low types.

After finding out your skill level, you will choose whether to earn a good grade, G or a bad grade, B. The bonus or penalty for each type for earning a good grade is shown on the board here. There is no bonus or penalty for earning a bad grade.

After everyone chooses what grades to earn, each player receives his or her test score – either Pass, P, or Fail, F. Your test score does not depend on your grade(, but it does depend on your skill level. High types always pass, Low types always fail, and Medium types pass 50% of the time and fail 50% of the time.)[. Each type passes 50% of the time and fails 50% of the time.] After test scores have been released, the outcomes for the entire group will be posted. There will be no information on individuals, but you will see how many players of each skill level there were in each of the 4 combinations of grades and test scores.

Then, each player will be hired by a firm. The firm that hires you does not know your skill level, but it does have access to the posted information and does know your grade and test score. Your salary for the round will be the average of the productivities of all players with the same grade and test score as you, including yourself. The productivities of the respective skill levels is shown on the board here. Your gross payoff for the round is your salary. Your net payoff for the round is your salary, plus or minus your bonus or penalty for earning a good grade (if any).

You have each been given a record sheet with spaces to write your decision and outcome for each round. At the beginning of each round, you will be given a piece of paper with one end folded and sealed. Please do not break the seal until you are asked to do so. On the outside of the paper is your skill level for the round and a space to write your grade. When you receive your paper, write your player number on the piece of paper and write your skill level in the record sheet. Then, choose what grade you will earn this round and write this in your record sheet and on the paper. Also, write down your bonus or penalty, if you earned a good grade, or 0, if you earned a bad grade, for the round. After everyone's grades are written down, players should break the seals on their papers to find out their test scores. At this time, write your test score in your record sheet. Then, the papers will be collected and results will be posted. When you find out your salary, write this down under "gross payoff" in your record sheet. Then, calculate your net payoff. After this, we will go on to the next round.

You will each receive a 5 dollar payment for completing this session. In addition, one round will be randomly chosen, and you will each receive your net payoff for that round, in addition to the 5 dollars. The exchange rate is 100 points = 1 dollar. Since every round has a chance of being chosen, you should choose your actions carefully in each round.

Are there any questions?

We will now play a practice round. This round will not be the round chosen to determine your payment, so do not worry about which action you choose in this round. We are playing this practice round to make everyone familiar with the way the experiment will be run.

References

Akerlof, G. A. (1970), 'The Market for 'Lemons:' Quality Uncertainty and the Market Mechanism,' *Quarterly Journal of Economics*, 84, 488–500.

Archibald, G. (1998), 'Coasian Bargaining Under Incomplete Information: Theories and Experiments,' Ph.D. thesis, University of Houston.

Archibald, G. and N. Wilcox (1999), 'Predicting the 'Winner's Curse' in a Coasian Bargaining Game with Incomplete Information,' Working paper, University of Houston.

Ball, S. B. (1991), 'Experimental Evidence on the Winner's Curse in Negotiations,' Ph.D. thesis, Northwestern University.

Ball, S. B., M. H. Bazerman and J. S. Carroll (1991), 'An Evaluation of Learning in the Bilateral Winner's Curse,' *Organizational Behavior and Human Decision Processes*, 48, 1–22.

Banks, J. S. and J. Sobel (1987), 'Equilibrium Selection in Signaling Games,' *Econometrica*, 55, 647–662.

Bhattacharyya, N. (1998), 'Good Managers Work More and Pay Less Dividends: A Screening Model of Dividend Policy,' Mimeo.

Brown, S. and W. Hamilton (1996), 'Autumn Colors: A Role for Aphids?' *Proceedings of XXth International Congress of Entomology*, p. 232.

Camerer, C. (1988), 'Gifts as Economic Signals and Social Symbols,' *American Journal of Sociology*, 94, S180–214.

Camerer, C. (1995), 'Individual Decision Making,' in J. Kagel and A. E. Roth, eds., 'Handbook of Experimental Economics,' pp. 587–703, Princeton University Press, Princeton, NJ.

Cho, I.-K. and D. M. Kreps (1987), 'Signaling games and stable equilibria,' *Quarterly Journal of Economics*, 102, 179–221.

Cho, I.-K. and J. Sobel (1990), 'Strategic Stability and Uniqueness in Signaling Games,' *Journal of Economic Theory*, 50, 381–413.

Cifuentes, L. A. and S. Sunder (1991), 'Some Further Evidence of the Winner's Curse,' Mimeo, Carnegie Mellon University.

Cooper, D. J., S. Garvin and J. H. Kagel (1997a), 'Adaptive Learning Vs. Equilibrium Refinements in an Entry Limit Pricing Game,' *Economic Journal*, 107, 553–575.

Cooper, D. J., S. Garvin and J. H. Kagel (1997b), 'Signalling and Adaptive Learning in an Entry Limit Pricing Game,' *Rand Journal of Economics*, 28, 662–683.

Engers, M. (1987), 'Signalling with Many Signals,' *Econometrica*, 55, 663–674.

Fudenberg, D. and J. Tirole (1991), *Game Theory*, MIT Press, Cambridge, MA.

Garvin, S. and J. H. Kagel (1994), 'Learning in Common–Value Auctions: Some Initial Observations,' *Journal of Economic Behavior and Organization*, 25, 351–372.

Haiman, J. (1998), *Talk is Cheap: Sarcasm, Alienation, and the Evolution of Language*, Oxford University Press.

Harbaugh, W. T. (1998), 'What Do Donations Buy? A Model of Philanthropy Based on Prestige and Warm Glow,' *Journal of Public Economics*, 67(2), 269–84.

Hertzendorf, M. N. (1993), 'I'm not a High-Quality Firm – but I Play One on TV,' *Rand Journal of Economics*, 24, 236–247.

Hubler, A. (1983), *Understatements and Hedges in English*, John Benjamins.

Huck, S., H. T. Normann and J. Oechssler (1999), 'Learning in Cournot Oligopoly: An Experiment,' *Economic Journal*, 109, C80–95.

Hvide, H. K. (1999), 'The Informational Role of Education in the Allocation of Talent,' Norwegian School of Economics working paper.

Kagel, J. H. and D. Levin (1986), 'The Winner's Curse and Public Information in Common Value Auctions,' *American Economic Review*, 76(5), 894–920.

Nasar, S. (1998), *A Beautiful Mind: A Biography of John Forbes Nash, Jr.*, Simon and Schuster, New York.

Naylor, R., J. Smith and A. McKnight (1998), 'Occupational Earnings: Evidence for the 1993 UK University Graduate Population from the University Student Record,' University of Warwick working paper.

Nelson, P. (1974), 'Advertising as Information,' *Journal of Political Economy*, 82, 729–754.

Pesendorfer, W. (1995), 'Design Innovation and Fashion Cycles,' *American Economic Review*, 85, 771–792.

Prendergast, C. and L. Stole (1996), 'Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning,' *Journal of Political Economy*, pp. 1105–1134.

Quinzii, M. and J.-C. Rochet (1985), 'Multidimensional Signalling,' *Journal of Mathematical Economics*, 14, 261–284.

Ross, S. A. (1977), 'The Determination of Financial Structure: The Incentive-Signalling Approach,' *Bell Journal of Economics*, 8, 23–40.

Roth, A. E., V. Prasnikar, M. Okuno-Fujiwara and S. Zamir (1991), 'Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study,' *American Economic Review*, 81, 1068–1095.

Seigel, S. and N. J. Castellan, Jr. (1988), *Non-Parametric Statistics for the Behavioral Sciences*, McGraw-Hill, New York.

Smith, J. and R. Naylor (1998), 'Determinants of Individual Degree Performance: Evidence for the 1993 UK University Graduate Population from the University Student Record,' University of Warwick working paper.

Spence, A. M. (1973a), 'Job Market Signaling,' *Quarterly Journal of Economics*, 87, 355–374.

Spence, A. M. (1973b), 'Time and Communication in Economics and Social Interaction,' *Quarterly Journal of Economics*, 87, 651–660.

Spence, A. M. (1974a), 'Competitive and Optimal Responses to Signals: An Analysis of Efficiency and Distribution,' *Journal of Economic Theory*, 7, 296–332.

Spence, A. M. (1974b), *Market Signaling*, Harvard University Press.

Spier, K. E. (1992), 'Incomplete Contracts and Signalling,' *Rand Journal of Economics*, 23, 432–43.

Teoh, S. H. and C. Y. Hwang (1991), ‘Nondisclosure and Adverse Disclosure as Signals of Firm Value,’ *Review of Financial Studies*, 4(2), 283–313.

Veblen, T. (1899), *The Theory of the Leisure Class*, Macmillan, New York.

Zahavi, A. (1975), ‘Mate Selection – a Selection for a Handicap,’ *Journal of Theoretical Biology*, 53, 205–214.

Department of Economics, University of Houston, Houston, TX 77204–5882, USA.

E-mail address: nfelt@bayou.uh.edu

Claremont McKenna College and Claremont Graduate University, Claremont, CA 91711, USA.

E-mail address: rick_harbaugh@mckenna.edu

Centre for the Study of Globalisation and Regionalisation, University of Warwick, Coventry CV4 7AL, UK.

E-mail address: T.To@warwick.ac.uk